DISSERTATION

LINEAR PREDICTION AND PARTIAL TAIL CORRELATION FOR EXTREMES

Submitted by

Jeongjin Lee

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Daniel Cooley

Piotr Kokoszka
Jay Breidt
Ali Pezeshki

ABSTRACT

LINEAR PREDICTION AND PARTIAL TAIL CORRELATION FOR EXTREMES

This dissertation consists of three main studies for extreme value analyses: linear prediction for extremes, uncertainty quantification for predictions, and investigating conditional relationships between variables at their extreme levels. We employ multivariate regular variation to provide a framework for modeling dependence in the upper tail, which is assumed to be a direction of interest. Cooley and Thibaud [2019] consider transformed-linear operations to define a vector space on the nonnegative orthant and show regular variation is preserved by these transformed-linear operations.

Extending the approach of Cooley and Thibaud [2019], we first consider the problem of performing prediction when observed values are at extreme levels. This linear approach is motivated by the limitation that traditional extreme value models have difficulties fitting a high dimensional extreme value model. We construct an inner product space of nonnegative random variables from transformed-linear combinations of independent regularly varying random variables. Rather than fully characterizing extremal dependence in high dimensions, we summarize tail behavior via a matrix of pairwise tail dependencies. The projection theorem yields the optimal transformed-linear predictor, which has a similar form to the best linear unbiased predictor in non-extreme prediction.

We then quantify uncertainty for the prediction of extremes by using information contained in the tail pairwise dependence matrix. We create the 95% prediction interval based on the geometry of regular variation. We show that the prediction intervals have good coverage in a simulation study as well as in two applications: prediction of high NO_2 air pollution levels, and prediction of large financial losses. We also compare prediction intervals with a linear approach to ones with a parametric approach.

Lastly, we develop the novel notion of partial tail correlation via projection theorem in the inner product space. Partial tail correlations are the analogue of partial correlations in non-extreme statistics but focus on extremal dependence. Partial tail correlation can be represented by the inner product of prediction errors associated with the previously defined best transformed-linear prediction for extremes. We find a connection between the partial tail correlation and the inverse matrix of tail pairwise dependencies. We then develop a hypothesis test for zero elements in the inverse extremal matrix. We apply the idea of partial tail correlation to assess flood risk in application to extreme river discharges in the upper Danube river basin. We compare the extremal graph constructed from the idea of the partial tail correlation to physical flow connections on the Danube.

ACKNOWLEDGEMENTS

DEDICATION

to Yoonjung

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

1.1   Motivation

In 2020, Colorado experienced its largest wildfire, the Cameron Peak Fire, that had burned more than 200,000 acres west of the city of Fort Collins. The fire's rapid spread was caused by extreme temperatures, heavy winds, and low humidity. The fire had destroyed a total of 469 structures including 220 outbuildings and 42 primary residences and forced the evacuation of over 16,000 people. Risks of such extreme events pose great hazards to the environment and to society. Extreme value methods are critical to assess risk associated with such extreme events across multiple disciplines, including hydrology, atmosphere science, engineering, finance, and insurance. Extreme value models can also be utilized in a number of purposes such as detecting and predicting extremes, and quantifying uncertainty in extremes.

The extremes philosophy is to "let the tail speak for itself". Consequently, statistical approaches for extremes focus only on large values as it is assumed that information gleaned from the distribution's bulk may not be useful for characterizing tail behavior. Extreme value theory (EVT) provides asymptotic characterizations of a distribution's tail, which provides a theoretical foundation for inference that extrapolates further into the tail beyond the data's range. The earliest focus of EVT was on describing limiting univariate distributions for block maxima and threshold exceedances. Methods for univariate extremes are well developed, although research continues on improving estimation and inference. Much recent research in extremes has been focused on describing and modeling dependence in the tail. There are several related frameworks that are commonly used for modeling multivariate extremes: the framework provided by characterizations of the multivariate extreme value distributions (De Haan and Ferreira [2007]), that provided by the multivariate generalized Pareto distributions, or that provided by multivariate regular variation.

One approach to model multivariate extreme value distributions is the block maxima approach. As the normalized block maxima converge in distribution to a non-degenerate distribution function, the limiting distribution can be approximated by a parametric family. There are several parametric families such as a standard class of the logistic family [Coles et al., 2001], the bilogistic family [Joe et al., 1992], or the Dirichlet family [Coles and Tawn, 1991]. Another approach is through a threshold exceedances approach. The classical asymptotic model for excesses over a high threshold is the generalized Pareto distribution [Rootzén and Tajvidi, 2006]. In the multivariate regular variation, the limiting measure has a radial measure and an angular measure of which they are independent. The angular measure describes extremal dependence. Some parametric angular measures are a Dirichlet model [Coles and Tawn, 1991], a pairwise beta model [Cooley et al., 2010], and mixture models [Boldi and Davison, 2007]. However, the main challenge in traditional extreme value analyses is that it is hard to fit multivariate extreme value models in high dimensions.

In this dissertation, we do not aim to fully characterize the tail behavior via a specified model for an extreme value distribution, a multivariate GPD, or the angular measure of a regular varying random vector. Rather, we extend the approach of Cooley and Thibaud [2019] who summarize tail behavior via a matrix of pairwise tail dependencies. Cooley and Thibaud [2019] rely on multivariate regular variation to provide a framework for modeling dependence in the upper tail, which is assumed to be a direction of interest. Cooley and Thibaud [2019] employ transformed-linear operations to define a vector space on the nonnegative orthant and show these transformed-linear operations preserve regular variation on the nonnegative orthant. It leads to a new framework that is tied to well-established linear models in traditional statistics.

Under the framework, we consider the linear prediction that relies on the matrix of pairwise tail dependencies. The advantage of the linear approach is that it can be fit in high dimensions. Using the information in the extremal matrix, we quantify uncertainty for the prediction of extremes. We also consider exploring relationships between variables given all other variables at their extreme levels. We develop the novel notion of partial tail correlation to investigate conditional relationships at their extreme levels.

This dissertation is organized as follows. In the remainder of Chapter 1, we provide essential concepts of EVT to understand the extreme value methods that we develop in the following chapters. We briefly review univariate extremes for block maxima and threshold exceedances and then introduce multivariate extremes. We introduce the framework of multivariate regular variation and describe how it is connected to multivariate extreme value models or threshold exceedances.

In Chapter 2, we consider the problem of performing prediction when observed values are at their extreme levels. Focusing on modeling dependence in the upper tail, we construct an inner product space constructed from transformed-linear combinations of independent regularly varying random variables. Instead of fully characterizing the asymptotic dependence in high dimensions, we summarize pairwise tail dependencies in a matrix. The projection theorem yields the optimal transformed-linear predictor, which has the same form as the best linear unbiased predictor in non-extreme prediction.

In Chapter 3, we quantify uncertainty for the prediction of extremes. We construct not only the 95% joint prediction region but the 95% conditional prediction interval by using the information in the tail pairwise dependence matrix. These intervals are constructed in the polar geometry of regular variation. We apply the linear approach to both nitrogen dioxide air pollution and financial losses. For the air pollution application, as a parametric approach for angular measures can be easily applied to five monitoring sites, we compare prediction intervals with the linear approach to ones with a parametric approach. Chapters 2 and 3 are based on the manuscript in a preprint on arXiv [Lee and Cooley, 2021].

In Chapter 4, we develop the novel notion of partial tail correlation in the inner product space constructed from the previous chapters. The projection theorem provides a natural way of defining the partial tail correlation in terms of the inner product of prediction errors. We find a connection between the partial tail correlation and the inverse of the tail pairwise dependence matrix by matrix inversion. Similarly to Gaussian cases, zero elements in the inverse extremal matrix imply conditional independence-like behaviors between variables at their extreme levels. We use the idea of partial tail correlation to explore partially correlated variables in extremes. We develop a

hypothesis test for zero elements in the inverse extremal matrix. The idea of partial tail correlation is applied to assess flood risk in application to extreme river discharges in the upper Danube river basin. Lastly, we conclude the dissertation with a summary and future work.

1.2   Univariate Extremes

We introduce two main approaches to modeling univariate extremes; block maxima approach and threshold excess approach. These reviews are by no means exhaustive. We refer to books that cover classical extreme value models (e.g., Coles et al. [2001], Beirlant et al. [2004]).

1.2.1   Block Maxima Approach

Classical extreme value theory focuses on the asymptotic behavior of maximum

$$M_n = \max\{X_1, ..., X_n\},$$

where $X_1, ..., X_n$ is a sequence of independent variables having the same distribution F. A typical example of M_n is the annual maxima when n is a number of measurements in a year,

In the setting of iid cases, we can derive the exact distribution of M_n for all n,

$$P(M_n \leq x) = P(X_1 \leq x) \times \cdots P(X_n \leq x)$$
$$= [F(x)]^n$$

However, the distribution function $F(x)$ is unknown. Plugging an empirical estimate into $F(x)$ is also problematic because a small difference in the estimate can cause a huge difference for $F(x)^n$. Let x^+ be an upper end-point for $F(x)$. It is shown that the distribution of M_n is degenerate on the upper end-point of $F(x)$ since $F(x)^n \to 0$, as $x \to \infty$ in (Resnick [2007]).

An alternative way is to find asymptotic families of models for F^n by renormalizing the variable M_n,

$$\frac{M_n - a_n}{b_n},$$

4

where $\{b_n > 0\}$ and $\{a_n\}$ are sequences of constants. By the extremal types theorem (Fisher and Tippett [1928]; Gnedenko [1943]), if there exist $\{b_n > 0\}$ and $\{a_n\}$ such that

$$P\Big(\frac{M_n - a_n}{b_n}\Big) = F^n(b_n x + a_n) \rightarrow G(x), \tag{1.1}$$

where $G(x)$ is a non-degenerate distribution function, then $G(x)$ belongs to one of the three possible families

$$\text{Type I (Gumbel)} : G(x) = \exp(-\exp(-x)), \ x \in \mathbb{R} \tag{1.2}$$

$$\text{Type II (Fréchet)} : G(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \tag{1.3}$$

$$\text{Type III (Weibull)} : G(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \leq 0 \\ 1, & x \geq 0 \end{cases} \tag{1.4}$$

where the Fréchet and Weibull families have a shape parameter $\alpha > 0$. The remark on this theorem is that the normalized M_n with suitable sequences $\{b_n > 0\}$ and $\{a_n\}$ is stabilized and its limiting distribution must belong to one of the three families above.

It may be appealing to choose one of the three families for the data and then perform statistical inference. However, the problem is that it does not allow us to capture uncertainty for such a decision, which is important in extreme value analyses. We can circumvent this issue by reformulating the models in the extreme types theorem. We combine the Gumbel, Fréchet, and Weibull families into a single family having a distribution form as

$$G(x) = \exp\Big\{-\Big[1 + \xi\Big(\frac{x - \mu}{\sigma}\Big)\Big]_+^{-1/\xi}\Big\}, \tag{1.5}$$

where $\{x \in \mathbb{R} : 1 + \xi(x - \mu)/\sigma > 0\}$, a location parameter $\mu \in \mathbb{R}$, a scale parameter $\sigma > 0$, and a shape parameter $\xi \in \mathbb{R}$. This family is called the generalized extreme value (GEV) family of

distributions. The shape parameter ξ is of primary interest in extreme value analyses. The Gumbel family corresponds to $\xi = 0$ (taken as a limit) and the tail is considered light. The Weibull family corresponds to $\xi < 0$ and the tail is bounded. The Fréchet family corresponds to $\xi > 0$ and the tail is considered heavy. With this unification of the three families of extreme value distributions, we do not need to make a decision about which family to use and we can also quantify uncertainty associated with the choice of the shape parameter.

1.2.2 Threshold Exceedances Approach

A potential drawback of the block maxima approach is that this method could be viewed as wasteful of data if data besides the block maxima are informative of the tail. The threshold exceedance approach is an alternative method which sets a suitably high threshold and utilizes all data which exceeds the threshold. The classical asymptotic model for excesses over a high threshold is called the genaralized Pareto distribution (GPD) (Balkema and De Haan [1974];Pickands III [1975]). Let X have distribution F and let $\bar{F}$ denote the survival function of X. Then, the distribution above the threshold u can be described conditionally:

$$P(X - \mu \leq x | X > u) = \frac{F(u + x)}{\bar{F}(u)}.$$

Since F is unknown, we approximate this conditional distribution by using the approximate distribution of M_n. From (1.5), for large enough n, we know $F^n(x) \approx \exp\{-[1 + \xi(\frac{x-\mu}{\sigma})_+^{-1/\xi}]\}$. Taking logs of both sides yields

$$n \log F(x) \approx -\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}. \tag{1.6}$$

For large x, $-\log F(x) \approx (1 - F(x))$ by a Taylor approximation. Substituting this into (1.6) gives

$$1 - F(u) \approx \frac{1}{n}\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}. \tag{1.7}$$

and for $x > 0$,

$$1 - F(u + x) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u + x - \mu}{\sigma} \right) \right]_+^{-1/\xi}. \tag{1.8}$$

Hence, for large u,

$$\begin{aligned}
P(X > u + x | X > u) &= \frac{1 - F(u + x)}{1 - F(u)} \\
&\approx \frac{\left[1 + \xi \left(\frac{u + x - \mu}{\sigma} \right) \right]_+^{-1/\xi}}{\left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi}} \\
&= \left(1 + \frac{\xi}{\sigma_u} x \right)_+^{-1/\xi},
\end{aligned} \tag{1.9}$$

where $\sigma_u = \sigma + \xi(u - \mu)$. Over a high threshold u, the conditional distribution of $X - u | X > u$ is approximately,

$$P(X - \mu \le x | X > u) \approx H(x) := 1 - \left(1 + \frac{\xi}{\sigma^*} x \right)_+^{-1/\xi}, \tag{1.10}$$

where the distribution is defined on $\{x : x > 0, 1 + \frac{\xi}{\sigma^*} x > 0\}$, $\sigma^* = \sigma + \xi(u - \mu)$, and $H(x)$ is called the generalized Pareto family.

The remarkable feature of the generalized Pareto distribution is that its parameters are uniquely determined by those of the GEV distribution. The shape parameter ξ is identical to that of the GEV distribution. The shape parameter ξ describes the tail behavior as it does for the GEV distribution.

To fit the GPD, it requires the threshold choice. There are several ways to select the threshold. Scarrott and MacDonald [2012] provide an overview of existing methods for threshold selection. One graphical diagnostic method is to use the fact that $E(X - u | X > u)$ is a linear function of u. If the generalized Pareto distribution is approximated well over the high threshold u, then the empirical estimate for $E(X - u | X > u)$ should be linear in u. We can choose the threshold at which the mean residual life plot is approximately linear in u. See more details in Coles et al. [2001].

1.3 Multivariate Extremes

Similarly to univariate extremes, the tail behavior of the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ can be described by two approaches: componentwise block maxima and threshold exceedances. We review these classical multivariate extreme models and its link with regular variation.

1.3.1 Multivariate Extreme Value Distributions

We consider the limiting distribution of the vector of renormalized componentwise maxima. Let $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ be a d-dimensional random vector with a distribution function $F(\boldsymbol{x})$. Suppose $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^T$ is iid copies of $\boldsymbol{X}$. We define the vector of componentwise maxima as

$$\boldsymbol{M}_n = \left(\vee_{i=1}^n X_{i,1}, ..., \vee_{i=1}^n X_{i,d} \right)^T$$

$\boldsymbol{M}_n$ is a vector composed of the maximum of each component in the sample. Note that M_n does not necessarily correspond to an observed vector. In other words, maxima do not necessarily occur simultaneously. Similarly to the univariate block maxima approach, if there exist sequences of vectors $\boldsymbol{b}_n > \boldsymbol{0}$ and $\boldsymbol{a}_n \in \mathbb{R}^d$ such that

$$P\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \le x\right) = F^n(\boldsymbol{a}_n\boldsymbol{x} + \boldsymbol{b}_n) \xrightarrow{d} G(\boldsymbol{x}), \tag{1.11}$$

as $n \to \infty$, where vector operations are applied componentwisely, then $G(\boldsymbol{x})$ is called a multivariate extreme value distribution (MVEVD). That is, the family of multivariate extreme value distributions are the limiting distributions of componentwise block maxima. The renormalized random vectors converge only if all marginals converge,

$$F_i^n(\boldsymbol{b}_n\boldsymbol{x} + \boldsymbol{a}_n) \xrightarrow{d} G_i(\boldsymbol{x}),$$

for $i = 1, \ldots, d$ and where F_i and G_i are the ith marginals of F and G, respectively. Note that each marginal of $G_j(\boldsymbol{x})$ is the generalized EVD. It is shown that the MVEVD is represented by $G(\boldsymbol{x}) = \exp(-\nu[\boldsymbol{0}, \boldsymbol{x}]^c])$ where ν is called the exponent measure in Resnick [1987].

For analyzing the tail dependence, we typically estimate the marginal distributions F_j to assume a common margin. For simplicity, we assume the vector $\boldsymbol{X}$ is normalized to standard Pareto margins. The normalized vector $\boldsymbol{X}$ is said to be in the max-domain of attraction of $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ if

$$\lim_{n \to \infty} P(\vee_{i=1}^n X_{i,1} \leq n z_1, ..., \vee_{i=1}^n X_{i,d} \leq n z_d) = P(\boldsymbol{Z} \leq \boldsymbol{z}),$$

where $\boldsymbol{Z}$ is a max-stable with unit Fréchet margins $P(Z_j \leq z) = \exp(-1/z)$, $z \geq 0$ for $j = 1, \ldots, d$. In this case, $P(\boldsymbol{Z} \leq \boldsymbol{z}) = \exp\{-\nu(\boldsymbol{z})\} := \exp\{-\nu([\boldsymbol{0}, \boldsymbol{z}]^c)\}$, where $\boldsymbol{z} \geq \boldsymbol{0}$ and the exponent measure ν is defined on the cone $E := [0, \infty)^d \setminus \{\boldsymbol{0}\}$. The tail dependence structure is characterized by the exponent measure ν. This exponent measure can be linked to regular variation described in the next subsection.

Similarly to threshold exceedances approach in univariate case, the idea of thresholding can also be used for modeling multivariate extremes. The challenge of this approach for multivariate models is how to define exceedances over multivariate thresholds of a random vector. Rootzén and Tajvidi [2006] consider the multivariate GPD where a vector $\boldsymbol{x} \in \mathbb{R}^d$ exceeds a multivariate threshold $\boldsymbol{u} \in \mathbb{R}^d$ if at least one component exceeds a high threshold. Furthermore, Rootzén et al. [2018] provide a wide range of parametrizations, properties for the multivariate Pareto distribution.

1.3.2 Multivariate Regular Variation

Informally, a multivariate regularly varying random variable has a distribution which is jointly heavy tailed. Regular variation is closely tied to classical extreme value analysis [De Haan and Ferreira, 2007, Appendix B], and Resnick [2007] gives a comprehensive treatment. Let $\boldsymbol{X}$ be a p-dimensional random vector that takes values in $\mathbb{R}_+^p = [0, \infty]^p$. $\boldsymbol{X}$ is regularly varying (denoted $RV_+^p(\alpha)$) if there exists a function $b(s) \to \infty$ as $s \to \infty$ and a non-degenerate limit measure $\nu_{\boldsymbol{X}}$

9

for sets in $E := [0, \infty)^p \setminus \{\mathbf{0}\}$ such that

$$sPr(b(s)^{-1}\mathbf{X} \in \cdot) \xrightarrow{v} \nu_{\mathbf{X}}(\cdot) \tag{1.12}$$

as $s \to \infty$, where $\xrightarrow{v}$ indicates vague convergence in the space of non-negative Radon measures on $[0, \infty]^p \setminus \{\mathbf{0}\}$. The normalizing function is of the form $b(s) = U(s)s^{1/\alpha}$ where $U(s)$ is a slowly varying function, and α is termed the tail index. For any set $C \subset [0, \infty]^p \setminus \{\mathbf{0}\}$ and $k > 0$, the measure has the scaling property $\nu_{\mathbf{X}}(kC) = k^{-\alpha}\nu_{\mathbf{X}}(C)$. This scaling property implies regular variation can be more easily understood in a polar geometry.

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

$$sPr((b(s)^{-1}R, \mathbf{W}) \in \cdot) \xrightarrow{\nu} c\nu_\alpha \times H_{\mathbf{X}}, \tag{1.13}$$

where ν_α is the measure on $(0, \infty]$. The right hand side of (1.13) is a product measure, implying that the radial and angular measure are independent of each other in the limit. The measure $H_{\mathbf{X}}$ is called the angular or spectral measure on Θ_{p-1}^+. The angular measure $H_{\mathbf{X}}$ fully describes tail dependence in the limit; however, modeling $H_{\mathbf{X}}$ even in moderate dimensions is difficult. For a set $C(r, B) = \{\mathbf{x} \in \mathbb{R}_+^p : \|\mathbf{x}\| > r, \mathbf{x}/\|\mathbf{x}\| \in B\}$ with Borel set $B \subset \Theta_{p-1}^+ = \{\mathbf{w} \in E : \|\mathbf{w}\| = 1\}$, the scaling property of $H_{\mathbf{X}}$ implies $\nu_{\mathbf{X}}(C(r, B)) = cr^{-\alpha}H_{\mathbf{X}}(B)$. The measure's intensity function in terms of polar coordinates is $\nu_{\mathbf{X}}(dr \times d\mathbf{w}) = \alpha r^{-\alpha-1}drdH_{\mathbf{X}}(\mathbf{w})$. There are many possible normalizations $b(s)$. But, all normalizations are asymptotically equivalent and yield only different constants by Resnick [2007]. We apply the same normalization $b(s)$ for all components of $\mathbf{X}$ and assume $\mathbf{X}$ has common marginal distributions throughout.

Now we describe the link between multivariate extreme value distributions and regular variation. Given any norm $\| \cdot \|$ and H, for the set $A = [\mathbf{0}, \mathbf{z}]^c$, the exponent measure ν corresponding

to the MVEVD can be expressed in Cartesian coordinates,

$$\nu(\{A\}) = \int_A \alpha r^{-\alpha-1} dr dH(\boldsymbol{w})$$

$$= \int_{\Theta_+^{d-1}} \int_{r=\wedge_{i=1}^d \frac{z_i}{w_i}}^{\infty} \alpha r^{-\alpha-1} dr dH(\boldsymbol{w})$$

$$= \int_{\Theta_+^{d-1}} \bigvee_{i=1}^d \left(\frac{z_i}{w_i}\right)^{-\alpha} H(d\boldsymbol{w}).$$

Due to the max operation, the tail dependence structure can be more easily understood through the polar geometry of regular variation.

We describe two important opposite ends of dependence in extremes analyses from Resnick [2007]: asymptotic independence and asymptotic full dependence. Suppose $\boldsymbol{X}$ is a $\mathbb{R}_+^d$-valued regularly varying random vector with tail index α such that

$$nP\left[\frac{\boldsymbol{Z}}{b_n} \in \cdot\right] \xrightarrow{\nu} \nu,$$

where

$$b_n = \left(\frac{1}{1-F}\right)^{\leftarrow}(n).$$

And $\nu(\cdot)$ is

$$\nu(dx^{(1)}, \ldots, dx^{(d)}) = \sum_{i=1}^d \delta(dx^{(1)}) \times \cdots \times \delta(dx^{(i-1)}) \times \nu(dx^{(i)}) \times \cdots \times \delta(dx^{(d)}).$$

11

Then, $\boldsymbol{X}$ has asymptotic independence. It means that the measure ν puts all masses on the axes and there is no mass in the interior of the positive orthant. For $d = 2$ and a constant $c \in \mathbb{R}$,

$$
\begin{aligned}
\lim_{t \to \infty} P[Z^{(2)} > t | Z^{(1)} > t] &= \lim_{t \to \infty} \frac{P[Z^{(1)} > t, Z^{(2)} > t]}{Z^{(1)} > t} \\
&= \lim_{n \to \infty} \frac{P[Z^{(1)} > b_n, Z^{(2)} > b_n]}{[Z^{(1)} > b_n]} \\
&= \lim_{n \to \infty} cn P[Z^{(1)} > b_n, Z^{(2)} > b_n] \\
&= c\nu(\boldsymbol{1}, \boldsymbol{\infty}] = 0
\end{aligned}
$$

This means if one component is large, the probability that the other component is also large is zero.

In contrast, suppose $\boldsymbol{X}$ is a $\mathbb{R}_+^d$-valued regularly varying random vector whose tail distribution concentrates on the positive orthant. If the limit measure ν has the form of

$$
\nu([\boldsymbol{0}, \boldsymbol{x}]^c) = \nu(\cup_{i=1}^d \{\boldsymbol{y} : y^{(i)} > x^{(i)}\}) = (\wedge_{i=1}^d x^{(i)})^{-\alpha},
$$

then $\boldsymbol{X}$ has asymptotic full dependence. In this case, the limit measure ν puts all masses on $\boldsymbol{1}/||\boldsymbol{1}||$.

Chapter 2

Transformed-linear Prediction for Extremes

2.1 Motivation

Prediction of unobserved quantities is a common objective of statistical analyses. Figure 2.1 shows the one-hour maximum measurements of the air pollutant nitrogen dioxide (NO_2) in parts per billion for four monitoring stations in the Washington DC area on January 23, 2020. Given these measurements, it is natural to ask what the predicted level would be at a nearby unmonitored location, such as Alexandria VA, which is marked "Alx" in Figure 2.1 and which had NO_2 monitoring prior to 2015. Linear prediction methods offer a straightforward answer by simply applying weights to each of the observations. In spatial statistics, linear prediction is called kriging, which provides the best (in terms of mean square prediction error) linear unbiased prediction (BLUP) weights given a covariance structure between the observed and unobserved measurements. Uncertainty is often summarized by the mean-square prediction error and prediction intervals are commonly based on Gaussian assumptions.

On this particular day however, the measurements are at very high levels–each exceeds each station's empirical 0.98 quantile for this year, and the Arlington station (Arl) is recording its highest measurement for the year. The aforementioned standard approach uses the 'usual' behavior of the data to describe relationships between sites and would not take into account that the observed measurements are at extreme levels. Covariance could be a poor descriptor of dependence in this distribution's joint upper tail, and Gaussian assumptions may be poorly suited to describe uncertainty in the tail. Moreover, public interest is likely highest when pollution levels are high.

In this work, we propose a method which is similar in spirit to familiar linear prediction, but which is specifically designed for prediction when variables are at extreme levels. We will analyze only data which are extreme to summarize dependence in the joint tail. To provide a framework for modeling dependence in the upper tail, we rely on regular variation on the positive orthant. Mod-
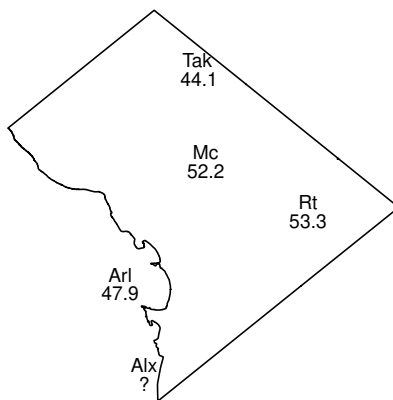
Figure 2.1: Maximum NO_2 measurements for January 23, 2020. All observations are above the empirical .98 quantile for each location.

eling in the positive orthant allows our method to focus only on the upper tail, which is assumed to be the direction of interest; in this example we are interested in predicting when pollution levels are high. On the way to developing our prediction method, we will construct a vector space of non-negative regularly-varying random vectors constructed from transformed-linear operations. We summarize pairwise tail dependencies in a matrix which has properties analogous to a covariance matrix. Our transformed-linear predictor has a similar form to the BLUP in non-extreme linear prediction. Uncertainty quantification is quite different because the geometry of regular variation is quite different than the elliptical geometry underlying standard linear prediction settings. We will show that our method has good coverage when applied to air pollution data and also when applied to a larger financial data set.

2.2   Background

2.2.1   Transformed Linear Operations

In order to perform linear-like operations for vectors in the positive orthant, Cooley and Thibaud [2019] defined transformed linear operations. Consider $x \in \mathbb{R}_+^p = [0, \infty)^p$, let t be a monotone bijection mapping from $\mathbb{R}$ to $\mathbb{R}_+$, with t^{-1} its inverse. For $y \in \mathbb{R}^p$, $t(y)$ applies the transform

componentwise. For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 \in \mathbb{R}_+^p = [0, \infty)^p$, define vector addition as $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 = t\{t^{-1}(\boldsymbol{x}_1) + t^{-1}(\boldsymbol{x}_2)\}$, and define scalar multiplication as $a \circ \boldsymbol{x}_1 = t\{at^{-1}(\boldsymbol{x}_1)\}$ for $a \in \mathbb{R}$. We define the additive identity as $\boldsymbol{0}_{\boldsymbol{x}} = t(\boldsymbol{0})$ and the additive inverse of any $\boldsymbol{x} \in \mathbb{R}_+^p$ as $-\boldsymbol{x} = t\{-t^{-1}(\boldsymbol{x})\}$. It is straightforward to show that $\mathbb{R}_+^p$ with these transformed-linear operations is a vector space as it is isomorphic to $\mathbb{R}^p$ with standard operations. For $x_j \in \mathbb{R}_+$ and $a_j \in \mathbb{R}$, $j = 1, ..., q$, a transformed-linear combination is defined as

$$a_1 \circ x_1 \oplus \cdots \oplus a_q \circ x_q = t\Big\{ \sum_{j=1}^{q} a_j t^{-1}(x_j) \Big\}. \tag{2.1}$$

Let $A = (\boldsymbol{a}_1, ..., \boldsymbol{a}_q)$ be a $p \times q$ matrix of real numbers. We define matrix multiplication 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 the $p \times p$ identity matrix I, $I \circ x = t\{It^{-1}(x)\} = x$. 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. To apply transformed linear operations to non-negative regularly-varying random vectors, Cooley and Thibaud [2019] consider the specific transform $t : \mathbb{R} \to (0, \infty)$,

$$t(y) = \log\{\exp(y) + 1\}. \tag{2.2}$$

and its inverse $t^{-1} = \log\{\exp(x) - 1\}$. This transform is the 1-Lipschitz function widely used in neural networks. The important property of this transform is $\lim_{y \to \infty} t(y)/y = \lim_{x \to \infty} t^{-1}(x)/x = 1$, which implies that the transform negligibly affects large values in Figure 2.2. We extend t such that $t(-\infty) = 0$, $t^{-1}(0) = -\infty$, and $t(\infty) = t^{-1}(\infty) = \infty$ so that $t : \bar{\mathbb{R}}^p \to \bar{\mathbb{R}}_+^p$, where $\bar{\mathbb{R}}^p = [-\infty, \infty]^p$ and $\bar{\mathbb{R}}_+^p = [0, \infty]^p$. Note that the additive zero vector in $\bar{\mathbb{R}}_+^p$ is a vector of components $t(0) = \log 2$. So long as $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{i,p})^T \in RV_+^p(\alpha)$ meets the lower tail condition

$$sPr\{X_{i,j} \le \exp(-kb(s))\} \to 0, \quad k > 0, j = 1, \cdots, p, \tag{2.3}$$

15
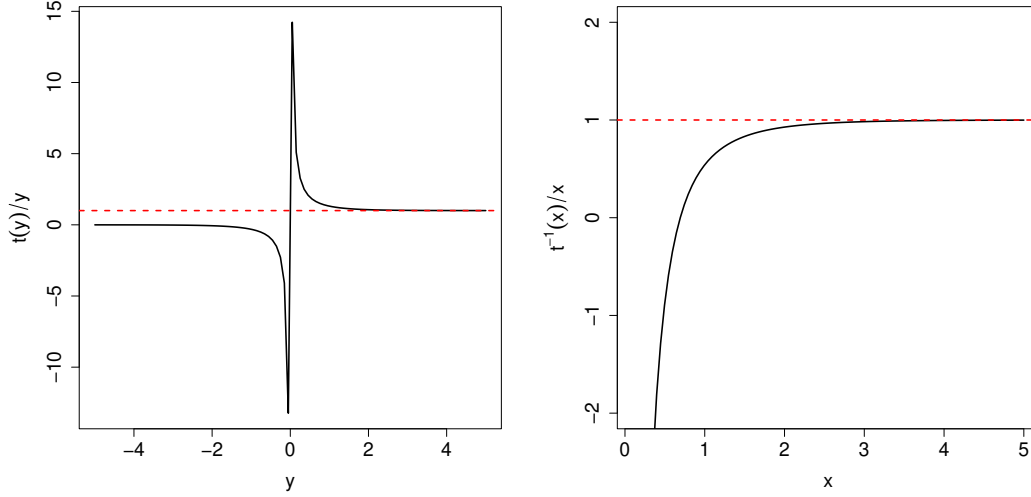
Figure 2.2: The limit ratio of transform t and its inverse t^{-1}.

as $s \to \infty$, transformed linear operations 'preserve' regular variation. This boundary condition ensures that none of marginals have enough non-zero mass at 0. Condition (2.3) is met by standard regularly varying distributions such as the Pareto and the Fréchet.

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

Proposition 2.2.1. *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 2.2.2. *Let* $sPr(b(s)^{-1}\boldsymbol{X} \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}}(\cdot)$, *then for* $a \in \mathbb{R}$,

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

In contrast to Max-linear approaches (e.g., Strokorb and Schlather [2015]), Cooley and Thibaud [2019] link regular variation to traditional linear algebra via a vector space arising from a transformation of $\mathbb{R}^p$. Propositions above provide a new way for describing extremal dependence and constructing extremes models in the upper tail.

16

Cooley and Thibaud [2019] go on to construct $\boldsymbol{X} \in RV_+^p(\alpha)$ 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 a matrix A to a vector of independent regularly varying random variables $\boldsymbol{Z}$. This construction method is analogous to that of constructing a Gaussian random vector $\boldsymbol{X} \sim N(\boldsymbol{0}, \Sigma)$, where Σ is a covariance matrix. By applying the matrix A such that $AA^T = \Sigma$ to a Gaussian random vector of $\boldsymbol{Z} \sim N(\boldsymbol{0}, I_{q \times q})$, we can construct $\boldsymbol{X} = A\boldsymbol{Z}$.

Corollary 2.2.1. *Let $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)$ with $\max_{i=1,\ldots p} a_{ij} > 0$ for all $j = 1, \ldots, q$, where $\boldsymbol{a}_j \in \mathbb{R}^p$ and hence $A \in \mathbb{R}^{p \times q}$. Let*

$$\boldsymbol{X} = A \circ \boldsymbol{Z} = t(At^{-1}(\boldsymbol{Z})), \tag{2.4}$$

where $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ is a vector of independent and identically distributed regularly varying random variables meeting $s Pr(b(s)^{-1} Z_j > z) \to z^{-\alpha}$ for $j = 1, \ldots, q$. Then, $\boldsymbol{X} \in RV_+^p(\alpha)$, and when normalized by $b(s)$, its angular measure is

$$H_{\boldsymbol{X}} = \sum_{j=1}^q \|\boldsymbol{a}_j^{(0)}\|^\alpha \delta_{\boldsymbol{a}_j^{(0)}/\|\boldsymbol{a}_j^{(0)}\|}(\cdot), \tag{2.5}$$

where δ is the Dirac mass function. The zero operation $a^{(0)} := \max(a, 0)$ will be important throughout, and is understood to be componentwise when applied to vectors or matrices. The angular measure $H_{\boldsymbol{X}}$ is consistent with point masses corresponding to the normalized columns of A.

Note that different matrices can yield the same limiting angular measure. Let $A^{(0)} = [a_{ij}^{(0)}]$. If $A \neq A'$ but $A^{(0)} = A'^{(0)}$, then $H_{A \circ Z} = H_{A^{(0)} \circ Z}$. When a regularly varying random vector $\boldsymbol{X}$ is produced by the construction method in Corollary 2.2.1, the class of regularly varying random vectors is analogous to the one of random vectors defined by max-linear combinations of independent regularly varying random variables (e.g., Schlather and Tawn [2002]; Fougères et al. [2013]). If we construct $A \times_{\max} Z = (\max_{j=1,\ldots,q} a_{1j} Z_j, \ldots, \max_{j=1,\ldots q} a_{pj} Z_j)^T$, then one can show that $H_{A \times_{\max} Z} = H_{A \circ Z}$. But, there is an important distinction between the two constructions. Large

17

realizations of the max-linear combination tend to have angular components that correspond exactly to the discrete locations on which the angular masses are placed whereas large realizations of the transformed-linear construction have angular components close but not exactly located at these discrete locations.

Proposition 2.2.3 below shows that as $q \to \infty$ the class of angular measures resulting from this construction method is dense in the class of possible angular measures. We only need to consider a nonnegative matrix A to construct the dense class.

Proposition 2.2.3. *Given any angular measure H, there exists a sequence of nonnegative matrices $\{A_q\}, q = 1, 2, ...$ such that $H_{A_q \circ Z_q} \xrightarrow{w} H$.*

2.2.2 Tail Pairwise Dependence Matrix

If p is even moderately large, it is challenging to describe the angular measure $H_{\boldsymbol{X}}$ for $\boldsymbol{X} \in RV_+^p(\alpha)$. Rather than fully characterize $H_{\boldsymbol{X}}$, we will summarize tail dependence via a matrix of pairwise summary measures. Many bivariate dependence measures have been suggested for extremes; we choose one which has properties similar to covariance.

Let $\alpha = 2$ and let $\boldsymbol{X} \in RV_+^p(2)$ have angular measure $H_{\boldsymbol{X}}$. Let $\Sigma_{\boldsymbol{X}} = \{\sigma_{\boldsymbol{X}_{ij}}\}_{i,j=1,\cdots,p}$ be the $p \times p$ matrix where

$$\sigma_{\boldsymbol{X}_{ij}} = \int_{\Theta_{p-1}^+} w_i w_j dH_{\boldsymbol{X}}(w), \tag{2.6}$$

and $\Theta_{p-1}^+ = \{\boldsymbol{w} \in \mathbb{R}_+^{p-1} : ||\boldsymbol{w}||_2 = 1\}$. Each element $\sigma_{\boldsymbol{X}_{ij}}$ is essentially the extremal dependence measure of Larsson and Resnick [2012]; however unlike Larsson and Resnick [2012], we require that $\alpha = 2$ and the L_2 norm which together make $\Sigma_{\boldsymbol{X}}$ have properties analogous to a covariance matrix. Specifically, $\Sigma_{\boldsymbol{X}}$ can be shown to be positive semi-definite [Cooley and Thibaud, 2019]. Following Cooley and Thibaud [2019], we call $\Sigma_{\boldsymbol{X}}$ the tail pairwise dependence matrix. This should not be confused with the 'tail dependence matrix' of Shyamalkumar and Tao [2020] which is a matrix of alternate extremal dependence measures χ_{ij} [Coles et al., 1999] and which is not guaranteed to be positive definite.

Larsson and Resnick [2012] also assume $H_{\boldsymbol{X}}$ is a probability measure, giving their extremal dependence measure a fixed range of values analogous to correlation. We do not require $H_{\boldsymbol{X}}$ to be a probability measure, and like a covariance matrix the diagonal elements $\sigma_{\boldsymbol{X}\,ii}$ reflect the relative magnitudes of the respective elements X_i. Regular variation implies $\lim_{s\to\infty} sPr(b(s)^{-1}X_i > c) = c^{-2}\sigma_{\boldsymbol{X}\,ii}$. Letting $x = cU(s)s^{1/2}$, there is a corresponding slowly varying function such that the relation can be rewritten as

$$\lim_{x\to\infty} \frac{Pr(X_i > x)}{x^{-2}L(x)} = \sigma_{\boldsymbol{X}\,ii}. \tag{2.7}$$

So the 'magnitude' of the elements of $\boldsymbol{X}$ described by the diagonal elements of the TPDM is in terms of suitably-normalized tail probabilities rather than variance. The presence of the slowly varying function $L(x)$ in the denominator means it is ambiguous to discuss the 'scale' of a regularly varying random variable, as scale information is in both the normalizing sequence and the angular measure (and consequently, TPDM). Because the notion of 'scale' is inherent in principal component analysis, Cooley and Thibaud [2019] further assumed that $\boldsymbol{X}$ was Pareto-tailed making $L(x)$ a constant that was pushed into the angular measure $H_{\boldsymbol{X}}$ and subsequently into the $\Sigma_{\boldsymbol{X}}$. Here, we will not require a Pareto tail, and the random variables we will construct in Section 2.3 will have a natural normalizing function. The sum of diagonal elements is the total mass of the angular measure since $\sum_{i=1}^{p} \sigma_{\boldsymbol{X}\,ii} = \int_{\Theta_{p-1}^+} dH_{\boldsymbol{X}}(\boldsymbol{w})$. $\sigma_{\boldsymbol{X}_{ij}} = 0$ implies asymptotic independence of the components X_i and X_j since $\sigma_{\boldsymbol{X}_{ij}} = 0$ if and only if $H_{\boldsymbol{X}}(\{\boldsymbol{w} \in \Theta_{p-1}^+ : w_i > 0, w_j > 0\}) = 0$.

An additional property of the TPDM that is not generally true for covariance matrices is that it is completely positive. That is, 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 not known, and A_* is not unique.

If $\boldsymbol{X} = A \circ \boldsymbol{Z}$ as in (2.4), the TPDM of the resulting vector is $\Sigma_{\boldsymbol{A}\circ\boldsymbol{Z}} = A^{(0)}A^{(0)T}$. More specifically, we assume $q \geq p$. Using the form of the angular measure in Corollary 2.2.1, we can denote the (i, k)th element of $\Sigma_{A\circ\boldsymbol{Z}}$ as

$$\sigma_{A\circ\boldsymbol{Z}_{ik}} = \int_{\Theta_{p-1}^+} w_i w_k dH_{A\circ\boldsymbol{Z}}(\boldsymbol{\theta}) = \sum_{j=1}^{q} \left(\frac{a_{ij}^{(0)}}{||\boldsymbol{a}_j^{(0)}||}\right)\left(\frac{a_{kj}^{(0)}}{||\boldsymbol{a}_j^{(0)}||}\right)||\boldsymbol{a}_j^{(0)}||_2^2 = \sum_{j=1}^{q} a_{ij}^{(0)} a_{kj}^{(0)}, \tag{2.8}$$

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 $\Sigma_{\boldsymbol{X}_*} = \Sigma_{\boldsymbol{X}}$. In Section 3, we will use this completely positive decomposition to create prediction intervals.

2.3 Inner Product Space and Prediction

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

We consider a space of regularly varying random variables constructed from a transformed-linear combinations. Let $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ be a vector of independent $Z_j \in RV_+^1(2)$ meeting lower tail condition $sPr(b(s)^{-1}Z_j > z) \to z^{-\alpha}$ and which have a common normalizing function $\lim_{z \to \infty} \frac{P(Z_j > z)}{z^{-2}L(z)} = 1$ for $j = 1, \ldots, q$. For $\boldsymbol{a} \in \mathbb{R}^q$, consider the space

$$\mathcal{V}^q = \left\{ X; X = \boldsymbol{a}^T \circ \boldsymbol{Z} = a_1 \circ Z_1 \oplus \cdots \oplus a_q \circ Z_q \right\}. \tag{2.9}$$

$\mathcal{V}^q \subset RV_+^1(2)$. If $X_1 = \boldsymbol{a}_1^T \circ \boldsymbol{Z}$ and $X_2 = \boldsymbol{a}_2^T \circ \boldsymbol{Z}$, then $X_1 \oplus X_2 = (\boldsymbol{a}_1 + \boldsymbol{a}_2)^T \circ \boldsymbol{Z}$. Also, $c \circ X_1 = c\boldsymbol{a}_1 \circ \boldsymbol{Z}$ for $c \in \mathbb{R}$. Any $X \in \mathcal{V}^q$ is uniquely identifiable by its vector of coefficients $\boldsymbol{a}$, thus $\mathcal{V}^q$ is isomorphic to $\mathbb{R}^q$. It is straightforward to see that $\mathcal{V}^q$ is a vector space; for completeness, the vector space conditions are shown in Appendix A. $\mathcal{V}^q$ differs from the vector space in Cooley and Thibaud [2019] which was a non-stochastic vector space for $\mathbb{R}_+^p$.

The inner product of $X_1 = \boldsymbol{a}_1^T \circ \boldsymbol{Z}$ and $X_2 = \boldsymbol{a}_1^T \circ \boldsymbol{Z}$ 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 say $X_1, X_2 \in \mathcal{V}^q$ are orthogonal if $\langle X_1, X_2 \rangle = 0$. The norm $\|X\|_{\mathcal{V}^q} = \sqrt{\langle X, X \rangle}$ defines the metric $d(X_1, X_2) = \|X_1 \ominus X_2\|_{\mathcal{V}^q}$. We use the subscript $\mathcal{V}^q$ to indicate that the norm is based on the coefficients which determine the random variable. We will further describe the meaning of this metric in Section 2.4.

We will consider vectors $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ where $X_i = \boldsymbol{a}_i \circ \boldsymbol{Z} \in \mathcal{V}^q$ for $i = 1, \ldots, p$. $\boldsymbol{X} \in RV_+^p(2)$ and is of the form $A \circ \boldsymbol{Z}$ in (2.4). We denote the matrix of inner products

$$\Gamma_{\boldsymbol{X}} = \langle X_i, X_j \rangle_{i,j=1,\ldots p} = AA^T. \tag{2.10}$$

2.3.2 Transformed-linear Prediction

As $\mathcal{V}^q$ is isomorphic to Hilbert space $\mathbb{R}^q$, the best transformed-linear predictor follows similarly. Assume $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z} \in \mathcal{V}^q$ for $i = 1, \ldots, p+1$. Let $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$. We aim to find $\boldsymbol{b} \in \mathbb{R}^p$ such that $d(\boldsymbol{b}^T \circ \boldsymbol{X}_p, X_{p+1})$ is minimized. Writing in matrix form

$$\begin{bmatrix} X_{p+1} \\ \boldsymbol{X}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{p+1}^T \\ A_p \end{bmatrix} \circ \boldsymbol{Z}_q,$$

where $A_p = (\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_p^T)^T$. The matrix of inner products of $(X_{p+1}, \boldsymbol{X}_p^T)^T$ is

$$\Gamma_{(X_{p+1}, \boldsymbol{X}_p^T)^T} = \begin{bmatrix} \boldsymbol{a}_{p+1}^T \boldsymbol{a}_{p+1} & \boldsymbol{a}_{p+1}^T A_p^T \\ A_p \boldsymbol{a}_{p+1} & A_p A_p^T \end{bmatrix} := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}. \tag{2.11}$$

Minimizing $d(\boldsymbol{b}^T \circ \boldsymbol{X}, X_{p+1})$ is equivalent to minimizing $\|A_p^T \boldsymbol{b} - \boldsymbol{a}_{p+1}\|_2^2$. Taking derivatives with respect to $\boldsymbol{b}$ and setting equal to zero, the minimizer $\hat{\boldsymbol{b}}$ solves $(A_p A_p^T)\hat{\boldsymbol{b}} = A_p \boldsymbol{a}_{p+1}$. If $A_p A_p^T$ is invertible, then the solution $\hat{\boldsymbol{b}}$ is,

$$\hat{\boldsymbol{b}} = (A_p A_p^T)^{-1} A_p \boldsymbol{a}_{p+1} = \Gamma_{22}^{-1} \Gamma_{21}. \tag{2.12}$$

An equivalent way to think of the best transformed-linear prediction is through the projection theorem. $\hat{X}_{p+1}$ is such that $X_{p+1} \ominus \hat{X}_{p+1}$ is orthogonal to the plane spanned by $X_1, \cdots, X_p$. The orthogonality condition can be stated as $< X_{p+1} \ominus \hat{X}_{p+1}, X_i >= 0$, for $i = 1, \cdots, p$. This

condition can equivalently be expressed with the matrix notation.

$$\left[< X_{p+1}, X_i >\right]_{i=1}^{p} = \left[< X_i, X_j >\right]_{i,j=1}^{p} \left[b_i\right]_{i=1}^{p} = \left[\sum_{k=1}^{q} a_{ik} a_{jk}\right]_{i,j=1}^{p} \left[b_i\right]_{i=1}^{p} \qquad (2.13)$$

Using the matrix notation from above, we require $\hat{\boldsymbol{b}}$ to satisfy $A_p \boldsymbol{a}_{p+1} = A_p A_p^T \hat{\boldsymbol{b}}$ as above.

2.4   Subset $\mathcal{V}_+^q$

We have employed transformed linear operations to construct regularly-varying random vectors $\boldsymbol{X} = A \circ \boldsymbol{Z}$ that take values in the positive orthant, and we have tied these vectors' elements to the vector space $\mathcal{V}^q$. It is essential that the elements of the coefficient vectors $\boldsymbol{a}$ are allowed to be negative for $\mathcal{V}^q$ to be a vector space. However, negative values in $\boldsymbol{a}$ do not influence tail behavior. Recalling that if regularly varying Z_1, Z_2 are independent, $P(Z_1 + Z_2 > z) \sim P(Z_1 > z) + P(Z_2 > z)$ as $z \to \infty$ [cf. Jessen and Mikosch, 2006, Lemma 3.1], we can discuss the magnitude of $X \in \mathcal{V}^q$ (as in (2.7)) in terms of the common tail behavior of the generating Z_j's. We call

$$TR(X) := \lim_{z \to \infty} \frac{P(X > z)}{P(Z_1 > z)} = \sum_{j=1}^{q} \left(a_j^{(0)}\right)^2 \qquad (2.14)$$

the tail ratio of X and only the positive elements of $\boldsymbol{a}$ contribute. $X = \boldsymbol{a} \circ \boldsymbol{Z} \in \mathcal{V}^q$ and $X_+ = \boldsymbol{a}^{(0)} \circ \boldsymbol{Z}$ have the same tail ratio. Furthermore, if $\boldsymbol{X} = A \circ \boldsymbol{Z}$, both it and $\boldsymbol{X}_+ = A^{(0)} \circ \boldsymbol{Z}$ have the same angular measure: $H_{\boldsymbol{X}} = H_{\boldsymbol{X}_+} = \sum_{j=1}^{q} \|a_j^{(0)}\|^\alpha \delta_{a_j^{(0)}/\|a_j^{(0)}\|}(\cdot)$. $\boldsymbol{X}$ and $\boldsymbol{X}_+$ are indistinguishable in terms of their tail behavior.

In terms of modeling, it seems reasonable to restrict our attention to the subset

$$\mathcal{V}_+^q = \left\{X; X = \boldsymbol{a}^T \circ \boldsymbol{Z} = a_1 \circ Z_1 \oplus \cdots \oplus a_q \circ Z_q\right\}, \qquad (2.15)$$

where $a_j \in [0, \infty)$, and $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ as in (2.9). Considering inference for a random vector $\boldsymbol{X} \in RV_+^p$, we assume that $\boldsymbol{X} = A \circ \boldsymbol{Z}$ for some unknown $p \times q$ matrix A because it is a simple and useful modeling framework. Recall such constructions are dense in RV_+^p. Inference for $\boldsymbol{X}$

will focus on its tail behavior, and since this is indistinguishable from that of $\boldsymbol{X}_+$, it is reasonable to assume $a_{ij} \geq 0$ for $i = 1, \ldots, p$, and $j = 1, \ldots, q$, and thus $X_i \in \mathcal{V}_+^q$ for $i = 1, \ldots, p$.

Continuing with inference, if p is even of moderate size, then estimating $H_{\boldsymbol{X}}$ is challenging, so we focus on summarizing dependence via the TPDM. If $\boldsymbol{X} = A \circ \boldsymbol{Z}$ and all $a_{ij} > 0$, then $\Sigma_{\boldsymbol{X}} = \Gamma_{\boldsymbol{X}} = AA^T$. Furthermore, if inference focuses on the TPDM, then q, the number of independent Z_j's from which $\boldsymbol{X}$ is generated, does not need to be specified.

Turning our attention toward prediction, it seems reasonable to assume that the elements of $(X_{p+1}, \boldsymbol{X}_p^T)^T$ are in $\mathcal{V}_+^q$, and prediction can be done in terms of the TPDM. Considering predictors of the form $\boldsymbol{b}^T \circ \boldsymbol{X}_p$ and letting $\Sigma_{(X_{p+1}, \boldsymbol{X}_p^T)^T}$ be partitioned as in (2.11), $\hat{X}_{p+1} = \hat{\boldsymbol{b}}^T \circ \boldsymbol{X}_p$ where $\hat{\boldsymbol{b}} = \Sigma_{22}^{-1} \Sigma_{21}$ will minimize $\|X_{p+1} \ominus \hat{X}_{p+1}\|_{\mathcal{V}^q}$. Because $\hat{\boldsymbol{b}}$ is not required to consist of nonnegative elements, the predictor $\hat{X}_{p+1}$ is not necessarily in $\mathcal{V}_+^q$.

We can now better discuss the meaning of the metric $d(X_1, X_2) = \|X_1 \ominus X_2\|_{\mathcal{V}^q}$. Note

$$TR(X_1 \ominus X_2) = \sum_{j=1}^{q} \left((a_{1j} - a_{2j})^{(0)} \right)^2 \neq TR(X_2 \ominus X_1) = \sum_{j=1}^{q} \left((a_{2j} - a_{1j})^{(0)} \right)^2.$$

However, because $P(\max(Z_1, Z_2) > z) \sim P(Z_1 > z) + P(Z_2 > z)$ as $z \to \infty$,

$$TR\left(\max((X_1 \ominus X_2), (X_2 \ominus X_1)) \right) = \sum_{j=1}^{q} (a_{1j} - a_{2j})^2 = d^2(X_1, X_2). \tag{2.16}$$

From the probabilistic prospective, the metric $d(X_1, X_2) = \|X_1 \ominus X_2\|_{\mathcal{V}^q}$ based on tail probabilities penalizes the worst possible tail prediction error on the positive orthant. See more details about the metric in Appendix B.

Since random vectors with elements in $\mathcal{V}^q$ are indistinguishable from those with elements in $\mathcal{V}_+^q$ we will assume $X_1, \ldots, X_p, X_{p+1} \in \mathcal{V}_+^q$, and we will use the TPDM for prediction.

Chapter 3

Prediction Error

3.1 Analogue to Mean Square Prediction Error

In the non-extreme setting, the error associated with linear prediction is quantified by the mean square prediction error (MSPE). As MSPE corresponds to the conditional variance under a Gaussian assumption, it is used to generate Gaussian-based prediction intervals. Our transformed linear prediction has an analogous quantity

$$
\begin{aligned}
||X_{p+1} \ominus \hat{X}_{p+1}||^2_{\mathcal{V}^q} :\; &=< X_{p+1} \ominus \hat{X}_{p+1}, X_{p+1} \ominus \hat{X}_{p+1} > \\
&= (\boldsymbol{a}^T_{p+1} - \hat{\boldsymbol{b}}^T A_p)(\boldsymbol{a}^T_{p+1} - \hat{\boldsymbol{b}}^T A_p)^T \qquad (3.1) \\
&= \Sigma_{11} - \Sigma_{12}\Sigma^{-1}_{22}\Sigma_{21} := K.
\end{aligned}
$$

Importantly, K can be calculated directly from the (estimated) TPDM. Unlike MSPE, K is not understood via expectation, but instead via tail probabilities as

$$
TR\left(\max((X_{p+1} \ominus \hat{X}_{p+1}), (\hat{X}_{p+1} \ominus X_{p+1}))\right) = K.
$$

The quantity K is meaningful to minimize and has a probabilistic interpretation, but seems not very useful for constructing prediction intervals. To illustrate, we simulate $n = 20,000$ four dimensional vectors $\boldsymbol{X}$ and obtain $\hat{X}_4$ predicted on $(X_1, X_2, X_3)^T$. $\boldsymbol{X}$ is generated from a 4×10 matrix A applied to a vector $\boldsymbol{Z}$ comprised of 10 independent $RV_+(2)$ random variables; the elements of A are drawn from a uniform$(0, 5)$ distribution. Using the known TPDM to obtain $K = 0.224$ and known tail behavior of the Z_j's, we calculate $P(D \leq 2.99) \approx 0.95$ where $D = \max((X_{p+1} \ominus \hat{X}_{p+1}), (\hat{X}_{p+1} \ominus X_{p+1})) \in RV^1_+(2)$. We observe 0.952 of the simulated D values are in fact below this bound. However, Figure 3.1 shows that knowledge of K is not useful for constructing prediction intervals. Unlike the Gaussian case where the variance of the conditional

24

distribution does not depend on the predicted value $\hat{X}_{p+1}$, in the polar geometry of regular varia-tion, the magnitude of the error is related to the size of the predicted value. In the next sections we use the polar geometry of regular variation to construct meaningful prediction intervals when $\hat{X}_{p+1}$ is large.
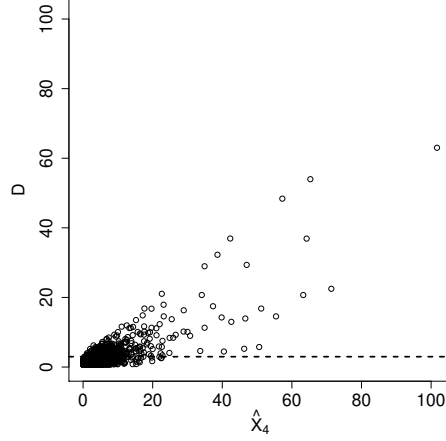
Figure 3.1: Left panel: The plot of $D = \max(\hat{X}_4 \ominus X_4, X_4 \ominus \hat{X}_4)$ against $\hat{X}_4$. The horizontal dashed line indicates the approximate 0.95 quantile for D.

3.2   Prediction TPDM and Completely Positive Decomposition

The vector $(X_{p+1}, \hat{X}_{p+1})^T \in RV_+^2(2)$, and this vector's tail dependence is characterized by $H_{(X_{p+1}, \hat{X}_{p+1})^T}$. While this angular measure is not readily available, its dependence is summarized by the 2×2 'prediction' TPDM

$$
\Sigma_{(X_{p+1}, \hat{X}_{p+1})^T} = \begin{bmatrix} \boldsymbol{a}_{p+1}^{(0)^T} \\ (\hat{\boldsymbol{b}}^T \boldsymbol{A}_p)^{(0)} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{p+1}^{(0)} & (\boldsymbol{A}_p^T \hat{\boldsymbol{b}})^{(0)} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{bmatrix}. \tag{3.2}
$$

We propose a method of using the information in this TPDM to construct a potential angular mea-sure, and using this to quantify the dependence between $\hat{X}_{p+1}$ and X_{p+1}. Because $\Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$ is completely positive, there exists a $q_* < \infty$ and nonnegative $2 \times q_*$ matrix B such that $\Sigma_{(X_{p+1}, \hat{X}_{p+1})^T} = BB^T$. The matrix B is not unique. In general, finding a completely positive decomposition for a

25

p-dimensional matrix can be challenging and q_* can be quite large. In this case since $p = 2$, it is relatively simple. Given any $q_* \geq 2$ and 2×2 completely positive matrix $\Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$, there exist nonegative matrices $B \in \mathbb{R}^{2 \times q_*}$ such that $BB^T = \Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$, and there exist procedures [Groetzner and Dür, 2020] to find examples of these matrices. See section 3.2.1 below for more details. Since our goal is to obtain a potential angular measure $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$ with TPDM $\Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$, there would seem to be incentive to set q_* large, thereby distributing the total mass of the angular measure $H_{B \circ \mathbf{Z}}$ into q_* point masses. On the other hand, as q_* grows, the procedures for obtaining B require more computation. We take a practical approach. We choose q_* to be of moderate size, but apply the procedure repeatedly, obtaining nonegative $B^{(k)}, k = 1, \ldots, n_{decomp}$ such that $B^{(k)} B^{(k)^T} = \Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$ for all k. We then set $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T} = n_{decomp}^{-1} \sum_{k=1}^{n_{decomp}} H_{A^{(k)} \circ \mathbf{Z}}$, and $\Sigma_{\hat{H}} = n_{decomp}^{-1} \sum_{k=1}^{n_{decomp}} B^{(k)} B^{(k)^T} = \Sigma_{(X_{p+1}, \hat{X}_{p+1})^T}$ as required. $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$ consists of $n_{decomp} q_*$ point masses.

We use a simulation study to illustrate. We again begin by generating a matrix A whose elements are drawn from a uniform(0,5) distribution; however this time the dimension of A is 7×400 thus the true angular measure consists of 400 point masses. We draw $60,000$ random realizations of $\mathbf{X} = A \circ \mathbf{Z}$, and use the first $40,000$ as a training set. The largest 1% of this training set is used to estimate the seven-dimensional TPDM, from which we obtain $\hat{\mathbf{b}}$ and additionally the estimated prediction TPDM $\hat{\Sigma}_{(X_{p+1}, \hat{X}_{p+1})^T}$. We then use the completely positive decomposition to obtain 2×9 matrices $B^{(k)}, k = 1, \ldots, 51$, resulting in an estimated angular measure $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$ consisting of 459 point masses in Figure 3.2. We obtain a 95% joint region by drawing bounds at the 0.025 and 0.975 quantiles of $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$. The right panel of Figure 3.2 shows the scatterplot of the 20000 remaining test points $\hat{X}_{p+1}$ and X_{p+1} and the 95% joint region. Thresholding at the 0.95 quantile of $\|(X_{p+1}, \hat{X}_{p+1})\|_{\mathcal{V}^q}$, we find that 0.963 of the large values fall within the joint region. The scatterplot shows asymmetry in the distribution of $\hat{X}_{p+1}$ and X_{p+1}. It is not a surprise that the joint region does not reflect the asymmetry seen in the scatterplot of $\hat{X}_{p+1}$ and X_{p+1}, as any information about asymmetry is not contained in $\hat{\Sigma}_{(X_{p+1}, \hat{X}_{p+1})^T}$.

26
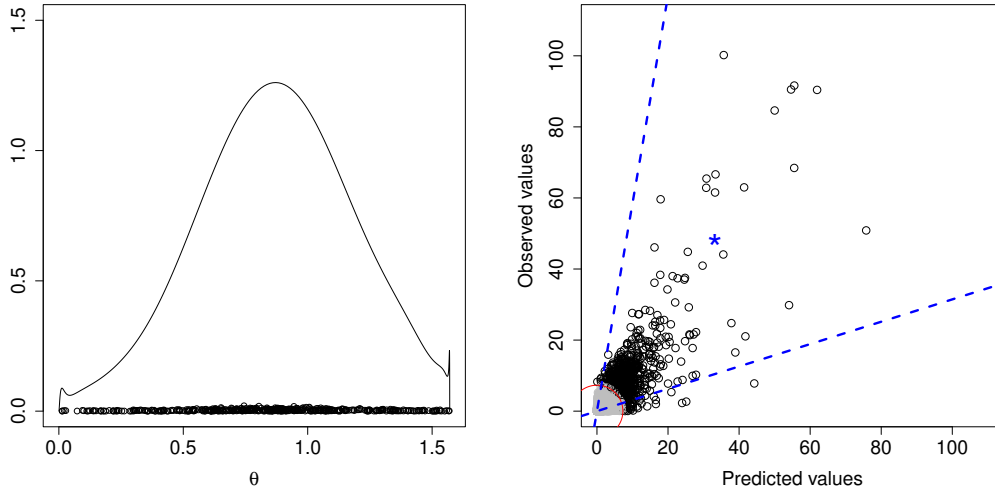
Figure 3.2: (Left) The weighted kernel density estimate in terms of angles. (Right) The estimated joint 95% joint prediction region based on the approximated angular measure $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$. The star indicates a particular observation which has a predicted value of 33.17 and an observed value of 48.15.

3.2.1 Completely Positive Decomposition

Groetzner and Dür [2020] devise a completely positive factorization problem as a nonconvex feasibility problem. They propose an alternative projection method by avoiding a second order cone problem (SOCP) to reduce the computational cost. We briefly introduce their method below without considering bounds on cp-ranks of a matrix.

While the usual goal of a non-negative factorization of a $p \times p$ matrix Σ is to find a non-negative $p \times q$ matrix B with the least dimension q such that $\Sigma = BB^T$, we need the matrix B with a large q to better approximate the angular measure. We start with an arbitrary initial factorization such as the Eigenvalue decomposition or the Cholesky decomposition. The obtained B such that $\Sigma = BB^T$ is not necessarily non-negative definite. The initial factorization normally provides a square matrix B. We augment the matrix B by adding additional m columns. In the Cholesky decomposition, the first positive column can be decomposed into $[\frac{1}{\sqrt{m}}\boldsymbol{b}_1, \frac{1}{\sqrt{m}}\boldsymbol{b}_1, ..., \frac{1}{\sqrt{m}}\boldsymbol{b}_1]$, where $\boldsymbol{b}_1$ is the first column vector of B. The augmented matrix $\hat{B}$ becomes a $p \times (q + m)$ matrix $\hat{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_{q-1}, \frac{1}{\sqrt{m}}\boldsymbol{b}_1, ..., \frac{1}{\sqrt{m}}\boldsymbol{b}_1]$.

The key idea of completely positive decomposition is to find an orthogonal matrix Q that rotates the factor B until it satisfies $BQ \geq 0$. Groetzner and Dür [2020] propose the alternative way of

27

projecting the matrix Q onto the intersection of $\mathcal{P}$ and $\mathcal{Q}_q$, where $\mathcal{P} = \{Q \in \mathbb{R}^{q \times q} | BQ \geq 0\}$ and $\mathcal{Q}_q$ is the set of $q \times q$ orthogonal matrices. By theorem 4.2 in Groetzner and Dür [2020], given an initial matrix $Q_0 \in \mathcal{Q}_q$, we keep projecting the kth matrix Q_k onto the space $\mathcal{Q}_q$ and $\mathcal{P}$ alternatively until the convergence of the matrix $Q_k \rightarrow Q^* \in \mathcal{P} \cap \mathcal{Q}_q$ as $k \rightarrow \infty$ is achieved and it meets $\Sigma = (BQ^*)(BQ^*)^T$. The initial matrix $Q_0 \in \mathcal{Q}_q$ is obtained by projecting an arbitrary matrix $M \in \mathbb{R}^{q \times q}$ onto $\mathcal{Q}_q$. For example, the arbitrary matrix M can consist of random draws from the standard normal distribution. The projection of M onto $\mathcal{Q}_q$ is obtained by the polar decomposition meaning that there exists a positive semi-definite matrix T and an orthogonal matrix Q such that $M = TQ$. See full details in Groetzner and Dür [2020]. The projection of Q onto the space $\mathcal{P}$ is originally tied to solving a second order cone problem (SOCP). But, Groetzner and Dür [2020] project the matrix BQ onto the nonnegative orthant space $D = max(BQ, 0)$ rather than project Q onto $\mathcal{P}$ so that it reduces the computational cost of solving SOCP. If $D = BQ$, then $BQ \geq 0$. Hence, $Q \in \mathcal{P}$. If $D \neq BQ$, then they use the matrix $\hat{P} := B^+D + (I - B^+B)Q$ where B^+ is the Moore-Penrose-Inverse of B to approximate the projection of Q onto $\mathcal{P}$.

3.3 Prediction Intervals for X_{p+1} Given Large $\hat{X}_{p+1}$

The region obtained in the previous section describes the joint behavior of $\hat{X}_{p+1}$ and X_{p+1}, but the quantity of interest is the conditional behavior of X_{p+1} given a specific large value of $\hat{X}_{p+1}$. Cooley et al. [2012] use the limiting intensity function of regular variation to get an approximate density of X_{p+1} given large $\boldsymbol{X}_p$. More specifically, they derive an approximate density of X_{p+1} given large $\boldsymbol{X}_p$ using the point process argument under L_1-norm. The sequence of normalized point processes, $N_n = \{\boldsymbol{X}_i/b_n\}$, converges to the limiting point process N with the intensity measure,

$$\Lambda(d\boldsymbol{x}) = ||\boldsymbol{x}||_1^{-(p+1)} h(\boldsymbol{x}/||\boldsymbol{x}||_1) d\boldsymbol{x}. \tag{3.3}$$

This intensity function is obtained by transforming L_1 polar coordinates to Cartesian coordinates with the Jacobian of the transformation. A joint density of $\boldsymbol{X}/b_n$ given $||\boldsymbol{X}||/b_n > r^*$ is approximately the product of some high threshold and the intensity function in terms of Cartesian

coordinates,

$$f_{\boldsymbol{X}/b_n}(\boldsymbol{x}, r^*) \to r^* ||\boldsymbol{x}||_1^{-(p+1)} h(\boldsymbol{x}/||\boldsymbol{x}||_1), \quad \text{for } ||\boldsymbol{X}||_{-p} > r^*, \tag{3.4}$$

where r^* is some high threshold. After integrating out the variables, $\boldsymbol{X}_{-p} = (X_1, \cdots, X_{p-1})^T$, we can derive the approximate conditional density by dividing by the normalizing constant. They fit a parametric model for $H_{(\boldsymbol{X}_p, X_{p+1})^T}$ and transform from polar form to obtain $\nu_{(\boldsymbol{X}_p, X_{p+1})^T}(d\boldsymbol{x})$. Cooley et al. [2012] applied their method in moderate dimension ($p = 4$). Applying their approach in higher dimensions would require fitting a high dimensional angular measure model. We adapt the method of Cooley et al. [2012] to model the relationship between X_{p+1} and $\hat{X}_{p+1}$. Regardless of p, we only need to describe this bivariate relationship.

Changing from polar coordinates to Cartesian, a bivariate regularly varying random vector (X_1, X_2) with $\alpha = 2$ and angular density $h_{(X_1, X_2)}$ defined on Θ_1^+ has limiting measure $\nu(dx_1, dx_2) = 2||\boldsymbol{x}||_2^{-5} x_2 h(\boldsymbol{x}||\boldsymbol{x}||_2^{-1})$. Following Cooley et al. [2012], the conditional density of $X_2|X_1 = x_1$ if x_1 is large is approximately

$$f_{X_2|X_1}(x_2|x_1) = 2c^{-1} ||(x_1, x_2)||_2^{-5} x_2 h\left(\frac{(x_1, x_2)}{||(x_1, x_2)||_2^{-1}}\right), \tag{3.5}$$

where $c = \int_0^\infty 2||(x_1, x_2)||_2^{-5} x_2 h\left(\frac{(x_1, x_2)}{||(x_1, x_2)||_2^{-1}}\right) dx_2$.

We use (3.5) to obtain an estimate of the conditional density of X_{p+1} given large $\hat{X}_{p+1}$. Since (3.5) requires an angular density, we use a kernel density estimate of $\hat{H}_{(X_{p+1}, \hat{X}_{p+1})^T}$. We use the adjusted boundary bias approach of Marron and Ruppert [1994] for the kernel density estimation since the support of $H_{(X_{p+1}, \hat{X}_{p+1})^T}$ is bounded. See more details in Appendix D. We then take the 0.025 and 0.975 quantiles of this estimated conditional density to obtain a 95% prediction interval in Figure 3.3. The left panel of Figure 3.3 illustrates the 95% prediction interval for a particular realization from the aforementioned simulation study where $\hat{X}_{p+1} = 33.17$ and with actual value $X_{p+1} = 48.15$ denoted by the blue star. Because the conditional density changes with $\hat{X}_{p+1}$, the units of the horizontal axis are the predicted values and the units of the conditional density are omitted. The right panel shows a scatterplot of the largest 5% (by $\hat{X}_{p+1}$) of the test set from the

aforementioned simulation along with the upper and lower bounds from the conditional density approximation. The coverage rate of these intervals is 0.947.
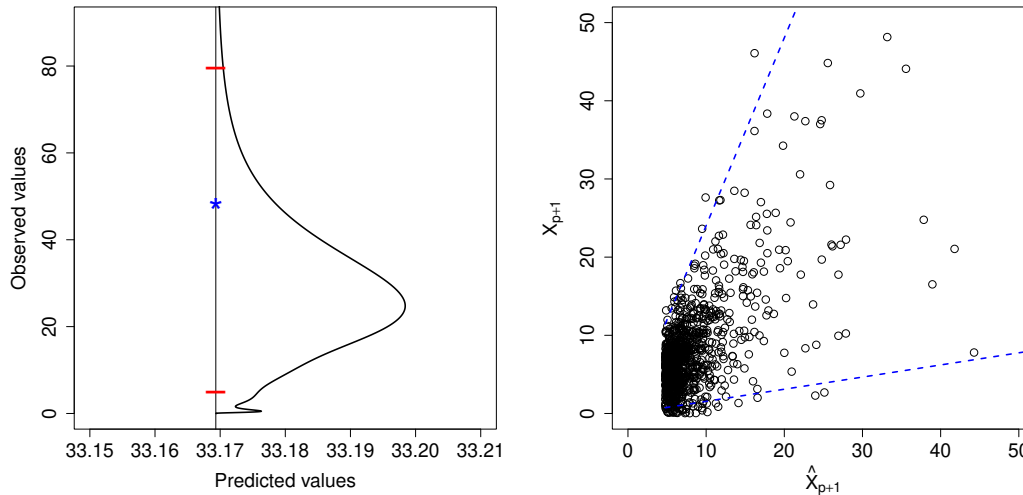
Figure 3.3: (Left) The approximate conditional density $f_{X_{p+1}|\hat{X}_{p+1}}(X_{p+1}|\hat{x}_{p+1} = 33.17)$ for a particular realization. The horizontal segments indicate the 95% conditional prediction interval, and the star denotes the actual value of 48.15. Because the conditional density changes with $\hat{X}_{p+1}$, the units of the horizontal axis are the predicted values and the units of the conditional density are omitted. (Right) the scatter plot of $\hat{X}_{p+1}$ and X_{p+1} with 95% conditional prediction intervals given each large value of $\hat{X}_{p+1}$.

3.4 Applications

3.4.1 Nitrogen Dioxide Air Pollution.

NO_2 is one of six air pollutants for which the US Environmental Protection Agency (EPA) has national air quality standards. We analyze daily EPA NO_2 data[1] from five locations in the Washington DC metropolitan area (see Figure 2.1). The first four stations (McMillan 11-001-0043, River Terrace 11-001-0041, Takoma 11-001-0025, Arlington 51-013-0020) have long data records spanning 1995-2020. Alexandria does not have observations after 2016. We will perform prediction at Alexandria given data at the other four locations. Observations in Alexandria actually come from two different stations: 51-510-0009 which has measurements from January1995 to

---

[1]https://www.epa.gov/outdoor-air-quality-data/download-daily-data

August 2012 and 51-510-00210 from August 2012 to April 2016. Exploratory analysis did not indicate any detectable change point in the Alexandria data either with respect to the marginal distribution or with dependence with other stations, so we treat this data as coming from a single station. There are 5163 days between 1995 and 2016 where all five locations have measurements. Because NO_2 levels have decreased over the study period, we detrend at each location using a moving average mean and standard deviation with window of 901 days to center and scale.

Our inner product space assumes each $X_i \in RV_+^1(\alpha = 2)$, and the detrended NO_2 data must be transformed to meet this assumption. In fact, it is unclear whether the NO_2 data are even heavy tailed. Nevertheless, we believe the regular variation framework is useful for describing the tail dependence for this data after marginal transformation. Characterizing dependence after marginal transformation is justified by Sklar's theorem (Sklar [1959], see also [Resnick, 1987, Proposition 5.15]), and such transformations are regularly used in multivariate extremes studies. After viewing standard diagnostic plots, we fit a generalized Pareto distribution above each location's 0.95 quantile and obtain the marginal estimated cdf's $\hat{F}_i$ which are the empirical cdf below the 0.95 quantile and the fitted generalized Pareto above. Letting $X_i^{(orig)}$ denote the random variable for detrended NO_2 at location i, we define $X_i = 1/\sqrt{(1 - \hat{F}_i(X_i^{(orig)}))} - \delta$ obtaining a 'shifted' Pareto distribution for $i = 1, \ldots, 5$. Each $X_i \in RV_+(\alpha = 2)$ and the shift $\delta = 0.9352$ is such that $E[t^{-1}(X_i)]$ = 0. This shift moves the transformed data closer to the axes which we found reduced bias in the estimation of the TPDM. We assume $\boldsymbol{X} = (X_1, \ldots, X_5)^T \in RV_+^5(\alpha = 2)$. Further, we let $\boldsymbol{X}_t$ denote the random vector of observations on day t, which we assume to be iid copies of $\boldsymbol{X}$. This is a simplifying assumption as there is temporal dependence in the NO_2 data, but it seems less informative that the spatial dependence exhibited by each day's observations.

We first predict during the period prior to 2015 in order that we can use the observed data at Alexandria to assess performance. Indices are randomly drawn to divide the data set into training and test sets consisting of 3442 and 1721 observations respectively, and both sets cover the entire observational period. Using the training set, the five-dimensional TPDM $\hat{\Sigma}_{\boldsymbol{X}}$ is estimated as follows. Let $\boldsymbol{x}_t$ denote the observed measurements on day t. For each $i \neq j$ in $1, \ldots, 5$, 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 $\hat{\sigma}_{ij} = 2n_{exc}^{-1} \sum_{t=1}^{n} w_{t,i} w_{t,j} \mathbb{I}(r_{t,ij} > r_{ij}^*)$, where $n_{exc} = \sum_{t=1}^{n} \mathbb{I}(r_{t,ij} > r_{ij}^*)$. We choose r_{ij}^* to correspond to the 0.95 quantile. The constant 2 arises from knowledge that the tail ratio of each X_i is one due to the marginal transformation. This pairwise estimation of the TPDM differs from the method in Cooley and Thibaud [2019] who used the entire vector norm as the radial component. Mhatre and Cooley [2021] show that the TPDM is equivalent whether it is defined in terms of the angular measure of the entire vector or the angular measure corresponding to the two-dimensional marginals.

From $\hat{\Sigma}_{\boldsymbol{X}}$, we obtain $\hat{X}_{t,5} = \hat{\boldsymbol{b}}^T \circ \boldsymbol{X}_{t,4}$, where $\hat{\boldsymbol{b}} = (-0.047, 0.177, 0.192, 0.482)^T$. We note that the largest weighted component is Arlington, which is closest to Alexandria. Interestingly, McMillan has a slightly negative weight. We calculate $\hat{X}_{t,5}$ for all t, but only consider those for which $\hat{X}_{t,5}$ exceeds the 0.95 quantile. The left panel of Figure 3.4 shows the scatterplot of the values $x_{t,5}$ versus $\hat{x}_{t,5}$. By taking the inverse of the marginal transformation, multiplying by the moving average standard deviation and adding the moving average mean, the predicted value can be put on the scale of the original data. The center panel of Figure 3.4 shows the scatterplot on the original scale.

For each large predicted value $\hat{x}_{t,5}$, we use the method described in Section 3.3 to create 95% prediction intervals. We chose the matrix B arising from the completely positive decomposition to again be 2×9. On the Pareto scale, these prediction intervals are linear with $\hat{x}_{t,5}$ and are shown in the left panel of Figure 3.4. The coverage rate of these intervals is 0.965. The intervals can similarly be back-transformed to be on the original scale as shown in the center panel of Figure 3.4. The lack of monotonicity in these intervals with respect to the predicted value is due to the trend in the data over the observation period.

For comparison to standard method, we find the BLUP based on the estimated covariance matrix from the entire data set, and create Gaussian-based 95% confidence intervals from the estimated MSPE. When done on the original data, we obtain a coverage rate of 0.88, and when done on square-root transformed data to account for the skewness, we obtain a coverage rate of 0.78.

We then apply our prediction method to five dates in 2019 and 2020 when observed values at the four recording stations were large and no observation was taken at Alexandria. Here, we use the entire period from 1995-2016 to estimate the TPDM, and we obtain a slightly different estimate $\hat{\boldsymbol{b}} = (0.026, 0.153, 0.118, 0.461)^T$. The right panel of Figure 3.4 shows the point estimate and 95% prediction intervals from our transformed-linear approach in solid line. On the Pareto scale (not shown), the positive skewness of the conditional distribution (as shown in the left panel of Figure 3.3) is preserved and the point estimate is lower than the center of the prediction interval. Transformation back to the original NO_2 scale results in the prediction intervals shown in Figure 3.4; the trend at Arlington was used for the unobserved trend at Alexandria. For comparison, covariance matrix-based BLUP's and MSPE-based 95% prediction intervals for these dates are shown in dashed line.
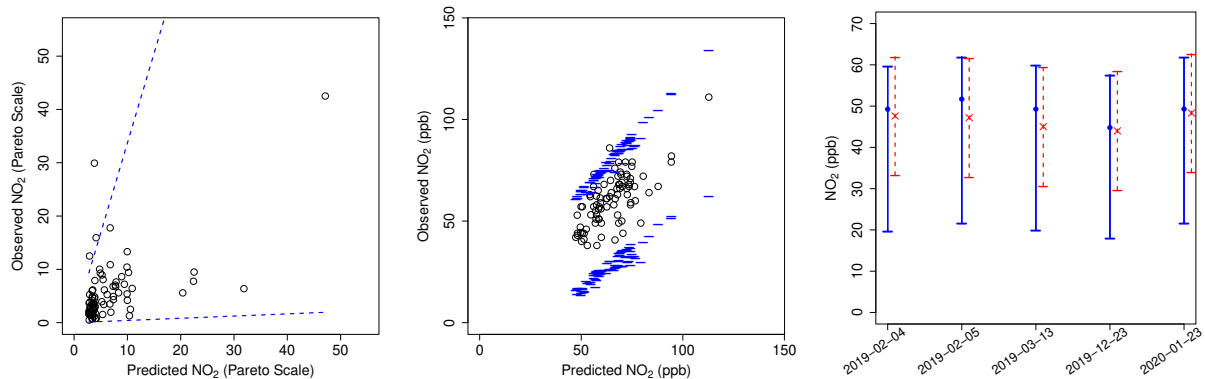
Figure 3.4: (Left) Scatterplot of $\hat{X}_5$ and X_5 with the 95% prediction intervals on the Pareto scale. (Center) Scatterplot and 95% prediction intervals after transformation back to the original scale of the NO_2 data. (Right) Comparison of the point predictions and 95% prediction intervals using the transformed linear approach (solid line) and a Gaussian-based approach with the covariance matrix (dashed line) for five recent dates when Alexandria is not observed.

3.4.2 Industry Portfolios.

We apply the transformed-linear prediction method to a higher dimensional financial data set. The data set obtained from the Kenneth French Data Library[2] contains the value-averaged daily

---

[2]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

returns of 30 industry portfolios. We analyze data for 1950-2020, consisting of $n = 17911$ observations. Since our interest is in extreme losses, we negate the returns, set negative returns to zero so that data is in the positive orthant, and apply the transform t to each variable to meet the lower tail bound condition. Although these data appear to be heavy-tailed, it still requires marginal transformation so that $\alpha = 2$ can be assumed. Let $\boldsymbol{X}^{(orig)}$ denote the random vector of the value-averaged daily returns. For simplicity we use the empirical CDF to perform the marginal transformation $X_i = 1/\sqrt{(1 - \hat{F}_i(X_i^{(orig)}))} - \delta$, which is applied to each industry's data so that X_i follows the same shifted Pareto distribution as before. We again assume $\boldsymbol{X}_t$, the random vector denoting the observations on day t, are iid copies of $\boldsymbol{X}$. The data set is randomly split into two sets. The training set consists of two-thirds of the data ($n_{train} = 11940$) to estimate the TPDM and obtain the vector $\hat{\boldsymbol{b}}$. The test set consists of the last one-third of the data ($n_{test} = 5970$) to assess coverage rates.

Following similar steps in the previous application, the 30×30 TPDM $\Sigma_{\boldsymbol{X}}$ is estimated first in the training set. We focus on performing the linear prediction for extreme losses of coal, beer, and paper. The three largest coefficients in $\hat{\boldsymbol{b}}_{coal}$ are $(0.42, 0.36, 0.20)$ and correspond to fabricated products and machinery, steel, and oil respectively. The three largest coefficients $\hat{\boldsymbol{b}}_{beer}$ are $(0.52, 0.24, 0.12)$ and correspond to food products, retail, and consumer goods (household). The three largest coefficients for $\hat{\boldsymbol{b}}_{paper}$ are $(0.21, 0.11, 0.08)$ and correspond to chemicals, consumer goods (household), and construction materials. Figure 3.5 shows the scatterplot of the observed daily returns versus predicted daily returns on the Pareto scale for coal, beer, and paper, respectively. The assessed coverage rates of our transformed linear 95% prediction intervals for coal, beer, and paper are 97.9%, 96.3%, and 98%, respectively.

For the purpose of comparison, we also assessed coverage rates of the MSPE-based 95% prediction intervals. Because the data are strongly non-Gaussian, we use the empirical CDF to transform the marginals to be standard normal before estimating the covariance matrix. The coverage rates of MSPE-based 95% prediction intervals are 79.3%, 66.6%, and 51.2% for coal, beer, and paper respectively.
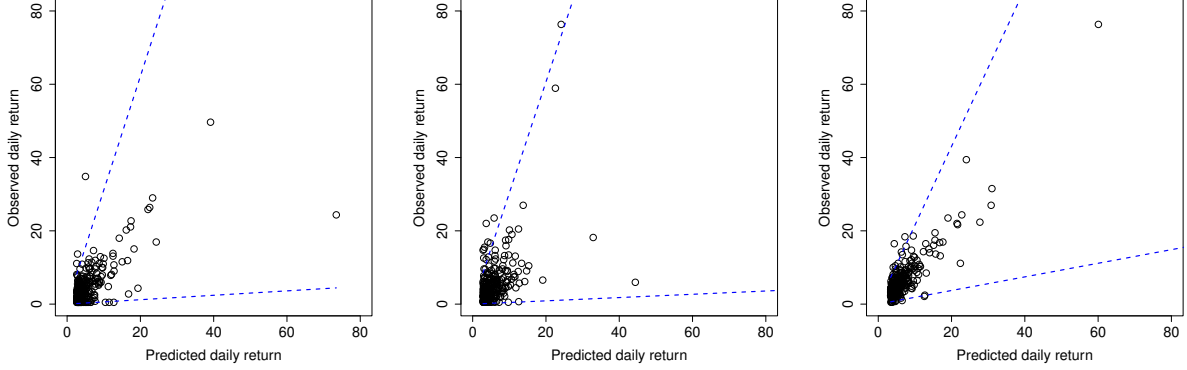
Figure 3.5: Scatterplot of observed daily returns and predicted daily returns with the 95% prediction intervals on the Pareto scale for coal, beer, and paper from left to right.

3.5 Comparison between Prediction Intervals with Linear Approach versus Parametric Approach

We aim to compare prediction intervals with linear approach to ones with a parametric angular approach. We use the point process argument in Section 3.3 to approximate the conditional angular density given large observed values. For the NO_2 air pollution application, it is feasible to fit a parametric angular measure model because only five monitoring sites are considered. Let $\{H_{\boldsymbol{X}}(\theta), \theta \in \Theta\}$ be a parametric class of angular measures. There are a number of possible parametric angular measures such as the bilogistic model proposed by Joe et al. [1992], the Dirichlet model by Coles and Tawn [1994], or the pairwise beta distribution by Cooley et al. [2010].

We choose the pairwise beta model by Cooley et al. [2010] for comparison because the model also describes all pairwise dependencies between variables at extreme levels. Cooley et al. [2010] construct the pairwise beta distribution by a geometric interpretation of the requirements of angular measure. They first consider the pairwise beta function as a symmetric beta density between angular components on the unit ball $\Theta_{p-1} = \{\boldsymbol{w} \in \mathbb{R}^{p-1} : ||\boldsymbol{w}|| = 1\}$,

$$h_{i,j}(\boldsymbol{w}; \beta_{i,j}) = \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j}-1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j}-1}, \tag{3.6}$$

35

where $\boldsymbol{w} \in \Theta_{p-1}$ and $\beta_{i,j} > 0$. The magnitude of the parameter $\beta_{i,j}$ describes the dependence between the ith and jth elements.

For each pair of angular components, simply adding up each of the $\binom{p}{2}$ pairwise beta functions is not suitable because one of the pairwise beta functions containing a strong dependence can yield some minimal level of dependence across all of the components. To resolve the issue, Cooley et al. [2010] include an additional parameter α to control the overall dependence in the model.

$$h(\boldsymbol{w}; \alpha, \boldsymbol{\beta}) = K_p(\alpha) \sum_{1 \leq i < j \leq p} h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}), \tag{3.7}$$

where

$$h_{i,j}(\boldsymbol{w}; \alpha, \beta_{i,j}) = (w_i + w_j)^{2\alpha - 1}(1 - (w_i + w_j))^{\alpha(p-2) - p + 2} \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left(\frac{w_i}{w_i + w_j}\right)^{\beta_{i,j} - 1} \left(\frac{w_j}{w_i + w_j}\right)^{\beta_{i,j} - 1},$$

and

$$K_p(\alpha) = \frac{2(p-3)!}{p(p-1)\sqrt{p}} \frac{\Gamma(\alpha p + 1)}{\Gamma(2\alpha + 1)\Gamma(\alpha(p-2))}.$$

After preprocessing the data, we implement the maximum likelihood approach to estimate parameters in the training set. For the largest 5% data (in terms of $||x_{p+1}||$) in the test set, resulting in 86 observations, we approximate the conditional density $f_{X_{t,5}|\boldsymbol{X}_{t,-5}}(X_{t,5}|\boldsymbol{X}_{t,-5})$ for $t = 1, \ldots, 86$, with the fitted pairwise beta angular density. In contrast, our linear approach approximates the conditional angular density $f_{X_{t,5}|\hat{X}_t}(X_{t,5}|\hat{X}_t)$ for $t = 1, \ldots, 86$, via completely positive decomposition for the prediction TPDM. Creating the 95% prediction intervals for each approach, we transform the prediction intervals back to the original scale for comparison.

The assessed coverage rate of 95% prediction intervals is 0.965 and 0.953 for the linear approach and the pairwise beta approach, respectively. To compare the transformed-linear approach to the parametric approach, we take the width of the prediction intervals for each approach. Let U and L denote the upper bound and the lower bound of the prediction interval, respectively. We let $U_{t,pair} - L_{t,pair}$ and $U_{t,linear} - L_{t,linear}$, $t = 1, ..., 86$, be the widths for a pairwise beta density and a linear approach, respectively. Rather than taking an average of widths for each approach, we take

the ratio of the two widths for each prediction interval, $\frac{U_{t,pair}-L_{t,pair}}{U_{t,linear}-L_{t,linear}}$. We then take the average of these ratios so that this average is robust to outliers. The average ratio of 1.04 implies that the prediction interval based on the best linear approach is slightly narrower on average.

3.6   Summary and Discussion

We have constructed an inner product space of nonnegative random variables arising from transformed linear combinations of independent regularly varying random variables. The elements of the TPDM correspond to these inner products if one is willing to assume that these random variables in $\mathcal{V}_+^q$. The projection theorem yields the optimal transformed linear predictor. Our method for obtaining prediction intervals shows very good performance both in a simulation study and in two applications. The method is simple and is based only on the TPDM which is estimable in high dimensions.

We restrict to nonnegative regularly varying random variables to focus attention on the upper tail. Relaxing this restriction could allow one to use standard linear operations. Even when the data can be negative, we believe there is value in focusing in one direction. In the financial application, tail dependence for extreme losses can be different than for gains, and this information is lost when dependence is summarized with a single number.

The random vectors $\boldsymbol{X} = A \circ \boldsymbol{Z}$ comprised of elements of our vector space have a simple angular measure consisting of q point masses where q is the number of columns of A. Previous models with angular measures consisting of discrete point masses have been criticized as being overly simple. A difference here is that we do not have to specify q to use this framework to perform prediction, or more generally, we do not have to really believe that our data arise from such a simple model. Rather, if we are comfortable with the information contained in the TPDM, then we can use its information to easily obtain a point prediction and sensible prediction intervals that reflect the information contained.

In many applications, dependence cannot be measured between the observed values and the value to be predicted. In kriging for example, a spatial process model is first fit so that covariance

between any two locations can be quantified. One can imagine modeling the extremal pairwise dependence as a function of distance before applying the methods here to perform prediction for extreme levels.

Chapter 4

Partial Tail Correlation

4.1 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 $\boldsymbol{X}_{\backslash(i,j)}$) 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 $\Sigma_{i,j}^{-1} = 0$, then X_i and X_j are conditionally independent given $\boldsymbol{X}_{\backslash(i,j)}$.

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 $\boldsymbol{X}_p = (\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})$, where $\boldsymbol{X}^{(1)} = (X_i, X_j)^T$ and $\boldsymbol{X}^{(2)} = \boldsymbol{X}_{\backslash(i,j)}^T$, 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

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

where $\hat{\boldsymbol{X}} = (\hat{X}_i, \hat{X}_j)^T$ is the vector of best linear predictors. The partial correlation between X_i and X_j given $\boldsymbol{X}_{\backslash(i,j)}$ is

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

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 4.1 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 4.1: 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 4.1 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 chapter, we develop a novel method for characterizing and investigating conditional extremal relationships between pairs of variables. We continue to rely on multivariate regular vari-

ation 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 (2.9), 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.

We then turn our attention to inference. We connect the matrix of inner products to the TPDM. 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, and we apply the method to assess flood risk in application to extreme river discharges in the upper Danube basin which was studied in Engelke and Hitz [2020].

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

4.2.1 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 in Section 2. For $\boldsymbol{a} \in \mathbb{R}^q$, the vector space is,

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

where $Z_j \in RV_+^1(2)$, $j = 1, \ldots, q$, are independent regularly varying random variables meeting lower tail condition $sPr(b(s)^{-1}Z_j > z) \to z^{-\alpha}$ with a common normalization $\lim_{z \to \infty} \frac{P(Z_j > z)}{z^{-2}L(z)} = 1$.

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 $d(X_1, X_2) = \|X_1 \ominus X_2\|_{\mathcal{V}^q} = \sqrt{\sum_{i=1}^{q}(a_{1i} - a_{2i})^2}$.

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

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

and note that 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 $z \to \infty$. 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 4.4.

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 $A \circ \boldsymbol{Z}$, where $A = (\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \ldots, \boldsymbol{a}_q^T)^T$ by Corollary 2.2.1. We denote the matrix of inner products by

$$\Gamma_{\boldsymbol{X}_p} = \langle X_i, X_j \rangle_{i,j=1,\ldots p} = AA^T. \tag{4.1}$$

The angular measure associated with $\boldsymbol{X}_p$ is given by (2.5).

4.2.2 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. Furthermore, $\mathcal{V}^q$ is complete and is therefore a Hilbert space. We show in Appendix C that $\mathcal{V}^q$ is isomorphic to $\mathbb{R}^q$ and complete. 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,

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

For completeness, 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 $\mathcal{V}_A$.

In Chapter 2, we 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 is 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 $\alpha = 2$.

Theorem 4.2.1. *(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 $X_i = \sum_{j=1}^{q} a_{ij} \circ Z_j \in \mathcal{V}^q, i = 1, \ldots, p,$ 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 $X_i = \sum_{j=1}^{q} a_{ij} \circ Z_j, i = 1, \ldots, p,$ and $X = \sum_{j=1}^{q} a_j^* \circ Z_j$ in $\mathcal{V}^q$. For $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$, consider $\boldsymbol{b}^T \circ \boldsymbol{X}_p \in \mathcal{V}_A$. $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||_{\mathcal{V}^q}^2 = \sum_{j=1}^{q} (a_j^* - \boldsymbol{b}^T \boldsymbol{a}_{\cdot j})^2$ where $\boldsymbol{a}_{\cdot j}$ is the jth column vector of $\boldsymbol{A}$. We assume $\text{Rank}(\boldsymbol{A}) = p$. 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. $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||_{\mathcal{V}^q}^2 = \sum_{j=1}^{q} f_j$ 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 $Y \in \mathcal{V}_A$,

$$\|X \ominus Y\|_{\mathcal{V}^q}^2 = \langle (X \ominus \hat{X}) \oplus (\hat{X} \ominus Y), (X \ominus \hat{X}) \oplus (\hat{X} \ominus Y) \rangle$$

$$= \|X \ominus \hat{X}\|_{\mathcal{V}^q}^2 + \|\hat{X} \ominus Y\|_{\mathcal{V}^q}^2$$

$$\geq \|X \ominus \hat{X}\|_{\mathcal{V}^q}^2,$$

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

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

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

$$\|X \ominus \tilde{X}\|_{\mathcal{V}^q}^2 = \langle X \ominus \hat{X} \oplus \hat{X} \ominus \tilde{X}, X \ominus \hat{X} \oplus \hat{X} \ominus \tilde{X} \rangle$$

$$= \|X \ominus \hat{X}\|_{\mathcal{V}^q}^2 + a^2 \circ \frac{1}{\|Y\|_{\mathcal{V}^q}^2} + 2\langle X \ominus \hat{X}, \hat{X} \ominus \tilde{X} \rangle$$

$$= \|X \ominus \hat{X}\|_{\mathcal{V}^q}^2 - a^2 \circ \frac{1}{\|Y\|_{\mathcal{V}^q}^2}$$

$$< \|X \ominus \hat{X}\|_{\mathcal{V}^q}^2.$$

$\square$

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 4.2.1.

Proposition 4.2.1. *(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 indicates $\alpha \circ P_{\mathcal{V}_A} X \oplus \beta \circ P_{\mathcal{V}_A} Y$ 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 $Y \ominus P_{\mathcal{V}_A} X$, $||Y \ominus P_{\mathcal{V}_A} X||_{\mathcal{V}^q}^2 = 0$ since $Z \ominus (I - P_{\mathcal{V}_A}) X \in \mathcal{V}_A^\perp$. Hence $Y = P_{\mathcal{V}_A} X$ and $Z = (I - P_{\mathcal{V}_A}) X$. $\square$

Theorem 4.2.1 shows that $\hat{X} \in \mathcal{V}_A$ is the unique element closest to X such that

$$\langle X \ominus \hat{X}, Y \rangle = 0 \tag{4.2}$$

for all $Y \in \mathcal{V}_A$. The equation (4.2) is called the prediction equation and makes $\hat{X}$ as the best predictor of $X \in \mathcal{V}^q$. Briefly returning to the context of Chapter 2, suppose we observe $X_1, \ldots, X_p$ and want to predict an unobserved X_{p+1} by using the transformed-linear combination of $(X_1, \ldots, X_p)$. 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.

$$\left[<X_{p+1}, X_i> \right]_{i=1}^{p} = \left[<X_i, X_j> \right]_{i,j=1}^{p} \left[b_i \right]_{i=1}^{p} = \left[\sum_{k=1}^{q} a_{ik}a_{jk} \right]_{i,j=1}^{p} \left[b_i \right]_{i=1}^{p} \qquad (4.3)$$

This in turn yields the form of the best transformed linear predictor given in Equation (2.12).

4.2.3 Inner Product Matrix of Prediction Errors

Here, we change focus from the setting where $\boldsymbol{X}_p$ is observed and X_{p+1} is unobserved. We continue to assume $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$ where $X_i \in \mathcal{V}^q$, for $i = 1, \ldots, p,$, 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

$$\Gamma_{(\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T} = \begin{bmatrix} A^{(1)}A^{(1)^T} & A^{(1)}A^{(2)^T} \\ A^{(2)}A^{(1)^T} & A^{(2)}A^{(2)^T} \end{bmatrix} := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}. \qquad (4.4)$$

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 $\|\boldsymbol{b}^T A^{(2)} - A^{(1)}\|_{\mathcal{V}^q}^2$. 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,

$$\hat{\boldsymbol{b}} = (A^{(2)}A^{(2)^T})^{-1}A^{(2)}A^{(1)^T} = \Gamma_{22}^{-1}\Gamma_{21}. \qquad (4.5)$$

48

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.

$$
\begin{aligned}
\|\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}\|_{\mathcal{V}^q}^2 :&= \langle \boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}, \boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}} \rangle \\
&= (A^{(1)} - \hat{\boldsymbol{b}}^T A^{(2)})(A^{(1)} - \hat{\boldsymbol{b}}^T A^{(2)})^T \\
&= \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}.
\end{aligned}
\tag{4.6}
$$

4.3 Partial Tail Correlation

4.3.1 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 4.2.1 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 4.3.1. 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

$$\rho_{ij}^E = \frac{< X_i \ominus P_{\mathcal{V}_{A_2}} X_i, X_j \ominus P_{\mathcal{V}_{A_2}} X_j >}{\|X_i \ominus P_{\mathcal{V}_{A_2}} X_i\|_{\mathcal{V}^q} \|X_j \ominus P_{\mathcal{V}_{A_2}} X_j\|_{\mathcal{V}^q}}, \qquad (4.7)$$

where the superscript E in ρ_{ij}^E 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 $\rho_{ij}^E = 0$, which we denote by $X_i \perp X_j | \boldsymbol{X}^{(2)}$.

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

$$\Gamma_{1|2} := < \boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}} \boldsymbol{X}^{(1)}, \boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}} \boldsymbol{X}^{(1)} >$$

$$= \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}.$$

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

$$\rho_{i,j}^E = \frac{a_{ij}}{\sqrt{a_{ii} a_{jj}}}, \quad i, j = 1, 2, \qquad (4.8)$$

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

4.3.2 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 is related 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 4.3.1. *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,

$$\hat{X}_{p+1} = P_{\mathcal{V}_A} X_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} \oplus P_{\mathcal{V}_{A_1^\perp}} X_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} \oplus c \circ (X_1 \ominus P_{\mathcal{V}_{A_1^\perp}} X_1), \qquad (4.9)$$

where $c = \frac{<X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1>}{\|X_1 \ominus P_{\mathcal{V}_{A_1}} X_1\|^2} = \frac{<X_{p+1} \ominus P_{\mathcal{V}_{A_1}} X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1>}{\|X_1 \ominus P_{\mathcal{V}_{A_1}} X_1\|^2}$ 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 (4.9), $\hat{X}_{p+1} = c \circ X_1 \oplus \left(\sum_{i=1}^{p-1} (d_i - ce_i) \circ X_{i+1} \right)$.
By matching the coefficient c of X_1 in (4.9) 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,

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

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

We can also understand $P_{\mathcal{V}_{A_1^\perp}} X_{p+1}$ in the regression setting. Let's consider a simple linear regression with no intercept, $Y = X\beta + \epsilon$. The projection of Y onto the space spanned by X is $\hat{Y} = P_X Y = X\hat{\beta}$ where $\beta = (X^T X)^{-1} X^T y$. Note that $\hat{Y}$ can be expressed as the inner products, $\hat{Y} = \frac{<X,y>}{<X,X>} X$. By replacing Y and X with X_{p+1} and $X_1 \ominus P_{\mathcal{V}_{A_1}} X_1$ respectively, $\hat{X}_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} = \frac{<X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1>}{\|X_1 \ominus P_{\mathcal{V}_{A_1}} X_1\|^2} (X_1 \ominus P_{\mathcal{V}_{A_1}} X_1)$.

4.3.3 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) contains 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 $\boldsymbol{X}_{\backslash(i,j)}$ is related to the (i,j)th element of the precision matrix. Specifically, $\Sigma_{ij}^{-1} = 0 \Leftrightarrow X_i \perp X_j | \boldsymbol{X}_{\backslash\{X_i, X_j\}}$. Analogously, we connect the idea of partial tail correlation in (4.7) 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 4.3.1, partition $\boldsymbol{X}_p$ into two subvectors $\boldsymbol{X}^{(1)} := (X_i, X_j)^T$ and $\boldsymbol{X}^{(2)} := \boldsymbol{X}_{\backslash(i,j)}^T$. Recall the block form of the inner product matrix of $\boldsymbol{X}_p = (\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T$

$$
\Gamma_{\boldsymbol{X}_p} := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \tag{4.10}
$$

By the matrix inversion in block form,

$$
\Gamma_{\boldsymbol{X}_p}^{-1} = \begin{bmatrix} \Gamma_{1|2}^{-1} & -\Gamma_{1|2}^{-1}\Gamma_{12}\Gamma_{22}^{-1} \\ -\Gamma_{22}^{-1}\Gamma_{21}\Gamma_{1|2}^{-1} & \Gamma_{22}^{-1} + \Gamma_{22}^{-1}\Gamma_{21}\Gamma_{1|2}^{-1}\Gamma_{12}\Gamma_{22}^{-1}, \end{bmatrix} \tag{4.11}
$$

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, we can easily find the relation between $\Gamma_{1|2}$ and $\Gamma_{1|2}^{-1}$ by the inversion formula,

$$
\Gamma_{1|2}^{-1} = \frac{1}{|\Gamma_{1|2}|} \begin{bmatrix} [\Gamma_{1|2}]_{22} & -[\Gamma_{1|2}]_{12} \\ -[\Gamma_{1|2}]_{21} & [\Gamma_{1|2}]_{11} \end{bmatrix}, \tag{4.12}
$$

52

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

$$\rho_{i,j}^E = \frac{[\Gamma_{1|2}]_{12}}{\sqrt{[\Gamma_{1|2}]_{11}[\Gamma_{1|2}]_{22}}} = \frac{-[\Gamma_{1|2}^{-1}]_{12}}{\sqrt{[\Gamma_{1|2}^{-1}]_{11}[\Gamma_{1|2}^{-1}]_{22}}}. \tag{4.13}$$

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 $[\Gamma_{1|2}^{-1}]_{12} = 0$, then it implies that X_i and X_j given $\boldsymbol{X}_{\backslash(i,j)}$ 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,

$$\Gamma_{(X_{p+1},\boldsymbol{X}_p^T)^T} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \tag{4.14}$$

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 $\boldsymbol{X}_p$. Assume Γ_{22} is invertible, then by the inverse formula,

$$\Gamma_{(X_{p+1},\boldsymbol{X}_p^T)^T}^{-1} = \begin{bmatrix} \frac{1}{k} & -\frac{1}{k}\Gamma_{12}\Gamma_{22}^{-1} \\ -\frac{1}{k}\Gamma_{22}^{-1}\Gamma_{21} & (\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12})^{-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{k} & -\frac{1}{k}\boldsymbol{b}^T \\ -\frac{1}{k}\boldsymbol{b} & (\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12})^{-1} \end{bmatrix} \tag{4.15}$$

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 $b_i = 0$.

Again for illustration, we now consider the transformed-linear model

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \phi & 1 & 0 & 0 \\ \phi^2 & \phi & 1 & 0 \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} \circ \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} \tag{4.16}$$

53

where $\{Z_t\}$ 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 $\{X_t\}$. The sequential generating equation is

$$X_t = \phi \circ X_{t-1} \oplus Z_t, \text{ for } i = 1, 2, 3, 4,$$

where $X_0 = 0 \ a.s.$

By the matrix inversion in (4.15), the inverse IPM has a sparse pattern.

$$\Sigma^{-1}_{(X_{p+1}, \boldsymbol{X}_p^T)^T} = \begin{bmatrix} 1 & -\phi & 0 & 0 \\ -\phi & 1 + \phi^2 & -\phi & 0 \\ 0 & -\phi & 1 + \phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{bmatrix}, \tag{4.17}$$

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

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

In chapter 2, we introduced the vector space $\mathcal{V}^q$ arising from the transformed linear combination of independent regularly-varying random variables where $X = \boldsymbol{a}^T \circ \boldsymbol{Z} \in \mathcal{V}^q$ takes values in the positive orthant. It is essential that the elements of the coefficient vectors $\boldsymbol{a}$ are allowed to be negative for $\mathcal{V}^q$ to be a vector space. However, we know 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 affect tail behavior. In addition, $\boldsymbol{X}_p = A \circ \boldsymbol{Z}_q$ and $\boldsymbol{X}_p^+ = A^{(0)} \circ \boldsymbol{Z}$ have the same angular measure: $H_{\boldsymbol{X}} = H_{\boldsymbol{X}_p^+} = \sum_{j=1}^q \|a_j^{(0)}\|^\alpha \delta_{a_j^{(0)}/\|a_j^{(0)}\|}(\cdot)$. Hence, $\boldsymbol{X}_p$ and $\boldsymbol{X}_p^+$ are indistinguishable in terms of their tail behavior.

In terms of modeling, it is reasonable to restrict attention to the 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 $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ as in (2.9).

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 for all the work heretofore in Chapter 4 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. 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].

4.5   Hypothesis Testing for Zero Elements in the Inverse TPDM

4.5.1   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

$$nP(n^{-1/2}\boldsymbol{X}_p \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}_p}(\cdot),$$

where $\nu_{\boldsymbol{X}_p}(dr \times dw) = 2r^{-3}drdH_{\boldsymbol{X}_p}(w)$. Let T denote the polar coordinate transformation so that $T : (\boldsymbol{X}_p) \mapsto (R, \boldsymbol{W}_p)$, where $(R, \boldsymbol{W}_p) = (||\boldsymbol{X}_p||_2, \boldsymbol{X}_p/||\boldsymbol{X}_p||_2)$. The TPDM of $\boldsymbol{X}_p$ is

$$\Sigma_{\boldsymbol{X}_p} := \int_{\Theta_{p-1}^+} \boldsymbol{w}_p\boldsymbol{w}_p^T dH_{\boldsymbol{X}}(\boldsymbol{w}),$$

where $H_{\boldsymbol{X}_p}$ is the angular measure on the L_2 unit ball $\Theta_{p-1}^+ = \{\boldsymbol{w} \in \mathbb{R}_+^{p-1} : ||\boldsymbol{w}||_2 = 1\}$. The (i, j)th element of the TPDM

$$\sigma_{ij} = \int_{\Theta_{p-1}^+} w_i w_j dH_{\boldsymbol{X}}(\boldsymbol{w}),$$

where σ_{ij} corresponds to the extremal dependence measure defined by Larsson and Resnick [2012] if $H_{\boldsymbol{X}_p}$ is assumed to be a probability measure. Because we do not make that assumption, the scale of the components of $\boldsymbol{X}_p$ is retained in the angular measure. If $\sigma_{ij} = 0$, then the angular measure $H_{\boldsymbol{X}_p}$ puts all masses on the coordinate axes. In other words, $\sigma_{ij} = 0$ iff $H_{\boldsymbol{X}_p}(\{\boldsymbol{w} \in \Theta_{p-1}^+ : w_i > 0, w_j > 0\}) = 0$ implies asymptotic independence of (X_i, X_j).

Larsson and Resnick [2012] provide an equivalent form of the extremal dependence measure as

$$\sigma_{ij} = \lim_{x \to \infty} m E[W_i W_j | R > x],$$

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. Proposition 4 in Larsson and Resnick [2012] provides a useful interpretation of the extremal dependence measure as the limit of conditional cross moments of X_1 and X_2.

This definition gives 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

$$\hat{\sigma}_{ij}(n) = \frac{\hat{m}}{k} \sum_{\ell=1}^n w_{\ell,i} w_{\ell,j} \mathbb{I}[r_\ell > r_{(k)}], \tag{4.18}$$

where $\hat{m}$ is an estimate of $H_{\boldsymbol{X}_p}(\Theta_{p-1}^+)$, and $k := k(n)$ is such that $\lim_{n\to\infty} k/n = 0$ and $r_{(k)}$ is the kth 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 $\hat{m} = (r_{(k)}^2/n)k$ 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.

$$\lim_{n\to\infty} \sqrt{k}\Big(\frac{n}{k}mE[W_iW_j\mathbb{I}[R/b(n/k) \geq t^{-\gamma}]] - mE[W_iW_j\frac{n}{k}\bar{F}(b(n/k)t^{-\gamma}]\Big) = 0, \qquad (4.19)$$

holds locally uniformly for $t \in [0,\infty)$, and assume that $\tau^2 = m^2Var(W_iW_j) > 0$. Larsson and Resnick [2012] notes that $\tau^2 = 0$ implies asymptotic independence and 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 $b(n/k)$. With this condition, we do not need to make a second-order regular variation condition. By using an order statistic, it is not required to know the normalization $b(\cdot)$. Under the condition (4.19), the estimator $\hat{\sigma}_{ij}(n)$ is asymptotically normal by Larsson and Resnick [2012].

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

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

Suppose $\boldsymbol{X}_p \in RV_p^+(2)$ is constructed by Corollary 2.2.1. 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 $H_{\boldsymbol{X}}$ on the L_2 unit ball given in (2.5). Let $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z}$ 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

$$\sigma_{ij} = \int_{\Theta_{p-1}^+} w_iw_j dH_{A\circ\boldsymbol{Z}}(\boldsymbol{w}) = \sum_{l=1}^{q} a_{il}^{(0)}a_{jl}^{(0)},$$

where $H_{\boldsymbol{X}}(\cdot) = \sum_{l=1}^{q} ||\boldsymbol{a}_l^{(0)}||^2\delta_{\boldsymbol{a}_l/||\boldsymbol{a}_l||}$. To find $\tau_{ij}^2 = m^2Var(W_iW_j)$, 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 $N_{\boldsymbol{X}}(\cdot) = m^{-1}H_{\boldsymbol{X}}(\cdot)$ is a probability measure. $E[W_i W_j] = \frac{1}{m}\sum_{l=1}^{q} a_{il}^{(0)} a_{jl}^{(0)}$ and $E[W_i^2 W_j^2] = \frac{1}{m}\sum_{l=1}^{q} \frac{a_{il}^{(0)2}}{||\boldsymbol{a}_l^{(0)}||} \frac{a_{jl}^{(0)2}}{||\boldsymbol{a}_l^{(0)}||}$.

$$Var(\hat{\sigma}_{ij}) = Var\left(\frac{m}{k}\sum_{l=1}^{n} W_{il} W_{jl} \mathbb{I}[R_l > R_{(k)}]\right)$$
$$= \frac{m^2}{k} Var[W_{i,1} W_{j,1}] \quad \text{since} \quad (W_{i,l}, W_{j,l}) \text{'s are iid}$$

We obtain an estimate of $\widehat{Var}(\hat{\sigma}_{ij})$ by estimating $Var(W_{i1} W_{j1})$. To get $\widehat{Var}(W_{i1} W_{j1})$, a natural estimate for $E[W_i W_j]$ and $E[W_i^2 W_j^2]$ are $\hat{E}[W_i W_j] = \frac{1}{k-1}\sum_{l=1}^{k} w_{i,l} w_{j,l} \mathbb{I}[R_l > R_{(k)}]$ and $\hat{E}[W_i^2 W_j^2] = \frac{1}{k-1}\sum_{l=1}^{k} w_{il}^2 w_{jl}^2 \mathbb{I}[R_l > R_{(k)}]$, respectively.

4.5.2 Residuals and Asymptotic Normality of the Conditional Inner Product Matrix

The main 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

$$\hat{\Gamma}_{1|2} = [\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}], \tag{4.20}$$

where $\hat{\Sigma}_{ij}$ for $i, j = 1, 2$ are sample block matrices as in (4.10). However, the distribution of $\hat{\Gamma}_{1|2}$ is not straightforward to obtain from (4.20).

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 $\Gamma_{1|2}$ is not necessarily completely postive. Unlike the original data where we can assume away the importance of any negative coefficients as described in Section 4.4, 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 (4.3.1) 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 $\Sigma_{1|2}$ is of primary interest because it is tied to the idea of the partial tail correlation ρ_{ij}^E. Following similar steps above, let $\boldsymbol{U}_\ell = (U_{\ell,1}, U_{\ell,2})$, $\ell = 1, \ldots, n$, 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 kth upper order statistic.

Since $\boldsymbol{U}$ is a linear combination of independent Z_j's, its angular measure is discrete and σ_{12}^u 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

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

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

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

$$\hat{\sigma}_{12,n}^u = \tilde{m} \int_{\Theta_1} w_1 w_2 d\hat{N}_{\boldsymbol{U}}(\boldsymbol{w}) = \frac{\tilde{m}}{k} \sum_{l=1}^n w_{1l} w_{2l} \mathbb{I}[R_l > R_{(k)}], \tag{4.21}$$

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 $\lim_{n \to \infty} k/n = 0$ and $R_{(k)}$ is the kth upper order statistic in the sample of size n.

Under the condition (4.19), the scaled estimator $\hat{\sigma}_{12,n}^u$ is asymptotically normal by Larsson and Resnick [2012],

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

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 $Var(W_1 W_2)$,

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

60

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

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

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

$$\frac{\hat{\sigma}^u_{12,n}}{\sqrt{\hat{\tau}^{u^2}/k}} \sim T_{k-1}, \tag{4.22}$$

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.

4.5.3 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 (4.3.3) with generating equation

$$X_t = \phi \circ X_{t-1} \oplus Z_t,$$

where $\{Z_t\}$ 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 4.2 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 4.2 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 4.5.1, we obtain estimated variances for τ^{u^2}. From the equation in (4.22), we construct a 95% confidence interval for each iteration and achieve a coverage rate of 0.95.
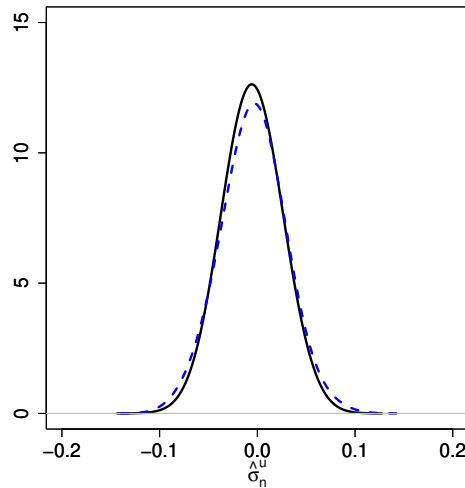
Figure 4.2: The kernel density based on the residuals (solid line) versus the kernel density from the partition of the TPDM (dashed line).

4.6 Asymptotic Distribution for the Vector $\hat{\boldsymbol{b}}$ and $\hat{\Sigma}_{1|2}$ when Predicting X_{p+1}

Especially in the case where we predict X_{p+1} using the best transformed-linear predictor $\hat{X} = \hat{\boldsymbol{b}}^T \circ \boldsymbol{X}_p$, the asymptotic distribution of the estimator for both the optimal weights $\boldsymbol{b}$ and the conditional TPDM $\Sigma_{1|2}$ could be of interest. In $\mathcal{V}_+^q$, the vector $\boldsymbol{b} = \Sigma_{22}^{-1}\Sigma_{12}$ itself is meaningful, and it is also tied to the inverse of $\Sigma_{(X_{p+1}, \boldsymbol{X}_p^T)^T}$ in (4.15). In addition, the conditional TPDM $\Sigma_{1|2}$ plays an role of uncertainty quantification for the prediction of extremes.

First, we derive the asymptotic normality for $\hat{\boldsymbol{b}}$ via Slutsky's theorem and asymptotic normality results by Larsson and Resnick [2012]. Let $\Sigma_{\boldsymbol{X}_{p+1}} = [\sigma_{ij}]_{i,j=1,\dots,p+1}$ be the TPDM of $\boldsymbol{X}_{p+1}$. Under the assumption in (4.19), a natural estimator for each element of the TPDM is asymptotically normal $\sqrt{k}(\hat{\sigma}_{ij}(n) - \sigma_{ij}) \sim N(0, \tau_{ij}^2)$, $i, j = 1, \dots, p+1$ and $\hat{\sigma}_{ij}(n) \xrightarrow{p} \sigma_{ij}$ by continuity mapping.

Let $\Sigma_{\boldsymbol{X}_{p+1}}$ be partitioned into block matrices,

$$\Sigma_{\boldsymbol{X}_{p+1}} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where Σ_{11} is the tail ratio of X_{p+1} and Σ_{22} is the TPDM of $\boldsymbol{X}_p$.

Minimizing the tail metric $\|X_{p+1} \ominus \hat{X}_{p+1}\|_{\mathcal{V}^q}$, we obtain the optimized weights $\boldsymbol{b}$ and its natural estimator is $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$. To specify the form of $\hat{\boldsymbol{b}}$, we start with the form of Σ_{21}. We define $R_{(k,p+1)} = \|(X_k, X_{p+1})\|_2$ and $(W_k, W_{p+1}) = (X_k/R_{(k,p+1)}, X_{p+1}/R_{(k,p+1)})$ for $k = 1, \dots, p$.

$$\Sigma_{21} = \begin{bmatrix} \int_{\Theta_1^+} w_1 w_{p+1} dH_{(X_1, X_{p+1})}(\boldsymbol{w}) \\ \int_{\Theta_1^+} w_2 w_{p+1} dH_{(X_2, X_{p+1})}(\boldsymbol{w}) \\ \vdots \\ \int_{\Theta_1^+} w_p w_{p+1} dH_{(X_p, X_{p+1})}(\boldsymbol{w}) \end{bmatrix}, \tag{4.23}$$

Let $R_{(k,p+1),i} = ||(X_{k,i}, X_{p+1,i})||$ and let $(W_{k,i}, W_{p+1,i}) = (X_{k,i}/R_{(k,p+1),i}, X_{p+1,i}/R_{(k,p+1),i})$ be iid copies of (W_k, W_{p+1}) for $k = 1, \ldots, p$, $i = 1, \ldots, n$. The natural estimator for Σ_{21} is

$$
\hat{\Sigma}_{21} = \begin{bmatrix} \frac{1}{k} \sum_{l=1}^{n} w_{1,l} w_{p+1,l} \mathbb{I}[R_{(1,p+1),l} > R_{(k)}] \\ \frac{1}{k} \sum_{l=1}^{n} w_{2,l} w_{p+1,l} \mathbb{I}[R_{(2,p+1),l} > R_{(k)}] \\ \vdots \\ \frac{1}{k} \sum_{l=1}^{n} w_{p,l} w_{p+1,l} \mathbb{I}[R_{(p,p+1),l} > R_{(k)}] \end{bmatrix}, \tag{4.24}
$$

where $R_{(k)}$ is the kth upper order statistic. Let $\tilde{T} = [\tau_{ij}^2]_{i,j=1,\ldots,p}$ be a covariance matrix of $\hat{\Sigma}_{21}$. If $\tilde{T}$ exists, then $\sqrt{k}(\hat{\Sigma}_{21} - \Sigma_{21}) \xrightarrow{d} (\mathbf{0}, \tilde{T})$. The asymptotic normality of the estimator $\hat{\Sigma}_{21}$ leads to the asymptotic normality for $\hat{b}$ via Slutsky's theorem.

Theorem 4.6.1. *Let $X_i \in \mathcal{V}^q$, $i = 1, \ldots, p+1$, and $\Sigma_{(X_{p+1}, \mathbf{X}_p^T)^T}$ be the TPDM of $(X_{p+1}, \mathbf{X}_p^T)^T$. We partition the TPDM as in (4.15). Let $\tilde{T}$ be a covariance matrix of $\hat{\Sigma}_{21}$. Assume $\tilde{T}$ exists and Σ_{22} is invertible. Under the condition (4.19),*

$$
\sqrt{k}(\hat{b} - b) \xrightarrow{d} (0, \tilde{\Sigma}),
$$

where k is the number of exceedances, $k := k(n)$ is such that $k/n \to 0$ as $n \to \infty$, $\hat{b} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$ and $\tilde{\Sigma} := \Sigma_{22}^{-1} \tilde{T} \Sigma_{22}^{-1^T}$ is a covariance matrix of $\hat{b}$.

Proof. We use the fact that $\hat{\Sigma}_{22}^{-1} \xrightarrow{p} \Sigma_{22}^{-1}$ since $\hat{\Sigma}_{22} \xrightarrow{p} \Sigma_{22}$ by continuity mapping and $\sqrt{k}(\hat{\Sigma}_{21} - \Sigma_{21}) \xrightarrow{d} (\mathbf{0}, \tilde{T})$ under the assumption in (4.19). By Slutsky's theorem, $\sqrt{k}(\hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} - \Sigma_{22}^{-1} \Sigma_{21}) \xrightarrow{d} (\mathbf{0}, \tilde{\Sigma})$ holds. As $\mathbf{X}_{p+1} = A \circ \mathbf{Z} \in \mathcal{V}_+^q$ arises from the matrix multiplication, we specify the form of the covariance matrix $\tilde{T}$ in terms of the matrix $A^{(0)}$. Let $\mathbf{X}_{p+1} = A \circ \mathbf{Z}$ and hence

$H_{\boldsymbol{X}_{p+1}} = A^{(0)}A^{(0)T}$ where each $X_i \in \mathcal{V}^q$, $i = 1, \ldots, p+1$. The $(1,1)$ element of $\tilde{T}$ is

$$
\begin{aligned}
Var(W_1 W_{p+1}) &= \int_{\Theta_1^+} [w_1 w_{p+1} - E[W_1 W_{p+1}]]^2 dH_{A \circ Z_{(1,p+1)}}(w) \\
&= \int_{\Theta_1^+} w_1^2 w_{p+1}^2 dH_{A \circ Z_{(1,p+1)}}(w) - \left[\int_{\Theta_1^+} w_1 w_{p+1} dH_{A \circ Z_{(1,p+1)}}(w) \right]^2 \\
&= \sum_{j=1}^{q} \frac{a_{1,j}^{(0)2}}{||\boldsymbol{a}_{(1,p+1),j}^{(0)}||^2} \frac{a_{p+1,j}^{(0)2}}{||\boldsymbol{a}_{(1,p+1),j}^{(0)}||^2} - \left[\sum_{j=1}^{q} a_{1,j}^{(0)} a_{p+1,j}^{(0)} \right]^2
\end{aligned}
$$

The $(1,2)$ element of $\tilde{T}$ is

$$
\begin{aligned}
Var(W_1 W_2 W_{p+1}^2) &= \int_{\Theta_2^+} [w_1 w_2 w_{p+1}^2 - E[W_1 W_2 W_{p+1}^2]]^2 dH_{A \circ Z_{(1,2,p+1)}}(\boldsymbol{w}) \\
&= \int_{\Theta_2^+} w_1^2 w_2^2 w_{p+1}^4 dH_{A \circ Z_{(1,2,p+1)}}(\boldsymbol{w}) - \left[\int_{\Theta_2^+} w_1 w_2 w_{p+1}^2 dH_{A \circ Z_{(1,2,p+1)}}(\boldsymbol{w}) \right]^2 \\
&= \sum_{j=1}^{q} \frac{a_{1,j}^{(0)2}}{||\boldsymbol{a}_{(1,2,p+1),j}^{(0)}||^2} \frac{a_{2,j}^{(0)2}}{||\boldsymbol{a}_{(1,2,p+1),j}^{(0)}||^2} \frac{a_{p+1,j}^{(0)4}}{||\boldsymbol{a}_{(1,2,p+1),j}^{(0)}||^2} \\
&\quad - \left[\sum_{j=1}^{q} \frac{a_{1,j}^{(0)}}{||\boldsymbol{a}_{(1,2,p+1),j}^{(0)}||_2} \frac{a_{2,j}^{(0)}}{||\boldsymbol{a}_{(1,2,p+1),j}^{(0)}||_2} a_{p+1,j}^{(0)2} \right]^2
\end{aligned}
$$

We can specify other elements in a similar way. $\qquad \square$

We can also employ the Slutsky's theorem to derive the asymptotic distribution for the sample conditional TPDM $\hat{\Sigma}_{1|2} = \hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$. Under some conditions, we use the facts that $\hat{\Sigma}_{11} \xrightarrow{p} \Sigma_{11}$, $\hat{\Sigma}_{22}^{-1} \xrightarrow{p} \Sigma_{22}^{-1}$, and $\sqrt{k}(\hat{\Sigma}_{21} - \Sigma_{21}) \xrightarrow{d} (\boldsymbol{0}, \tilde{T})$ by Larsson and Resnick [2012]. We show that the sample conditional TPDM is asymptotically a non-central Chi-square distribution,

$$
\hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \xrightarrow{d} \Sigma_{11} - \chi^2(\boldsymbol{c}),
$$

where $\boldsymbol{c}$ is a vector of noncentrality parameters. As $\sqrt{k}(\hat{\Sigma}_{21} - \Sigma_{21}) \xrightarrow{d} N(\boldsymbol{0}, \tilde{T})$, we show that the quadratic form of $Q := \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$ is asymptotically distributed to a non-central Chi-square distribution by Mohsenipour [2012]. By Slutsky's theorem, $\hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \xrightarrow{d} \Sigma_{11} - \chi^2(\boldsymbol{c})$.

Specifically, we decompose the covariance matrix $\tilde{T}$ into $\tilde{T} = \tilde{T}^{1/2}\tilde{T}^{1/2}$, where $\tilde{T}^{1/2} = P\Lambda^{1/2}P^T$ and P is the orthogonal matrix composed of eigenvectors of $\tilde{T}$ and $\Lambda = [\lambda_{ij}]_{i,j=1,\ldots,p}$ is a diagonal matrix of eigenvalues of $\tilde{T}$. Let

$$z = \tilde{T}^{-1/2}(\hat{\Sigma}_{21} - \Sigma_{21}) \sim N(\mathbf{0}, I).$$

$$\tilde{T}^{1/2}(z + \tilde{T}^{-1/2}\Sigma_{21}) = \hat{\Sigma}_{21}$$

$$\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21} = (z + \tilde{T}^{-1/2}\Sigma_{21})^T\tilde{T}^{1/2}\hat{\Sigma}_{22}^{-1}\tilde{T}^{1/2}(z + \tilde{T}^{-1/2}\Sigma_{21})$$

Let $\tilde{T}^{1/2}\hat{\Sigma}_{22}^{-1}\tilde{T}^{1/2}$ be decomposed into $P^*\Lambda^*P^{*^T}$. Then,

$$\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21} = (P^{*^T}z + P^{*^T}\tilde{\Sigma}^{-1/2}\Sigma_{21})^T\Lambda^*(P^{*^T}z + P^{*^T}\tilde{\Sigma}^{-1/2}\Sigma_{21})$$

Let $U = P^{*^T}z \sim N(\mathbf{0}, I)$ and $c = P^{*^T}\tilde{T}^{-1/2}\Sigma_{21}$. We can express $\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ in terms of U,

$$\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21} = (U + c)^T\Lambda^*(U + c)$$

That is, the quadratic form Q becomes a linear combination of independent non-central Chi-square variables.

$$Q = \sum_{j=1}^{p} \lambda_j^*(U_j + c_j)^2,$$

where c_j, $j = 1, \ldots, p$ is a noncentrality parameter.

4.7 Application

4.7.1 Danube River Basin

We apply the notion of partial tail correlation to learn 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[3]. 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 4.3 shows the river network in the upper Danube basin where the path $10 \rightarrow \cdots \rightarrow 1$ is the main channel and the 21 other locations are on tributaries.
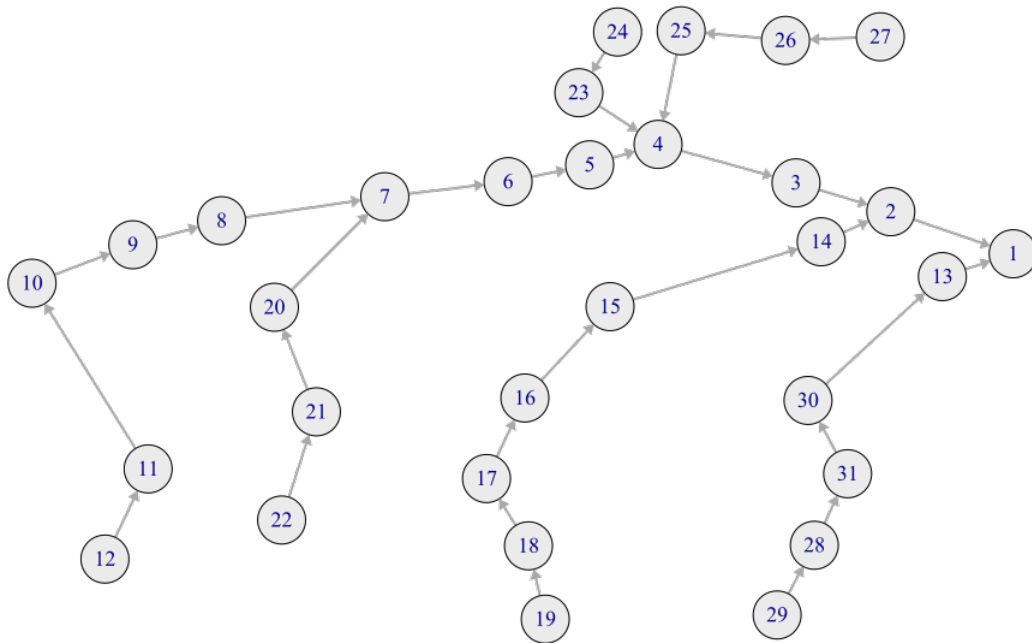
Figure 4.3: Physical flow connections in the upper Danube river basin

We analyze average daily river discharges from 31 gauging stations for 1960-2009. The data are available in the Bavarian Environmental Agency[4]. 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.

---

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

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

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 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.

Our inner product space assumes each $X_i \in RV_+^1(2)$. 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. 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. 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, the first step is to 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. 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 4.1. Our aim can now be described as trying to assess if each off-diagonal element is significantly different from 0.

Table 4.1: 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 $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$. We compute $\hat{\boldsymbol{X}}_t$ for all t but only consider those for which $\hat{\boldsymbol{X}}_t$ exceeds the 0.98 quantile. We then 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.

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.

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 because the Tukey's exact procedure is well-suited for all pairwise comparisons where the number of exceedances is equal across all pairwise compar-

isons. 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 (4.2). If $|t| < 5.847$, then we fail to reject the null hypothesis that $\rho_{ij}^E = 0$.

Table 4.2: 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	-

Let $\mathcal{G} = (V, E)$ be an undirected graphic with nodes $V = \{1, \ldots, 10\}$ and 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 4.4. Each circle indicates a node. Extreme discharges at nearby stations tend to be 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, which is a noteworthy reduction from the $\binom{10}{2} = 45$ possible edges.

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 in the appendix E.1. 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. Figure 4.5 shows that partial
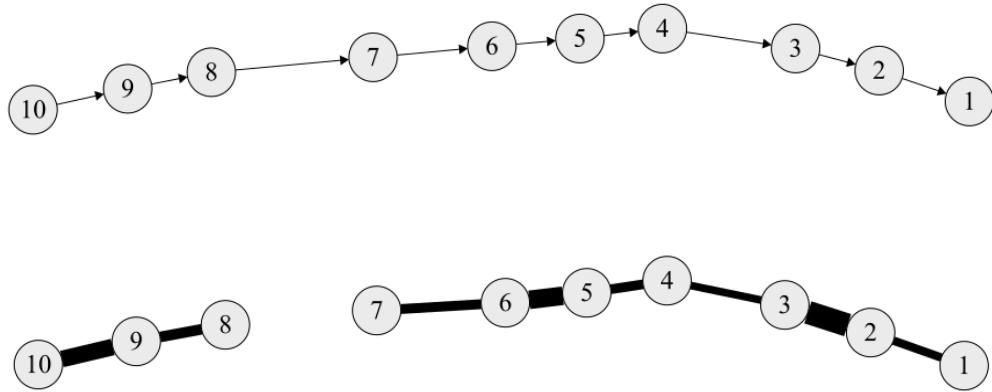
Figure 4.4: The known physical flow connections (above) versus the extremal graph induced by partial tail correlation (below) for the main path $10 \rightarrow \cdots \rightarrow 1$ in the upper Danube river basin. The thickness of edges corresponds to absolute values of test statistics being greater than 5.847.

tail correlation graph 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 4.4. We believe that including too many variables may introduce noise when estimating the TPDM.

4.7.2 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 2.1). We analyze 5163 daily NO_2 data between 1995 and 2016 where all five stations have measurements. We follow the same preprocessing process described in Section 3.4 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
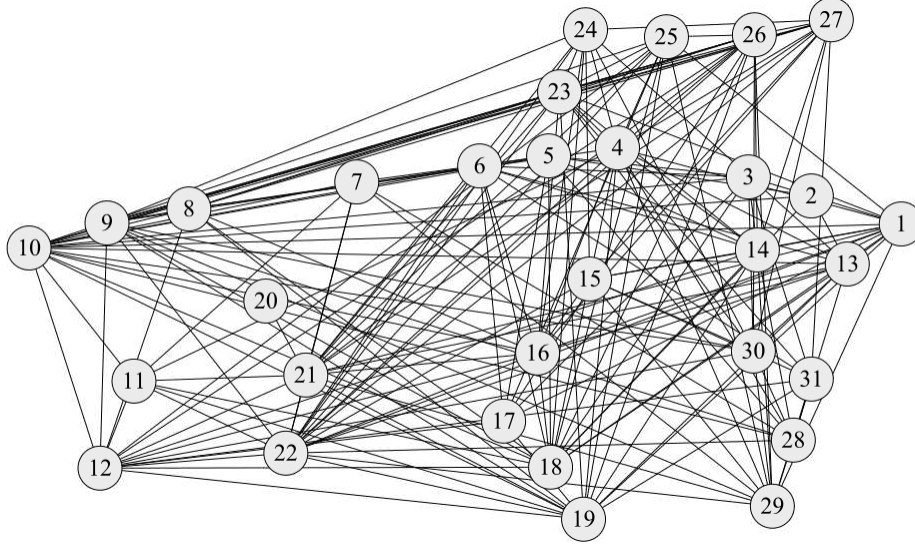
Figure 4.5: An undirected graphical model for all gauging stations in the upper Danube river basin.

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 $i = 1, \ldots 5$. We assume $\boldsymbol{X} = (X_1, \ldots, X_5)^T \in RV_+^5(2)$. 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, which we treat as iid copies of $\boldsymbol{X}$.

We aim to test whether or not extreme NO$_2$ levels between each pair of stations exhibit partial tail correlation. To perform a hypothesis test, we first estimate the TPDM $\hat{\Sigma}_{\boldsymbol{X}}$. Let $\boldsymbol{x}_t$ be 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 $\hat{\sigma}_{ij} = 2k^{-1} \sum_{i=1}^k w_{t,i} w_{t,j} \mathbb{I}[r_{t,ij} > r_{i,j}^*]$, 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. 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 $\hat{\Sigma}^{-1}$ given in Table 4.3. The goal is now to assess if each off-diagonal element is significantly different from 0.

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 $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$. We compute $\hat{\boldsymbol{X}}_t$ for all t but only consider those for which $\hat{\boldsymbol{X}}_t$ exceeds the 0.98 quantile. We then find residual vectors $\boldsymbol{U}_t = t^{-1}(\boldsymbol{X}_t^{(1)}) - t^{-1}(\hat{\boldsymbol{X}}_t) \in RV^2(2)$.

Table 4.3: 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 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 $\tilde{m}\sqrt{Var(W_i W_j)}$. We follow the Tukey's exact procedure to adjust for multiple 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 a matrix (Table 4.4). If $|t| < 4.797$, then we fail to reject the null hypothesis that $\rho^E_{ij} = 0$.

Table 4.4: 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	-

Let $\mathcal{G} = (V, E)$ be an undirected graphic with nodes $V = \{1, \ldots, 5\}$ and edge set E. We create an undirected graphical model for five stations by assuming partial tail correlation implies conditional relationships in Figure 4.6. Each circle indicates a node. 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.
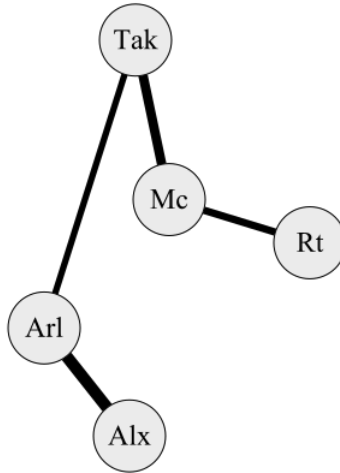
Figure 4.6: The extremal graph induced by partial tail tail correlation for five stations. The thickness of edges corresponds to absolute values of test statistics being greater than 4.797.

4.8   Summary and Discussion

Our inner product space is constructed from transformed linear combinations of independent regularly varying random variables. In contrast to max-linear models, our framework is well suited to linear models in traditional statistics. The projection theorem provides a natural way for defining partial tail correlation in extremes. By matrix inversion, zero elements in the inverse of the TPDM are naturally connected to the partial tail correlation so that we have the analogue of partial correlations for extremes.

Using the fact that the natural estimator for the conditional TPDM is asymptotically normal, we develop a hypothesis test for zero elements in the inverse extremal matrix. We note that since our framework does not assume densities, it does not lead to causality directly. However, partially uncorrelated variables in their extreme levels can provide a meaningful interpretation for extremal

relationships between variables in extremes. We observed that the extremal relationships between variables are sensibly described by the partial tail correlation in both river discharges application and an extreme AR(1) simulation. This method allows one to use the idea of the partial tail correlation to explore structural relationships between variables' extremes as an exploratory analysis.

Since we do not admit densities, finding suitable variable selection criteria or decomposible graphs would be important for future research. Considering directions between variables in their tail behaviors would also be important. For the main path $10 \rightarrow \cdots \rightarrow 1$, in the upper Danube basin, $|t| > 5.847$ between (X_1, X_3) and (X_1, X_4) say (X_1, X_3, X_4) are related to each other. However, their opposite signs may imply (X_1, X_3) and (X_1, X_4) are in fact not partially correlated in terms of tail behaviors and this apparent relationship may come up because of the confounding effect.

Bibliography

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

A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of probability*, 2(5): 792–804, 1974.

J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications*, volume 558. John Wiley & Sons, 2004.

M.-O. Boldi and A. C. Davison. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229, 2007.

P. J. Brockwell, R. A. Davis, and S. E. Fienberg. *Time series: theory and methods: theory and methods*. Springer Science & Business Media, 1991.

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

D. B. Cline and P. J. Brockwell. Linear prediction of arma processes with infinite variance. *Stochastic Processes and their applications*, 19(2):281–296, 1985.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analysis. *Extremes*, 2:339–365, 1999.

S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392, 1991.

S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: an application to structural design. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):1–31, 1994.

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

D. Cooley, R. A. Davis, and P. Naveau. The pairwise beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, 101(9):2103–2117, 2010.

D. Cooley, R. A. Davis, and P. Naveau. Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. *The Annals of Applied Statistics*, 6(4):1406–1429, 2012.

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

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

W. Feller. An introduction to probability theory and its applications. 1957.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

A.-L. Fougères, 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.

B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.

P. Groetzner and M. Dür. A factorization method for completely positive matrices. *Linear Algebra and its Applications*, 591:1–24, 2020.

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

H. A. Jessen and T. Mikosch. Regularly varying functions. *Publications de L'institut Mathematique*, 80(94):171–192, 2006.

H. Joe, R. L. Smith, and I. Weissman. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1):171–183, 1992.

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

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

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

J. S. Marron and D. Ruppert. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(4):653–671, 1994.

N. Mhatre and D. Cooley. Transformed-linear models for time series extremes, 2021.

A. A. Mohsenipour. On the distribution of quadratic expressions in various types of random vectors. 2012.

J. Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

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

S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, 1987.

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

H. Rootzén and N. Tajvidi. Multivariate generalized pareto distributions. *Bernoulli*, 12(5):917–930, 2006.

H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate generalized pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131, 2018.

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

F. Röttger, S. Engelke, and P. Zwiernik. Total positivity in multivariate extremes, 2021. URL https://arxiv.org/abs/2112.14727.

C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical journal*, 10(1):33–60, 2012.

M. Schlather and J. Tawn. Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes*, 5(1):87–102, 2002.

N. D. Shyamalkumar and S. Tao. On tail dependence matrices. *Extremes*, pages 1–41, 2020.

M. Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

K. Strokorb and M. Schlather. An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli*, pages 276–302, 2015.

Appendix A

Vector Space

We show that the $\mathcal{V}_q$ is the vector space in a rigorous way.

$$\mathcal{V}^q = \{X; X = a_1 \circ Z_1 \oplus \cdots \oplus a_q \circ Z_q\},$$

where Z_j are independent and identically distributed regularly varying random variables with the unit scale and meeting the lower tail condition. Let $X_i := t(\sum_{j=1}^q a_{ij} t^{-1}(Z_j)) \in \mathcal{V}^q$, for $i = 1, 2, \ldots$,

1. Closure under addition: For $X_1, X_2 \in \mathcal{V}^q$, $X_1 \oplus X_2 = t\left(\sum_{j=1}^q (a_{1j} + a_{2j}) t^{-1}(Z_j)\right) \in \mathcal{V}^q$ as the sum of coefficients is finite.

2. Closure under scalar multiplication: For any $b \in \mathbb{R}$, $X \in \mathcal{V}^q$, $b \circ X = t(bt^{-1}(X)) = t\left(\sum_j^q ba_j t^{-1}(Z_j)\right) \in \mathcal{V}^q$.

3. Commutative property: For $X_1, X_2 \in \mathcal{V}^q$, $X_1 \oplus X_2 = t\left(\sum_{j=1}^q (a_{1q} + a_{2q}) t^{-1}(Z_q)\right) = t\left(\sum_{j=1}^q (a_{2j} + a_{1j}) t^{-1}(Z_j)\right) = X_2 \oplus X_1$.

4. Associative property: For $X_1, X_2, X_3 \in \mathcal{V}^q$, $(X_1 \oplus X_2) \oplus X_3 = \left\{ t\left(\sum_{j=1}^q (a_{1j} + a_{2j}) t^{-1}(Z_j)\right) \right\}$
   $\oplus X_3 = t\left(\sum_{j=1}^q (a_{1j} + a_{2j}) t^{-1}(Z_j) + \sum_{j=1}^q a_{3j} t^{-1}(Z_j)\right) = t\left(\sum_{q=1}^q (a_{1q} + a_{2q} + a_{3q}) t^{-1}(Z_q)\right) = t\left(\sum_{j=1}^q a_{1j} t^{-1}(Z_j) + \sum_{j=1}^q (a_{2j} + a_{3j}) t^{-1}(Z_j)\right) = X_1 \oplus (X_2 \oplus X_3)$.

5. Associative property: For $\alpha, \beta \in \mathbb{R}$, $X \in \mathcal{V}^q$, $(\alpha\beta) \circ X = t\left((\alpha\beta) \sum_{j=1}^q a_j t^{-1}(X_j)\right) = t\left(\alpha \sum_{j=1}^q (\beta a_j) t^{-1}(X_j)\right) = \alpha \circ (\beta X)$.

6. Distributive properties: For $\alpha \in \mathbb{R}$ and $X_1, X_2 \in \mathcal{V}^q$, $\alpha \circ (X_1 \oplus X_2) = \alpha \circ (t(\sum_{j=1}^q (a_{1j} + a_{2j}) t^{-1}(Z_j)) = t(\alpha \sum_{j=1}^q (a_{1j} + a_{2j}) t^{-1}(Z_j)) = t(\sum_{j=1}^q \alpha a_{1q} t^{-1}(Z_j) + \sum_{j=1}^q \alpha a_{2q} t^{-1}(Z_j)) = \alpha \circ X_1 \oplus \alpha \circ X_2$.

7. Additive identity: For $X \in \mathcal{V}^q$, let $\mathbf{0} := 0 \circ Z_1 \oplus \cdots \oplus 0 \circ Z_n = t(\mathbf{0})$. Then, $X \oplus \mathbf{0} = t(\sum_{j=1}^{q} a_j t^{-1}(Z_j) + \sum_{j=1}^{q} 0 \cdot t^{-1}(Z_j)) = t\left(\sum_{j=1}^{q} (a_j + 0) t^{-1}(Z_j)\right) = X$.

8. Additive inverse: For $X \in \mathcal{V}^q$, let $-X := t(-\sum_{j=1}^{q} a_j t^{-1}(Z_j))$. Then, $X \circ (-X) = t(\sum_{j=1}^{q} a_j t^{-1}(Z_j) - \sum_{j=1}^{q} a_j t^{-1}(Z_j)) = t(\mathbf{0}) = \mathbf{0}$.

9. Multiplicative identity: For $X \in \mathcal{V}^q$, $1 \circ X = t(1 \cdot \sum_{j=1}^{q} a_j t^{-1}(Z_j)) = t(\sum_{j=1}^{q} a_j t^{-1}(Z_j)) = X$.

Appendix B

Tail Metric

Let $\hat{X}_{p+1} = \boldsymbol{b}^T \circ \boldsymbol{X}_p$ where $X_p = A \circ \boldsymbol{Z}$ be the best transformed-linear predictor and $X_{p+1} \in \mathcal{V}^q$ be predictand, where $c_{1j} = [b^T A_p]_j$ is the jth element of $b^T A_p$ and $c_{2j} = [\boldsymbol{a}_{p+1}^T]_j$ is the jth element of $\boldsymbol{a}_{p+1}^T$ for $j = 1, \cdots, q$. We show that the tail metric of $\hat{X}_{p+1}$ and X_{p+1} can be expressed in terms of tail probabilities. Let $Y_1 := t^{-1}(\hat{X}_{p+1})$ and $Y_2 := t^{-1}(X_{p+1})$ be its preimage. The following proof is written as the modified version in the context of our framework from Cline and Brockwell [1985] and Feller [1957]. We start with the tail probabilities of $|t^{-1}(\hat{X}_{p+1}) - t^{-1}(X_{p+1})|$. If the case of $q = 2$ is true, then the general case can be extended by induction. For any $\delta > 0$, let $d_1 = (c_{11} - c_{21})$ and $d_2 = (c_{21} - c_{22})$.

$$
\begin{aligned}
d^2(t^{-1}(\hat{X}_{p+1}), t^{-1}(X_{p+1})) &= \lim_{z \to \infty} \frac{P(|t^{-1}(\hat{X}_2 \ominus X_2)| > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&= \lim_{z \to \infty} \frac{P(|(c_{11} - c_{21})t^{-1}(Z_1) + (c_{12} - c_{22})t^{-1}(Z_2)| > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&\geq \lim_{z \to \infty} \frac{P(|d_1 t^{-1}(Z_1)| > (1+\delta)t^{-1}(z), |d_2 t^{-1}(Z_2)| \leq \delta t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&+ \lim_{z \to \infty} \frac{P(|d_1 t^{-1}(Z_1)| \leq \delta t^{-1}(z)), |d_2 t^{-1}(Z_2)| > (1+\delta)t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&= \lim_{z \to \infty} \frac{P(|t^{-1}(Z_1)| > \frac{(1+\delta)}{|d_1|}t^{-1}(z), |t^{-1}(Z_2)| \leq \frac{\delta}{|d_2|}t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&+ \lim_{z \to \infty} \frac{P(|t^{-1}(Z_1)| \leq \frac{\delta}{|d_1|}t^{-1}(z), |t^{-1}(Z_2)| > \frac{(1+\delta)}{|d_2|}t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&= (1+\delta)^{-\alpha}|d_1|^\alpha + (1+\delta)^{-\alpha}|d_2|^\alpha,
\end{aligned}
$$

$$d^2(t^{-1}(\hat{X}_{p+1}), t^{-1}(X_{p+1})) = \lim_{z \to \infty} \frac{P(|t^{-1}(\hat{X}_{p+1} \ominus X_{p+1})| > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$\leq \lim_{z \to \infty} \frac{P(|d_1 t^{-1}(Z_1)| + |d_2 t^{-1}(Z_2)| > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$= \lim_{z \to \infty} \frac{P(|t^{-1}(Z_1)| > t^{-1}(z)/|d_1|)}{P(t^{-1}(Z) > t^{-1}(z))} + \lim_{z \to \infty} \frac{P(|t^{-1}(Z_2)| > t^{-1}(z)/|d_2|)}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$= |d_1|^\alpha + |d_2|^\alpha.$$

Thus, $d^2(t^{-1}(\hat{X}_{p+1}), t^{-1}(X_{p+1})) = \sum_{i=1}^2 |d_i|^\alpha$ for arbitrary δ. By induction,

$$d^2(\hat{X}_{p+1}, X_{p+1}) = \lim_{z \to \infty} \frac{P(|t^{-1}(\hat{X}_{p+1} \ominus X_{p+1})| > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$= \lim_{z \to \infty} \frac{P(|\boldsymbol{b}^T \boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T| \sum_{j=1}^q t^{-1}(Z_j) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$= \lim_{z \to \infty} \frac{P(|\boldsymbol{b}^T \boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T| \sum_{j=1}^q t^{-1}(Z_j) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))}$$

$$= (\boldsymbol{b}^T \boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T)(\boldsymbol{b}^T \boldsymbol{A}_p - \boldsymbol{a}_{p+1})^T$$

$$= \sum_{j=1}^q (c_{1j} - c_{2j})^2$$

Now the tail metric of $D = max(\hat{X}_{p+1} \ominus X_{p+1}, X_{p+1} \ominus \hat{X}_{p+1})$ is

$$
\begin{aligned}
d^2(D) &= \lim_{z\to\infty} \frac{P(D > z)}{P(Z > z)} \\
&= \lim_{z\to\infty} \frac{P(max(\hat{X}_{p+1} \ominus X_{p+1}, X_{p+1} \ominus \hat{X}_{p+1}) > z)}{P(Z > z)} \\
&= \lim_{z\to\infty} \frac{P(\hat{X}_{p+1} \ominus X_{p+1}) > z)}{P(Z > z)} + \lim_{z\to\infty} \frac{P(X_{p+1} \ominus \hat{X}_{p+1}) > z)}{P(Z > z)} \\
&\left(\text{since } \lim_{z\to\infty} \frac{P(\hat{X}_{p+1} \ominus X_{p+1}, X_{p+1} \ominus \hat{X}_{p+1} > z)}{P(Z > z)} = 0 \right) \\
&= \lim_{z\to\infty} \frac{P(t^{-1}(\hat{X}_{p+1} \ominus X_{p+1}) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z)} + \lim_{z\to\infty} \frac{P(t^{-1}(X_{p+1} \ominus \hat{X}_{p+1}) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&= \lim_{z\to\infty} \frac{P((\boldsymbol{b}^T\boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T)\sum_{j=1}^q t^{-1}(Z_j) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&\quad + \frac{P((\boldsymbol{a}_{p+1}^T - \boldsymbol{b}^T\boldsymbol{A}_p)\sum_{j=1}^q t^{-1}(Z_j) > t^{-1}(z))}{P(t^{-1}(Z) > t^{-1}(z))} \\
&= (\boldsymbol{b}^T\boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T)^{(0)}(\boldsymbol{b}^T\boldsymbol{A}_p - \boldsymbol{a}_{p+1}^T)^{(0)^T} + (\boldsymbol{a}_{p+1}^T - \boldsymbol{b}^T\boldsymbol{A}_p)^{(0)}(\boldsymbol{a}_{p+1}^T - \boldsymbol{b}^T\boldsymbol{A}_p)^{(0)^T} \\
&= \sum_{j=1}^q \left[(c_{1j} - c_{2j})^{(0)}\right]^2 + \sum_{j=1}^q \left[(c_{2j} - c_{1j})^{(0)}\right]^2 = \sum_{j=1}^q (c_{1j} - c_{2j})^2
\end{aligned}
$$

Appendix C

$\mathcal{V}^q$ is isomorphic to $\mathbb{R}^q$ and Complete

Theorem C.0.1. *Let $\mathcal{V}^q$ be a vector space defined in (2.9). For any $X = a_1 \circ Z_1 \oplus \cdots a_q \circ Z_q \in \mathcal{V}^q$, let $\Psi : \mathcal{V}^q \to \mathbb{R}^q$ be defined by $\Psi(X) = \boldsymbol{a}$. Then, $\mathcal{V}^q$ is isomorphic to $\mathbb{R}^q$.*

Proof. We need to show Ψ is a linear transformation, one to one, and onto. For scalars $c_1, c_2 \in \mathbb{R}$ and variables $X_1, X_2 \in \mathcal{V}^q$, let $X_1 = a_{11} \circ Z_1 \oplus \cdots a_{1q} \circ Z_q$ and $X_2 = a_{21} \circ Z_1 \oplus \cdots a_{2q} \circ Z_q$.

$$
\begin{aligned}
\Psi(c_1 \circ X_1 \oplus c_2 \circ X_2) &= \Psi\big(t(c_1 a_{11} t^{-1}(Z_1) + \cdots + c_1 a_{1q} t^{-1}(Z_q) + c_2 a_{21} t^{-1}(Z_1) + \cdots \\
&\qquad + c_2 a_{2q} t^{-1}(Z_q))\big) \\
&= \Psi\big(t((c_1 a_{11} + c_2 a_{21}) t^{-1}(Z_1) + \cdots + (c_1 a_{1q} + c_2 a_{2q}) t^{-1}(Z_q))\big) \\
&= \big((c_1 a_{11} + c_2 a_{21}), \cdots, (c_1 a_{1q} + c_2 a_{2q})\big)^T \\
&= c_1 (a_{11}, \cdots, a_{1q})^T + c_2 (a_{21}, \cdots, a_{2q})^T \\
&= c_1 \Psi(X_1) + c_2 \Psi(X_2).
\end{aligned}
$$

Thus, Ψ is linear. To show that Ψ is one to one, let $X_1, X_2 \in \mathcal{V}^q$. Assume $\Psi(X_1) = \Psi(X_2)$. Then, $(a_{11}, \ldots, a_{1q})^T = (a_{21}, \ldots, a_{2q})^T$ and $a_{1j} = a_{2j}$ for $j = 1, \ldots, q$. That is, $X_1 = X_2$ and hence Ψ is one to one. To show that Ψ is onto, for $\boldsymbol{c} = (c_1, \ldots, c_q)^T \in \mathbb{R}^q$, we can find $X \in \mathcal{V}^q$ such that $X = c_1 \circ Z_q \oplus \cdots, \oplus c_q \circ Z_q$. That is, $\Psi(X) = \boldsymbol{c}$. Thus, Ψ is onto. $\square$

To prove that $\mathcal{V}^q$ is complete, we need to show every Cauchy sequence $\{X_n\}$ converges in norm to $X \in \mathcal{V}^q$. As the metric defined in Section 2.3 is based on the coefficients, we can follow the proof in classical statistics [Brockwell et al., 1991].

Theorem C.0.2. *$\mathcal{V}^q$ defined in (2.9) is complete.*

Proof. For $X_n = \boldsymbol{a}_n^T \circ \boldsymbol{Z} \in \mathcal{V}^q$, if X_n is such that

$$\|X_n \ominus X_m\|_{\mathcal{V}^q}^2 = \sum_{j=1}^{q} |a_{n,j} - a_{m,j}|^2 \to 0$$

as $m, n \to \infty$, then each component $j = 1, \ldots, q$, must satisfy $|a_{n,j} - a_{m,j}| \to 0$ as $m, n \to \infty$. There exists $a_j \in \mathbb{R}$ such that $|a_{n,j} - a_j| \to 0$ as $n \to \infty$ by the completeness of $\mathbb{R}$. Hence, for $X \in \mathcal{V}^q$,

$$\|X_n \ominus X\|_{\mathcal{V}^q} \to 0$$

as $n \to \infty$. Thus, $\mathcal{V}^q$ is complete. $\qquad\square$

Appendix D

Transformation of data method

The transformation of data method is one of the simple ways to address boundary effects in kernel density estimation [Marron and Ruppert, 1994]. The idea is to map the bounded support of data to the real line by a suitable transformation and to fit the regular kernel density to the transformed data. Our goal is to estimate the angular density $h(w)$ evaluated at the bounded support, $w \in [0, 1]$. Let $H(w) = F(t(w))$, $w \in [0, 1]$ be the distribution of the transformed data. For the support of $w \in [0, 1]$, the Probit transformation $t(w) = \Phi^{-1}(w)$ is one option where $\Phi(w)$ is the standard normal distribution mapping $w \in [0, 1]$ to $\mathbb{R} = [-\infty, \infty]$. By taking derivatives with respect to w, $h(w) = f(t(w))t'(w)$, where $t'(w) = \frac{\mathrm{d}}{\mathrm{d}w}\Phi^{-1}(w) = \phi^{-1}(\Phi^{-1}(w))$, the adjusted angular density evaluated at w is $g(w) = f(\Phi^{-1}(w))\frac{1}{\phi(\Phi^{-1}(w))}$, where $\Phi^{-1}(w)$ is the inverse CDF of the standard normal distribution.

Appendix E

Test statistics table on the whole river network

Table E.1: Test statistics for all pairs of gauging stations $i \neq j$ for $i, j = 1, \ldots, 31$ in the upper Danube river basin.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	-	11.29	-0.36	7.49	-0.45	5.65	-5.47	0.84	-6.73	-30.75	4.67	-5.60	14.29	-17.88	9.52	-8.33	-14.30	70.42	-13.66	-5.44	-18.51	-15.44	11.08	-3.74	-8.29	4.17	-2.19	-5.67	9.86	-0.85	0.17
2	11.29	-	22.27	4.77	-170.88	6.48	0.53	2.45	-1.22	-16.62	-0.23	-4.43	-9.07	4.61	0.21	1.41	2.80	-14.00	0.38	-1.09	-1.67	3.34	2.51	-3.21	-6.36	4.78	0.38	2.10	-2.50	6.86	-6.20
3	-0.36	22.27	-	-5.57	-16.68	-7.84	-0.86	-0.53	-1.81	269.06	1.56	0.46	1.59	-8.21	0.21	3.17	-11.90	33.31	-2.45	-1.35	4.55	-5.27	1.62	-0.19	-9.13	-1.18	-3.98	0.70	-1.66	-17.06	3.15
4	7.49	4.77	-5.57	-	5.88	-5.37	2.69	12.34	6.16	1.93	-8.63	3.37	-15.42	-7.30	10.09	17.40	-5.45	6.16	-7.43	-3.83	-0.93	7.39	23.10	-16.52	-2.09	15.84	-28.06	-40.67	-8.09	-6.51	18.94
5	-0.45	-170.88	-16.68	5.88	-	18.59	-7.28	5.71	-18.59	-35.27	1.40	-5.26	3.21	1	-12.55	1.33	-5.76	-62.18	1.83	0.46	-3.52	4.70	-1.49	0.57	-5.67	2.61	-2.83	-8.39	6.66	-3.06	9.56
6	5.65	6.48	-7.84	-5.37	18.59	-	5.71	1.83	31.12	-6.24	6.59	6.99	5.36	0.59	5.60	6.37	0.52	8.05	4.16	0.85	-7.81	11.05	2.16	-6.43	4.92	4.51	0.52	8.35	-3.06	-5.25	1.24
7	-5.47	0.53	-0.86	2.69	-7.28	5.71	-	9.39	-17.61	-5.13	-8.76	-6.41	0.78	-2.68	-3.14	0.52	-7.81	6.37	-6.43	0.52	4.92	-1.41	-5.95	-1.52	-14.81	-14.35	3.14	-3.42	-4.34	-5.25	-0.01
8	0.84	2.45	-0.53	12.34	5.71	1.83	9.39	-	12.34	15.77	3.52	-6.41	15.77	0.78	-2.68	2.69	31.12	0.53	-5.95	0.53	-3.14	-5.95	-10.27	-14.81	15.77	11.96	-3.48	-8.70	-9.38	-1.10	-0.01
9	-6.73	-1.22	-1.81	6.16	-18.59	31.12	-17.61	12.34	-	-6.24	22.12	20.09	3.63	-29.10	-2.17	0.45	7.90	364.86	3.12	-3.61	-11.93	-58.14	15.31	4.33	-264.74	26.93	-18.48	10.81	11.29	-13.34	-5.04
10	-30.75	-16.62	269.06	1.93	-35.27	-6.24	-5.13	15.77	-6.24	-	20.09	-29.10	-4.40	-40.94	-28.04	5.64	5.78	126.49	5.44	2.64	21.74	16.03	-7.16	-18.48	-49.09	21.70	-6.74	10.81	-49.09	-14.68	-17.30
11	4.67	-0.23	1.56	-8.63	1.40	6.59	-8.76	3.52	22.12	20.09	-	15.29	18.69	-3.31	3.45	8.48	-4.20	16.49	-2.59	2.69	21.74	11.78	1.75	5.26	8.91	3.76	3.14	5.83	2.22	-0.32	-8.86
12	-5.60	-4.43	0.46	3.37	-5.26	6.99	-6.41	-6.41	20.09	-29.10	15.29	-	-4.40	18.69	5.43	1.10	4.95	18.90	-5.03	2.75	17.41	16.53	-10.27	-4.03	-7.09	-9.37	0.52	-3.42	-4.51	6.14	-14.68
13	14.29	-9.07	1.59	-15.42	3.21	5.36	-6.41	15.77	3.63	-4.40	18.69	-4.40	-	5.43	1.10	8.48	1.84	12.22	2.77	3.74	-9.40	18.90	6.87	3.86	8.47	-5.76	8.47	-8.70	8.73	7.80	-17.30
14	-17.88	4.61	-8.21	-7.30	1	0.59	0.78	0.78	-29.10	-40.94	-3.31	18.69	5.43	-	17.49	-4.36	-7.48	12.22	3.74	1.60	-7.48	12.22	6.87	4.49	3.58	-11.32	-9.38	-4.34	6.14	-9.09	4.62
15	9.52	0.21	0.21	10.09	-12.55	5.60	-2.68	-2.68	-2.17	-28.04	3.45	5.43	1.10	17.49	-	-18.10	-1.38	8.41	2.88	9.94	8.41	-2.05	-9.21	3.58	1.43	4.71	8.47	4.57	8.73	2.95	7.88
16	-8.33	1.41	3.17	17.40	1.33	6.37	-3.14	2.69	0.45	5.64	8.48	1.10	8.48	-4.36	-18.10	-	-1.38	1.84	-1.71	-2.05	4.95	1.60	-8.44	-18.10	18.10	-11.32	0.52	-8.70	8.47	4.90	4.90
17	-14.30	2.80	-11.90	-5.45	-5.76	0.52	-7.81	31.12	7.90	5.78	-4.20	4.95	1.84	-7.48	-1.38	-1.38	-	8.41	2.64	-1.33	-1.55	17.41	1.75	-9.21	-0.88	4.71	3.14	-9.38	-1.54	-1.54	4.27
18	70.42	-14.00	33.31	6.16	-62.18	8.05	6.37	0.53	364.86	126.49	16.49	18.90	12.22	12.22	8.41	1.84	8.41	-	-3.51	2.75	-1.55	-1.46	6.63	0.22	2.64	16.33	6.74	2.80	8.73	2.95	3.51
19	-13.66	0.38	-2.45	-7.43	1.83	4.16	-6.43	-5.95	3.12	5.44	-2.59	-5.03	2.77	3.74	2.88	-1.71	2.64	-3.51	-	-3.85	-14.68	16.16	-13.62	4.42	-1.19	-4.95	10.06	-1.37	-8.04	4.06	2.95
20	-5.44	-1.09	-1.35	-3.83	0.46	0.85	0.52	0.53	-3.61	2.64	2.69	2.75	3.74	1.60	9.94	-2.05	-1.33	2.75	-3.85	-	7.69	22.26	-1.19	-0.08	14.68	1.51	1.97	-4.32	-4.95	1.02	4.27
21	-18.51	-1.67	4.55	-0.93	-3.52	-7.81	4.92	-3.14	-11.93	21.74	21.74	17.41	-9.40	-7.48	8.41	4.95	-1.55	-1.55	-14.68	7.69	-	2.49	10.29	-11.16	19.19	-9.44	-9.68	-4.32	-4.09	1.02	-1.08
22	-15.44	3.34	-5.27	7.39	4.70	11.05	-1.41	-5.95	-58.14	16.03	11.78	16.53	18.90	12.22	-2.05	1.60	17.41	-1.46	16.16	22.26	2.49	-	141.04	4.42	21.94	12.90	18.73	8.63	-4.39	16.33	-6.22
23	11.08	2.51	1.62	23.10	-1.49	2.16	-5.95	-10.27	15.31	-7.16	1.75	-10.27	6.87	6.87	-9.21	-8.44	1.75	6.63	-13.62	-1.19	10.29	141.04	-	-141.04	-10.07	16.33	-14.93	2.80	6.26	10.75	7.88
24	-3.74	-3.21	-0.19	-16.52	0.57	-6.43	-1.52	-14.81	4.33	-18.48	5.26	-4.03	3.86	4.49	3.58	-18.10	-9.21	0.22	4.42	-0.08	-11.16	4.42	-141.04	-	3.55	-36.26	10.06	-1.02	24.86	2.48	-6.22
25	-8.29	-6.36	-9.13	-2.09	-5.67	4.92	-14.81	15.77	-264.74	-49.09	8.91	-7.09	8.91	1.43	1.43	18.10	-0.88	2.64	-1.19	14.68	19.19	21.94	-10.07	3.55	-	26.47	-3.82	-1.02	-7.31	2.48	1.28
26	4.17	4.78	-1.18	15.84	2.61	4.51	-14.35	11.96	26.93	21.70	3.76	-9.37	-5.76	-11.32	4.71	-11.32	4.71	16.33	-4.95	1.51	-9.44	12.90	16.33	-36.26	26.47	-	8.00	6.74	10.91	-21.99	5.50
27	-2.19	0.38	-3.98	-28.06	-2.83	0.52	3.14	-3.48	-18.48	-6.74	3.14	0.52	8.47	-9.38	3.14	0.52	3.14	6.74	10.06	1.97	-9.68	18.73	-14.93	10.06	-3.82	8.00	-	6.74	-0.32	10.75	-8.86
28	-5.67	2.10	0.70	-40.67	-8.39	8.35	-3.42	-8.70	10.81	10.81	5.83	-3.42	-8.70	-4.34	4.57	-8.70	-9.38	2.80	-1.37	-4.32	8.63	8.63	2.80	-1.02	-7.31	6.74	6.74	-	-34.63	36.02	-14.68
29	9.86	-2.50	-1.66	-8.09	6.66	-3.06	-4.34	-9.38	11.29	-49.09	2.22	-4.51	8.73	6.14	8.73	8.47	-1.54	8.73	-8.04	-4.09	-4.09	-4.39	6.26	24.86	-7.31	10.91	-0.32	-34.63	-	-21.19	26.97
30	-0.85	6.86	-17.06	-6.51	-3.06	-5.25	-5.25	-1.10	-13.34	-14.68	-0.32	6.14	7.80	-9.09	2.95	4.90	-1.54	2.95	4.06	1.02	1.02	16.33	10.75	2.48	2.48	-21.99	10.75	36.02	-21.19	-	-1.05
31	0.17	-6.20	3.15	18.94	9.56	1.24	-0.01	-0.01	-5.04	-17.30	-8.86	-14.68	-17.30	4.62	7.88	4.90	4.27	3.51	2.95	4.27	-1.08	-6.22	7.88	-6.22	1.28	5.50	-8.86	-14.68	26.97	-1.05	-

88