Partial Tail Correlation Coefficient Applied to Extremal Network Learning

Yan Gong, Peng Zhong, Thomas Opitz, Raphaël Huser

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

ext stat Extreme Statistics Research Group

Extremal networks: Literature review

In the context of modeling extremal dependence using graphs:

- Huang et al. (2019) provide exploratory tools, such as the χ-network for extremal dependence modeling, used in their analysis of the maximum precipitation during the hurricane season at the US Gulf Coast and surrounding areas;
- Engelke and Hitz (2020) introduce a new notion of conditional independence adapted to multivariate Pareto distributions, and propose parametric graphical models for extremes;
- Gissibl (2018) and Klüppelberg and Krali (2021) propose max-linear model-based frameworks;

Extremal networks: Literature review

- Tran et al. (2021) propose QTree, a simple and efficient algorithm to solve the Latent River Problem;
- For the important case of tree models, Engelke and Volgushev (2020) develop a data-driven methodology for learning the graphical structure;
- Röttger et al. (2021) propose Hüsler–Reiss graphical models under multivariate total positivity of order 2 (MTP_2);
- Engelke and Ivanovs (2021) review the recent developments of graphical models for extremes.

Existing graphical models for extremes rely on asymptotically justified models, sometimes yielding dimensionality issues and/or the lack of general graph structures.

Our goal

Complementing these approaches by

- defining the new concept of "partial tail correlation" (by analogy to the usual partial correlation);
- defining a new coefficient that can be used to estimate general extremal networks under minimal modeling assumptions.

Although "partial tail uncorrelatedness" is a weaker assumption than "conditional tail independence", it can still

- provide interesting insights into extremal dependence structures,
- guide modeling choices in exploratory data analyses.

Regular variation framework and transformed linear algebra

To introduce the partial tail correlation coefficient (PTCC), we first briefly review the multivariate regular variation framework and the foundations of transformed-linear algebra (Cooley and Thibaud, 2019), which

- is within the framework of regular variation;
- uses a transformation to define a vector space on the positive orthant;
- employs transformed-linear operations applied to regularly-varying random vectors that preserve regular variation;
- summarizes tail dependence via a matrix of pairwise tail dependence metrics.

Inner product space construction

Let t be a bijection (transformation) from $\mathbb{R}$ onto some open set $\mathbb{X}$, and t^{-1} be its inverse. For a p-dimensional vector $\boldsymbol{y} \in \mathbb{R}^p$, $\boldsymbol{x} = t(\boldsymbol{y}) \in \mathbb{X}^p$ is defined componentwise.

Define

- vector addition: $x_1 \oplus x_2 = t\{t^{-1}(x_1) + t^{-1}(x_2)\}$,
- scalar multiplication: $a \circ x = t\{a t^{-1}(x)\}$, for $a \in \mathbb{R}$,
- additive identity: $0_{\mathbb{X}^p} = t(0)$,
- additive inverse: $x_1 \ominus x_2 = t\{t^{-1}(x_1) - t^{-1}(x_2)\}$, for any $x_1, x_2 \in \mathbb{X}^p$.

A particular inner product space on the positive orthant whose operations have a negligible effect on large values can be defined using a specific transformation $t : \mathbb{R} \mapsto (0, \infty)$, given by

$$t(y) = \log\{1 + \exp(y)\},$$

which may be used to perform transformed-linear operations.

Inner product space construction

Using the transformation, regularly varying random vectors can be constructed on the positive orthant. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_q)^T$ be a vector of i.i.d. regularly varying random variables with tail index α, i.e., there exists a sequence $\{b_n\}$ such that, as $n \to \infty$,

$$n\Pr(Z_j > b_n z) \to z^{-\alpha}, \quad n\Pr\{Z_j \leq \exp(-kb_n)\} \to 0,$$

for any $k > 0$ and for all j.

Construct p-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ as

$$\boldsymbol{X} = \bigoplus_{j=1}^{q} \boldsymbol{a}_j \circ Z_j.$$

Let $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q) \in \mathbb{R}_+^{p \times q}$, where $\boldsymbol{a}_j \in \mathbb{R}_+^p$, then, we have $\boldsymbol{X} = A \circ \boldsymbol{Z} \in \mathrm{RV}_+^p(\alpha)$.

Inner product space construction

This construction guarantees the regular variation preservation, according to Cooley and Thibaud (2019), Corollary 1. Furthermore, we require A to have full row-rank.

Following Lee and Cooley (2021), we further define the inner product of $< X_i, X_k >$ as follows:

$$< X_i, X_k > = \boldsymbol{a}_i^T \boldsymbol{a}_k = \sum_{j=1}^{q} a_{ij} a_{kj},$$

and the norm is $||X_i|| = \sqrt{< X_i, X_i >}$. The metric induced by the inner product is

$$d(X_i, X_k) = ||X_i \ominus X_k|| = \Big[\sum_{j=1}^{q} (a_{ij} - a_{kj})^2 \Big]^{1/2}, \text{ for } i, k = 1, \dots, p.$$

Tail pairwise dependence matrix (TPDM)

The TPDM is inspired from statistical practice in nonextreme settings, where the covariance matrix is widely used to summarize dependence information in multivariate distributions, sometimes even when they are non-Gaussian.

Choose $\alpha = 2$, and employ *L_2-norm* when making radial/angular transformation. Then the TPDM is defined as

$$\sigma_{\boldsymbol{X}_{ik}} := \int_{\Theta_{p-1}^{+}} w_i w_k dH_{\boldsymbol{X}}(\boldsymbol{w}), \quad \Sigma_{\boldsymbol{X}} = (\sigma_{ik})_{i,k=1,\dots,p},$$

where $H_{\boldsymbol{X}}$ is a Radon measure on the $\Theta_{p-1}^{+} = \{\boldsymbol{w} \in \mathbb{R}_{+}^{p} : ||\boldsymbol{w}||_2 = 1\}$, which gives useful but incomplete dependence information.

TPDM properties

- Diagonals describe scale (variance)

$$\lim_{n\to\infty} n\Pr\left(\frac{X_i}{\sqrt{n}} > x\right) = x^{-2}\sigma_{ii}.$$

- The sum of the diagonal elements is equal to the total mass of the angular measure.
- Asymptotic independence: $\sigma_{ik} = 0$ (Sibuya, 1960; Ledford and Tawn, 1996).
- $\Sigma_{\boldsymbol{X}}$ is positive semi-definite and completely positive,
  i.e., there exists a finite $p \times q$ matrix A with non-negative entries $\Sigma_{\boldsymbol{X}} = AA^T$.

We further request $\Sigma_{\boldsymbol{X}}$ to be positive definite which guarantees the existence of the inverse.

Definition of the PTCC

Remove X_i and X_k from the random vector $\boldsymbol{X}$, and the remaining $(p-2)$-dimensional vector is denoted as $\boldsymbol{X}_{\text{-}ik}$. $A_{\text{-}ik}$ denotes the matrix without A's i-th and k-th columns.

Write vectors in the form $\boldsymbol{X}_{ik} = (X_i, X_k)^T$ and re-order the columns of $\boldsymbol{X}$ as

$$\boldsymbol{X}' = (\boldsymbol{X}_{ik}^T, \boldsymbol{X}_{\text{-}ik}^T)^T = (A_{ik}, A_{\text{-}ik}) \circ \boldsymbol{Z}.$$

The best linear unbiased predictor of $\boldsymbol{X}_{ik}$ given $\boldsymbol{X}_{\text{-}ik}$ can be expressed as

$$\widehat{\boldsymbol{X}}_{ik} = \boldsymbol{b}^T \circ \boldsymbol{X}_{-ik}$$

such that $d(\boldsymbol{X}_{ik}, \widehat{\boldsymbol{X}}_{\text{-}ik})$ attains its minimum.

Definition of the PTCC

Suppose that the TPDM of $\boldsymbol{X}'$ is of the form

$$\Sigma_{\boldsymbol{X}'} = \begin{bmatrix} \Sigma_{hh} & \Sigma_{hl} \\ \Sigma_{lh} & \Sigma_{ll} \end{bmatrix}, \text{ and } \Sigma_{hh} = \text{TPDM}(\boldsymbol{X}_{ik}) = \begin{bmatrix} \Sigma_{ii} & \Sigma_{ik} \\ \Sigma_{ki} & \Sigma_{kk} \end{bmatrix},$$

where $\Sigma_{hl} = \Sigma_{lh}^T = (\Sigma_{il}, \Sigma_{kl})^T$. Then, based on the projection theorem, we have

$$\widehat{\boldsymbol{b}} = \Sigma_{ll}^{-1}\Sigma_{lh} = \arg\min_{\boldsymbol{b}} d(\boldsymbol{X}_{ik}, \widehat{\boldsymbol{X}}_{\text{-}ik}).$$

Then, we can show that the prediction error $\boldsymbol{e} = \boldsymbol{X}_{ik} \ominus \widehat{\boldsymbol{X}}_{\text{-}ik}$ has the following TPDM:

$$\text{TPDM}(\boldsymbol{e}) = \Sigma_{hh} - \Sigma_{hl}\Sigma_{ll}^{-1}\Sigma_{lh} = \begin{bmatrix} \Sigma_{ii} - \Sigma_{il}\Sigma_{ll}^{-1}\Sigma_{li} & \Sigma_{ik} - \Sigma_{il}\Sigma_{ll}^{-1}\Sigma_{lk} \\ \Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} & \Sigma_{kk} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{lk} \end{bmatrix}. \quad (1)$$

Definition of the PTCC

Definition

The **PTCC** *of two random variables X_i and X_k is defined as the TPDM of the residuals with linear dependence of all other random variables removed, as expressed in (1).*

Definition

Let $\mathbf{X}_{ik} = (X_i, X_k)^T$ and $\mathbf{X}_{\text{-}ik}$ be defined as above. Given $\mathbf{X}_{\text{-}ik}$, X_i and X_k are **partially tail uncorrelated**, *if the PTCC of X_i and X_k (given $\mathbf{X}_{\text{-}ik}$) equals zero, i.e., if $\Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} = 0$ according to (1).*

Notice that the residuals of two partially tail uncorrelated random variables are necessarily asymptotically independent.

Properties

Proposition

If a matrix is represented as

$$\Sigma_{\boldsymbol{X}'} = TPDM\{(X_i, X_k, \boldsymbol{X}_{\text{-}ik}^T)^T\} = \begin{bmatrix} \Sigma_{ii} & \Sigma_{ik} & \Sigma_{il} \\ \Sigma_{ki} & \Sigma_{kk} & \Sigma_{kl} \\ \Sigma_{li} & \Sigma_{lk} & \Sigma_{ll} \end{bmatrix}, \quad (2)$$

which is a 3×3 block matrix. Then the following statements are equivalent (Proposition 1 in Speed and Kiiveri (1986)):

$$(1)\ \Sigma_{ki} - \Sigma_{kl}\Sigma_{ll}^{-1}\Sigma_{li} = 0, \qquad (2)\ (\Sigma^{-1})_{ik} = 0.$$

Corollary

Denote the inverse matrix $Q = \Sigma^{-1}$, where Σ is the TPDM of random vector $\boldsymbol{X}$.
$Q_{ik} = 0 \iff PTCC_{ik} = 0$, where $PTCC_{ik}$ is the PTCC of components X_i and X_k.
Recall that zero PTCC means partial tail uncorrelatedness.

Graphical structures for extremes

Let $G = (V, E)$ be a graph, where $V = \{1, \ldots, d\}$ represents the node set and $E \in V \times V$ represents the edge set. G is called an undirected graph, if for two nodes $i, j \in V$, edge $(i, j) \in E$ if and only if edge $(j, i) \in E$.

In this study, we estimate graphical structures for extremes for any type of undirected graphs, where an edge (i, j) is missing when the variables X_i and X_j are partially tail uncorrelated given all the other variables in the graph, written as

$$X_i \perp_p X_j \mid \boldsymbol{X}_{\setminus\{i,j\}}.$$

Graphical structures for extremes

Our methods work for general undirected graph structures: For example, a tree, a decomposable graph, and a non-decomposable graph.

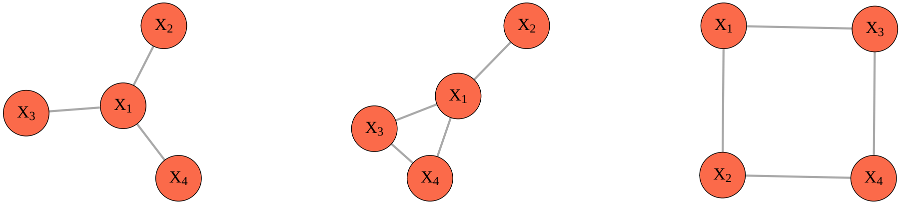

Figure 1: Undirected graph structures: a tree, a decomposable graph, and a non-decomposable graph (from the left to the right).

Sparse representation of high-dimensional extreme networks

For high-dimensional extremes, a graphical representation of the extremal dependence structure is desired for reasons of parsimony and interpretability.

We now introduce two inference methods to learn extremal networks from high-dimensional data via the PTCC based on state-of-the-art graph theories:

- extremal graphical Lasso;
- structured graph learning method via Laplacian spectral constraints.

These two methods both work efficiently in high-dimensional settings, while returning sparse partial tail correlation structures.

Extremal graphical Lasso

For an empirical TPDM estimator, $\widehat{\Sigma}$, and a tuning parameter, $\lambda \geq 0$, the extremal graphical Lasso is expressed as follows (Friedman et al., 2008):

$$\widehat{\Theta}_\lambda = \arg\max_{\Theta \succeq 0} \left[\log\det\Theta - \mathrm{tr}(\widehat{\Sigma}\Theta) - \lambda \sum_{i \neq j} |\Theta_{ij}| \right]$$

where $\widehat{\Theta}_\lambda$ contains the information on the extremal graph structure. Larger λ enforces more zeros in $\widehat{\Theta}_\lambda$ and hence fewer edges in the graph.

Choosing an ideal value for λ is essential.

- On the one hand, we want to enforce sparsity in the graph, where the significant connections are mainly identified in the network.
- On the other hand, $\widehat{\Theta}_\lambda$ should be well-defined and estimated stably.

Structured graph learning via Laplacian spectral constraints

Structured Graph Laplacian (SGL) (Kumar et al., 2019) can control the sparsity and also conveniently preserve the graph connectedness using Laplacian spectral constraints. The optimization problem can be approximated as follows:

$$\max_{w,\lambda,U} \log \operatorname{dget}(\operatorname{Diag}(\lambda)) - \operatorname{tr}(\widehat{\Sigma} Lw) + \alpha ||Lw||_1 + \frac{\beta}{2} ||Lw - U \operatorname{Diag}(\lambda) U^T||_F^2,$$

$$\text{subject to } \boldsymbol{w} \geq 0, \lambda \in S_\lambda, \text{and } U^T U = I,$$

where S_λ denotes the set of vectors that constrains the eigenvalues of the Laplacian matrix, and a Laplacian matrix $\Theta = Lw$, where L is a linear operator that maps a non-negative set of edge weights w into Θ.

This method can be seen as an extension of the graphical Lasso.

Similarly, a larger α increases the sparsity level of the graph, while β additionally controls the level of connectedness, and a larger β enforces a higher level of the connectedness of the estimated graph structure.

Simulation results

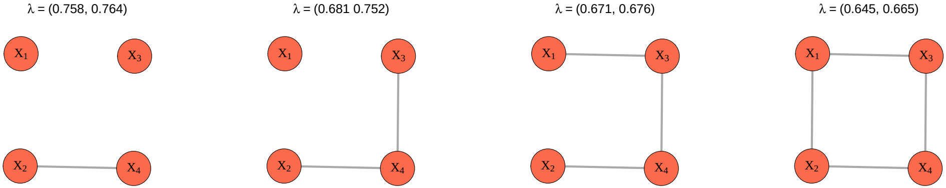

Figure 2: Extremal graph learning using extremal graphical Lasso method for Case 3.

Applications

Risk networks are critical for quantitative risk management to elucidate the complex extremal dependence of random variables. We show two examples in both environmental and financial risk analysis.

- First, we study the river discharge data of the upper Danube basin, which is a benchmark example from the recent literature, and the true underlying physical river flow network is available.
- Second, we apply our method to exploring the historical global currency exchange rate data from different historical periods, including different two economic cycles, COVID-19, and the 2022 Russian invasion of Ukraine (2022.02.24–2022.04.15).

Extremal network estimation for Danube river network

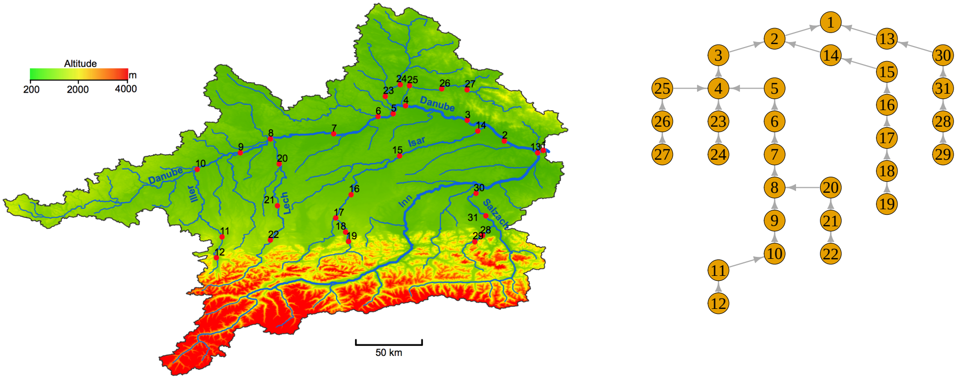

Figure 3: (a) Topographic map of the upper Danube basin, showing 31 sites of gauging stations (red circles) and the altitudes of the region. (b) The true physical river flow connections; the arrows show the flow directions.

Graph structure learning using extremal graphical Lasso

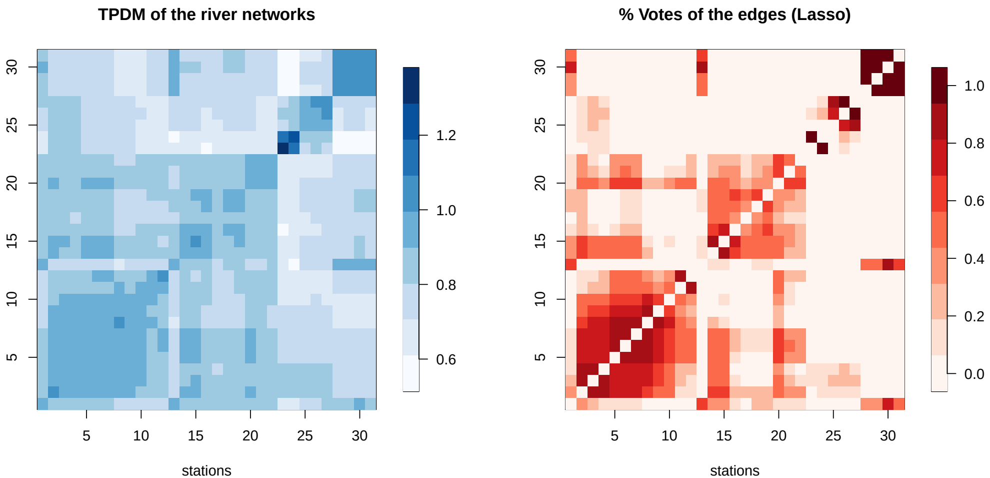

Figure 4: (a) Estimated TPDM of the river discharge from the upper Danube basin. (b) Votes (%) of the edges based on the extremal graphical Lasso method.

Graph structure learning using the SGL method

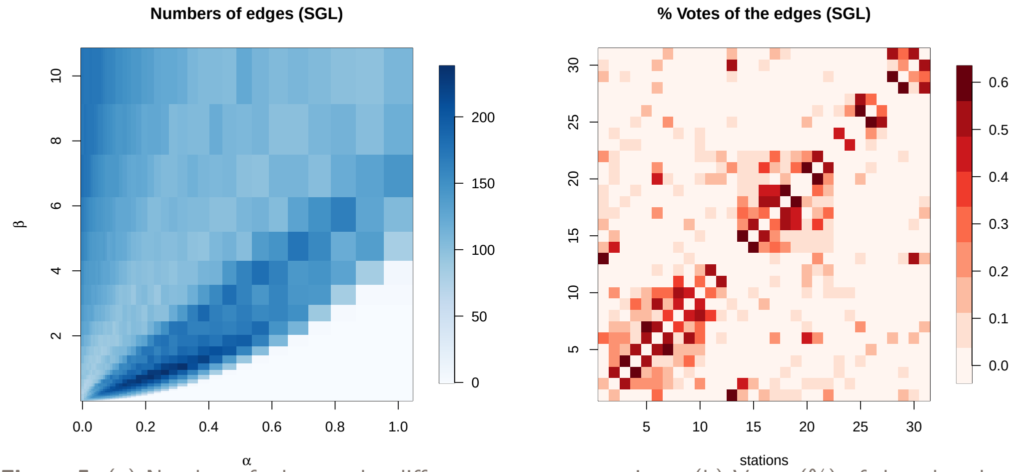

Figure 5: (a) Number of edges under different parameter settings. (b) Votes (%) of the edges based on the SGL method.

Estimated extremal river discharge networks

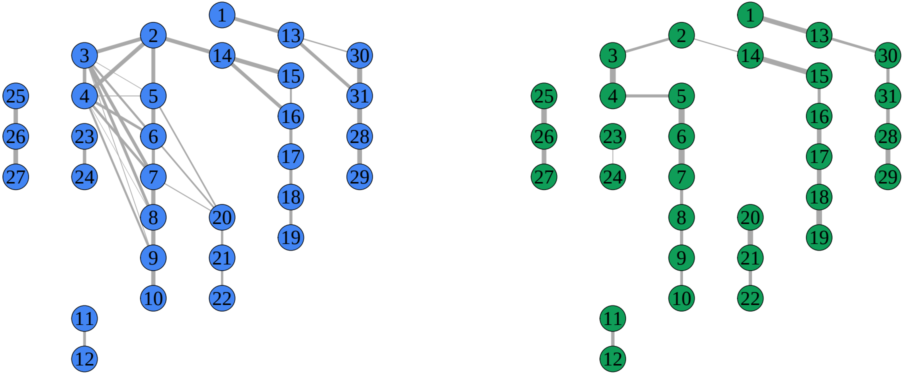

Figure 6: Estimated extremal river discharge networks using the (a) extremal graphical Lasso and (b) SGL methods. The thickness of the edges is proportional to the votes (in percentage). The edges are drawn until no nodes are left isolated.

Extremal network estimation for global currency exchange rate network

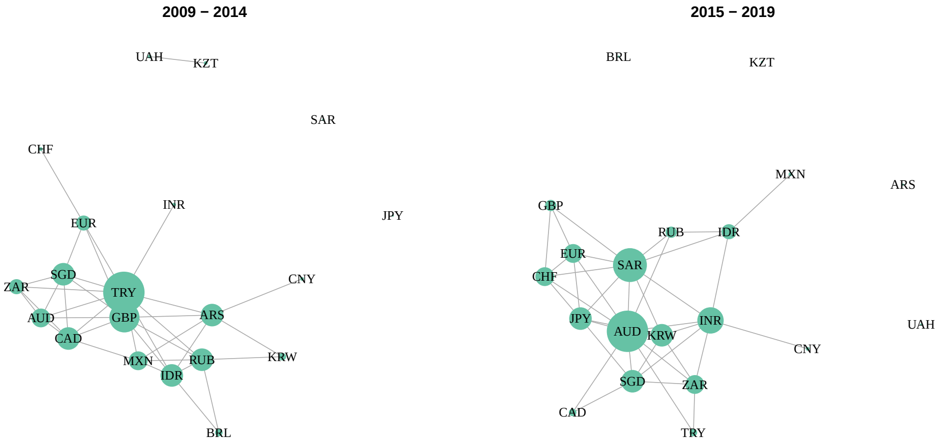

Figure 7: Estimated extremal currency exchange rate risk networks (a) 2009–2014 and (b) 2015–2019.

Extremal network estimation for global currency exchange rate network

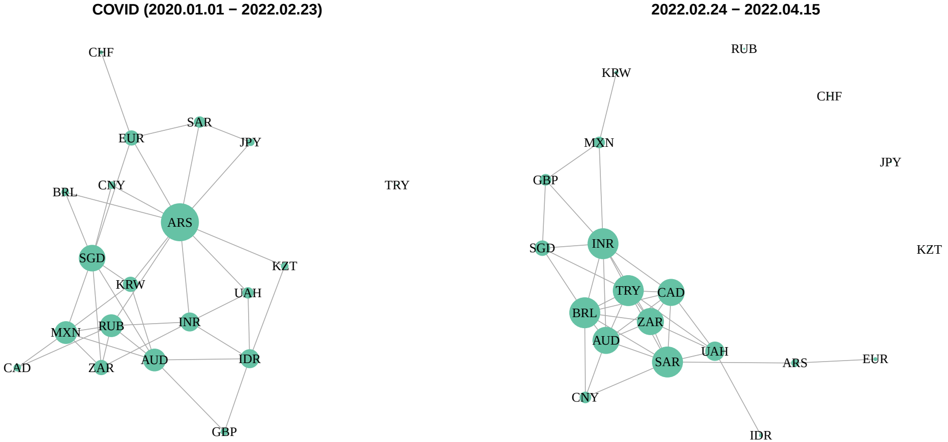

Figure 8: Estimated extremal currency exchange rate risk networks (a) COVID-19 and (b) the 2022 Russian invasion of Ukraine.

Thanks for your attention!

Any comments?

References I

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

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

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8:241–270.

Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. *arXiv preprint arXiv:2012.06179*.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Gissibl, N. (2018). *Graphical modeling of extremes: Max-linear models on directed acyclic graphs*. PhD thesis, Universitätsbibliothek der TU München.

References II

Huang, W. K., Cooley, D. S., Ebert-Uphoff, I., Chen, C., and Chatterjee, S. (2019). New exploratory tools for extremal dependence: χ networks and annual extremal networks. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):484–501.

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

Kumar, S., Ying, J., de Miranda Cardoso, J. V., and Palomar, D. (2019). Structured graph learning via laplacian spectral constraints. *Advances in neural information processing systems*, 32.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

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

References III

Röttger, F., Engelke, S., and Zwiernik, P. (2021). Total positivity in multivariate extremes. *arXiv preprint arXiv:2112.14727*.

Sibuya, M. (1960). Bivariate extreme statistics, i. *Annals of the Institute of Statistical Mathematics*, 11(3):195–210.

Speed, T. P. and Kiiveri, H. T. (1986). Gaussian markov distributions over finite graphs. *The Annals of Statistics*, pages 138–150.

Tran, N. M., Buck, J., and Klüppelberg, C. (2021). Estimating a latent tree for extremes. *arXiv preprint arXiv:2102.06197*.