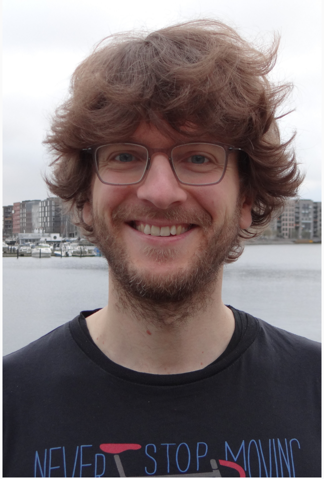
# Causal discovery in heavy-tailed models

Nicola Gnecco — Research Center for Statistics, University of Geneva

BIRS Workshop, Kelowna, June 26 – July 1, 2022

Nicolai Meinshausen
ETH Zurich



Jonas Peters
University of Copenhagen

Sebastian Engelke
University of Geneva

# Goal

- Retrieve causal information from observational data, using the signal in tail of the distribution.

# Goal

- Retrieve causal information from observational data, using the signal in tail of the distribution.
- Often, causal relationships manifest themselves in extreme events.

# Goal

- Retrieve causal information from observational data, using the signal in tail of the distribution.
- Often, causal relationships manifest themselves in extreme events.

**Example**
Earth system science, climate science, finance, etc.

# Generative model

# Linear SCM

- Consider the linear structural causal model, or SCM,

$$X_j := \sum_{k \in \mathrm{pa}(j,G)} \beta_{jk} X_k + \varepsilon_j, \quad j \in V,$$

  where $G = (V, E)$ is the underlying DAG with $V = \{1, \ldots, p\}$, $\mathrm{pa}(j, G)$ are the graphical parents of $j \in V$ in $G$, and $\beta_{jk} > 0$.

# Linear SCM

- Consider the linear structural causal model, or SCM,

$$X_j := \sum_{k \in \mathrm{pa}(j,G)} \beta_{jk} X_k + \varepsilon_j, \quad j \in V,$$

  where $G = (V, E)$ is the underlying DAG with $V = \{1, \ldots, p\}$, $\mathrm{pa}(j, G)$ are the graphical parents of $j \in V$ in $G$, and $\beta_{jk} > 0$.
- Assume that the noise $\varepsilon_j$, $j \in V$, are regularly varying with index $\alpha > 0$, i.e.,

$$P(\varepsilon_j > x) \sim \ell(x) x^{-\alpha}, \quad x \to \infty.$$

# Related work

- The literature that combines causality and extreme value theory is sparse.

# Related work

- The literature that combines causality and extreme value theory is sparse.
- [Gissibl and Klüppelberg, 2018] consider the class of max-linear models

  $$X_j := \max\left(\max_{k \in \mathrm{pa}(j,G)} \beta_{jk} X_k, \varepsilon_j\right), \quad j \in V.$$

  In this framework, [Gissibl et al., 2020] show how to recover a certain class of DAGs and the related edge weights.

# Related work

- The literature that combines causality and extreme value theory is sparse.
- [Gissibl and Klüppelberg, 2018] consider the class of max-linear models

$$X_j := \max\left( \max_{k \in \mathrm{pa}(j,G)} \beta_{jk} X_k, \varepsilon_j \right), \quad j \in V.$$

  In this framework, [Gissibl et al., 2020] show how to recover a certain class of DAGs and the related edge weights.
- [Naveau et al., 2018] try to answer the counterfactual question ”what the Earth’s climate might have been” without anthropogenic interventions, by studying extreme climate events.

# Causal tail coefficient

# Causal tail coefficient

- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j$, $j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

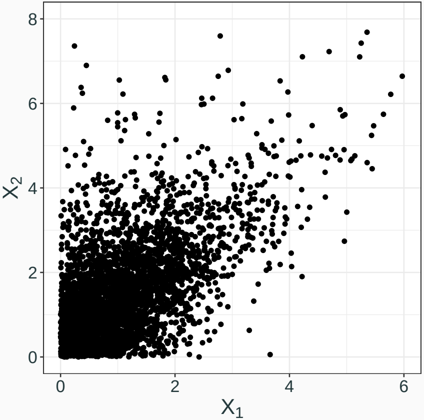
# Causal tail coefficient

- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j, j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

**Example**

$$X_1 := \varepsilon_1,$$
$$X_2 := X_1 + \varepsilon_2$$



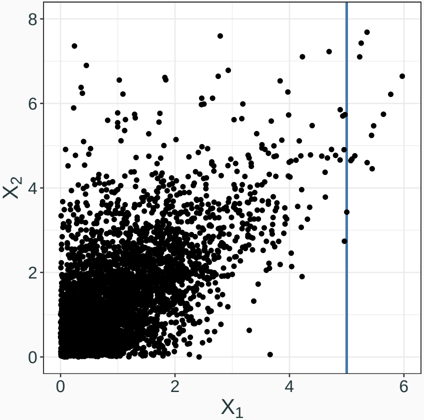
# Causal tail coefficient

- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j$, $j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

**Example**

$$X_1 := \varepsilon_1,$$
$$X_2 := X_1 + \varepsilon_2$$



# Causal tail coefficient

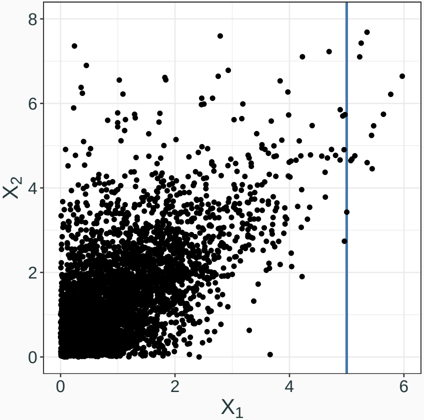
- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j, j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

**Example**

$$X_1 := \varepsilon_1,$$
$$X_2 := X_1 + \varepsilon_2$$

$$\Gamma_{12} = 1$$



# Causal tail coefficient

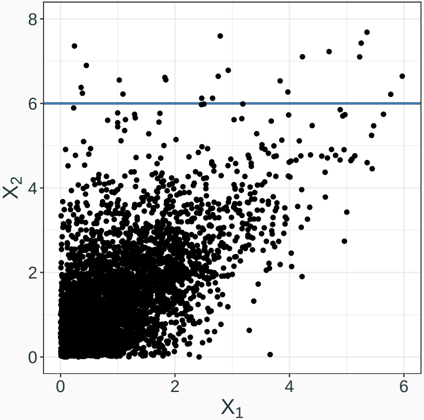
- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j$, $j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

**Example**

$$X_1 := \varepsilon_1,$$
$$X_2 := X_1 + \varepsilon_2$$

$$\Gamma_{12} = 1$$



# Causal tail coefficient

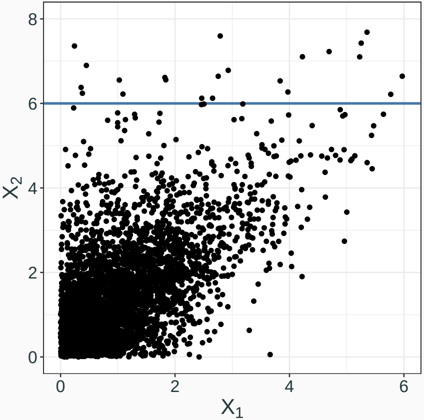
- Consider $X_1$ and $X_2$, two variables of the above SCM with cdfs $F_j, j = 1, 2$.
- We define the causal tail coefficient

$$\Gamma_{12} = \lim_{q \to \infty} \mathrm{E}\left[F_2(X_2) \mid X_1 > q\right].$$

**Example**

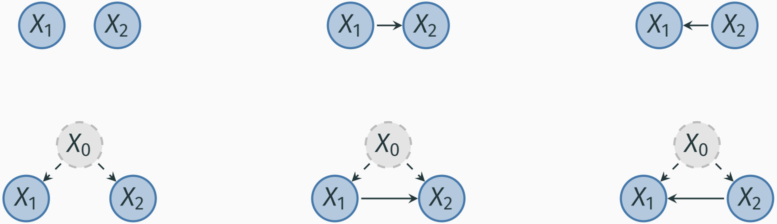
$$X_1 := \varepsilon_1,$$
$$X_2 := X_1 + \varepsilon_2$$

$$\Gamma_{12} = 1$$
$$\Gamma_{21} < 1$$



# A collection of causal models

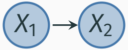
- We try to distinguish the following causal structures.



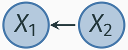
# Causal tail coefficient

- Knowledge of $\Gamma_{12}$ and $\Gamma_{21}$ allows us to distinguish the following cases:

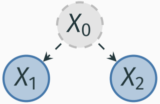
$\Gamma_{12} = \Gamma_{21} = 1/2$



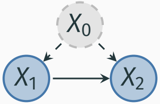
$\Gamma_{12} = 1, \Gamma_{21} < 1$



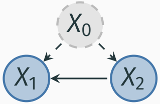
$\Gamma_{12} < 1, \Gamma_{21} = 1$



$\Gamma_{12} < 1, \Gamma_{21} < 1$



$\Gamma_{12} = 1, \Gamma_{21} < 1$



$\Gamma_{12} < 1, \Gamma_{21} = 1$

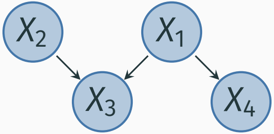
# Extremal Ancestral SEarch (EASE)

# Causal discovery: EASE

- Our goal is to recover the causal order $\pi$ of a heavy-tailed linear SCM over $p$ variables.

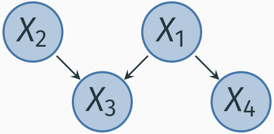
# Causal discovery: EASE

- Our goal is to recover the causal order $\pi$ of a heavy-tailed linear SCM over $p$ variables.



# Causal discovery: EASE

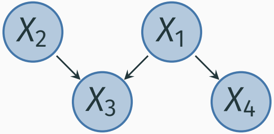
- Our goal is to recover the causal order $\pi$ of a heavy-tailed linear SCM over $p$ variables.



$$\pi = (1, 4, 2, 3)$$

# Causal discovery: EASE

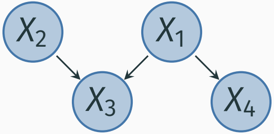
- Our goal is to recover the causal order $\pi$ of a heavy-tailed linear SCM over $p$ variables.
- The input is a matrix $\Gamma \in \mathbb{R}^{p \times p}$ of causal tail coefficients $\Gamma_{jk}$, $j, k \in V$.



$$\pi = (1, 4, 2, 3)$$

# Causal discovery: EASE

- Our goal is to recover the causal order $\pi$ of a heavy-tailed linear SCM over $p$ variables.
- The input is a matrix $\Gamma \in \mathbb{R}^{p \times p}$ of causal tail coefficients $\Gamma_{jk}, j, k \in V$.
- The algorithm is named Extremal Ancestral SEarch (EASE).



$$\pi = (1, 4, 2, 3)$$

# Oracle property

- Consider a heavy-tailed linear SCM over $p$ variables, with an underlying DAG $G$.
- Denote by $\Pi_G$ the set of causal orders of $G$.

**Proposition**

(i) If $\Gamma$ is the input to EASE, then the algorithm returns a permutation $\pi \in \Pi_G$.

# Oracle property

- Consider a heavy-tailed linear SCM over $p$ variables, with an underlying DAG $G$.
- Denote by $\Pi_G$ the set of causal orders of $G$.

**Proposition**

(i) If $\Gamma$ is the input to EASE, then the algorithm returns a permutation $\pi \in \Pi_G$.

(ii) This remains true in the case with hidden confounders.

# Estimator and asymptotics

# Causal tail coefficient estimator

- Let $X_{i1}$ and $X_{i2}$, $i = 1, \ldots, n$, be independent copies of $X_1$ and $X_2$, respectively, where $X_1$ and $X_2$ are two of the $p$ variables of a heavy-tailed linear SCM.

# Causal tail coefficient estimator

- Let $X_{i1}$ and $X_{i2}$, $i = 1, \dots, n$, be independent copies of $X_1$ and $X_2$, respectively, where $X_1$ and $X_2$ are two of the $p$ variables of a heavy-tailed linear SCM.
- We propose the non-parametric estimator

$$\widehat{\Gamma}_{12} = \frac{1}{k} \sum_{i=1}^{n} \widehat{F}_2(X_{i2}) \mathbf{1}\{X_{i1} > X_{(n-k),1}\},$$

  where
  - $\widehat{F}_2$ is the empirical cdf of $X_2$,
  - $X_{(n-k),1}$ denotes the $(n-k)$-th order statistic of $X_{i1}$, $i = 1, \dots, n$,
  - $k = k_n$ depends on the sample size $n$.

# Asymptotic properties

**Proposition**

Let $k_n \in \mathbb{N}$ be an intermediate sequence with

$$k_n \to \infty \quad \text{and} \quad k_n/n \to 0, \quad n \to \infty.$$

Then the estimator $\widehat{\Gamma}_{12}$ is consistent, as $n \to \infty$, i.e.,

$$\widehat{\Gamma}_{12} \xrightarrow{P} \Gamma_{12}.$$

# Asymptotic properties of the algorithm

$$X \in \mathbb{R}^{n \times p} \xrightarrow{k_n} \widehat{\Gamma} \in \mathbb{R}^{p \times p} \xrightarrow{\text{EASE}} \widehat{\pi}.$$

# Asymptotic properties of the algorithm

$$X \in \mathbb{R}^{n \times p} \xrightarrow{k_n} \widehat{\Gamma} \in \mathbb{R}^{p \times p} \xrightarrow{\text{EASE}} \widehat{\pi}.$$

**Proposition**

Let $k_n \in \mathbb{N}$ be an intermediate sequence with

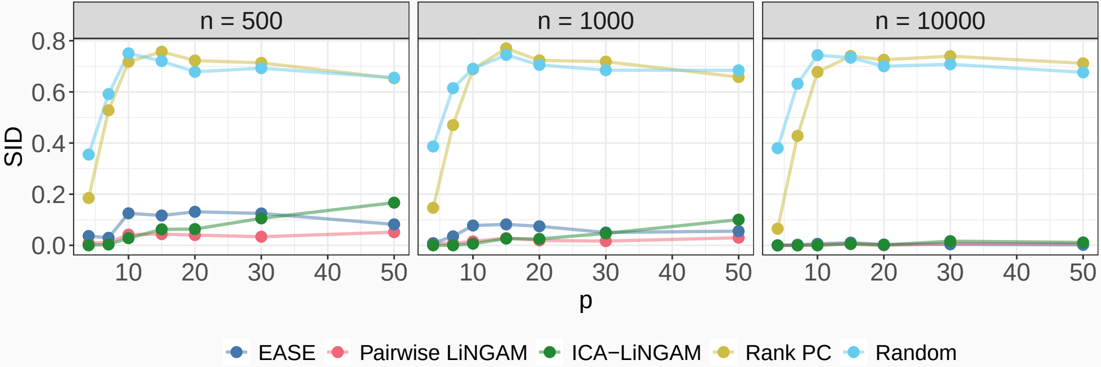
$$k_n \to \infty \quad \text{and} \quad k_n/n \to 0, \quad n \to \infty.$$

Then, as $n \to \infty$, EASE is consistent, i.e.,

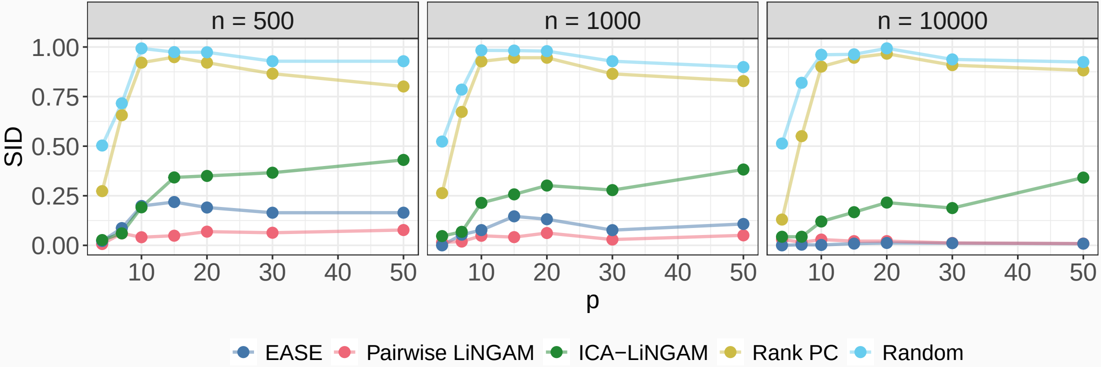
$$\Pr\left(\widehat{\pi} \notin \Pi_G\right) \to 0, \quad \text{as } n \to \infty.$$

# Simulations

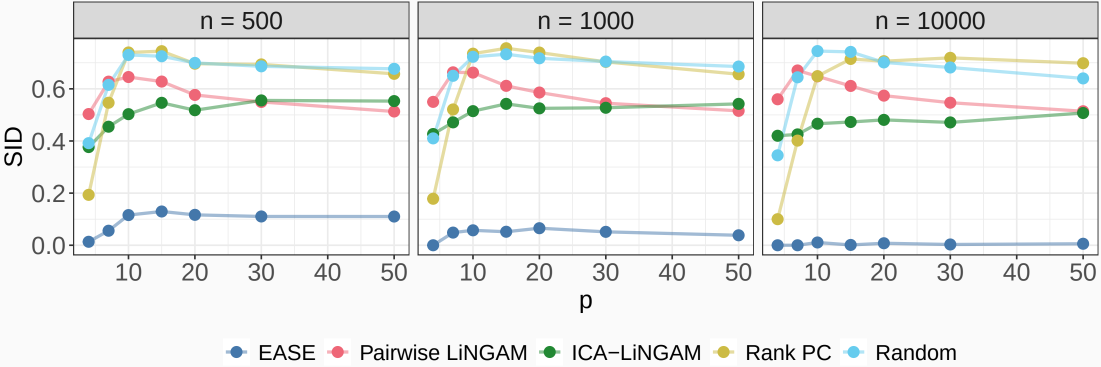
# Setting 1 — Linear SCMs with $\alpha = 1.5$



# Setting 2 — Hidden confounders with $\alpha = 1.5$



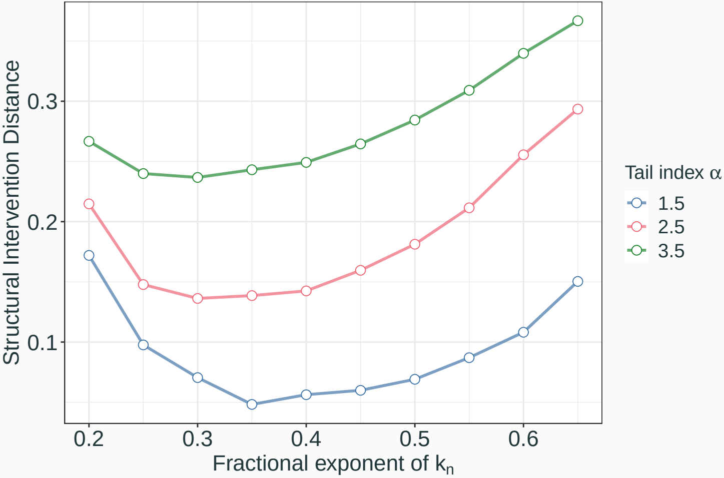
# Setting 3 — Marginal transformation with $\alpha = 1.5$



$$\widetilde{X}_j = F_j(X_j), \quad j \in V.$$

# How to choose $k$?

Performance of EASE in the linear SCM, for different fractional exponents $x \in [0.2, 0.7]$, where $k_n = n^x$.



# Application: river data

# River structure

- Dataset from [Asadi et al., 2015].

# River structure

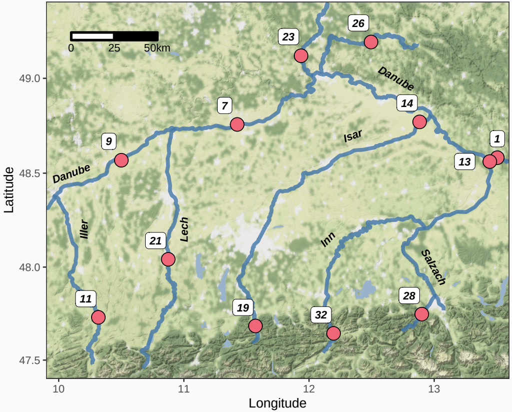
- Dataset from [Asadi et al., 2015].
- Daily data, from 1960 to 2009, with $n = 4600$.

# River structure

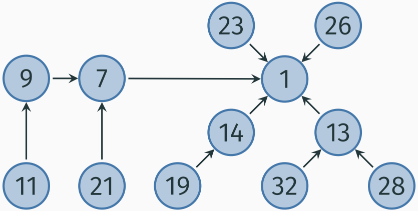
- Dataset from [Asadi et al., 2015].
- Daily data, from 1960 to 2009, with $n = 4600$.
- Only summer months, due to heavy rainfalls.

# River structure

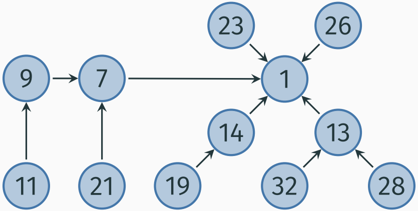
- Dataset from [Asadi et al., 2015].
- Daily data, from 1960 to 2009, with $n = 4600$.
- Only summer months, due to heavy rainfalls.



# River structure



# River structure



Causal order from EASE: $\widehat{\pi} = (23, 32, 26, 28, 19, 21, 11, 9, 7, 14, 13, 1).$

# Conclusions

# Takeaway

- We retrieve causal order from heavy-tailed variables by looking at extreme events.

# Takeaway

- We retrieve causal order from heavy-tailed variables by looking at extreme events.
- EASE is a fast algorithm that is consistent as $n \to \infty$.

# Takeaway

- We retrieve causal order from heavy-tailed variables by looking at extreme events.
- EASE is a fast algorithm that is consistent as $n \to \infty$.
- The results hold also with hidden confounders.

# Takeaway

- We retrieve causal order from heavy-tailed variables by looking at extreme events.
- EASE is a fast algorithm that is consistent as $n \to \infty$.
- The results hold also with hidden confounders.

Gnecco, N., Meinshausen, N., Peters, J., and Engelke, S. (2021).
**Causal discovery in heavy-tailed models.**
*Annals of Statistics, 49(3):1755-1778.*

# Thank You!

# References

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

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

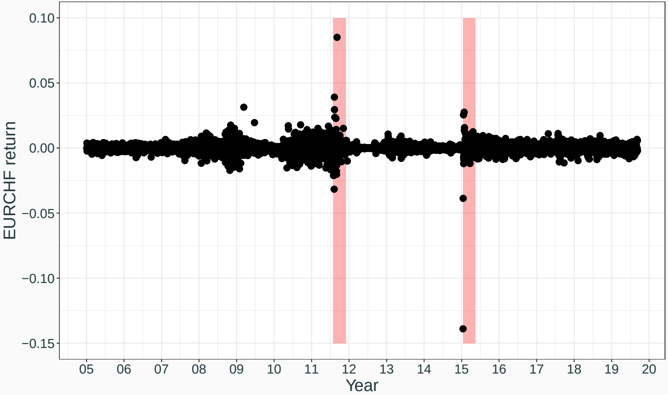
Gissibl, N., Klüppelberg, C., and Lauritzen, S. (2020).
**Identifiability and estimation of recursive max-linear models.**
*Scandinavian Journal of Statistics*.

Naveau, P., Ribes, A., Zwiers, F., Hannart, A., Tuel, A., and Yiou, P. (2018).
**Revising return periods for record events in a climate event attribution context.**
*Journal of Climate*, 31(9):3411–3422.

# Application: financial data

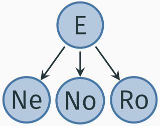
# EURCHF exchange rate

- Return of Euro-Swiss franc exchange rates (EURCHF), and the returns of Nestlé, Novartis and Roche in a period from January 2005 to September 2019.
- How do shocks in the exchange rate influence returns of large Swiss stocks?



# Cause and effects

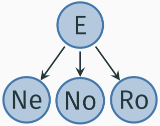
- Estimates of $\Psi \in [1/2, 1]$ (a two-sided version of $\Gamma$) show that the extremes of the exchange rate drive large changes in the stocks' returns, but not *vice versa*.



$$\widehat{\Psi} = \begin{array}{c c} \begin{array}{cccc} E & Ne & No & Ro \end{array} & \\ \begin{pmatrix} \cdot & 0.86 & 0.90 & 0.90 \\ 0.72 & \cdot & 0.85 & 0.87 \\ 0.72 & 0.94 & \cdot & 0.81 \\ 0.71 & 0.94 & 0.86 & \cdot \end{pmatrix} & \begin{array}{c} E \\ Ne \\ No \\ Ro \end{array} \end{array}$$

# Cause and effects

- Estimates of $\Psi \in [1/2, 1]$ (a two-sided version of $\Gamma$) show that the extremes of the exchange rate drive large changes in the stocks' returns, but not *vice versa*.

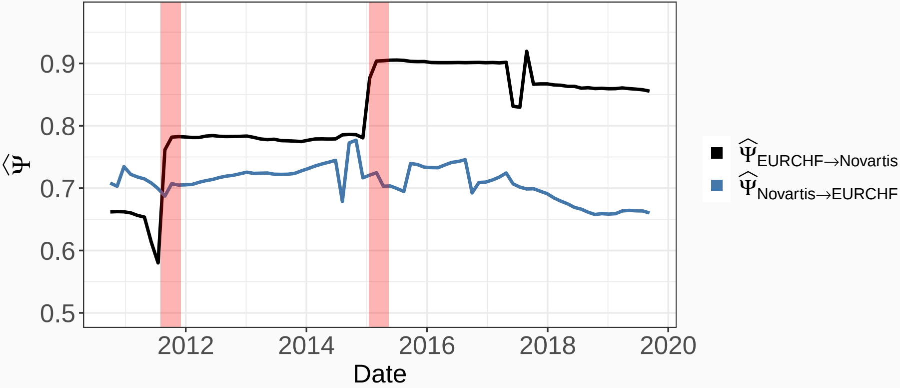


$$\widehat{\Psi} = \begin{array}{cc} & \begin{array}{cccc} E & Ne & No & Ro \end{array} \\ & \left(\begin{array}{cccc} \cdot & 0.86 & 0.90 & 0.90 \\ 0.72 & \cdot & 0.85 & 0.87 \\ 0.72 & 0.94 & \cdot & 0.81 \\ 0.71 & 0.94 & 0.86 & \cdot \end{array}\right) \begin{array}{c} E \\ Ne \\ No \\ Ro \end{array} \end{array}$$

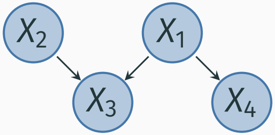
Causal order from EASE: $\widehat{\pi} = (E, No, Ro, Ne)$.

# Causal coefficient over time

- Plot of estimates $\widehat{\Psi}$ between EURCHF and Novartis using a rolling window of 1500 days.

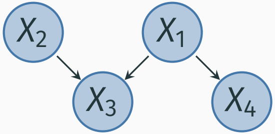


# EASE



$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

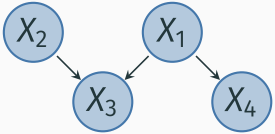
# EASE



$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V$.

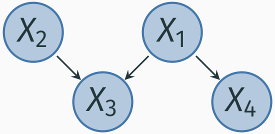
# EASE



$$\Gamma = \begin{bmatrix} & 0.50 & \boxed{1} & \boxed{1} \\ 0.50 & & 1 & 0.50 \\ \boxed{0.59} & \boxed{0.68} & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V$.

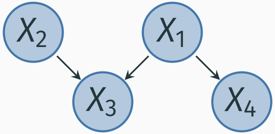
# EASE



$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ \boxed{0.59} & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 1$.

# EASE

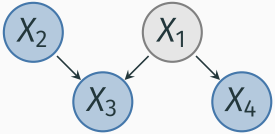


$$\pi = (1, \ , \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 1$.
- Update causal order.

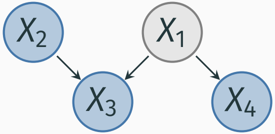
# EASE



$$\pi = (1, \; , \; , \; )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

# EASE

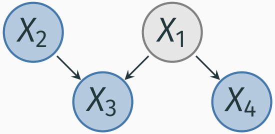


$$\pi = (1, \;, \;, \;)$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1\}$.

# EASE

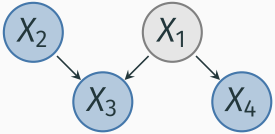


$$\pi = (1, \ , \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & \boxed{1} & 0.50 \\ 0.59 & \boxed{0.68} & & \boxed{0.59} \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1\}$.

# EASE

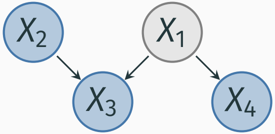


$$\pi = (1, \ , \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & \boxed{0.59} \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1\}$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 4$.

# EASE

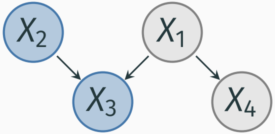


$$\pi = (1, 4, \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1\}$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 4$.
- Update causal order.

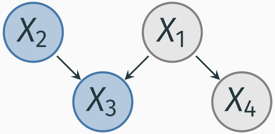
# EASE



$$\pi = (1, 4, \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

# EASE

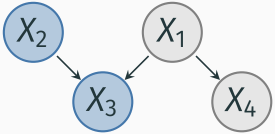


$$\pi = (1, 4, \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1, 4\}$.

# EASE

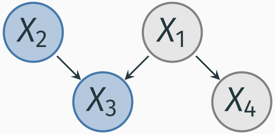


$$\pi = (1, 4, \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & \boxed{1} & 0.50 \\ 0.59 & \boxed{0.68} & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1, 4\}$.

# EASE

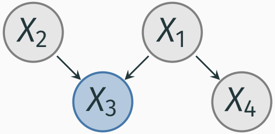


$$\pi = (1, 4, \ , \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & \boxed{0.68} & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1, 4\}$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 2$.

# EASE

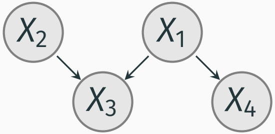


$$\pi = (1, 4, 2, \ )$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1, 4\}$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 2$.
- Update causal order.

# EASE



$$\pi = (1, 4, 2, 3)$$

$$\Gamma = \begin{bmatrix} & 0.50 & 1 & 1 \\ 0.50 & & 1 & 0.50 \\ 0.59 & 0.68 & & 0.59 \\ 0.54 & 0.50 & 0.54 & \end{bmatrix}$$

- Take the column-wise maxima $M_j$, $j \in V \setminus \{1, 4\}$.
- Choose the column $j$ with the smallest $M_j$. In this case $j = 2$.
- Update causal order.