ERNSI Workshop 2019
Kasteel Vaeshartelt, Maastricht

Causal Inference from Multivariate Time Series: Principles and Problems

Michael Eichler

Department of Quantitative Economics

Maastricht University

http://researchers-sbe.unimaas.nl/michaeleichler

22 September 2019

Outline

- **Causality concepts**
- Graphical representation
  - Definition
  - Markov properties
  - Extension: systems with latent variables
- Causal learning
  - Basic principles
  - Identification from empirical relationships
- Non-Markovian constraints
  - Trek-separation in graphs
  - Tetrad representation theorem
  - Testing for tetrad constraints
- Open problems and conclusions

Concepts of causality for time series

If X precedes Y then X is the cause of Y?

We consider two variables X and Y measured at discrete times $t \in \mathbb{Z}$:

$$X = \left(X_t\right)_{t\in\mathbb{Z}}, \quad Y = \left(Y_t\right)_{t\in\mathbb{Z}}.$$

Question: When is it justified to say that X causes Y?

Various approaches:

- Intervention causality (Pearl, 1993; Eichler & Didelez 2007, 2010)
- Structural causality (White and Lu, 2010)
- Granger causality (Granger, 1967, 1980, 1988)
- Sims causality (Sims, 1972)

Intervention causality

Idea: consider primitive interventions

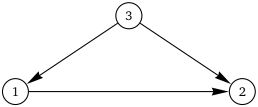
- sets value of target variable X_t to specified value x^*
- yields new probability distribution $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid \mathrm{do}(X_t = x^*))$

Intervention causality

Idea: consider primitive interventions

- sets value of target variable X_t to specified value x^*
- yields new probability distribution $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid \mathrm{do}(X_t = x^*))$

Observational regime $\mathbb{P}$

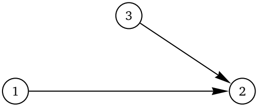


Intervention causality

Idea: consider primitive interventions

- sets value of target variable X_t to specified value x^*
- yields new probability distribution $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid do(X_t = x^*))$

Intervention regime $\mathbb{P}^*$



Intervention causality

Question: relation between ***interventional*** and ***observational regimes***?

Notation: $\boldsymbol{V}_t = (X_t, Y_t, Z_t)$ and $\boldsymbol{V}^t = \{\boldsymbol{V}_s | s \leq t\}$.

- distributions of X^{t-1}, Y^t, Z^t

$$\mathbb{P}(A \,|\, \mathrm{do}(X_t = x^*)) = \mathbb{P}(A) \qquad \forall A \in \sigma\{X^{t-1}, Y^t, Z^t\}$$

- conditional distributions of $\boldsymbol{V}_{t+1}, \ldots, \boldsymbol{V}_{t'}$ given $\boldsymbol{V}^t$

$$\mathbb{P}(A \,|\, \boldsymbol{V}^t, \mathrm{do}(X_t = X^*)) = \mathbb{P}(A \,|\, \boldsymbol{V}^t) \qquad \forall A \in \sigma\{\boldsymbol{V}^{t+h}\} \ \forall h > 0$$

- distribution of X_t

$$\mathbb{P}(X_t = x \,|\, \mathrm{do}(X_t = x^*)) = 1_{\{x = x^*\}}$$

Note: this is an assumption about the nature of the primitive intervention and about the process V_t (sufficiently rich).

Intervention causality

Average causal effect of setting $X_t = x^*$ on $Y_{t'}$

$$\text{ACE}_{x^*} = \mathbb{E}(Y_{t'} \mid \text{do}(X_t = x^*)) - \mathbb{E}(Y_{t'})$$

Noncausality: X_t has no causal effect on $Y_{t'}$ if

$$\mathbb{P}(Y_{t'} \mid \text{do}(X_t = x^*)) = \mathbb{P}(Y_{t'})$$

Structural causality

Assumption: X_t and Y_t structurally generated by

$$X_t = q_{x,t}(X^{t-1}, Y^{t-1}, Z^{t-1}, U_{x,t})$$
$$Y_t = q_{y,t}(X^{t-1}, Y^{t-1}, Z^{t-1}, U_{y,t})$$

Direct structural causality:

X does not directly structurally cause Y if

$$q_{y,t}(x^{t-1}, y^{t-1}, z^{t-1}, u_{y,t}) = q_{y,t}(y^{t-1}, z^{t-1}, u_{y,t}).$$

Relation to intervention causality:

If X does not directly structurally cause Y then

$$\mathbb{P}(Y_{t+1} | \mathrm{do}(X_t = x^*)) = \mathbb{P}(Y_{t+1})$$

Granger causality

Two fundamental princples:

- The cause precedes its effect in time.
- The causal series contains special information about the series being caused that is not available otherwise.

Granger causality

Two fundamental princples:

- The cause precedes its effect in time.
- The causal series contains special information about the series being caused that is not available otherwise.

This leads us to consider two information sets:

- $\mathscr{F}^*(t)$ - all information in the universe up to time t
- $\mathscr{F}^*_{-X}(t)$ - this information except the values of X

Granger causality

Two fundamental princples:

- The cause precedes its effect in time.
- The causal series contains special information about the series being caused that is not available otherwise.

This leads us to consider two information sets:

- $\mathscr{F}^*(t)$ - all information in the universe up to time t
- $\mathscr{F}^*_{-X}(t)$ - this information except the values of X

Granger's definition of causality (Granger 1969, 1980)

We say that Y causes X if the probability distributions of

- X_{t+1} given $\mathscr{F}^*(t)$ and
- X_{t+1} given $\mathscr{F}^*_{-Y}(t)$

are different.

Granger causality

Relation to direct structural causality

Suppose that V_t satisfies

$$Y_{t+1} \perp\!\!\!\perp \mathscr{F}_t^* \,|\, \boldsymbol{V}^t$$

Then

- X does not directly structurally cause Y

implies

- X does not Granger cause Y

The converse is not true: $U_{i,t} \sim \mathscr{N}(0,1)$

$$Y_t = q_t(X_{t-1}, U_{1,t}, U_{2,t}) = \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} U_{1,t} + \frac{1}{\sqrt{1+X_{t-1}^2}} U_{2,t},$$

Note: this causal relationship cannot be tested

Granger causality
Relation to intervention causality

Suppose that $\boldsymbol{V}_t$ satisfies

$$Y_{t+1} \perp\!\!\!\perp \mathscr{F}_t^* \,|\, \boldsymbol{V}^t$$

Then

- X does not Granger cause Y

implies

- $\mathbb{P}(Y_{t+1}|\mathrm{do}(X_t = x^*)) = \mathbb{P}(Y_{t+1})$

Granger causality

Problem: The definition cannot be used with actual data.

Granger causality

Problem: The definition cannot be used with actual data.

Suppose data consist of multivariate time series $\boldsymbol{V} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ and let

- $\{\boldsymbol{X}^t\}$ - information given by $\boldsymbol{X}$ up to time t
- similarly for $\boldsymbol{Y}$ and $\boldsymbol{Z}$

Definition: Granger non-causality

- $\boldsymbol{X}$ is *Granger-noncausal* for $\boldsymbol{Y}$ with respect to $\boldsymbol{V}$ if

$$Y_{t+1} \perp\!\!\!\perp \boldsymbol{X}^t \,|\, \boldsymbol{Y}^t, \boldsymbol{Z}^t.$$

- Otherwise we say that X *Granger-causes* Y with respect to $\boldsymbol{V}$.

Granger causality

Problem: The definition cannot be used with actual data.

Suppose data consist of multivariate time series $\boldsymbol{V} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ and let

- $\{\boldsymbol{X}^t\}$ - information given by $\boldsymbol{X}$ up to time t
- similarly for $\boldsymbol{Y}$ and $\boldsymbol{Z}$

Definition: Granger non-causality

- $\boldsymbol{X}$ is *Granger-noncausal* for $\boldsymbol{Y}$ with respect to $\boldsymbol{V}$ if

$$Y_{t+1} \perp\!\!\!\perp \boldsymbol{X}^t \,|\, \boldsymbol{Y}^t, \boldsymbol{Z}^t.$$

- Otherwise we say that X *Granger-causes* Y with respect to $\boldsymbol{V}$.

Additionally:

- X and Y are said to be *contemporaneously independent* w.r.t. $\boldsymbol{V}$ if

$$X_{t+1} \perp\!\!\!\perp Y_{t+1} \,|\, \boldsymbol{V}^t$$

Sims causality

Definition: Sims non-causality

$\boldsymbol{X}$ does not Sims-cause $\boldsymbol{Y}$ with respect to $\boldsymbol{V} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ if

$$\boldsymbol{Y}^{t+h} \perp\!\!\!\perp X_t \,|\, \boldsymbol{X}^{t-1}, \boldsymbol{Y}^t, \boldsymbol{Z}^t \qquad \text{for all } h > 0$$

Note:

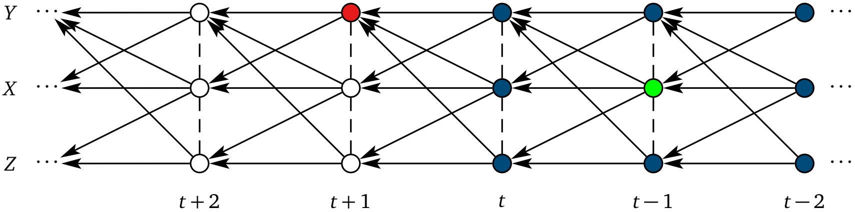
- Granger causality is a concept of direct causality
- Sims causality is a concept of total causality (direct and indirect pathways)

The following statistics are measures for Sims causality:

- impulse response function (time and frequency domain)
- direct transfer function (DTF)

Granger and Sims causality

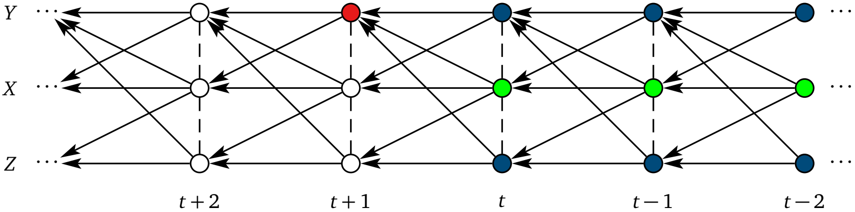
Granger causality: single lag



$$Y_{t+1} \perp\!\!\!\perp X_{t-k} \mid \boldsymbol{Y}^t, \boldsymbol{Z}^t, \boldsymbol{X}^t \setminus \{X_{t-k}\} \qquad \text{for all } k \geq 0$$

Granger and Sims causality

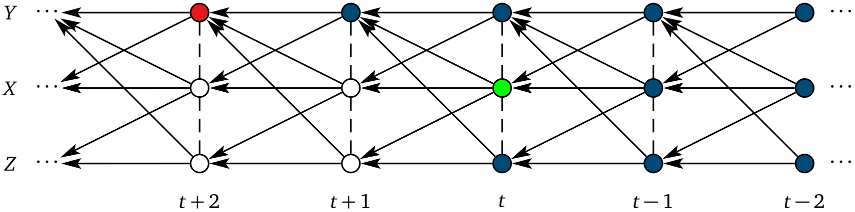
Granger causality: all lags



$Y_{t+1} \perp\!\!\!\perp X^t \mid Y^t, Z^t$

Granger and Sims causality

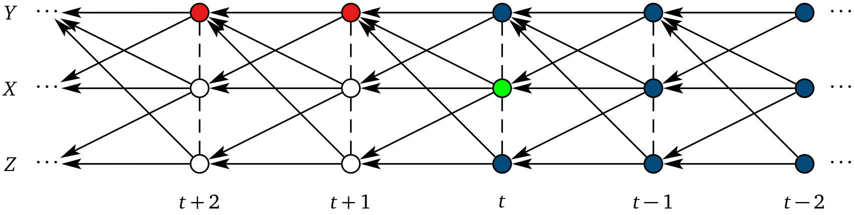
Sims causality: single lag



$Y_{t+h} \perp\!\!\!\perp X_t \mid \boldsymbol{Y}^{t+h-1}, \boldsymbol{Z}^t, \boldsymbol{X}^{t-1}$ for all $h > 0$

Granger and Sims causality

Sims causality: all lags



$$\boldsymbol{Y}^{t+h} \perp\!\!\!\perp X_t \mid \boldsymbol{Y}^t, \boldsymbol{Z}^t, \boldsymbol{X}^{t-1} \qquad \text{for all } h > 0$$

Vector autoregressive processes

Let $\boldsymbol{X}$ be a multivariate stationary Gaussian time series with vector autoregressive representation

$$\boldsymbol{X}_t = \sum_{k=1}^{\infty} \mathbf{A}_k \boldsymbol{X}_{t-k} + \boldsymbol{\varepsilon}_t$$

Granger non-causality in VAR models:

The following are equivalent:

- X_b does not Granger cause X_a with respect to $\boldsymbol{X}$;
- $A_{ab,k} = 0$ for all $k \in \mathbb{N}$.

Vector autoregressive processes

Let $\boldsymbol{X}$ be a multivariate stationary Gaussian time series with vector autoregressive representation

$$\boldsymbol{X}_t = \sum_{k=1}^{\infty} \mathbf{A}_k \boldsymbol{X}_{t-k} + \boldsymbol{\varepsilon}_t = \sum_{k=0}^{\infty} \mathbf{B}_k \, \boldsymbol{\varepsilon}_{t-k}$$

Granger non-causality in VAR models:

The following are equivalent:

- X_b does not Granger cause X_a with respect to $\boldsymbol{X}$;
- $A_{ab,k} = 0$ for all $k \in \mathbb{N}$.

Sims non-causality in VAR models:

The following are equivalent:

- X_b does not Sims cause X_a with respect to $\boldsymbol{X}$;
- $B_{ab,k} = 0$ for all $k \in \mathbb{N}$.

Outline

- Causality concepts
- **Graphical representation**
  - Definition
  - Markov properties
  - Extension: systems with latent variables
- Causal learning
  - Basic principles
  - Identification from empirical relationships
- Non-Markovian constraints
  - Trek-separation in graphs
  - Tetrad representation theorem
  - Testing for tetrad constraints
- Open problems and conclusions

Graphical models for time series

Basic idea: use graphs to encode conditional independences among variables

- nodes/vertices represent variables
- missing edge between two nodes implies conditional independence of the two variables

Application to time series:

- treat each variable at each time separately ($\rightsquigarrow$ time series chain graphs)
- treat each series as one variables (only one node in the graph)

Graphical models for time series
Granger causality graphs (Eichler 2007)

Idea: represent Granger-causal relations in $\boldsymbol{X}$ by *mixed graph* G:

- vertices $v \in V$ represent the variables (time series) X_v;

Graphical models for time series

Granger causality graphs (Eichler 2007)

Idea: represent Granger-causal relations in $\boldsymbol{X}$ by *mixed graph* G:

- vertices $v \in V$ represent the variables (time series) X_v;
- directed edges between the vertices indicate Granger-causal relationships;

Graphical models for time series

Granger causality graphs (Eichler 2007)

Idea: represent Granger-causal relations in $\boldsymbol{X}$ by *mixed graph* G:

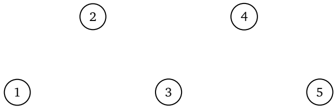
- vertices $v \in V$ represent the variables (time series) X_v;
- directed edges between the vertices indicate Granger-causal relationships;
- additionally undirected (dashed) edges indicate contemporaneous associations.

Graphical models for time series

Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$

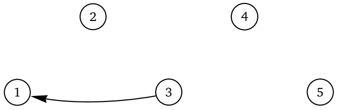


Graphical models for time series

Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



with

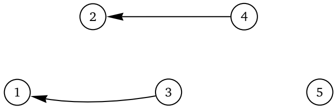
- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$

Graphical models for time series

Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



with

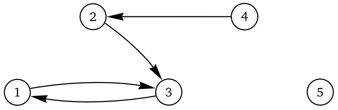
- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$
- $X_{2,t} = f_2(X_{4,t-1}) + \varepsilon_{2,t}$

Graphical models for time series

Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



with

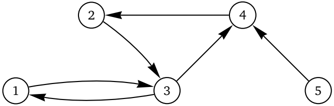
- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$
- $X_{2,t} = f_2(X_{4,t-1}) + \varepsilon_{2,t}$
- $X_{3,t} = f_3(X_{1,t-1}, X_{2,t-1}) + \varepsilon_{3,t}$

Graphical models for time series

Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



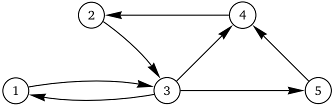
with

- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$
- $X_{2,t} = f_2(X_{4,t-1}) + \varepsilon_{2,t}$
- $X_{3,t} = f_3(X_{1,t-1}, X_{2,t-1}) + \varepsilon_{3,t}$
- $X_{4,t} = f_4(X_{3,t-1}, X_{5,t-1}) + \varepsilon_{4,t}$

Graphical models for time series
Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



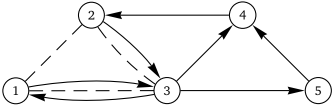
with

- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$
- $X_{2,t} = f_2(X_{4,t-1}) + \varepsilon_{2,t}$
- $X_{3,t} = f_3(X_{1,t-1}, X_{2,t-1}) + \varepsilon_{3,t}$
- $X_{4,t} = f_4(X_{3,t-1}, X_{5,t-1}) + \varepsilon_{4,t}$
- $X_{5,t} = f_5(X_{3,t-1}) + \varepsilon_{5,t}$

Graphical models for time series
Granger causality graphs (Eichler 2007)

Example: consider five-dimensional autoregressive process X_V

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}) + \boldsymbol{\varepsilon}_t$$



with

- $X_{1,t} = f_1(X_{3,t-1}) + \varepsilon_{1,t}$
- $X_{2,t} = f_2(X_{4,t-1}) + \varepsilon_{2,t}$
- $X_{3,t} = f_3(X_{1,t-1}, X_{2,t-1}) + \varepsilon_{3,t}$
- $X_{4,t} = f_4(X_{3,t-1}, X_{5,t-1}) + \varepsilon_{4,t}$
- $X_{5,t} = f_5(X_{3,t-1}) + \varepsilon_{5,t}$
- $\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t} \perp\!\!\!\perp \varepsilon_{4,t}, \varepsilon_{5,t}$
  $\varepsilon_{4,t} \perp\!\!\!\perp \varepsilon_{5,t}$

Markov properties

Objective: derive Granger-causal relationships for $\boldsymbol{X}_S$, $S \subseteq V$

Markov properties

Objective: derive Granger-causal relationships for $\boldsymbol{X}_S$, $S \subseteq V$

Idea: characterize pathways that induce associations

Markov properties

Objective: derive Granger-causal relationships for $\boldsymbol{X}_S$, $S \subseteq V$

Idea: characterize pathways that induce associations

Tool: concepts of separation in graphs

- DAGs: d-separation (Pearl 1988)
- mixed graphs: d-separation (Spirtes et al. 1998, Koster 1999) or m-separation (Richardson 2003)

Markov properties

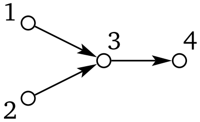
Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization

$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1)$$



Markov properties

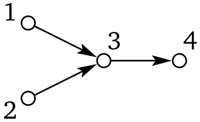
Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G

$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}(i)})$$



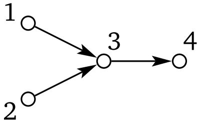
Markov properties

Acyclic directed graph (DAG):

- Graph $G = (V,E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G



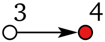
$$p(x) = \prod_{i=1}^{n} p(x_i|x_{i-1},\ldots,x_1) = \prod_{i=1}^{n} p(x_i|x_{\mathrm{pa}(i)})$$

Example:

$$p(x_4,x_3,x_2,x_1) = p(x_4|x_3)$$

Conditional Independences:

- $X_4 \perp\!\!\!\perp X_1,X_2 \,|\, X_3$



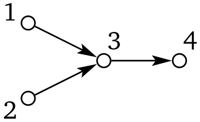
Markov properties

Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G



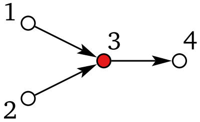
$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}(i)})$$

Example:

$$p(x_4, x_3, x_2, x_1) = p(x_4 | x_3)\, p(x_3 | x_2, x_1)$$

Conditional Independences:

- $X_4 \perp\!\!\!\perp X_1, X_2 \,|\, X_3$



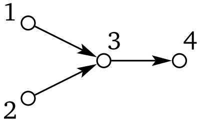
Markov properties

Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G



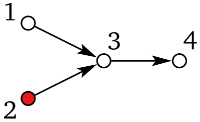
$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}(i)})$$

Example:

$$p(x_4, x_3, x_2, x_1) = p(x_4 | x_3)\, p(x_3 | x_2, x_1)\, p(x_2)$$

Conditional Independences:

- $X_4 \perp\!\!\!\perp X_1, X_2 \,|\, X_3$
- $X_2 \perp\!\!\!\perp X_1$



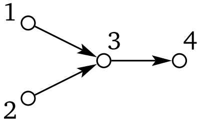
Markov properties

Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G



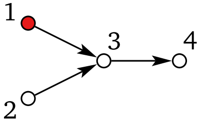
$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}(i)})$$

Example:

$$p(x_4, x_3, x_2, x_1) = p(x_4 | x_3) p(x_3 | x_2, x_1) p(x_2) p(x_1)$$

Conditional Independences:

- $X_4 \perp\!\!\!\perp X_1, X_2 \,|\, X_3$
- $X_2 \perp\!\!\!\perp X_1$



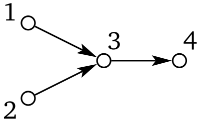
Markov properties

Acyclic directed graph (DAG):

- Graph $G = (V, E)$ with edges of type $\longrightarrow$
- No directed cylces $\quad i \longrightarrow \ldots \longrightarrow i$

Applications: expert systems, Bayesian networks

Recursive factorization with respect to DAG G



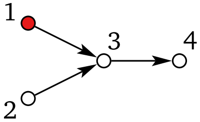
$$p(x) = \prod_{i=1}^{n} p(x_i | x_{i-1}, \ldots, x_1) = \prod_{i=1}^{n} p(x_i | x_{\mathrm{pa}(i)})$$

Example:

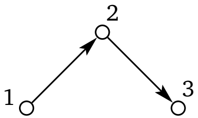
$$p(x_4, x_3, x_2, x_1) = p(x_4 | x_3)\, p(x_3 | x_2, x_1)\, p(x_2)\, p(x_1)$$

Conditional Independences:

- $X_4 \perp\!\!\!\perp X_1, X_2 \,|\, X_3$
- $X_2 \perp\!\!\!\perp X_1$
- $X_1 \perp\!\!\!\perp X_2 | X_3$?



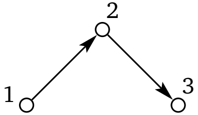
Markov properties



$$p(x) = p(x_3|x_2)p(x_2|x_1)p(x_1)$$

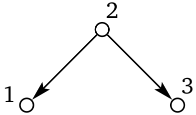
$$\Rightarrow X_3 \perp\!\!\!\perp X_1 \,|\, X_2$$

Markov properties



$$p(x) = p(x_3|x_2)p(x_2|x_1)p(x_1)$$

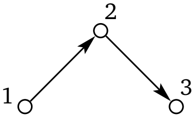
$$\Rightarrow X_3 \perp\!\!\!\perp X_1 \,|\, X_2$$



$$p(x) = p(x_1|x_2)p(x_3|x_2)p(x_2)$$

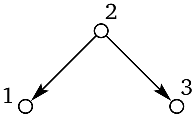
$$\Rightarrow X_3 \perp\!\!\!\perp X_1 \,|\, X_2$$

Markov properties



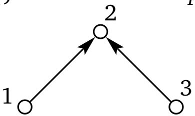
$$p(x) = p(x_3|x_2)p(x_2|x_1)p(x_1)$$

$$\Rightarrow X_3 \perp\!\!\!\perp X_1 \mid X_2$$



$$p(x) = p(x_1|x_2)p(x_3|x_2)p(x_2)$$

$$\Rightarrow X_3 \perp\!\!\!\perp X_1 \mid X_2$$



$$p(x) = p(x_2|x_1, x_3)p(x_3)p(x_1)$$

$$\not\Rightarrow X_3 \perp\!\!\!\perp X_1 \mid X_2$$

Global Granger-causal Markov property

Separation in mixed graphs

Question: What type of paths induce Granger causal relations between variables?

Note: Granger (non)causality is not symmetric

Idea: consider only paths ending with a directed edge $\longrightarrow$

Examples: $1 \longrightarrow 2 \longleftarrow 3 \longrightarrow 4$ entails

- X_1 does not Granger cause X_4 with respect to X_1, X_4
- X_1 does not Granger cause X_4 with respect to X_1, X_3, X_4
- X_1 does not Granger cause X_4 with respect to X_1, X_2, X_3, X_4

but **not**

- X_1 does not Granger cause X_4 with respect to X_1, X_2, X_4

Outline

- Causality concepts
- Graphical representation
  - Definition
  - Markov properties
  - Extension: systems with latent variables
- **Causal learning**
  - Basic principles
  - Identification from empirical relationships
- Non-Markovian constraints
  - Trek-separation in graphs
  - Tetrad representation theorem
  - Testing for tetrad constraints
- Open problems and conclusions

Principles of causal inference

Objective: identify causal structure of process X

Question: What to use in practise?

- Granger causality or Sims causality
- bivariate or fully multivariate analysis

Principles of causal inference

Objective: identify causal structure of process X

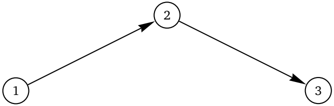
Question: What to use in practise?

- Granger causality or Sims causality
- bivariate or fully multivariate analysis

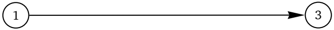
Answer:
For causal inference . . . ***all and more***.

Principles of identification

An example of indirect causality:

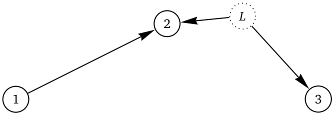


implies for the bivariate submodel

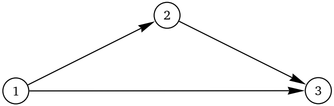


Principles of identification

An example of spurious causality:



implies for the trivariate and bivariate submodels



Principles of identification

Inverse problem:
What can we say about the full system based on observed Granger-noncausal relations for the observed (sub)process?

Suppose

- $X_a \longrightarrow X_c \ [X_S]$ for all $\{a, c\} \subseteq S \subseteq V$
- $X_c \longrightarrow X_b \ [X_S]$ for all $\{c, b\} \subseteq S \subseteq V$

Rules of causal inference

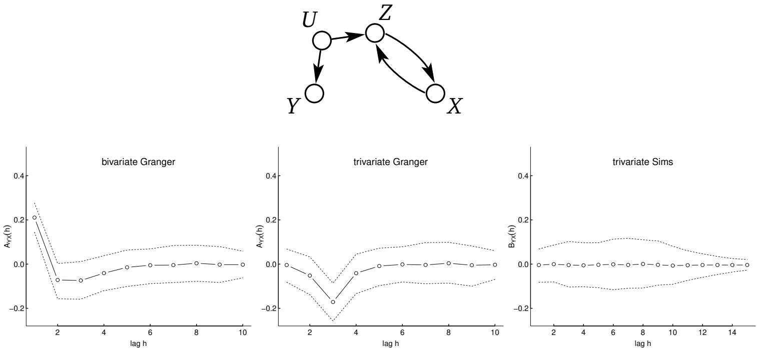
- ***Indirect causality rule:*** X_a truely causes X_b if

$$X_a \nrightarrow X_b \ [S] \qquad \text{for some } S \subseteq V \text{ with } c \in S$$

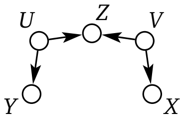
- ***Spurious causality rule:*** X_a is a spurious cause of X_b if

$$X_a \nrightarrow X_b \ [S] \qquad \text{for some } S \subseteq V \text{ with } c \notin S$$

Principles of causal inference



Principles of causal inference



Identification of causal structure

Algorithm: identification of adjacencies

- insert $a \dashrightarrow\!\!\!- b$ whenever X_a and X_b are not contemporaneously independent
- insert $a \dashrightarrow b$ whenever
  - $X_b \rightarrow X_a \ [X_S]$ for all $S \subseteq V$ with $a, b \in S$;
  - $X_a(t-k) \not\!\perp\!\!\!\perp X_b(t+1) \,|\, \mathscr{F}_{S_1}(t) \vee \mathscr{F}_{S_2}(t-k) \vee \mathscr{F}_a(t-k-1)$ for all $k \in \mathbb{N}$, $t \in \mathbb{Z}$, for all disjoint $S_1, S_2 \subseteq V$ with $b \in S_1$ and $a \notin S_1 \cup S_2$.

    (combination of Granger and Sims causality)

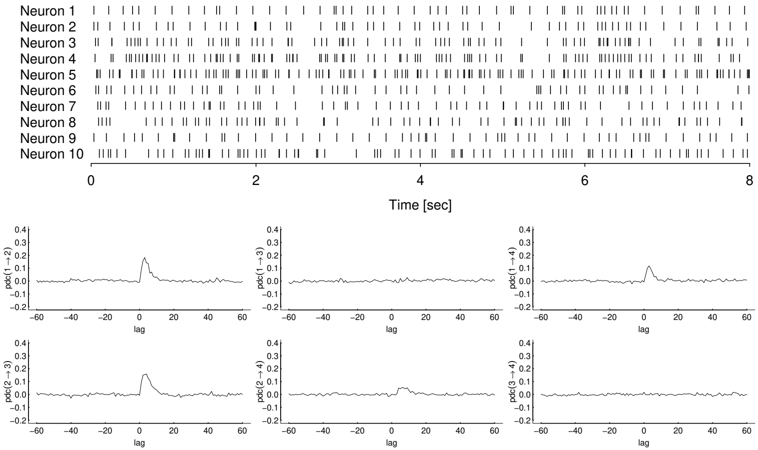
Identification of causal structure

Algorithm: identification of tails

- *colliders:*
  $a \cdots\!\!\rightarrow c \cdots\!\!\rightarrow b \in G$ and $X_a \nrightarrow X_b \;[X_S]$ for some S such that $c \notin S$
  $\Rightarrow \; c \cdots\!\!\rightarrow b \rightsquigarrow c \dashrightarrow b$
- *non-colliders:*
  $a \cdots\!\!\rightarrow c \cdots\!\!\rightarrow b \in G$ and $X_a \nrightarrow X_b \;[X_S]$ for some S such that $c \in S$
  $\Rightarrow \; c \cdots\!\!\rightarrow b \rightsquigarrow c \longrightarrow b$
- *ancestors:*
  $a \longrightarrow \ldots \longrightarrow b$ in $G \quad \Rightarrow \quad a \cdots\!\!\rightarrow b \rightsquigarrow a \longrightarrow b$
- *discriminating paths:* e.g. Ali et al. (2004)

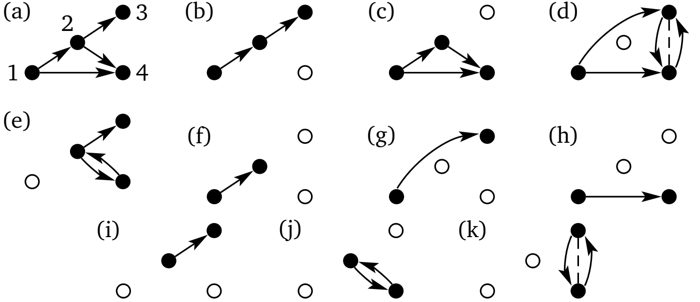
Identification of causal structure

Example: application to neural spike train data

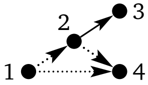


Identification of causal structure

Example:



Result:



Interpretation

Question: What can we learn from the graphical analysis about the true causal structure?

Interpretation

Question: What can we learn from the graphical analysis about the true causal structure?

- *Causal Markov assumption:*

  all observed dependencies are due to causal influences; key assumption underlying all approaches to causal inference from observational data

Interpretation

Question: What can we learn from the graphical analysis about the true causal structure?

- *Causal Markov assumption:*

  all observed dependencies are due to causal influences; key assumption underlying all approaches to causal inference from observational data
- *Faithfulness assumption:*

  the independencies observed are structural and not due to several influences exactly cancelling out ($\rightsquigarrow$ spurious causality of type I)

Interpretation

Question: What can we learn from the graphical analysis about the true causal structure?

- *Causal Markov assumption:*

  all observed dependencies are due to causal influences; key assumption underlying all approaches to causal inference from observational data
- *Faithfulness assumption:*

  the independencies observed are structural and not due to several influences exactly cancelling out ($\rightsquigarrow$ spurious causality of type I)
- *Invariance of edges:*

  only edges that are invariant in the Markov equivalence class are identified as
  - causes if the edge is $\longrightarrow$;
  - spurious causes if the edge is $\dashrightarrow$.

Interpretation

Question: What can we learn from the graphical analysis about the true causal structure?

- *Causal Markov assumption:*

  all observed dependencies are due to causal influences; key assumption underlying all approaches to causal inference from observational data
- *Faithfulness assumption:*

  the independencies observed are structural and not due to several influences exactly cancelling out ($\rightsquigarrow$ spurious causality of type I)
- *Invariance of edges:*

  only edges that are invariant in the Markov equivalence class are identified as
  - causes if the edge is $\longrightarrow$;
  - spurious causes if the edge is $\dashrightarrow$.
- *Ancestrality of graphs:*

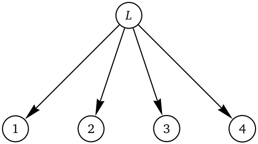
  Directed edges cannot be interpreted as direct causes, but signify a causal link.

Outline

- Causality concepts
- Graphical representation
  - Definition
  - Markov properties
  - Extension: systems with latent variables
- Causal learning
  - Basic principles
  - Identification from empirical relationships
- **Non-Markovian constraints**
  - Trek-separation in graphs
  - Tetrad representation theorem
  - Testing for tetrad constraints
- Open problems and conclusions

Problem

Example:



- X_1, X_2, X_3, X_4 are conditionally independent given L
- no conditional independences among $X_1, \ldots, X_4$.

Trek separation

Problem:

- conditional independences are not sufficient to describe processes that involve latent variables
- identification of such structures relies on sparsity that is often not given

Approach: Sullivant et al (2011) for multivariate Gaussian distributions

- new concept of separation in graphs
- encodes rank constraints on minors of covariance matrix
- generalizes other concepts of separation
- special case: conditional independences

Trek separation

A *trek* between nodes i and j is a path $\pi = (\pi_L, \pi_M, \pi_R)$ such that

- π_L is a directed path from some node k_L to i;
- π_R is a directed path from some node k_R to j;
- π_M is an undirected edge k_L --- k_R or a path of length zero ($k_L = k_R$).

Examples: $i \longleftarrow k_R$ --- $k_L \longrightarrow j$, $i \longleftarrow v \longleftarrow k \longrightarrow j$, $i \longrightarrow v \longrightarrow j$, i --- j

Definition (trek separation)

(C_L, C_R) *t-separates* sets A and B if for every trek (π_L, π_M, π_R)

- π_L contains a vertex in C_L or
- π_R contains a vertex in C_R.

Trek separation

Let X be a stationary Gaussian process with spectral matrix $\Sigma(\omega)$ satisfying

$$\Sigma(\omega) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \operatorname{cov}(X_t, X_{t-u})\, e^{-\mathrm{i}\, u\, \omega}.$$

Theorem

Let X be G-Markov. Then the following are equivalent:

- $\operatorname{rank}(\Sigma_{AB}(\omega)) \leq r$ for all $\omega \in [-\pi, \pi]$
- A and B are t-separated by some (C_L, C_R) with $|C_L| + |C_R| \leq r$.

Trek separation

Corollaries:

Let X be Gaussian stationary process. Then

$$X_A \perp\!\!\!\perp X_B \,|\, X_C \iff \operatorname{rank}(\Sigma_{A\cup C, B\cup C}) = |C|.$$

Furthermore the following are equivalent:

- $X_A \perp\!\!\!\perp X_B \,|\, X_C$ for all G-Markov processes X;
- (C_A, C_B) t-separates $A \cup C$ and $B \cup C$ for some partition $C = C_A \cup C_B$.

Tetrad representation theorem

Consider the class $\mathscr{M}(G)$ of all G-Markov stationary Gaussian processes

Proposition

The following are equivalent:

- The spectral matrices $\Sigma(\cdot)$ of processes in $\mathscr{M}(G)$ satisfy

$$\Sigma_{ik}(\omega)\,\Sigma_{jl}(\omega) - \Sigma_{il}(\omega)\,\Sigma_{jk}(\omega) = 0;$$

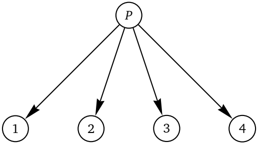
- $\{i,j\}$ and $\{k,l\}$ are t-separated by $(c,\varnothing)$ or $(\varnothing,c)$ for some node c in G

Tetrad representation theorem

If the spectral matrix $\Sigma(\omega)$ satisfies the tetrad constraints

$$\Sigma_{ik}(\omega)\Sigma_{jl}(\omega) - \Sigma_{il}(\omega)\Sigma_{jk}(\omega) = 0$$
$$\Sigma_{ij}(\omega)\Sigma_{kl}(\omega) - \Sigma_{il}(\omega)\Sigma_{kj}(\omega) = 0$$
$$\Sigma_{ik}(\omega)\Sigma_{lj}(\omega) - \Sigma_{ij}(\omega)\Sigma_{lk}(\omega) = 0$$

then there exists a node P such that X_i, X_j, X_k, and X_l are mutually conditionally independent given X_P.



Note: If no such X_P is among the observed variables, X_P must be a latent factor.

Latent variable models

Common identifiability constraint for factor models:

factors are uncorrelated/independent

But: in many applications (eg in neuroscience), we think of latent variables that are causally connected.

- EEG recordings measures neural activity in close cortical regions
- fMRI recordings measure hemodynamic responses which depend on underlying neural activity

Objective: recover latent processes and interrelations among them

Latent variable models

Suppose that $Y(t)$ can be partioned into $Y_{I_1}(t), \ldots, Y_{I_r}(t)$ such that

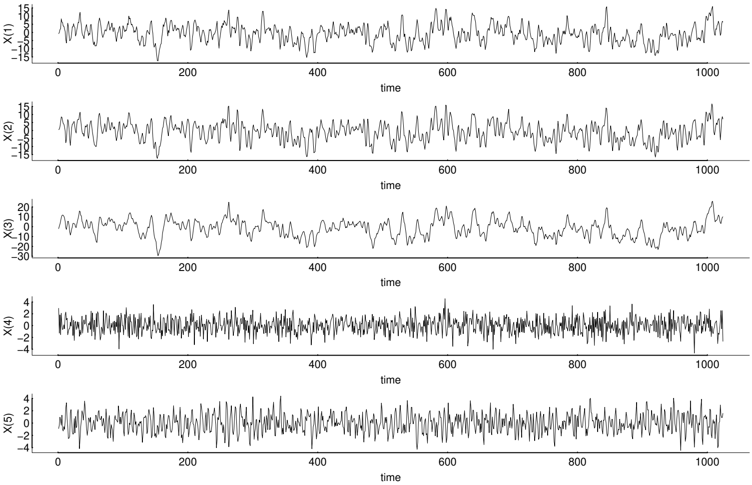
$$Y_{I_j}(t) = \Lambda_j X_j(t) + \varepsilon_{I_j}(t)$$

and $X(t)$ is a VAR(p) process.

Then the model can be fitted by the following steps:

- identify clusters of variables depending on one latent variable (based on tetrad rules)
- use PCA to determine latent variable processes $X_j(t)$
- fit VAR model to all latent variable processes jointly

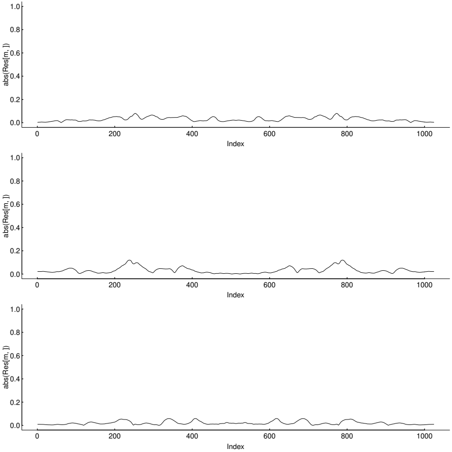
Latent variable models
Example



Latent variable models
Example

Set $\{1,2\}$ ***with:***

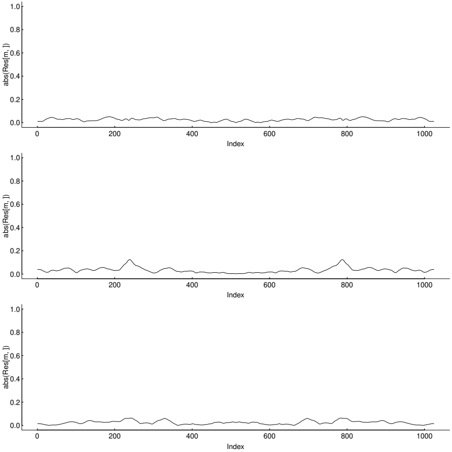
- $\{3,4\}$: $S = -0.98$
- $\{3,5\}$: $S = -0.31$
- $\{4,5\}$: $S = -1.4$



Latent variable models
Example

Set $\{1,3\}$ ***with:***

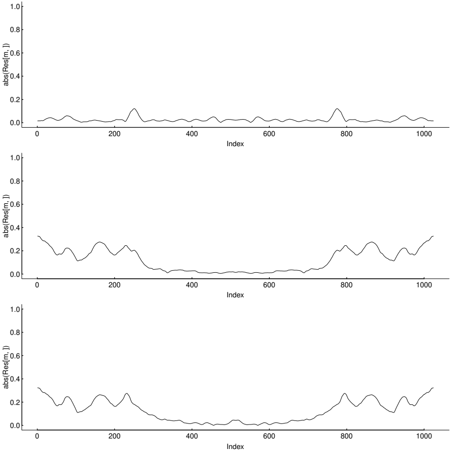
- $\{2,4\}$: $S = -1.37$
- $\{2,5\}$: $S = 0.76$
- $\{4,5\}$: $S = -0.44$



Latent variable models
Example

Set $\{1,4\}$ ***with:***

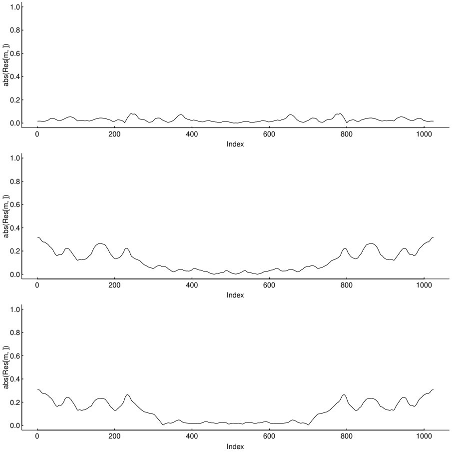
- $\{2,3\}$: $S = -1.19$
- $\{2,5\}$: $S = 6.54$
- $\{3,5\}$: $S = 6.55$



Latent variable models
Example

Set $\{1,5\}$ ***with:***

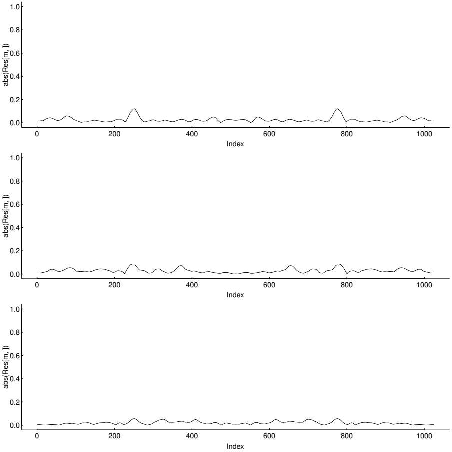
- $\{2,3\}$: $S = -1.22$
- $\{2,4\}$: $S = 5.43$
- $\{3,4\}$: $S = 5.77$



Latent variable models
Example

Set $\{2,3\}$ ***with:***

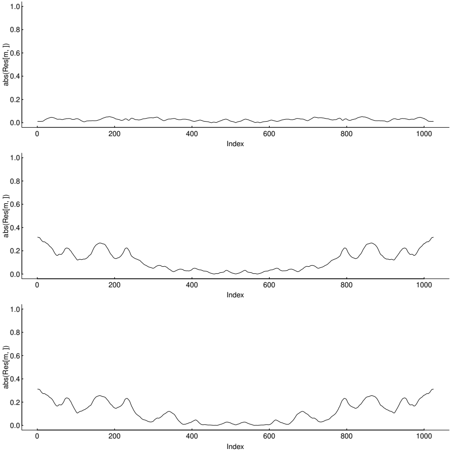
- $\{1,4\}$: $S = -1.18$
- $\{1,5\}$: $S = -1.21$
- $\{4,5\}$: $S = -1.58$



Latent variable models
Example

Set* $\{2,4\}$ *with:

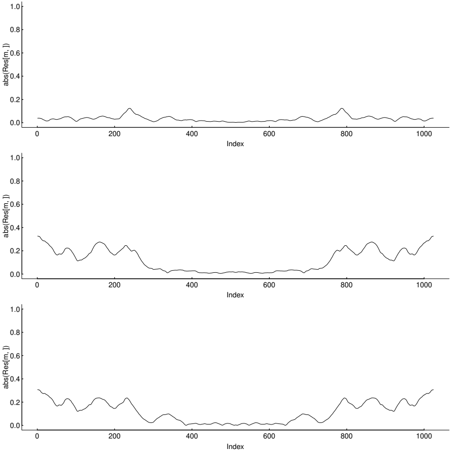
- $\{3,4\}$: $S = -1.36$
- $\{3,5\}$: $S = 5.43$
- $\{4,5\}$: $S = 5.66$



Latent variable models
Example

Set $\{2,5\}$ ***with:***

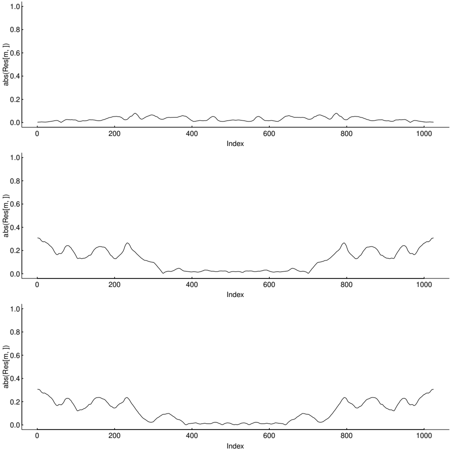
- $\{1,3\}$: $S = 0.76$
- $\{1,4\}$: $S = 6.55$
- $\{3,4\}$: $S = 5.73$



Latent variable models
Example

Set $\{3,4\}$ ***with:***

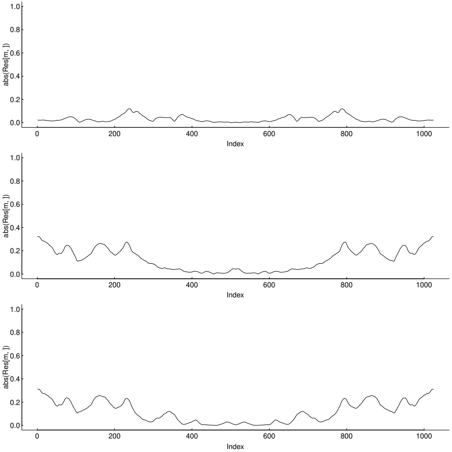
- $\{1,2\}$: $S = -0.98$
- $\{1,5\}$: $S = 5.77$
- $\{2,5\}$: $S = 5.73$



Latent variable models
Example

Set $\{3,5\}$ ***with:***

- $\{1,2\}$: $S = -0.31$
- $\{1,4\}$: $S = 6.54$
- $\{2,4\}$: $S = 5.66$

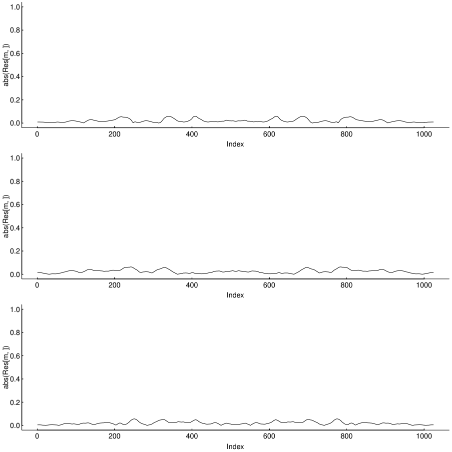


Latent variable models

Example

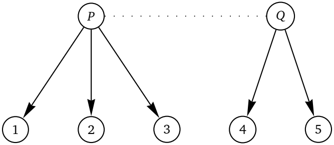
Set* $\{4, 5\}$ *with:

- $\{1, 2\}$: $S = -1.41$
- $\{1, 3\}$: $S = -0.44$
- $\{2, 3\}$: $S = -1.58$



Latent variable models

Example:



Conclusion

Causal Inference is a complex task

- requires modelling at all levels (bivariate to fully multivariate)
- requires Granger causality as well as other measures (e.g. Sims causality)
- definite results may be sparse without further assumptions
- latent variables induces further (non-Markovian) constraints on the distribution

Open Problems:

- merging of information about latent variables;
  development of algortihms for latent variables
- uncertainty in identification of Granger causal relationships
- instantaneous causality
- aggregation over time (distortion of identification only possible up to Markov equivalence
- non-stationarity and non-linearity

References

- E. (2007), Granger-causality and path diagrams for multivariate time series, *Journal of Econometrics* **137**, 334-353.
- E. (2008), Testing nonparametric and semiparametric hypotheses in vector stationary processes. *Journal of Multivariate Analysis* **99**, 968-1009.
- E. (2009), Causal inference from time series: what can be learned from Granger causality? In: G. Glymour, W. Wang, D. Westerståhl (eds), *Proceedings of the 13th International Congress of Logic, Methodology and Philosophy of Science*, College Publications, London.
- E. (2010), Graphical Modelling of multivariate time series with latent variables. *Journal of Machine Learning Research* **W&CP 9**
- E. (2012), Graphical modelling of multivariate time series. *Probability Theory and Related Fields* **153**, 233-268.
- E. (2012). Causal inference in time series analysis. In: C. Berzuini, A.P. Dawid, L. Bernardinelli (eds), *Causality: Statistical Perspectives and Applications*, Wiley, Chichester.
- E. (2013). Causal inference with multiple time series: principles and problems. *Philosophical Transaction of The Royal Society A* **371**, 20110613.