Exploratory Causal Analysis in Bivariate Time Series Data

J. M. McCracken
Department of Physics and Astronomy
George Mason University, Fairfax, VA

October 29, 2015

Outline

1. **Causality studies**
2. **Data causality**
3. **Exploratory causal analysis**
4. **Making an ECA guess**
   - Transfer entropy difference
   - Granger causality statistic
   - Pairwise asymmetric inference
   - Weighed mean observed leaning
   - Lagged cross-correlation difference
5. **Empirical examples**
   - Cooling/Heating System Data
   - Snowfall Data
6. **Times series causality as data analysis**

Causality studies

The study of causality is as old as science itself

- Modern historians credit Aristotle with both the first theory of causality ("four causes") and an early version of the scientific method
- The modern study of causality is broadly interdisciplinary; far too broad to review in a short talk.

Illari and Russo's textbook[1] provides an overview of causality studies

[1] Illari, P., & Russo, F. (2014). Causality: Philosophical theory meets scientific practice. Oxford University Press.

Towards a taxonomy of causal studies

Paul Holland identified four types of causal questions[2]:

- the ultimate meaningfulness of the notion of causality
- the details of causal mechanisms
- the causes of a given effect
- the effects of a given cause

[2] Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Towards a taxonomy of causal studies

Paul Holland identified four types of causal questions[2]:

- **the ultimate meaningfulness of the notion of causality**
- **the details of causal mechanisms**
- the causes of a given effect
- the effects of a given cause

Foundational causality "Is a cause required to precede an effect?" or "How are causes and effects related in space-time?"

[2] Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Towards a taxonomy of causal studies

Paul Holland identified four types of causal questions[2]:

- the ultimate meaningfulness of the notion of causality
- the details of causal mechanisms
- **the causes of a given effect**
- **the effects of a given cause**

Foundational causality "Is a cause required to precede an effect?" or "How are causes and effects related in space-time?"

Data causality "Does smoking cause lung cancer?" or "Are traffic accidents caused by rain storms?"

[2] Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)
- potential outcomes (Rubin's counterfactuals)

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)
- potential outcomes (Rubin's counterfactuals)
- directed acyclic graphs (DAGs) with structural equation models (SEMs); popularized by Pearl as "structural causal models (SCMs)"

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)
- potential outcomes (Rubin's counterfactuals)
- directed acyclic graphs (DAGs) with structural equation models (SEMs); popularized by Pearl as "structural causal models (SCMs)"
- time series causality

Data causality

Data causality is data analysis to draw causal inferences

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)
- potential outcomes (Rubin's counterfactuals)
- directed acyclic graphs (DAGs) with structural equation models (SEMs); popularized by Pearl as "structural causal models (SCMs)"
- time series causality

There is no consensus on the best approach to data causality

Many authors consider their favored approach to be the exclusive *correct* approach.

Data causality

$\boxed{\text{Data causality is data analysis to draw causal inferences}}$

Approaches to data causality studies include

- design of experiments (e.g., Fisher randomization)
- potential outcomes (Rubin's counterfactuals)
- directed acyclic graphs (DAGs) with structural equation models (SEMs); popularized by Pearl as "structural causal models (SCMs)"
- **time series causality**

There is no consensus on the best approach to data causality

Many authors consider their favored approach to be the exclusive *correct* approach.

Time series causality

$\boxed{\text{Time series causality is data causality with time series data}}$

Time series causality

$\boxed{\text{Time series causality is data causality with time series data}}$

Approaches to time series causality can be roughly divided into five categories,

- Granger (model based approaches)
- Information-theoretic
- State space reconstruction (SSR)
- Correlation
- Penchant

Exploratory causal analysis

Language

Exploring causal structures in data sets is distinct from **confirming** causal structures in data sets.

Causal language used in ECA should not be conflated with other typical uses; i.e., “cause”, “effect”, “drive”, etc. are used as technical terms with definitions unrelated to their common, everyday definitions.

Exploratory causal analysis

Language

Exploring causal structures in data sets is distinct from **confirming** causal structures in data sets.

Causal language used in ECA should not be conflated with other typical uses; i.e., "cause", "effect", "drive", etc. are used as technical terms with definitions unrelated to their common, everyday definitions.

$\rightarrow$ and $\leftarrow$ will be used as shorthand for causal statements,
e.g., A drives B will be written as $A \rightarrow B$.

Exploratory causal analysis
Assumptions

A cause always precedes an effect.

This assumption is required for the operational definitions of causality.

A driver may be present in the data being analyzed.

This assumption may lead to issues of confounding.

Exploratory causal analysis

ECA guess vector approach

We will not favor a specific operational definition of causality $\Rightarrow$ we do not favor any particular tool

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$,

Exploratory causal analysis

ECA guess vector approach

We will not favor a specific operational definition of causality $\Rightarrow$ we do not favor any particular tool

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$,

ECA guess vector

Define a vector $\vec{g}$ where each element g_i is defined as either 0 if $\mathbf{X} \rightarrow \mathbf{Y}$, 1 if $\mathbf{X} \leftarrow \mathbf{Y}$, or 2 if no causal inference can be made. The value of each g_i comes from a specific time series causality tool.

Exploratory causal analysis

ECA guess vector approach

We will not favor a specific operational definition of causality $\Rightarrow$ we do not favor any particular tool

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$,

ECA guess vector

Define a vector $\vec{g}$ where each element g_i is defined as either 0 if $\mathbf{X} \rightarrow \mathbf{Y}$, 1 if $\mathbf{X} \leftarrow \mathbf{Y}$, or 2 if no causal inference can be made. The value of each g_i comes from a specific time series causality tool.

ECA guess

The *ECA guess* is either $\mathbf{X} \rightarrow \mathbf{Y}$, $\mathbf{Y} \rightarrow \mathbf{X}$, or undefined, with $g_i = 0 \; \forall g_i \in \vec{g} \Rightarrow \mathbf{X} \rightarrow \mathbf{Y}$ and $g_i = 1 \; \forall g_i \in \vec{g} \Rightarrow \mathbf{Y} \rightarrow \mathbf{X}$.

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 transfer entropy difference
- g_2 Granger log-likelihood statistics
- g_3 pairwise asymmetric inference (PAI)
- g_4 average weighted mean observed leaning
- g_5 lagged cross-correlation difference

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 **transfer entropy difference** information-theoretic
  g_2 Granger log-likelihood statistics
  g_3 pairwise asymmetric inference (PAI)
  g_4 average weighted mean observed leaning
  g_5 lagged cross-correlation difference

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 **transfer entropy difference** information-theoretic
  g_2 **Granger log-likelihood statistics** Granger
  g_3 pairwise asymmetric inference (PAI)
  g_4 average weighted mean observed leaning
  g_5 lagged cross-correlation difference

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 **transfer entropy difference** information-theoretic
  g_2 **Granger log-likelihood statistics** Granger
  g_3 **pairwise asymmetric inference (PAI)** SSR
  g_4 average weighted mean observed leaning
  g_5 lagged cross-correlation difference

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 **transfer entropy difference** information-theoretic
  g_2 **Granger log-likelihood statistics** Granger
  g_3 **pairwise asymmetric inference (PAI)** SSR
  g_4 **average weighted mean observed leaning** penchant
  g_5 lagged cross-correlation difference

Making an ECA guess

Our focus is on time series, so each causal inference $g_i \in \vec{g}$ will be drawn from a tool in one of each of the five time series causality categories.

- g_1 **transfer entropy difference** information-theoretic
- g_2 **Granger log-likelihood statistics** Granger
- g_3 **pairwise asymmetric inference (PAI)** SSR
- g_4 **average weighted mean observed leaning** penchant
- g_5 **lagged cross-correlation difference** correlation

Transfer entropy (g_1)

Shannon entropy

The uncertainty that a random variable $\mathbf{X}$ takes some specific value X_n is given by the Shannon (or information) entropy,

$$H_X = -\sum_{n=1}^{N_X} P(\mathbf{X} = X_n) \log_2 P(\mathbf{X} = X_n)$$

Transfer entropy (g_1)

Shannon entropy

The uncertainty that a random variable **X** takes some specific value X_n is given by the Shannon (or information) entropy,

$$H_X = -\sum_{n=1}^{N_X} \boxed{P(\mathbf{X} = X_n)} \log_2 \boxed{P(\mathbf{X} = X_n)}$$

$P(\mathbf{X} = X_n)$ is the probability that **X** takes the specific value X_n

Transfer entropy (g_1)

Shannon entropy

The uncertainty that a random variable $\mathbf{X}$ takes some specific value X_n is given by the Shannon (or information) entropy,

$$H_X = -\boxed{\sum_{n=1}^{N_X}} P(\mathbf{X} = X_n) \log_2 P(\mathbf{X} = X_n)$$

The sum is over all possible values of X_n; $n = 1, 2, \ldots, N_X$

Transfer entropy (g_1)

Shannon entropy

The uncertainty that a random variable $\mathbf{X}$ takes some specific value X_n is given by the Shannon (or information) entropy,

$$H_X = -\sum_{n=1}^{N_X} P(\mathbf{X} = X_n) \boxed{\log_2} P(\mathbf{X} = X_n)$$

The base of the logarithm sets the entropy units, which is "bits" here

Transfer entropy (g_1)

Shannon entropy example

Binary example (to help with intuition)

Consider a coin **C** that take the value H with probability p_H and T with probability p_T. The Shannon entropy is

$$H_C = -\left(p_H \log_2 p_H + p_T \log_2 p_T\right)$$

Transfer entropy (g_1)

Shannon entropy example

Binary example (to help with intuition)

Consider a coin **C** that take the value H with probability p_H and T with probability p_T. The Shannon entropy is

$$H_C = -\left(p_H \log_2 p_H + p_T \log_2 p_T\right)$$

completely uncertain of outcome

Fair coin $\Rightarrow p_H = p_T = 0.5 \Rightarrow H_C = 1$

Transfer entropy (g_1)

Shannon entropy example

Binary example (to help with intuition)

Consider a coin **C** that take the value H with probability p_H and T with probability p_T. The Shannon entropy is

$$H_C = -\left(p_H \log_2 p_H + p_T \log_2 p_T\right)$$

completely uncertain of outcome

Fair coin $\Rightarrow p_H = p_T = 0.5 \Rightarrow H_C = 1$

completely certain of outcome

Always heads (or tails) $\Rightarrow p_{H(T)} = 0, p_{T(H)} = 1 \Rightarrow H_C = 0$

Transfer entropy (g_1)

Shannon entropy example

Binary example (to help with intuition)

Consider a coin **C** that take the value H with probability p_H and T with probability p_T. The Shannon entropy is

$$H_C = -\left(p_H \log_2 p_H + p_T \log_2 p_T\right)$$

completely uncertain of outcome

Fair coin $\Rightarrow p_H = p_T = 0.5 \Rightarrow H_C = 1$

completely certain of outcome

Always heads (or tails) $\Rightarrow p_{H(T)} = 0, p_{T(H)} = 1 \Rightarrow H_C = 0$

(Entropy calculations almost always assume $0 \log_2 0 := 0$.)

Transfer entropy (g_1)

Mutual information

A pair of random variables $(\mathbf{X}, \mathbf{Y})$ have some mutual information given by

$$\begin{aligned} I_{X;Y} &= H_X + H_Y - H_{X,Y} \\ &= \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) \log_2 \frac{P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)}{P(\mathbf{X} = X_n) P(\mathbf{Y} = Y_m)} \end{aligned}$$

Transfer entropy (g_1)

Mutual information

A pair of random variables $(\mathbf{X}, \mathbf{Y})$ have some mutual information given by

$$
\begin{aligned}
I_{X;Y} &= H_X + H_Y - H_{X,Y} \\
&= \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} \boxed{P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)} \log_2 \frac{P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)}{P(\mathbf{X} = X_n)P(\mathbf{Y} = Y_m)}
\end{aligned}
$$

$P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)$ is the probability that $\mathbf{X}$ takes the specific value X_n and $\mathbf{Y}$ takes the specific value Y_m

Transfer entropy (g_1)

Mutual information

A pair of random variables $(\mathbf{X}, \mathbf{Y})$ have some mutual information given by

$$
\begin{aligned}
I_{X;Y} &= H_X + H_Y - H_{X,Y} \\
&= \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) \log_2 \boxed{\frac{P(\mathbf{X}=X_n, \mathbf{Y}=Y_m)}{P(\mathbf{X}=X_n)P(\mathbf{Y}=Y_m)}}
\end{aligned}
$$

If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then
$P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) = P(\mathbf{X} = X_n)P(\mathbf{Y} = Y_m) \Rightarrow I_{X;Y} = 0$

Transfer entropy (g_1)

Mutual information

A pair of random variables $(\mathbf{X}, \mathbf{Y})$ have some mutual information given by

$$\begin{aligned} I_{X;Y} &= H_X + H_Y - H_{X,Y} \\ &= \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) \log_2 \frac{P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)}{P(\mathbf{X} = X_n) P(\mathbf{Y} = Y_m)} \end{aligned}$$

The mutual information is symmetric; i.e., $I_{X;Y} = I_{Y;X}$

Transfer entropy (g_1)

Mutual information

A pair of random variables $(\mathbf{X}, \mathbf{Y})$ have some mutual information given by

$$\begin{aligned} I_{X;Y} &= H_X + H_Y - H_{X,Y} \\ &= \sum_{n=1}^{N_X} \sum_{m=1}^{N_Y} P(\mathbf{X} = X_n, \mathbf{Y} = Y_m) \log_2 \frac{P(\mathbf{X} = X_n, \mathbf{Y} = Y_m)}{P(\mathbf{X} = X_n) P(\mathbf{Y} = Y_m)} \end{aligned}$$

Schreiber proposed an extension of the mutual information to measure "information flow" by making it conditional and including assumptions about the temporal behavior $\mathbf{X}$ and $\mathbf{Y}$.

Transfer entropy (g_1)

Information flow

Suppose **X** and **Y** are both Markov processes. The directed flow of information from **Y** to **X** is given by the transfer entropy,

$$T_{Y\rightarrow X} = \sum_{n=1}^{N_X}\sum_{m=1}^{N_Y} p_{n+1,n,m}\log_2\frac{p_{n+1|n,m}}{p_{n+1|n}}$$

with

- $p_{n+1,n,m} = P(\mathbf{X}(t+1) = X_{n+1}, \mathbf{X}(t) = X_n, \mathbf{Y}(\tau) = Y_m)$
- $p_{n+1|n,m} = P(\mathbf{X}(t+1) = X_{n+1}|\mathbf{X}(t) = X_n, \mathbf{Y}(\tau) = Y_m)$
- $p_{n+1|n} = P(\mathbf{X}(t+1) = X_{n+1}|\mathbf{X}(t) = X_n)$

Transfer entropy (g_1)

Information flow

There is no directed information flow from **Y** to **X** if **X** is conditionally independent of **Y**; i.e.,

$$p_{n+1|n,m} = p_{n+1|n} \Rightarrow T_{Y \to X} = 0$$

Transfer entropy (g_1)

Information flow

There is no directed information flow from **Y** to **X** if **X** is conditionally independent of **Y**; i.e.,

$$p_{n+1|n,m} = p_{n+1|n} \Rightarrow T_{Y \to X} = 0$$

Operational causality (information-theoretic)

X causes **Y** if the directed information flow from **X** to **Y** is higher than the directed information flow from **Y** to **X**; i.e.,

$$\begin{aligned} T_{X \to Y} - T_{Y \to X} > 0 &\quad \Rightarrow \quad \mathbf{X} \to \mathbf{Y} \\ T_{X \to Y} - T_{Y \to X} < 0 &\quad \Rightarrow \quad \mathbf{Y} \to \mathbf{X} \\ T_{X \to Y} - T_{Y \to X} = 0 &\quad \Rightarrow \quad \text{no causal inference} \end{aligned}$$

Granger causality (g_2)

Granger's axioms

Consider a discrete universe with two time series $\mathbf{X} = \{X_t \mid t = 1, \ldots, n\}$ and $\mathbf{Y} = \{Y_t \mid t = 1, \ldots, n\}$, where $t = n$ is considered the present time. All knowledge available in the universe at all times $t \leq n$ is denoted as Ω_n.

Granger causality (g_2)

Granger's axioms

Consider a discrete universe with two time series $\mathbf{X} = \{X_t \mid t = 1, \ldots, n\}$ and $\mathbf{Y} = \{Y_t \mid t = 1, \ldots, n\}$, where $t = n$ is considered the present time. All knowledge available in the universe at all times $t \leq n$ is denoted as Ω_n.

Axiom 1

The past and present may cause the future, but the future cannot cause the past.

Granger causality (g_2)

Granger's axioms

Consider a discrete universe with two time series $\mathbf{X} = \{X_t \mid t = 1, \ldots, n\}$ and $\mathbf{Y} = \{Y_t \mid t = 1, \ldots, n\}$, where $t = n$ is considered the present time. All knowledge available in the universe at all times $t \leq n$ is denoted as Ω_n.

Axiom 1

The past and present may cause the future, but the future cannot cause the past.

Axiom 2

Ω_n contains no redundant information, so that if some variable $\mathbf{Z}$ is functionally related to one or more other variables, in a deterministic fashion, then $\mathbf{Z}$ should be excluded from Ω_n.

Granger causality (g_2)

Granger's axioms

Consider a discrete universe with two time series $\mathbf{X} = \{X_t \mid t = 1, \ldots, n\}$ and $\mathbf{Y} = \{Y_t \mid t = 1, \ldots, n\}$, where $t = n$ is considered the present time. All knowledge available in the universe at all times $t \leq n$ is denoted as Ω_n.

Granger's definition of causality

Given some set A, $\mathbf{Y}$ causes $\mathbf{X}$ if

$$P(X_{n+1} \in A|\Omega_n) \neq P(X_{n+1} \in A|\Omega_n - \mathbf{Y})$$

Granger causality (g_2)

Granger's axioms

Consider a discrete universe with two time series $\mathbf{X} = \{X_t \mid t = 1, \ldots, n\}$ and $\mathbf{Y} = \{Y_t \mid t = 1, \ldots, n\}$, where $t = n$ is considered the present time. All knowledge available in the universe at all times $t \leq n$ is denoted as Ω_n.

Granger's definition of causality

Given some set A, $\mathbf{Y}$ causes $\mathbf{X}$ if

$$P(X_{n+1} \in A | \Omega_n) \neq P(X_{n+1} \in A | \Omega_n - \mathbf{Y})$$

Granger's original goal was to make this notion of causality "operational".

Granger causality (g_2)
VAR models

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. Suppose there is a vector autoregressive (VAR) model that describes the pair,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{11}^i & A_{12}^i \\ A_{21}^i & A_{22}^i \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

Granger causality (g_2)

VAR models

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. Suppose there is a vector autoregressive (VAR) model that describes the pair,

$$\boxed{\begin{pmatrix} X_t \\ Y_t \end{pmatrix}} = \sum_{i=1}^{n} \begin{pmatrix} A^i_{11} & A^i_{12} \\ A^i_{21} & A^i_{22} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

The current time step t of $\mathbf{X}$ and $\mathbf{Y}$

Granger causality (g_2)

VAR models

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. Suppose there is a vector autoregressive (VAR) model that describes the pair,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{11}^i & A_{12}^i \\ A_{21}^i & A_{22}^i \end{pmatrix} \boxed{\begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix}} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

The current time step t of $\mathbf{X}$ and $\mathbf{Y}$ is modeled as a sum of n past steps

Granger causality (g_2)

VAR models

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. Suppose there is a vector autoregressive (VAR) model that describes the pair,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{11}^i & A_{12}^i \\ A_{21}^i & A_{22}^i \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

The current time step t of $\mathbf{X}$ and $\mathbf{Y}$ is modeled as a sum of n past steps of $\mathbf{X}$ *and* $\mathbf{Y}$,

Granger causality (g_2)

VAR models

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. Suppose there is a vector autoregressive (VAR) model that describes the pair,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{11}^i & A_{12}^i \\ A_{21}^i & A_{22}^i \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \boxed{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}$$

The current time step t of $\mathbf{X}$ and $\mathbf{Y}$ is modeled as a sum of n past steps of $\mathbf{X}$ *and* $\mathbf{Y}$, plus uncorrelated noise terms.

Granger causality (g_2)

Comparison of VAR models

Consider two different VAR models for the pair $(\mathbf{X}, \mathbf{Y})$,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{xx,i} & A_{xy,i} \\ A_{yx,i} & A_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{pmatrix}$$

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A'_{xx,i} & 0 \\ 0 & A'_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon'_{x,t} \\ \varepsilon'_{y,t} \end{pmatrix}$$

Granger causality (g_2)

Comparison of VAR models

Consider two different VAR models for the pair $(\mathbf{X}, \mathbf{Y})$,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{xx,i} & A_{xy,i} \\ A_{yx,i} & A_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{pmatrix}$$

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A'_{xx,i} & 0 \\ 0 & A'_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon'_{x,t} \\ \varepsilon'_{y,t} \end{pmatrix}$$

The *G-causality log-likelihood* statistic is defined as

$$F_{Y \to X} = \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|}$$

Granger causality (g_2)

Comparison of VAR models

Consider two different VAR models for the pair $(\mathbf{X}, \mathbf{Y})$,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{xx,i} & A_{xy,i} \\ A_{yx,i} & A_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{pmatrix}$$

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A'_{xx,i} & 0 \\ 0 & A'_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon'_{x,t} \\ \varepsilon'_{y,t} \end{pmatrix}$$

The *G-causality log-likelihood* statistic is defined as

$$F_{Y \to X} = \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|}$$

Covariance of **X** model residuals given no dependence on **Y**

Granger causality (g_2)

Comparison of VAR models

Consider two different VAR models for the pair $(\mathbf{X}, \mathbf{Y})$,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A_{xx,i} & A_{xy,i} \\ A_{yx,i} & A_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{pmatrix}$$

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} A'_{xx,i} & 0 \\ 0 & A'_{yy,i} \end{pmatrix} \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon'_{x,t} \\ \varepsilon'_{y,t} \end{pmatrix}$$

The *G-causality log-likelihood* statistic is defined as

$$F_{Y \to X} = \ln \frac{|\Sigma'_{xx}|}{|\Sigma_{xx}|}$$

Covariance of **X** model residuals given a possible dependence on **Y**

Granger causality (g_2)

G-causality log-likelihood statistic

If both VAR models fit (or "forecast") the data equally well, then there is no G-causality; i.e.,

$$|\Sigma'_{xx}| = |\Sigma_{xx}| \Rightarrow F_{Y \to X} = 0$$

Granger causality (g_2)

G-causality log-likelihood statistic

If both VAR models fit (or "forecast") the data equally well, then there is no G-causality; i.e.,

$$|\Sigma'_{xx}| = |\Sigma_{xx}| \Rightarrow F_{Y \to X} = 0$$

Operational causality (Granger)

X causes **Y** if the **X**-dependent forecast of **Y** decreases the **Y** model residual covariance (as compared to the **X**-independent forecast) more than the **Y**-dependent forecast of **X** decreases the **X** model residual covariance (as compared to the **Y**-independent forecast); i.e.,

$$\begin{aligned}
F_{X \to Y} - F_{Y \to X} > 0 &\quad \Rightarrow \quad \mathbf{X} \to \mathbf{Y} \\
F_{X \to Y} - F_{Y \to X} < 0 &\quad \Rightarrow \quad \mathbf{Y} \to \mathbf{X} \\
F_{X \to Y} - F_{Y \to X} = 0 &\quad \Rightarrow \quad \text{no causal inference}
\end{aligned}$$

Pairwise asymmetric inference (g_3)

State space reconstruction

Consider an embedding of the time series $\mathbf{X} = \{x_t \mid t = 0, 1 \ldots, L-1, L\}$ constructed from delayed time steps as

$$\tilde{\mathbf{X}} = \{\tilde{x}_t \mid t = 1 + (E-1)\tau, \ldots, L\}$$

with

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}\right)$$

Pairwise asymmetric inference (g_3)

State space reconstruction

Consider an embedding of the time series $\mathbf{X} = \{x_t \mid t = 0, 1 \ldots, L-1, L\}$ constructed from delayed time steps as

$$\tilde{\mathbf{X}} = \{\tilde{x}_t \mid t = 1 + (E-1)\tau, \ldots, L\}$$

with

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}\right)$$

- τ is the delay time step

Pairwise asymmetric inference (g_3)

State space reconstruction

Consider an embedding of the time series $\mathbf{X} = \{x_t \mid t = 0, 1 \ldots, L-1, L\}$ constructed from delayed time steps as

$$\tilde{\mathbf{X}} = \{\tilde{x}_t \mid t = 1 + (E-1)\tau, \ldots, L\}$$

with

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau} \right)$$

- τ is the delay time step
- E is the embedding dimension

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, y_t\right)$$

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, y_t\right)$$

1. Find the n nearest neighbors to $\tilde{x}_t$ (in $\tilde{\mathbf{X}}$), where "nearest" means smallest Euclidean distance, d; i.e., $d_1 < d_2 < \ldots < d_n$

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, y_t)$$

2. Create weights, w, from the nearest neighbors as

$$w_i = \frac{e^{-\frac{d_i}{d_1}}}{\sum_{j=1}^{n} e^{-\frac{d_j}{d_1}}}$$

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, y_t\right)$$

3. Construct the *cross-mapped* estimate of $\mathbf{Y}$ using the weights as

$$\mathbf{Y}|\tilde{\mathbf{X}} = \left\{ Y_t|\tilde{\mathbf{X}} = \sum_{i=1}^{n} w_i Y_{\hat{t}_i} \mid t = 1 + (E-1)\tau, \ldots, L \right\}$$

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = \left(x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, y_t\right)$$

Each cross-mapped point in the estimate of $\mathbf{Y}$, i.e.,

$$Y_t|\tilde{\mathbf{X}} = \sum_{i=1}^{n} \frac{e^{-d_i/d_1}}{\sum_{j=1}^{n} e^{-d_j/d_1}} Y_{\hat{t}_i}$$

depends on comparisons of

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = \left(x_t, \boxed{x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}}, y_t\right)$$

Each cross-mapped point in the estimate of $\mathbf{Y}$, i.e.,

$$Y_t|\tilde{\mathbf{X}} = \sum_{i=1}^{n} \frac{e^{-d_i/d_1}}{\sum_{j=1}^{n} e^{-d_j/d_1}} Y_{\hat{t}_i}$$

depends on comparisons of the **pasts** of $\mathbf{X}$

Pairwise asymmetric inference (g_3)

Cross-mapping

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The *shadow manifold* of $\mathbf{X}$ (labeled $\tilde{\mathbf{X}}$) is constructed from the points

$$\tilde{x}_t = (\boxed{x_t}, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(E-1)\tau}, \boxed{y_t})$$

Each cross-mapped point in the estimate of $\mathbf{Y}$, i.e.,

$$Y_t|\tilde{\mathbf{X}} = \sum_{i=1}^{n} \frac{e^{-d_i/d_1}}{\sum_{j=1}^{n} e^{-d_j/d_1}} Y_{\hat{t}_i}$$

depends on comparisons of the **pasts** of $\mathbf{X}$ and the **presents** of **X and Y**.

Pairwise asymmetric inference (g_3)

Cross-mapped correlation

A good cross-mapped estimate is defined as one that is strongly correlated with the original times series. The cross-mapped correlation is

$$C_{YX} = \left[\rho(\mathbf{Y}, \mathbf{Y}|\tilde{\mathbf{X}})\right]^2$$

where $\rho(\cdot)$ is Pearson's correlation coefficient.

Pairwise asymmetric inference (g_3)

Cross-mapped correlation

A good cross-mapped estimate is defined as one that is strongly correlated with the original times series. The cross-mapped correlation is

$$C_{YX} = \left[\rho(\mathbf{Y}, \mathbf{Y}|\tilde{\mathbf{X}})\right]^2$$

where $\rho\left(\cdot\right)$ is Pearson's correlation coefficient.

Cross-mapping interpretation

If similar histories of $\mathbf{X}$ (i.e., nearest neighbors in the shadow manifold) capably estimate $\mathbf{Y}$ (i.e., lead to $C_{YX} \approx 1$, or at least $C_{YX} \neq 0$), then the presence (or action) of $\mathbf{Y}$ in the system has been recorded in $\mathbf{X}$.

Pairwise asymmetric inference (g_3)

Cross-mapping interpretation of causality

A time series pair $(\mathbf{X}, \mathbf{Y})$ will have two cross-mapped correlations, C_{YX} and C_{XY}.

Pairwise asymmetric inference (g_3)

Cross-mapping interpretation of causality

A time series pair $(\mathbf{X}, \mathbf{Y})$ will have two cross-mapped correlations, C_{YX} and C_{XY}.

Operational causality (SSR)

X causes **Y** if similar histories of **Y** estimate **X** better than similar histories of **X** estimate **Y**, where the "similar histories" of one time series are used to estimate another time series through shadow manifold nearest neighbor weighting (cross-mapping); i.e.,

$$\begin{aligned} C_{YX} - C_{XY} < 0 &\quad\Rightarrow\quad \mathbf{X} \to \mathbf{Y} \\ C_{YX} - C_{XY} > 0 &\quad\Rightarrow\quad \mathbf{Y} \to \mathbf{X} \\ C_{YX} - C_{XY} = 0 &\quad\Rightarrow\quad \text{no causal inference} \end{aligned}$$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P\left(E|C\right) - P\left(E|\bar{C}\right)$$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = \boxed{P(E|C)} - P\left(E|\bar{C}\right)$$

$P(E|C)$ is the probability of some effect E given some cause C

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P\left(E|C\right) - \boxed{P\left(E|\bar{C}\right)}$$

$P\left(E|\bar{C}\right)$ is the probability of some effect E given no cause C

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\boxed{\rho_{EC}} = P\left(E|C\right) - P\left(E|\bar{C}\right)$$

So, the penchant is the probability of an effect E given a cause C minus the probability of that effect without the cause

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P\left(E|C\right) - P\left(E|\bar{C}\right)$$

In the psychology/medical literature, the causal penchant is known as the *Eells measure of causal strength* or *probability contrast*.

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P\left(E|C\right) - P\left(E|\bar{C}\right)$$

$\boxed{\text{If } C \text{ drives } E\text{, then it is expected that } \rho_{EC} > 0.}$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P\left(E|C\right) - \boxed{P\left(E|\bar{C}\right)}$$

$P\left(E|\bar{C}\right)$ is often considered "unobservable". It can be eliminated from the penchant formula using Bayes theorem.

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P(E|C)\left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)}$$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P(E|C)\left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)}$$

If E and C are independent, then $P(E|C) = P(E)$, which implies

$$\rho_{EC} = P(E) + \frac{P(E)P(C) - P(E)}{1 - P(C)} = P(E) - P(E) = 0$$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = P(E|C)\left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)}$$

Example (to help with intuition)

Consider C and E to be two fair coins, c_1 and c_2, being "heads"; i.e., $P(c_1 = \text{``}heads'') = 0.5$ and $P(c_2 = \text{``}heads'') = 0.5$. If the coins are independent, then

$$P(c_2 = \text{``}heads''|c_1 = \text{``}heads'') = P(c_2 = \text{``}heads'') = 0.5 \Rightarrow \rho_{EC} = 0$$

If they are completely dependent then

$$P(c_2 = \text{``}heads''|c_1 = \text{``}heads'') = 1 \text{ or } 0 \Rightarrow \rho_{EC} = 1 \text{ or } -1$$

Weighted mean observed leaning (g_4)

Causal penchant

The causal penchant $\rho_{EC} \in [1, -1]$ is

$$\rho_{EC} = \boxed{P(E|C)} \left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)}$$

This formula has the additional benefit of only needing to estimate one conditional probability from the data.

Weighted mean observed leaning (g_4)

Causal leaning

A difference of penchants can be used to compare different cause-effect assignments (i.e., different assumptions of what should be considered a cause and what should be considered an effect). The leaning is

$$\lambda_{EC} = \rho_{EC} - \rho_{CE}$$

Weighted mean observed leaning (g_4)

Causal leaning

A difference of penchants can be used to compare different cause-effect assignments (i.e., different assumptions of what should be considered a cause and what should be considered an effect). The leaning is

$$\lambda_{EC} = \rho_{EC} - \rho_{CE}$$

Leaning interpretation

If $\lambda_{EC} > 0$, then C drives E more than E drives C.

Weighted mean observed leaning (g_4)

Usefulness of the leaning

The usefulness of the leaning depends on two things,

Weighted mean observed leaning (g_4)

Usefulness of the leaning

The usefulness of the leaning depends on two things,

1. Operational definitions of C and E (called the *cause-effect assignment*)

Weighted mean observed leaning (g_4)

Usefulness of the leaning

The usefulness of the leaning depends on two things,

1. Operational definitions of C and E (called the *cause-effect assignment*)
2. Estimations of $P(C)$, $P(E)$, $P(C|E)$, and $P(E|C)$ from the data

Weighted mean observed leaning (g_4)

Usefulness of the leaning

The usefulness of the leaning depends on two things,

1. Operational definitions of C and E (called the *cause-effect assignment*)
2. Estimations of $P(C)$, $P(E)$, $P(C|E)$, and $P(E|C)$ from the data

The primary cause-effect assignment will be the *l-standard assignment*,

l-standard assignment

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The l-standard assignment initially assumes the cause is the l lagged time step of $\mathbf{X}$ and the effect is the current time step of $\mathbf{Y}$; i.e., $\{C, E\} = \{x_{t-l}, y_t\}$.

Weighted mean observed leaning (g_4)

Usefulness of the leaning

The usefulness of the leaning depends on two things,

1. Operational definitions of C and E (called the *cause-effect assignment*)

2. Estimations of $P(C)$, $P(E)$, $P(C|E)$, and $P(E|C)$ from the data

The primary cause-effect assignment will be the *l-standard assignment*,

***l*-standard assignment**

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The l-standard assignment initially assumes the cause is the l lagged time step of $\mathbf{X}$ and the effect is the current time step of $\mathbf{Y}$; i.e., $\{C, E\} = \{x_{t-l}, y_t\}$.

Probabilities will estimated using data frequency counts.

Weighted mean observed leaning (g_4)

Leaning from the data

The cause-effect assignment must be specific if the probabilities are to be estimated with frequency counts and need to include *tolerance domains* to account for noise in the measurements.

Weighted mean observed leaning (g_4)

Leaning from the data

Consider the time series pair $(\mathbf{X}, \mathbf{Y})$. The penchant calculation depends on the conditional $P(y_t = a|x_{t-l} = b)$, where $a \in \mathbf{Y}$ and $b \in \mathbf{X}$. This conditional will be estimated as

$$P(y_t \in [a - \delta_y^L, a + \delta_y^R] | x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]) = \frac{n_{a \cap b}}{n_b}$$

Weighted mean observed leaning (g_4)

Leaning from the data

Consider the time series pair $(\mathbf{X}, \mathbf{Y})$. The penchant calculation depends on the conditional $P(y_t = a | x_{t-l} = b)$, where $a \in \mathbf{Y}$ and $b \in \mathbf{X}$. This conditional will be estimated as

$$P(y_t \in [a - \delta_y^L, a + \delta_y^R] | x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]) = \frac{\boxed{n_{a \cap b}}}{n_b}$$

$n_{a \cap b}$ is the number of times $y_t \in [a - \delta_y^L, a + \delta_y^R]$ and $x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]$ in $(\mathbf{X}, \mathbf{Y})$

Weighted mean observed leaning (g_4)

Leaning from the data

Consider the time series pair $(\mathbf{X}, \mathbf{Y})$. The penchant calculation depends on the conditional $P(y_t = a | x_{t-l} = b)$, where $a \in \mathbf{Y}$ and $b \in \mathbf{X}$. This conditional will be estimated as

$$P(y_t \in [a - \delta_y^L, a + \delta_y^R] | x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]) = \frac{n_{a \cap b}}{n_b}$$

n_b is the number of times $x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]$ in $\mathbf{X}$

Weighted mean observed leaning (g_4)

Leaning from the data

Consider the time series pair $(\mathbf{X}, \mathbf{Y})$. The penchant calculation depends on the conditional $P(y_t = a | x_{t-l} = b)$, where $a \in \mathbf{Y}$ and $b \in \mathbf{X}$. This conditional will be estimated as

$$P(y_t \in [a - \boxed{\delta_y^L}, a + \boxed{\delta_y^R}] | x_{t-l} \in [b - \boxed{\delta_x^L}, b + \boxed{\delta_x^R}]) = \frac{n_{a \cap b}}{n_b}$$

The tolerance domains are usually considered symmetric; i.e., $\delta_x^L = \delta_x^R$ and $\delta_y^L = \delta_y^R$

Weighted mean observed leaning (g_4)

Leaning from the data

Consider the time series pair $(\mathbf{X}, \mathbf{Y})$. The penchant calculation depends on the conditional $P(y_t = a | x_{t-l} = b)$, where $a \in \mathbf{Y}$ and $b \in \mathbf{X}$. This conditional will be estimated as

$$P(y_t \in [a - \delta_y^L, a + \delta_y^R] | x_{t-l} \in [b - \delta_x^L, b + \delta_x^R]) = \frac{n_{a \cap b}}{n_b}$$

The causal inference implied by the leaning calculations are dependent on both the cause-effect assignment and the tolerance domains.

Weighted mean observed leaning (g_4)

Weighted mean

Any time series pair $(\mathbf{X}, \mathbf{Y})$ will have many leanings; e.g., an l-standard assignment of $\{C, E\} = \{x_{t-l} = b \pm \delta_x, y_t = a \pm \delta_y\}$ will have a different leaning calculation for each $x_{t-1} \in [b - \delta_x, b + \delta_x]$ and $y_t \in [a - \delta_y, a + \delta_y]$.

Weighted mean observed leaning (g_4)

Weighted mean

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ and some cause-effect assignment $\{C, E\}$ for which reasonable tolerance domains have been defined.

Weighted mean observed leaning (g_4)

Weighted mean

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ and some cause-effect assignment $\{C, E\}$ for which reasonable tolerance domains have been defined.

Any penchant calculation for which the (estimated) conditional $P(E|C) \neq 0$ (or $P(C|E) \neq 0$) is called an observed penchant.

Weighted mean observed leaning (g_4)

Weighted mean

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ and some cause-effect assignment $\{C, E\}$ for which reasonable tolerance domains have been defined.

Any penchant calculation for which the (estimated) conditional $P(E|C) \neq 0$ (or $P(C|E) \neq 0$) is called an *observed* penchant.

The weighed mean observed penchant, $\langle \rho_{EC} \rangle_w$, is the weighed algebraic mean of the observed penchants.

Weighted mean observed leaning (g_4)

Weighted mean

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ and some cause-effect assignment $\{C, E\}$ for which reasonable tolerance domains have been defined.

Any penchant calculation for which the (estimated) conditional $P(E|C) \neq 0$ (or $P(C|E) \neq 0$) is called an *observed* penchant.

The *weighed mean observed penchant*, $\langle \rho_{EC} \rangle_w$, is the weighed algebraic mean of the observed penchants.

The weighed mean observed leaning, $\langle \lambda_{EC} \rangle_w$, is the difference of the weighed mean observed penchants; i.e., $\langle \lambda_{EC} \rangle_w = \langle \rho_{EC} \rangle_w - \langle \rho_{CE} \rangle_w$

Weighted mean observed leaning (g_4)

Causal inference

Operational causality (penchant)

X causes **Y** if the weighted mean observed leaning is positive given a cause-effect assignment (and reasonable tolerance domains) in which the assumed cause **X** precedes the assumed effect **Y**; i.e.,

$$\begin{aligned} \langle \lambda_{EC} \rangle_w > 0 &\Rightarrow \mathbf{X} \rightarrow \mathbf{Y} \\ \langle \lambda_{EC} \rangle_w < 0 &\Rightarrow \mathbf{Y} \rightarrow \mathbf{X} \\ \langle \lambda_{EC} \rangle_w = 0 &\Rightarrow \text{no causal inference} \end{aligned}$$

given $C \in \mathbf{X}$, $E \in \mathbf{Y}$, and C precedes E.

Lagged cross-correlation difference (g_5)

Cross-correlation

The *cross-correlation* between two time series **X** and **Y** is

$$\rho^{xy} = \frac{E\left[\left(x_t - \mu_X\right)\left(y_t - \mu_Y\right)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Lagged cross-correlation difference (g_5)

Cross-correlation

The *cross-correlation* between two time series **X** and **Y** is

$$\rho^{xy} = \frac{E\left[\left(x_t - \boxed{\mu_X}\right)(y_t - \mu_Y)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Every point in **X** is compared to the **mean** of **X**

Lagged cross-correlation difference (g_5)

Cross-correlation

The *cross-correlation* between two time series **X** and **Y** is

$$\rho^{xy} = \frac{E\left[(x_t - \mu_X)\left(y_t - \boxed{\mu_Y}\right)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Every point in **Y** is compared to the **mean** of **Y**

Lagged cross-correlation difference (g_5)

Cross-correlation

The *cross-correlation* between two time series **X** and **Y** is

$$\rho^{xy} = \frac{E\left[\left(x_t - \mu_X\right)\left(y_t - \mu_Y\right)\right]}{\sqrt{\boxed{\sigma_X^2 \sigma_Y^2}}}$$

The product of the individual **variances** of **X** and **Y** is used as a normalization

Lagged cross-correlation difference (g_5)

Cross-correlation

The *cross-correlation* between two time series **X** and **Y** is

$$\rho^{xy} = \frac{E\left[\left(x_t - \mu_X\right)\left(y_t - \mu_Y\right)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Example (to help with intuition)

$$\mathbf{X} = \mathbf{Y} \Rightarrow \rho^{xy} = \frac{E\left[\left(x_t - \mu_X\right)\left(y_t - \mu_Y\right)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{E\left[\left(x_t - \mu_X\right)^2\right]}{\sigma_X^2} = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

Lagged cross-correlation difference (g_5)

Lagged cross-correlation

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The past of $\mathbf{Y}$ may be compared to the present of $\mathbf{X}$ by introducing a lag l into the cross-correlation calculation,

$$\rho_l^{xy} = \frac{E\left[\left(x_t - \mu_X\right)\left(y_{t-l} - \mu_Y\right)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Lagged cross-correlation difference (g_5)

Lagged cross-correlation

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$. The past of $\mathbf{Y}$ may be compared to the present of $\mathbf{X}$ by introducing a lag l into the cross-correlation calculation,

$$\rho_l^{xy} = \frac{E\left[(x_t - \mu_X)(y_{t-l} - \mu_Y)\right]}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Operational causality (correlation)

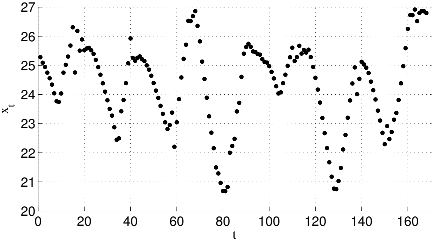
$\mathbf{X}$ causes $\mathbf{Y}$ (at lag l) if the past of $\mathbf{X}$ (i.e., $\mathbf{X}$ lagged by l time steps) is more strongly correlated with the present of $\mathbf{Y}$ than the past of $\mathbf{Y}$ (i.e., $\mathbf{Y}$ lagged by l time steps) is with the present of $\mathbf{X}$; i.e.,

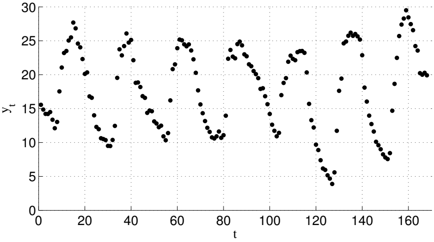
$$\begin{aligned}
|\rho_l^{xy}| - |\rho_l^{yx}| < 0 \quad &\Rightarrow \quad \mathbf{X} \rightarrow \mathbf{Y} \\
|\rho_l^{xy}| - |\rho_l^{yx}| > 0 \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \\
|\rho_l^{xy}| - |\rho_l^{yx}| = 0 \quad &\Rightarrow \quad \text{no causal inference}
\end{aligned}$$

Cooling/Heating System Data

Time series data

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ where $\mathbf{X}$ are indoor temperature measurements (in degrees Celsius) in a house with "experimental" environmental controls and $\mathbf{Y}$ is the temperature outside of that house, measured at the same time intervals (168 measurements in each series)[3]





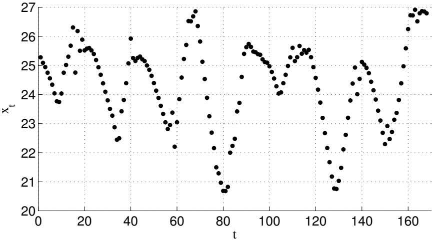
X **Y**

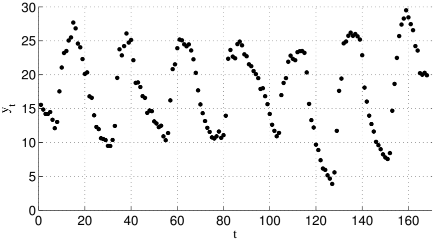
[3] This data was originally presented at a time series conference. The abstract is available here, http://www.osti.gov/scitech/biblio/5231321 . The data is also available as part of the UCI Machine Learning Repository.

Cooling/Heating System Data

Time series data

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ where $\mathbf{X}$ are indoor temperature measurements (in degrees Celsius) in a house with "experimental" environmental controls and $\mathbf{Y}$ is the temperature outside of that house, measured at the same time intervals (168 measurements in each series)[3]





X **Y**

$\boxed{\text{The intuitive causal inference is } \mathbf{Y} \to \mathbf{X}.}$

[3] This data was originally presented at a time series conference. The abstract is available here, http://www.osti.gov/scitech/biblio/5231321 . The data is also available as part of the UCI Machine Learning Repository.

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)
- cause-effect assignment and tolerance domains for g_4 (leaning)

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)
- cause-effect assignment and tolerance domains for g_4 (leaning)
- lags for g_5 (cross-correlation)

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)
- cause-effect assignment and tolerance domains for g_4 (leaning)
- lags for g_5 (cross-correlation)

The embedding dimension will be set (somewhat arbitrarily) to $E = 10$ and the time delay will be $\tau = 1$.

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)
- cause-effect assignment and tolerance domains for g_4 (leaning)
- lags for g_5 (cross-correlation)

The embedding dimension will be set (somewhat arbitrarily) to $E = 10$ and the time delay will be $\tau = 1$.

The tolerance domains will be the f-*width* tolerance domains; i.e., $\pm\delta_x = f(\max(\mathbf{X}) - \min(\mathbf{X}))$ and $\pm\delta_y = f(\max(\mathbf{Y}) - \min(\mathbf{Y}))$. For this example, $f = 1/4$.

Cooling/Heating System Data

ECA guess preliminaries

An ECA guess requires several parameters be set from the data, including

- embedding dimension and time delay for g_3 (PAI)
- cause-effect assignment and tolerance domains for g_4 (leaning)
- lags for g_5 (cross-correlation)

The embedding dimension will be set (somewhat arbitrarily) to $E = 10$ and the time delay will be $\tau = 1$.

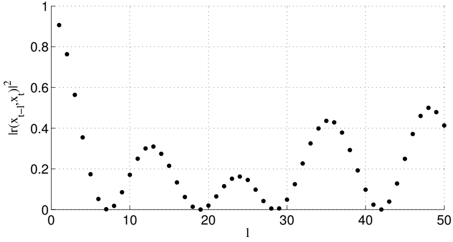
The tolerance domains will be the f-*width* tolerance domains; i.e., $\pm\delta_x = f(\max(\mathbf{X}) - \min(\mathbf{X}))$ and $\pm\delta_y = f(\max(\mathbf{Y}) - \min(\mathbf{Y}))$. For this example, $f = 1/4$.

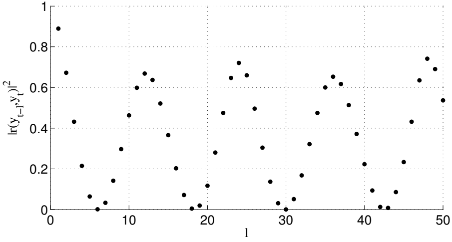
The cause-effect assignment will be the l-standard assignment, but **there is still the problem of determining relevant lags** l.

Cooling/Heating System Data

Autocorrelations

There are autocorrelations in both time series (only 50 lags are shown),





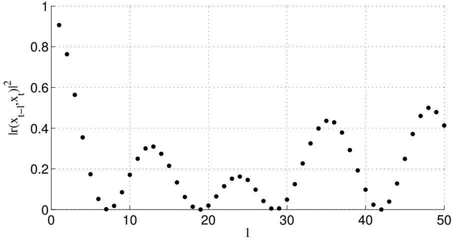
X **Y**

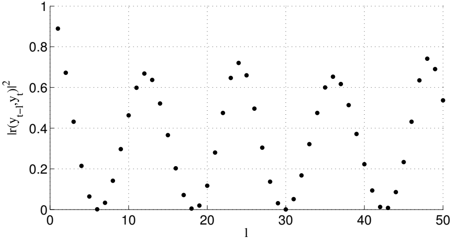
The autocorrelations appear cyclic and initially drop to zero around $l = 7$ for both time series.

Cooling/Heating System Data

Autocorrelations

There are autocorrelations in both time series (only 50 lags are shown),





X **Y**

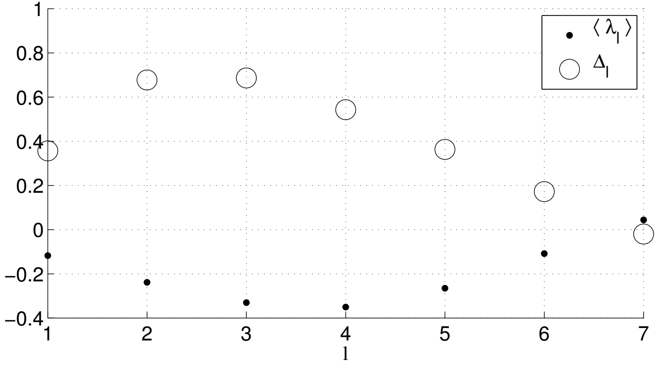
The autocorrelations appear cyclic and initially drop to zero around $l = 7$ for both time series.

This observation will be used justify using lags of $\boxed{l = 1, 2, \cdots, 7}$ for both g_4 (leaning) and g_5 (cross-correlation).

Cooling/Heating System Data

Lagged cross-correlations and leanings

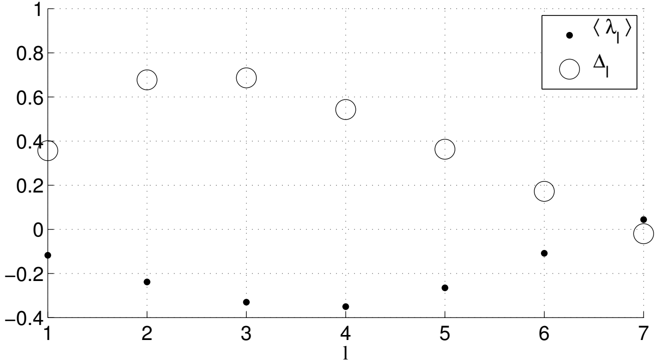
The lagged cross-correlations and leaning (using the *l*-standard assignment) can be plotted for each tested lag,



Cooling/Heating System Data

Lagged cross-correlations and leanings

The lagged cross-correlations and leaning (using the l-standard assignment) can be plotted for each tested lag,

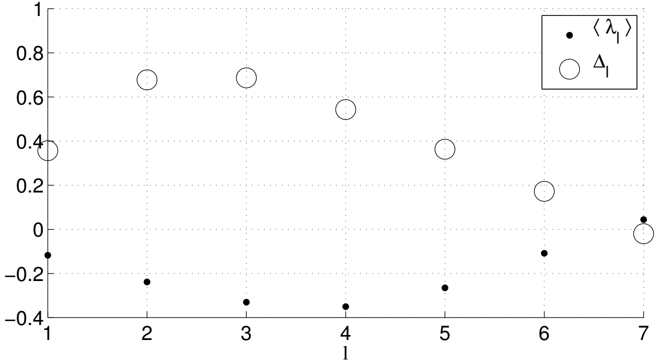


There are 7 different causal inferences in this plot, all of which agree except $l = 7$.

Cooling/Heating System Data

Lagged cross-correlations and leanings

The lagged cross-correlations and leaning (using the l-standard assignment) can be plotted for each tested lag,



There are 7 different causal inferences in this plot, all of which agree except $l = 7$. **A single causal inference (for each tool) will be found with the algebraic mean across all the tested lags.**

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$T_{X \to Y} - T_{Y \to X} = -0.14 \quad \Rightarrow \quad \mathbf{Y} \to \mathbf{X} \quad \Rightarrow \quad g_1 = 1$$

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{lcccc}
T_{X\to Y} - T_{Y\to X} = -0.14 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_1 = 1 \\
F_{X\to Y} - F_{Y\to X} = -0.35 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_2 = 1
\end{array}
$$

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{rcccl}
T_{X \to Y} - T_{Y \to X} = -0.14 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_1 = 1 \\
F_{X \to Y} - F_{Y \to X} = -0.35 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = 3.1 \times 10^{-4} & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_3 = 1
\end{array}
$$

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X\to Y} - T_{Y\to X} = -0.14 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_1 = 1 \\
F_{X\to Y} - F_{Y\to X} = -0.35 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = 3.1 \times 10^{-4} & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_3 = 1 \\
\langle\langle \lambda_{EC} \rangle_w \rangle = -0.20 & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_4 = 1
\end{array}
$$

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X\rightarrow Y} - T_{Y\rightarrow X} = -0.14 & \Rightarrow & \mathbf{Y}\rightarrow\mathbf{X} & \Rightarrow & g_1 = 1 \\
F_{X\rightarrow Y} - F_{Y\rightarrow X} = -0.35 & \Rightarrow & \mathbf{Y}\rightarrow\mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = 3.1\times 10^{-4} & \Rightarrow & \mathbf{Y}\rightarrow\mathbf{X} & \Rightarrow & g_3 = 1 \\
\langle\langle\lambda_{EC}\rangle_w\rangle = -0.20 & \Rightarrow & \mathbf{Y}\rightarrow\mathbf{X} & \Rightarrow & g_4 = 1 \\
\langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = 0.40 & \Rightarrow & \mathbf{Y}\rightarrow\mathbf{X} & \Rightarrow & g_5 = 1
\end{array}
$$

Cooling/Heating System Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

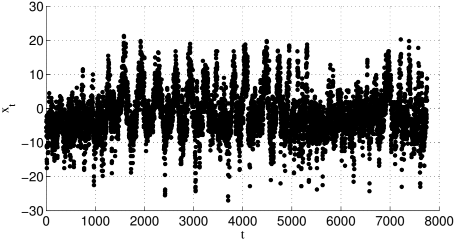
$$
\begin{aligned}
T_{X\rightarrow Y} - T_{Y\rightarrow X} = -0.14 \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \quad \Rightarrow \quad g_1 = 1 \\
F_{X\rightarrow Y} - F_{Y\rightarrow X} = -0.35 \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \quad \Rightarrow \quad g_2 = 1 \\
C_{YX} - C_{XY} = 3.1 \times 10^{-4} \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \quad \Rightarrow \quad g_3 = 1 \\
\langle\langle \lambda_{EC} \rangle_w\rangle = -0.20 \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \quad \Rightarrow \quad g_4 = 1 \\
\langle |\rho_l^{xy}| - |\rho_l^{yx}| \rangle = 0.40 \quad &\Rightarrow \quad \mathbf{Y} \rightarrow \mathbf{X} \quad \Rightarrow \quad g_5 = 1
\end{aligned}
$$

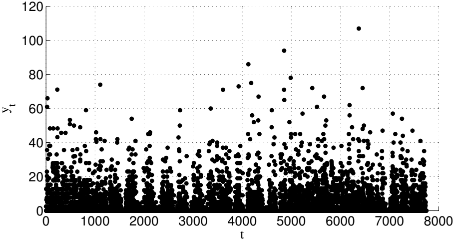
$\therefore$ the ECA guess is $\mathbf{Y} \rightarrow \mathbf{X}$, which agrees with intuition

Snowfall Data

Time series data

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ where $\mathbf{X}$ is the mean daily temperature (in degrees Celsius) at Whistler, BC, Canada, and $\mathbf{Y}$ is the total snowfall (in centimeters) (7,753 measurements in each series)[4]





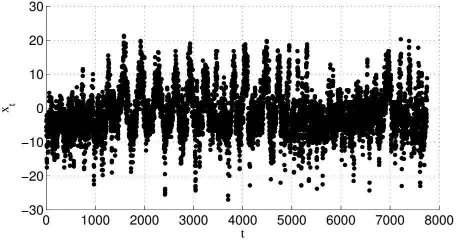
$\mathbf{X}$ $\mathbf{Y}$

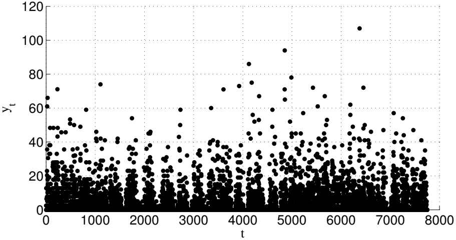
[3] This data is available as part of the UCI Machine Learning Repository. The data was recorded from July 1, 1972 to December 31, 2009.

Snowfall Data

Time series data

Consider a time series pair $(\mathbf{X}, \mathbf{Y})$ where $\mathbf{X}$ is the mean daily temperature (in degrees Celsius) at Whistler, BC, Canada, and $\mathbf{Y}$ is the total snowfall (in centimeters) (7,753 measurements in each series)[4]





X **Y**

$\boxed{\text{The intuitive causal inference is } \mathbf{X} \rightarrow \mathbf{Y}.}$

[3] This data is available as part of the UCI Machine Learning Repository. The data was recorded from July 1, 1972 to December 31, 2009.

Snowfall Data

ECA guess preliminaries

The ECA guess can be made with similar parameters as the previous example,

- The embedding dimension will be $E = 100$ with a time delay of $\tau = 1$
- The cause-effect assignment will be the l-standard assignment
- The tolerance domains will be the $1/4$-width domains
- The tested lags will be $l = 1, 2, \ldots, 20$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$T_{X \to Y} - T_{Y \to X} = 2.1 \times 10^{-2} \quad \Rightarrow \quad \mathbf{X} \to \mathbf{Y} \quad \Rightarrow \quad g_1 = 0$$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X\to Y} - T_{Y\to X} = 2.1 \times 10^{-2} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X\to Y} - F_{Y\to X} = -2.6 \times 10^{-3} & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_2 = 1
\end{array}
$$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X \to Y} - T_{Y \to X} = 2.1 \times 10^{-2} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X \to Y} - F_{Y \to X} = -2.6 \times 10^{-3} & \Rightarrow & \mathbf{Y} \to \mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = -3.4 \times 10^{-2} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_3 = 0
\end{array}
$$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X\rightarrow Y} - T_{Y\rightarrow X} = 2.1 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X\rightarrow Y} - F_{Y\rightarrow X} = -2.6 \times 10^{-3} & \Rightarrow & \mathbf{Y} \rightarrow \mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = -3.4 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_3 = 0 \\
\langle\langle \lambda_{EC} \rangle_w\rangle = 3.7 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_4 = 0
\end{array}
$$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{ccccc}
T_{X\rightarrow Y} - T_{Y\rightarrow X} = 2.1 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X\rightarrow Y} - F_{Y\rightarrow X} = -2.6 \times 10^{-3} & \Rightarrow & \mathbf{Y} \rightarrow \mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = -3.4 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_3 = 0 \\
\langle\langle \lambda_{EC} \rangle_w \rangle = 3.7 \times 10^{-2} & \Rightarrow & \mathbf{X} \rightarrow \mathbf{Y} & \Rightarrow & g_4 = 0 \\
\langle |\rho_l^{xy}| - |\rho_l^{yx}| \rangle = 2.3 \times 10^{-2} & \Rightarrow & \mathbf{Y} \rightarrow \mathbf{X} & \Rightarrow & g_5 = 1
\end{array}
$$

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$\begin{array}{rcccccc}
T_{X\to Y} - T_{Y\to X} = 2.1\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_1 = 0} \\
F_{X\to Y} - F_{Y\to X} = -2.6\times 10^{-3} & \Rightarrow & \mathbf{Y}\to\mathbf{X} & \Rightarrow & \boxed{g_2 = 1} \\
C_{YX} - C_{XY} = -3.4\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_3 = 0} \\
\langle\langle\lambda_{EC}\rangle_w\rangle = 3.7\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_4 = 0} \\
\langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = 2.3\times 10^{-2} & \Rightarrow & \mathbf{Y}\to\mathbf{X} & \Rightarrow & \boxed{g_5 = 1}
\end{array}$$

$\therefore$ the ECA guess is *undefined*

Snowfall Data

Making an ECA guess

Each of the five time series tools leads to a causal inference in the ECA guess vector,

$$
\begin{array}{cccccc}
T_{X\to Y} - T_{Y\to X} = 2.1\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_1 = 0} \\
F_{X\to Y} - F_{Y\to X} = -2.6\times 10^{-3} & \Rightarrow & \mathbf{Y}\to\mathbf{X} & \Rightarrow & g_2 = 1 \\
C_{YX} - C_{XY} = -3.4\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_3 = 0} \\
\langle\langle\lambda_{EC}\rangle_w\rangle = 3.7\times 10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & \boxed{g_4 = 0} \\
\langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = 2.3\times 10^{-2} & \Rightarrow & \mathbf{Y}\to\mathbf{X} & \Rightarrow & g_5 = 1
\end{array}
$$

$\therefore$ the ECA guess is *undefined*

The majority of the causal inferences agree with intuition.

Times series causality as data analysis

Objections to causal studies

Data analysis often ignores causality.

Times series causality as data analysis

Objections to causal studies

Data analysis often ignores causality.

Two primary objections to time series causality

1. Correlation is not causation
2. Confounding cannot be controlled

Times series causality as data analysis

Objections to causal studies

Data analysis often ignores causality.

Two primary objections to time series causality

1. **Correlation is not causation**
2. Confounding cannot be controlled

Many different tools have been developed that go beyond correlation and ignoring such tools means ignoring potentially useful inferences that can be drawn from the data.

Times series causality as data analysis

Objections to causal studies

Data analysis often ignores causality.

Two primary objections to time series causality

1. Correlation is not causation
2. **Confounding cannot be controlled**

Many different tools have been developed that go beyond correlation and ignoring such tools means ignoring potentially useful inferences that can be drawn from the data.

True, but this is an issue of defining "causality". **Exploring** potential causal relationships within data sets can be done with **operational** definitions of causality. These different causalities may provide deeper insight into the system dynamics.

BACK-UP

Impulse with linear response

Consider $\{\mathbf{X},\mathbf{Y}\} = \{\{x_t\},\{y_t\}\}$ where $t = 0, 1, \ldots, L$,

$$x_t = \begin{cases} 2 & t = 1 \\ A\eta_t & \forall\ t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall\ t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

and $y_t = x_{t-1} + B\eta_t$ with $y_0 = 0$, $A, B \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0,1)$. Specifically, consider $L = 500$, $A = 0.1$, and $B = 0.4$.

$$\begin{array}{rcccc} T_{X\to Y} - T_{Y\to X} = 5.3\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_1 = 0 \\ F_{X\to Y} - F_{Y\to X} = 4.5\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_2 = 0 \\ C_{YX} - C_{XY} = -8.3\times 10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_3 = 0 \\ \langle\langle\lambda_{EC}\rangle_w\rangle = 6.6\times 10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_4 = 0 \\ \langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = -2.8\times 10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_5 = 0 \end{array}$$

ECA guess is $\mathbf{X}\to\mathbf{Y}$, which agrees with intuition.

Cyclic driving with linear response

Consider $\{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\}$ where $t = 0, 1, \ldots, L$,

$$x_t = a \sin(bt + c) + A\eta_t$$

and

$$y_t = x_{t-1} + B\eta_t$$

with $y_0 = 0$, $A \in [0, 1]$, $B \in [0, 1]$, $\eta_t \sim \mathcal{N}(0, 1)$, and with the amplitude a, the frequency b, and the phase c all in the appropriate units. Specifically, consider $L = 500$, $A = 0.1$, $B = 0.4$, $a = b = 1$, and $c = 0$.

$$\begin{array}{cccccc}
T_{X \to Y} - T_{Y \to X} = 1.9 \times 10^{-1} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X \to Y} - F_{Y \to X} = 2.1 \times 10^{-1} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_2 = 0 \\
C_{YX} - C_{XY} = -9.8 \times 10^{-3} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_3 = 0 \\
\langle \langle \lambda_{EC} \rangle_w \rangle = 3.9 \times 10^{-3} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_4 = 0 \\
\langle |\rho_l^{xy}| - |\rho_l^{yx}| \rangle = -2.9 \times 10^{-2} & \Rightarrow & \mathbf{X} \to \mathbf{Y} & \Rightarrow & g_5 = 0
\end{array}$$

ECA guess is $\mathbf{X} \to \mathbf{Y}$, which agrees with intuition.

Cyclic driving with non-linear response

Consider$\{\mathbf{X},\mathbf{Y}\} = \{\{x_t\},\{y_t\}\}$ where $t = 0, 1, \ldots, L$,

$$x_t = a\sin(bt + c) + A\eta_t$$

and

$$y_t = Bx_{t-1}\left(1 - Cx_{t-1}\right) + D\eta_t,$$

with $y_0 = 0$, with $A, B, C, D \in [0,1]$, $\eta_t \sim \mathcal{N}(0,1)$, and with the amplitude a, the frequency b, and the phase c all in the appropriate units given $t = 0, f\pi, 2f\pi, 3f\pi, \ldots, 6\pi$ with $f = 1/30$, which implies $L = 181$. Specifically, consider $A = 0.1$, $B = 0.3$, $C = 0.4$, $D = 0.5$, $a = b = 1$, and $c = 0$.

$$\begin{array}{cccccc}
T_{X\to Y} - T_{Y\to X} = 2.7\times10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X\to Y} - F_{Y\to X} = 2.6\times10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_2 = 0 \\
C_{YX} - C_{XY} = -1.8\times10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_3 = 0 \\
\langle\langle\lambda_{EC}\rangle_w\rangle = 8.4\times10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_4 = 0 \\
\langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = -6.8\times10^{-2} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_5 = 0
\end{array}$$

ECA guess is $\mathbf{X}\to\mathbf{Y}$, which agrees with intuition.

Coupled logistic map

Consider $\{\mathbf{X},\mathbf{Y}\} = \{\{x_t\},\{y_t\}\}$ where $t = 0, 1, \ldots, L$,

$$x_t = x_{t-1}\left(r_x - r_x x_{t-1} - \beta_{xy} y_{t-1}\right)$$

and

$$y_t = y_{t-1}\left(r_y - r_y y_{t-1} - \beta_{yx} x_{t-1}\right)$$

where the parameters $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbb{R} \geq 0$. Specifically, consider $L = 500$, $\beta_{xy} = 0.5$, $\beta_{yx} = 1.5$, $r_x = 3.8$, and $r_y = 3.2$ with initial conditions $x_0 = y_0 = 0.4$.

$$\begin{array}{ccccc}
T_{X\to Y} - T_{Y\to X} = 4.9\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_1 = 0 \\
F_{X\to Y} - F_{Y\to X} = 5.4\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_2 = 0 \\
C_{YX} - C_{XY} = -3.9\times 10^{-3} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_3 = 0 \\
\langle\langle\lambda_{EC}\rangle_w\rangle = 2.7\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_4 = 0 \\
\langle|\rho_l^{xy}| - |\rho_l^{yx}|\rangle = -2.6\times 10^{-1} & \Rightarrow & \mathbf{X}\to\mathbf{Y} & \Rightarrow & g_5 = 0
\end{array}$$

ECA guess is $\mathbf{X}\to\mathbf{Y}$, which agrees with intuition.