Time Series

Seminar on causality and causal inference

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

10. Juli 2024

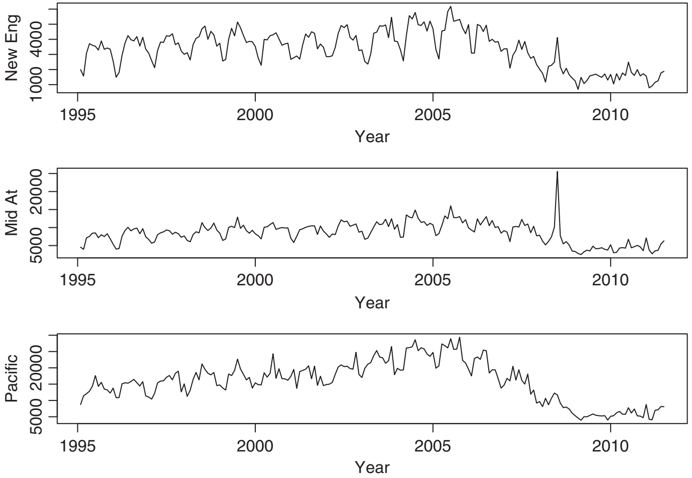


FIGURE 1.4 Time plots of the monthly housing starts for the New England, Middle Atlantic, and Pacific divisions of the United States from January 1995 to June 2011. The data are not seasonally adjusted.

The Figure 1.4 is from: Tsay, Ruey S..„Multivariate Time Series Analysis : With R and Financial Applications“

Preliminaries and Terminology

Here, we discuss causal inference in multivariate time series, that is, we have a d-variate time series $(\mathbf{X}_t)_{t\in\mathbb{Z}}$.

$$\mathbf{X}_t = \begin{pmatrix} X_t^1 \\ \vdots \\ X_t^d \end{pmatrix}$$

No Instantaneous Effects

The causal graph with nodes X_t^j contains only arrows from X_t^j to X_s^k for $t < s$ but not for $t = s$; see Figure 10.1.

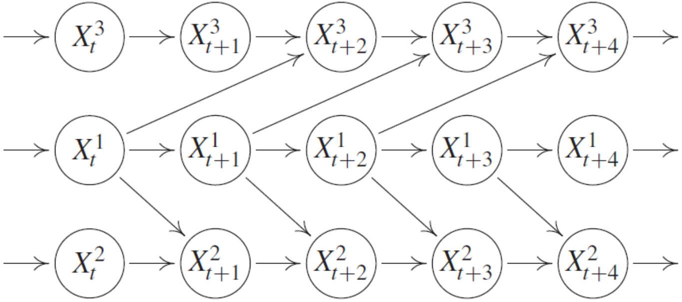


Figure 10.1: Example of a time series with no instantaneous effects.

Then we say there are no instantaneous effects.

Instantaneous Effects

The causal graph contains instantaneous effects, that is, arrows from X_t^j to X_t^k for some j and k in addition to arrows going from X_t^m to X_s^l with $t < s$ and some m and l, as shown in Figure 10.2

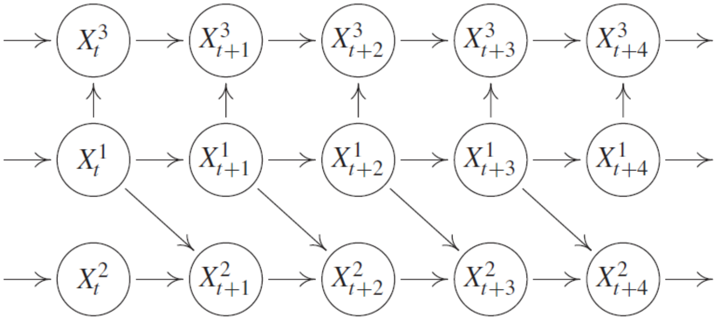
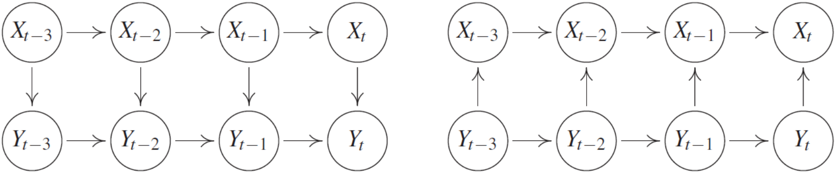


Figure 10.2: Example of a time series with instantaneous effects.

Purely Instantaneous

We call the causal structure purely instantaneous if for any $j \neq k$ and $h > 0$ the variable X_t^j may influence X_t^k and X_{t+h}^j but not X_{t+h}^k; see Figures 10.5(a) and 10.5(b).



(a) There are v-structures at all nodes of $(Y_t)_{t\in\mathbb{Z}}$.

(b) There are v-structures at all nodes of $(X_t)_{t\in\mathbb{Z}}$.

Figure 10.5: Two DAGs that are not Markov equivalent although they coincide up to instantaneous effects.

Further Definitions

- We define the full time graph as the DAG having X_t^i as nodes, as visualized in Figures 10.1 and 10.2. In contrast to previous chapters, the full time graph is a DAG with infinitely many nodes.
- The summary graph is the directed graph with nodes $X^1; \ldots ; X^d$ containing an arrow from X^j to X^k for $j \neq k$ whenever there is an arrow from X_t^j to X_s^k for some $t \leq s \in \mathbb{Z}$. Figure 10.3 shows the summary graph corresponding to the full time graphs depicted in Figures 10.1 and 10.2.

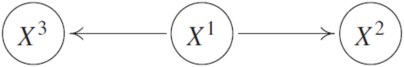


Figure 10.3: Summary graph of the full time graphs shown in Figures 10.1 and 10.2.

Note that the summary graph is a directed graph that may contain cycles, although we will assume that the full time graph is acyclic.

Further Notations

- For any $t \in \mathbb{Z}$, we denote by $\mathbf{X}_{past(t)}$ the set of of all $\mathbf{X}_s$ with $s < t$.
- $X^j_{past(t)}$ for the past of a specific component X^j.
- $(\mathbf{X}_t^{-j})_{t\in\mathbb{Z}}$ denotes the collection of time series $(X_t^k)_{t\in\mathbb{Z}}$ for all $k \neq j$.
- $(\mathbf{PA}_s^j)_{t-s}$ denotes the set of variables X^k_{t-s}, $k = 1, \dots, d$, that influence X_t^j. Note that $(\mathbf{PA}_s^j)_{t-s}$ may contain X^j_{t-s} for all $s > 0$, but not for $s = 0$.

Structural Causal Models

We assume that the stochastic process $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ admits a description by an SCM in which at most the past q values (for some q) of all variables occur:

$$X_t^j := f^j\left((\mathbf{PA}_q^j)_{t-q}, \dots, (\mathbf{PA}_1^j)_{t-1}, (\mathbf{PA}_0^j)_t, N_t^j\right)$$

where

$$\dots, N_{t-1}^1, \dots, N_{t-1}^d, N_t^1, \dots, N_t^d, N_{t+1}^1, \dots, N_{t+1}^d, \dots$$

are jointly independent noise terms. We assume the corresponding full time graph to be acyclic.

A popular special case is the class of vector autoregressive models (VAR) [Lütkepohl, 2007]:

$$X_t^j = \sum_{j=1}^{q} A_i^j \mathbf{X}_{t-i} + N_t^j$$

[Lütkepohl, 2007] H. Lütkepohl. New Introduction to Multiple Time Series Analysis. Springer, Berlin, Germany, 2007

Interventions

As in the i.i.d. setting, SCMs formalize the effect of interventions; more precisely, an intervention corresponds to replacing some of the structural assignments. Interventions may, for instance, consist in setting all values $(X_t^j)_{t \in \mathbb{Z}}$ for some j to certain values. Alternatively, one could also intervene on X_t^j only at one specific time instant t.

Identifiabilty in absence of instantaneous effects

Theorem 1

Assume that two full time graphs are induced by SCMs without instantaneous effects. If the full time graphs are Markov equivalent, then they are equal.

see Peters et al., [2013], Proof of Theorem 1.

Peters et al., [2013]: J. Peters, D. Janzing, and B. Schölkopf. Causal inference on time series using restricted structural equation models. In Advances in Neural Information Processing Systems 26 (NIPS), pages 154-162, 2013.

Identifiability for acyclic summary graphs

Theorem 2

Assume that two full time graphs are induced by SCMs, and that in both cases for each j, X_t^j is influenced by X_{t-s}^j for some $s \geq 1$. Assume further that the summary graphs are acyclic. If the full time graphs are Markov equivalent, then they are equal.

See Peters et al., [2013], Theorem 1.

Peters et al., [2013]: J. Peters, D. Janzing, and B. Schölkopf. Causal inference on time series using restricted structural equation models. In Advances in Neural Information Processing Systems 26 (NIPS), pages 154-162, 2013.

Justification of Granger causality

Theorem 3

Consider an SCM without instantaneous effects for the time series $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ such that the induced joint distribution is faithful with respect to the corresponding full time graph. Then the summary graph has an arrow from X^j to X^k if and only if there exists a $t \in \mathbb{Z}$ such that

$$X_t^k \not\perp\!\!\!\perp X_{past(t)}^j | \mathbf{X}_{past(t)}^{-j}$$

Detection of arrow $X \to Y$

Theorem 4

Consider an SCM for the bivariate time series $(X_t, Y_t)_{t\in\mathbb{Z}}$.

(i) If there is a $t \in \mathbb{Z}$ such that

$$Y_t \not\perp\!\!\!\perp X_{past(t)} | Y_{past(t)}$$

then the summary graph contains an arrow from X to Y.

(ii) Assume further that there are no instantaneous effects and the joint density of any finite subset of variables is strictly positive.
If for all $t \in \mathbb{Z}$, we have

$$Y_t \perp\!\!\!\perp X_{past(t)} | Y_{past(t)}$$

then the summary graph contains no arrow from X to Y.

Typical scenario, in which Theorem 4 holds

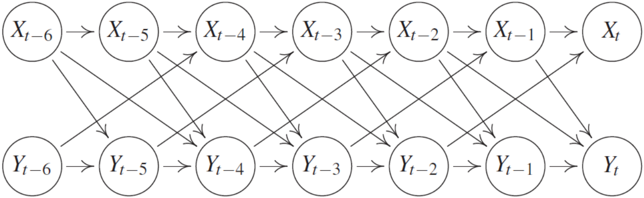


Figure 10.6: Typical scenario, in which Granger causality works: if all arrows from X to Y were missing, Y_t would be conditionally independent of the past values of X, given its own past. Here, Y_t does depend on the past values of X, given its own past. Thus, condition (10.4) proves the existence of an influence from X to Y.

Bivariate Granger Causality

Definition 1

$$X \; \textit{Granger} - \textit{causes} \; Y$$

$$\Longleftrightarrow$$

$$Y_t \not\perp\!\!\!\perp X_{past(t)} \,|\, Y_{past(t)}$$

This idea already goes back to Wiener [1956, pages 189-190], who argued that X has a causal influence on Y if the prediction of Y from its own past is improved by additionally accounting for X.

Wiener [1956]: N.Wiener. The theory of prediction. In E. Beckenbach, editor, Modern Mathematics for Engineers. McGraw-Hill, New York, NY, 1956.

Bivariate Granger Causality and linear prediction

Compare the following two linear regression models:

$$Y_t = \sum_{i=1}^{q} a_i Y_{t-i} + N_t \tag{1}$$

$$Y_t = \sum_{i=1}^{q} a_i Y_{t-i} + \sum_{i=1}^{q} b_i X_{t-i} + \tilde{N}_t \tag{2}$$

where $(N_t)_{t\in\mathbb{Z}}$ and $(\tilde{N}_t)_{t\in\mathbb{Z}}$ are assumed to be i.i.d. time series, respectively.

X is inferred to Granger-cause Y whenever the noise term $\tilde{N}_t$ (for predictions including X) has significantly smaller variance than the noise term N_t obtained without X.

Multivariate Granger Causality

Definition 2

$$X^j \text{ Granger} - \text{causes } X^k$$

$$\Longleftrightarrow$$

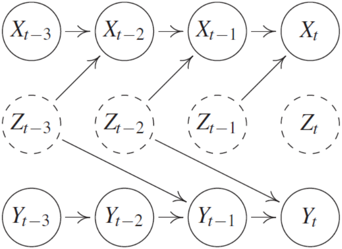
$$X_t^k \not\perp\!\!\!\perp X_{past(t)}^j | \mathbf{X}_{past(t)}^{-j}$$

Granger already emphasized that proper use of Granger causality would actually require to condition on all relevant variables in the world.

Limitations of Granger Causality I

Example 1

A bivariate time series $(X_t, Y_t)_{t\in\mathbb{Z}}$ is observed. Assume that both X_t and Y_t are influenced by previous instances of a hidden time series $(Z_t)_{t\in\mathbb{Z}}$. This is depicted in Figure 10.7(a) where Z influences X with a delay of 1, and Y with a delay of 2.



(a) Due to the hidden common cause Z, Granger causality erroneously infers causal influence from X to Y.

Limitations of Granger Causality II

Example 1

Assuming faithfulness, the d-separation criterion tells us

$$Y_t \not\perp\!\!\!\perp X_{past(t)} | Y_{past(t)},$$

while we have

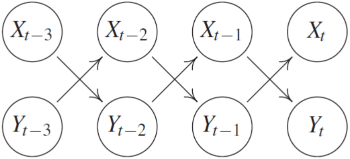
$$X_t \perp\!\!\!\perp Y_{past(t)} | X_{past(t)}.$$

Thus, naive application of Granger causality infers that X causes Y and Y does not cause X.

Limitations of Granger Causality III

Example 2

A second example for a scenario where Granger fails has been provided by Ay and Polani [2008] and is depicted in Figure 10.7(b).



(b) Granger causality erroneously infers neither causal influence from X to Y nor from Y to X if the influence from X_t on Y_{t+1} and the one from Y_t to X_{t+1} are deterministic.

Assume that X_{t-1} influences Y_t deterministically via a copy operation, that is, $Y_t := X_{t-1}$. Likewise, the value of Y_{t-1} is copied to X_t .

Limitations of Granger Causality IV

Example 2

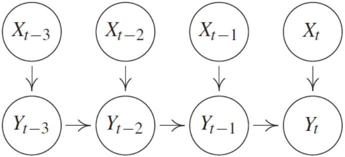
Then it is intuitively obvious that X and Y strongly influence each other in the sense that intervening on the value X_t changes all the values Y_{t+1+2k} for $k \in \mathbb{N}_0$. Likewise, intervening on Y_t changes all values X_{t+1+2k}. Nevertheless, the past of X is useless for predicting Y_t from its past, because Y_t can already be predicted perfectly from its own past.

Limitations of Granger Causality VI

Example 3a)

For Figure 10.8(a), d-separation yields

$$Y_t \perp\!\!\!\perp X_{past(t)} \big| Y_{past(t)}.$$



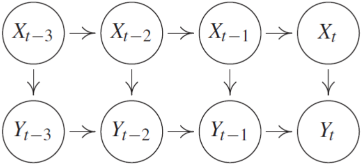
(a) Granger causality cannot detect the influence of X on Y because the past of X influences Y_t only via the past of Y.

Intuitively speaking, only the present value X_t would help for better predicting Y_t , but the past values $X_{t-1}, X_{t-2}, \ldots$ are useless and thus, Granger causality does not propose a link from X to Y.

Limitations of Granger Causality VII

Example 3b)

In Figure 10.8(b), however, Granger causality does detect the influence of X on Y (if we assume faithfulness) although it is still purely instantaneous, but the slight amount of improvement of the prediction does not properly account for the potentially strong influence of X_t on Y_t



(b) Here, the past of X is still helpful for predicting Y_t since X_{t-1} influences Y_t indirectly via X_t. Thus, Granger causality is still able to detect the influence of X on Y.