

CAUSAL INFERENCE FROM TIME SERIES: WHAT CAN BE LEARNED FROM GRANGER CAUSALITY?

MICHAEL EICHLER

*Department of Quantitative Economics, University of Maastricht
P.O. Box 616, 6200 MD Maastricht, The Netherlands*

December 24, 2007

ABSTRACT. In time series analysis, inference about cause-effect relationships among multiple time series is commonly based on the concept of Granger causality, which exploits temporal structure to achieve causal ordering of dependent variables. One major and well known problem in the application of Granger causality for the identification of causal relationships is the possible presence of latent variables that affect the measured components and thus lead to so-called spurious causalities. In this paper, we present a new graphical approach for describing and analysing Granger-causal relationships in multivariate time series that are possibly affected by latent variables. We show how such representations can be used for inductive causal learning from time series and discuss the underlying assumptions and their implications for causal learning.

1. INTRODUCTION

The notion of causality and the identification of new causal relationships play a central role in scientific research. In time series analysis, inference about cause-effect relationships is commonly based on the concept of Granger causality (Granger 1969), which is defined in terms of predictability and exploits the direction of the flow of time to achieve a causal ordering of dependent variables. This concept of causality does not rely on the specification of a scientific model and thus is particularly suited for empirical investigations of cause-effect relationships. On the other hand, it is commonly known that Granger causality basically is a measure of association between the variables and thus can lead to so-called spurious causalities if important relevant variables are not included in the analysis (Hsiao 1982). Since in most analyses involving time series data the presence of latent variables that affect the measured components cannot be ruled out, this raises the question whether and how the causal structure can be recovered from time series data.

Recent advances in the understanding of such latent variable structures were based on graphical models, which provide a general framework for describing and inferring causal relations (e.g. Pearl 2000, Lauritzen 2001). For time series, this graphical approach for the discussion of causal relationships in systems

E-mail address: m.eichler@ke.unimaas.nl (M. Eichler)

that are affected by latent variables has been first considered in Eichler (2005). Based on previously developed graphical representations of Granger-causal relationships in multivariate time series (Eichler 2001, 2007), a new class of path diagrams for the representation of the interrelationships in multivariate time series with latent variables has been introduced. These general path diagrams allow a more complete encoding of the conditional independencies if the system is affected by latent variables. In Eichler (2005), a multi-step procedure for the identification of such general path diagrams was proposed although no definite algorithm has been given.

In this paper, we discuss graphical representations for multivariate dynamic systems affected by latent variables in more detail. For this, we review in Section 2 the concept of Granger causality as originally introduced by Granger (1969, 1980) and in Section 3 related graphical representation for multivariate time series and their Markov properties. In Section 4, we discuss some properties of general path diagram for systems affected by latent variables and introduce dynamic ancestral graphs. An approach for causal learning from time series data is presented in Section 5. More details and proofs for the results presented will be provided in a longer technical version of the paper. Section 6 contains some concluding remarks.

2. CAUSALITY IN MULTIVARIATE TIME SERIES

While controlled experiments still provide the ideal framework for causal analysis, many complex phenomena such as information processing in the brain can only be studied from non-experimental or quasi-experimental data. It is therefore important to have an operational definition of causality that allows inference about cause-effect relationships also from observational studies. For multivariate time series, such a definition has been introduced by Granger (1969, 1980). In this section, we review this concept and problems related to it.

2.1. Granger causality

Suppose that $X = (X(t))_{t \in \mathbb{Z}}$ and $Y = (Y(t))_{t \in \mathbb{Z}}$ are two stationary time series that are statistically dependent on each other. When is it justified to say that the one series X causes the other series Y ? In order to come up with a general definition, Granger (1969, 1980) evokes the following two fundamental principles:

1. The effect does not precede its cause in time.
2. The causal series contains unique information about the series being caused that is not available otherwise.

The first principle of temporal precedence of causes is commonly accepted and has been also the basis for other probabilistic theories of causation (e.g., Good 1961, 1962, Suppes 1970). In contrast, the second principle is more subtle as it requires the separation of the special information provided by the former series X from any other possible information. To this end, Granger considers two information sets:

- $\mathcal{I}^*(t)$ - the set of all information in the universe up to time t ;

- $\mathcal{I}_{-X}^*(t)$ - the same information set except for the values of series X up to time t .

Here it is assumed that all variables in the universe are measured at equidistant points in time, namely $t \in \mathbb{Z}$. Now, if the series X causes series Y , we expect by the above principles that the probability distributions of $Y(t+1)$ conditionally on the two information sets $\mathcal{I}^*(t)$ and $\mathcal{I}_{-X}^*(t)$ differ.

Granger's definition of causality (1969, 1980). We say that the series X *causes* the series Y if

$$\mathbb{P}(Y(t+1) \in A | \mathcal{I}^*(t)) \neq \mathbb{P}(Y(t+1) \in A | \mathcal{I}_{-X}^*(t))$$

for some measurable set $A \subseteq \mathbb{R}$ and all $t \in \mathbb{Z}$.

It is important to note that this concept of causality covers only direct causal relationships. For example, if X affects Y only via a third series Z , then $\mathcal{I}_{-X}^*(t)$ comprises the past values of Z and $Y(t+1)$ is independent from the past values of X given $\mathcal{I}_{-X}^*(t)$. Implicitly, the separation of the two information sets $\mathcal{I}^*(t)$ and $\mathcal{I}_{-X}^*(t)$ is based on the assumption that the universe considered is discretized not only in time but also in space.

Because of the metaphysical character of the information set $\mathcal{I}^*(t)$, the above definition cannot be used with actual data. In practice, only the background knowledge available at time t can be incorporated into an analysis. Therefore, the definition must be modified to become operational.

In the sequel, suppose that we have observed variables X_v , $v \in V$, and let $\mathbf{X}_V = (\mathbf{X}_V(t))_{t \in \mathbb{Z}}$ be the corresponding multivariate time series. Substituting the information set $\mathcal{I}^*(t)$ by the information set $\mathcal{I}_V(t)$ generated by the values of \mathbf{X}_V up to time t and, similarly, $\mathcal{I}_{-X}^*(t)$ by $\mathcal{I}_{V \setminus \{b\}}(t)$, the information set generated by the values of $\mathbf{X}_{V \setminus \{b\}}$ up to time t , we obtain the following modified version of the above definition (Granger 1980, 1988).

Definition 2.1. Let $a, b \in V$.

- (i) The series X_b is said to be a *prima facie cause* of the series X_a with respect to \mathcal{I}_V if

$$\mathbb{P}(X_a(t+1) \in A | \mathcal{I}_V(t)) \neq \mathbb{P}(X_a(t+1) \in A | \mathcal{I}_{V \setminus \{b\}}(t))$$

for some measurable set $A \subseteq \mathbb{R}$ and all $t \in \mathbb{Z}$.

- (ii) The series X_b *does not cause* the series X_a with respect to \mathcal{I}_V if

$$\mathbb{P}(X_a(t+1) \in A | \mathcal{I}_V(t)) = \mathbb{P}(X_a(t+1) \in A | \mathcal{I}_{V \setminus \{b\}}(t))$$

for all measurable sets $A \subseteq \mathbb{R}$ and all $t \in \mathbb{Z}$.

The condition in (ii) is equivalent to that $X_a(t+1)$ and $\mathcal{I}_V(t)$ are independent conditionally on $\mathcal{I}_{V \setminus \{b\}}(t)$, abbreviated as $X_a(t+1) \perp\!\!\!\perp \mathcal{I}_V(t) | \mathcal{I}_{V \setminus \{b\}}(t)$. In that case, we say that X_b is *Granger-noncausal* for X_a with respect to \mathcal{I}_V ; otherwise we say that X_b *Granger-causes* X_a with respect to \mathcal{I}_V . We note that besides this *strong* version of Granger causality the notions of *Granger causality in mean* (Granger 1980, 1988) and *linear Granger causality* (Hosoya 1977, Florens and Mouchart 1985) exist.

It is clear from the general definition given above that Granger intended the information to be chosen as large as possible including all available and

possibly relevant variables. Despite of this, most (econometric) textbooks (e.g., Lütkepohl 1993) introduce Granger causality only in the bivariate case. This has lead to some confusion about a multivariate definition of Granger causality (e.g., Kamiński et al. 2001)

2.2. The problem of spurious causality

One major drawback of the above operational definition of causality is its dependence on the information set \mathcal{I}_V . If two or more variables are jointly affected by variables that are not included in the analysis and hence in the information set, this can induce conditional dependences among the observed variables that are wrongly interpreted as causal relationships. It therefore becomes necessary to find criteria to distinguish such spurious causalities from true cause-effect relationships.

A first step in this direction is the paper by Hsiao (1982), who discussed causal patterns for vector time series of three variables. The general idea is that direct causes—described by Granger’s general definition—persist regardless of the background information used for the analysis whereas indirect as well as spurious causes can be identified by either adding new variables to the analysis or removing already included variables from the information set.

Patterns of causality (Hsiao 1982)

- (a) X_1 is a *direct cause* of X_2 if X_1 Granger-causes X_2 with respect to $\mathcal{I}_{\{1,2\}}$ as well as with respect to $\mathcal{I}_{\{1,2,3\}}$.
- (b) X_1 is an *indirect cause* of X_2 if
 - X_1 Granger-causes X_2 with respect to $\mathcal{I}_{\{1,2\}}$ but not with respect to $\mathcal{I}_{\{1,2,3\}}$ and
 - X_1 Granger-causes X_3 and X_3 Granger-causes X_2 both with respect to $\mathcal{I}_{\{1,2,3\}}$.
- (c) X_1 is a *spurious cause of type II* for X_2 if
 - X_1 Granger-causes X_2 with respect to $\mathcal{I}_{\{1,2\}}$ but not with respect to $\mathcal{I}_{\{1,2,3\}}$ and
 - X_3 Granger-causes X_1 and X_2 with respect to $\mathcal{I}_{\{1,2,3\}}$.
- (d) X_1 is a *spurious cause of type I* for X_2 if
 - X_1 Granger-causes X_2 with respect to $\mathcal{I}_{\{1,2,3\}}$ but not with respect to $\mathcal{I}_{\{1,2\}}$ and
 - X_1 Granger-causes X_3 and X_3 Granger-causes X_2 both with respect to $\mathcal{I}_{\{1,2,3\}}$.

This characterization of causal patterns could be generalized to higher dimensions although the formulation of the appropriate conditions seems technical. In the following sections, we discuss an alternative approach based on graphical representations of Granger-causal relationships.

3. GRAPHICAL REPRESENTATIONS FOR MULTIVARIATE TIME SERIES

Let $\mathbf{X}_V = (\mathbf{X}_V(t))_{t \in \mathbb{Z}}$ with $\mathbf{X}_V(t) = (X_v(t))'_{v \in V}$ be the vector time series of interest. For simplicity, we assume that \mathbf{X}_V is stationary Gaussian process

with mean zero and covariances $\mathbf{\Gamma}(u) = \mathbb{E}\mathbf{X}_V(t)\mathbf{X}_V(t-u)'$. Throughout the paper, we make the following assumption.

Assumption 3.1. The spectral density matrix

$$\mathbf{f}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \mathbf{\Gamma}(u) e^{-i\lambda u}$$

of \mathbf{X}_V exists, and its eigenvalues are bounded and bounded away from zero uniformly for all $\lambda \in [-\pi, \pi]$.

Under this assumption, the process \mathbf{X}_V has a mean-square convergent autoregressive representation

$$\mathbf{X}_V(t) = \sum_{u=1}^{\infty} \mathbf{\Phi}(u) \mathbf{X}_V(t-u) + \boldsymbol{\varepsilon}_V(t), \quad (3.1)$$

where $\mathbf{\Phi}(u)$ is a square summable sequence of $V \times V$ matrices and $\{\boldsymbol{\varepsilon}_V(t)\}$ is a Gaussian white noise process with non-singular covariance matrix $\boldsymbol{\Sigma}$.

3.1. Path diagrams and Granger causality

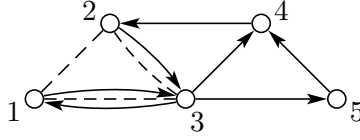
The autoregressive structure of the time series \mathbf{X}_V can be visualized by a path diagram, in which the vertices correspond to the variables of \mathbf{X}_V while the edges—arrows and lines—between vertices indicate non-zero coefficients in the autoregressive representation of \mathbf{X}_V (Eichler 2006a, 2007).

Definition 3.2. Let \mathbf{X}_V be a stationary Gaussian process with autoregressive representation (3.1). Then the *path diagram* associated with \mathbf{X}_V is the graph G with vertex set V and edge set such that for distinct $a, b \in V$

- (i) $a \rightarrow b$ not in G if and only if $\Phi_{ba}(u) = 0$ for all $u \in \mathbb{N}$;
- (ii) $a \text{ --- } b$ not in G if and only if $\Sigma_{ab} = \Sigma_{ba} = 0$.

Since path diagrams of this form may contain two types of edges, they will be referred to as mixed graphs. Furthermore, we note that, unlike in graphs commonly used for graphical modelling, two vertices a and b may be connected by up to three edges, namely $a \rightarrow b$, $a \leftarrow b$, and $a \text{ --- } b$. Similar path diagrams have been used to represent linear structural equation models (Wright 1934, Goldberger 1972, Koster 1999)¹.

¹In path diagrams for structural equation systems, correlated errors commonly are represented by bi-directed edges (\leftrightarrow) instead of dashed lines (---). Since in our approach directions are associated with temporal ordering, we prefer (dashed) undirected edges to indicate correlation between the error variables. Dashed edges with a similar connotation are used for covariance graphs (e.g. Cox and Wermuth 1996), whereas undirected edges — are commonly associated with nonzero entries in the inverse of the variance matrix (e.g. Lauritzen 1996).

FIGURE 3.1. Path diagram associated with the process \mathbf{X}_V in (3.2).

As an example, consider a vector autoregressive process given by the equations

$$\begin{aligned} X_1(t) &= \Phi_{11} X_1(t-1) + \Phi_{13} X_3(t-1) + \varepsilon_1(t), \\ X_2(t) &= \Phi_{22} X_2(t-1) + \Phi_{24} X_4(t-1) + \varepsilon_2(t), \\ X_3(t) &= \Phi_{33} X_3(t-1) + \Phi_{31} X_1(t-1) + \Phi_{32} X_2(t-1) + \varepsilon_3(t), \\ X_4(t) &= \Phi_{44} X_4(t-1) + \Phi_{43} X_3(t-1) + \Phi_{45} X_5(t-1) + \varepsilon_4(t), \\ X_5(t) &= \Phi_{55} X_5(t-1) + \Phi_{53} X_3(t-1) + \varepsilon_5(t). \end{aligned} \quad (3.2)$$

Additionally, we assume that $\varepsilon_{\{1,2,3\}}(t)$, $\varepsilon_4(t)$, and $\varepsilon_5(t)$ are pairwise uncorrelated. Then the path diagram associated with this process is given by the graph in Fig. 3.1.

It is well known that the pairwise Granger-causal relationships among the variables of a process \mathbf{X}_V are reflected in the autoregressive coefficients of the process and, thus, in the presence and absence of edges in the associated path diagram. More precisely, we have the following result.

Lemma 3.3. *Let G be the path diagram associated with a stationary Gaussian process \mathbf{X}_V satisfying Assumption 3.1. Then for distinct $a, b \in V$*

- (i) $a \rightarrow b$ not in G if and only if X_a is Granger-noncausal for X_b with respect to \mathcal{I}_V ;
- (ii) $a \text{ --- } b$ not in G if and only if X_a and X_b are contemporaneously independent with respect to \mathcal{I}_V , that is, $X_a(t+1) \perp\!\!\!\perp X_b(t+1) \mid \mathcal{I}_V(t)$ for all $t \in \mathbb{Z}$.

Because of this result, the path diagram associated with a process \mathbf{X}_V is also called the *Granger causality graph* of the process \mathbf{X}_V .

3.2. Markov properties

Under the assumptions imposed on \mathbf{X}_V , more general Granger-causal relationships than those in Lemma 3.3 can be derived from the path diagram associated with \mathbf{X}_V . This global Markov interpretation is based on a path-oriented concept of separating subsets of vertices in a mixed graph, which has been used previously to represent the Markov properties of linear structural equation systems (e.g. Spirtes et al. 1998, Koster 1999). Following Richardson (2003) we will call this notion of separation in mixed graphs *m-separation*.

Let G be a mixed graph and $a, b \in V$. A *path* π in G is a sequence $\pi = \langle e_1, \dots, e_n \rangle$ of edges e_i in G with an associated sequence of nodes v_0, \dots, v_n such that e_i is an edge between v_{i-1} and v_i . The vertices v_0 and v_n are the *endpoints* while v_1, \dots, v_{n-1} are the *intermediate vertices* of the path. Notice

that paths may be self-intersecting since we do not require that the vertices v_i are distinct.

An intermediate vertex c on a path π is said to be an m -collider on π if the edges preceding and succeeding c both have an arrowhead or a dashed tail at c (i.e. $\rightarrow c \leftarrow$, $\rightarrow c \text{ ---}$, $\text{---} c \leftarrow$, $\text{---} c \text{ ---}$); otherwise c is said to be an m -noncollider on π . A path π between a and b is said to be m -connecting given a set C if

- (i) every m -noncollider on π is not in C and
- (ii) every m -collider on π is in C ;

otherwise we say that π is m -blocked given C . If all paths between a and b are m -blocked given C , then a and b are said to be m -separated given C . Similarly, two sets A and B are said to be m -separated given C if for every pair $a \in A$ and $b \in B$, a and b are m -separated given C .

With this notion of separation, it can be shown that path diagrams for multivariate time series have a similar Markov interpretation as path diagrams for linear structural equation systems (cf Koster 1999). For disjoint subsets $A, B, C \subseteq V$, the subprocesses \mathbf{X}_A and \mathbf{X}_B are said to be *conditionally independent* given \mathcal{I}_C if $\mathbb{E}(g(\mathbf{X}_A)|\mathcal{I}_{B \cup C}) = \mathbb{E}(g(\mathbf{X}_A)|\mathcal{I}_C)$ for all real-valued measurable functions $g(\cdot)$ on $\mathbb{R}^{A \times \mathbb{Z}}$, where $\mathcal{I}_{B \cup C}$ (\mathcal{I}_B) is the informaton set (σ -algebra) generated by the complete series $\mathbf{X}_{B \cup C}$ (\mathbf{X}_B). In this case, we write $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathcal{I}_C$. Then separation in the path diagram can be translated into conditional independence among complete subprocesses of \mathbf{X}_V (Eichler 2007).

Proposition 3.4. *Let \mathbf{X}_V be a stationary Gaussian process that satisfies Assumption 3.1, and let G be its path diagram. Then, for all disjoint $A, B, S \subseteq V$,*

$$A \not\propto_m B | S \text{ in } G \text{ then } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathcal{I}_S.$$

Derivation of such conditional independence statements requires that all paths between two sets are m -blocked. For the derivation of Granger-causal relationships, it suffices to consider only a subset of these paths, namely those having an arrowhead at one endpoint. For a formal definition, we say that a path π between a and b is b -pointing if it has an arrowhead at the endpoint b ; furthermore, a path between sets A and B is said to be B -pointing if it is b -pointing for some $b \in B$. Then, to establish Granger noncausality from \mathbf{X}_A to \mathbf{X}_B , it suffices to consider only all B -pointing paths between A and B . Similarly, a graphical condition for contemporaneous correlation can be obtained based on bi -pointing path, which have an arrowhead at both endpoints.

Definition 3.5. A stationary Gaussian process \mathbf{X}_V is Markov for a graph G if, for all disjoint subsets $A, B, C \subseteq V$, the following two conditions hold:

- (i) if every B -pointing path between A and B is m -blocked given $B \cup C$, then \mathbf{X}_A is Granger-noncausal for \mathbf{X}_B with respect to $\mathcal{I}_{A \cup B \cup C}$;
- (ii) if the sets A and B are not connected by an undirected edge (---) and every bi -pointing path between A and B is m -blocked given $A \cup B \cup C$, then \mathbf{X}_A and \mathbf{X}_B are contemporaneously uncorrelated with respect to $\mathcal{I}_{A \cup B \cup C}$.

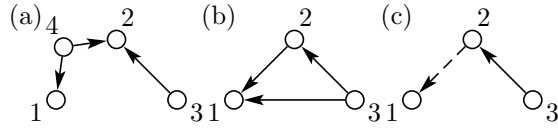


FIGURE 4.1. Graphical representations of the four-dimensional VAR(2) process in (4.1): (a) path diagram associated with $X_{\{1,2,3,4\}}$; (b) path diagram associated with $X_{\{1,2,3\}}$; (c) general path diagram for $X_{\{1,2,3\}}$.

With this definition, it can be shown that path diagrams for Gaussian vector autoregressive processes can be interpreted in terms of such global Granger-causal relationships (cf Eichler 2007).

Theorem 3.6. *Let \mathbf{X}_V be a stationary Gaussian process that satisfies Assumption 3.1, and let G be the associated path diagram. Then \mathbf{X}_V is Markov for G .*

4. REPRESENTATION OF SYSTEMS WITH LATENT VARIABLES

As pointed out above, the notion of Granger causality is based on the assumption that all relevant information is included in the analysis (Granger 1969, 1980). The omission of important variables can lead to temporal correlations among the observed components that are falsely detected as causal relationships. The detection of such so-called spurious causalities becomes a major problem when identifying the structure of systems that may be affected by latent variables.

Of particular interest will be spurious causalities of type I, where a Granger-causal relationship with respect to the complete process vanishes when only a subprocess is considered. Since the path diagrams from the previous section are defined, by Lemma 3.3, in terms of the pairwise Granger-causal relationships with respect to the complete process, they provide no means to distinguish such spurious causalities of type I from true causal relationships. To illustrate this remark, we consider the four-dimensional vector autoregressive process \mathbf{X}_V with components

$$\begin{aligned} X_1(t) &= \alpha X_4(t-2) + \varepsilon_1(t), \\ X_2(t) &= \beta X_4(t-1) + \gamma X_3(t-1) + \varepsilon_2(t), \\ X_3(t) &= \varepsilon_3(t), \\ X_4(t) &= \varepsilon_4(t), \end{aligned} \tag{4.1}$$

where $\varepsilon_i(t)$, $i = 1, \dots, 4$ are uncorrelated white noise processes with mean zero and variance one. The true dynamic structure of the process is shown in Fig. 4.1(a). In this graph, the 1-pointing path $3 \rightarrow 2 \leftarrow 4 \rightarrow 1$ is m -connecting given $S = \{2\}$, but not given the empty set. By Theorem 3.6, we conclude that \mathbf{X}_3 is Granger-noncausal for \mathbf{X}_1 in a bivariate analysis, but not necessarily in an analysis based on $\mathbf{X}_{\{1,2,3\}}$.

Now suppose that variable \mathbf{X}_4 is latent. Simple derivations show (cf Eichler 2005) that the autoregressive representation of $\mathbf{X}_{\{1,2,3\}}$ is given by

$$\begin{aligned} X_1(t) &= \frac{\alpha\beta}{1+\beta^2} X_2(t-1) + \frac{\alpha\beta\gamma}{1+\beta^2} X_3(t-2) + \tilde{\varepsilon}_1(t), \\ X_2(t) &= \gamma X_3(t-1) + \tilde{\varepsilon}_2(t), \\ X_3(t) &= \varepsilon_3(t), \end{aligned}$$

where $\tilde{\varepsilon}_2(t) = \varepsilon_2(t) + \beta X_4(t-1)$ and

$$\tilde{\varepsilon}_1(t) = \varepsilon_1(t) - \frac{\alpha\beta}{1+\beta^2} \varepsilon_2(t-1) + \frac{\alpha}{1+\beta^2} X_4(t-2).$$

The path diagram associated with $\mathbf{X}_{\{1,2,3\}}$ is depicted in Fig. 4.1(b). In contrast to the graph in Fig. 4.1(a), this path diagram contains an edge $3 \rightarrow 1$ and, thus, does not entail that \mathbf{X}_3 is Granger-noncausal for \mathbf{X}_1 in a bivariate analysis.

As a response to such situations, two approaches have been considered in the literature. One approach suggests to include all latent variables explicitly as additional nodes in the graph (e.g., Pearl 2000, Eichler 2007); this leads to models with hidden variables, which can be estimated, for example, by application of the EM algorithm (e.g., Boyen et al. 1999). For a list of possible problems with this approach, we refer to Richardson and Spirtes (2002, §1).

The alternative approach focuses on the conditional independence relations among the observed variables; examples of this approach include linear structural equations with correlated errors (e.g. Pearl 1995, Koster 1999) and the maximal ancestral graphs by Richardson and Spirtes (2002). In the time series setting, this approach has been discussed by Eichler (2005), who considered path diagrams in which dashed edges represent associations due to latent variables. For the trivariate subprocess $\mathbf{X}_{\{1,2,3\}}$ in the above example, such a path diagram is depicted in Fig. 4.1(c).

Following this latter approach, we consider mixed graphs that may contain three types of edges, namely undirected edges ($---$), directed edges (\rightarrow), and dashed directed edges (\dashrightarrow). For the sake of simplicity, we also use $a \leftrightarrow b$ as an abbreviation for the triple edge $a \overset{\dashrightarrow}{\leftrightarrow} b$. Unlike path diagrams for autoregressions, these graphs in general are not defined in terms of pairwise Granger-causal relationships, but only through the global Markov interpretation according to Definition 3.5. To this end, we simply extend the concept of m -separation introduced in the previous section by adapting the definition of m -noncolliders and m -colliders. Let π be a path in a mixed graph G . Then an intermediate vertex n is called an m -noncollider on π if at least one of the edges preceding and succeeding n on the path is a directed edge (\rightarrow) and has its tail at n . Otherwise, n is called an m -collider on π . With this extension, we leave all other definition such as m -separation or pointing paths unchanged.

4.1. Marginalization

The main difference between the class of mixed graphs with directed (\rightarrow) and undirected ($---$) edges and the more general class of mixed graphs that has

TABLE 4.1. Creation of edges by marginalizing over i .

Subpath π in G	Associated edge e_π in $G^{\{i\}}$
$a \rightarrow i \rightarrow b$	$a \rightarrow b$
$a \dashrightarrow i \rightarrow b$	$a \dashrightarrow b$
$a \cdashrightarrow i \rightarrow b$	$a \dashrightarrow b$
$a \leftarrow i \rightarrow b$	$a \leftrightarrow b$
$a \leftarrow i \dashrightarrow b$	$a \leftrightarrow b$

been just introduced is that the latter class is closed under marginalization. This property makes it suitable for representing systems with latent variables.

Let G be a mixed graph and $i \in V$. For every subpath $\pi = \langle e_1, e_2 \rangle$ of length 2 between vertices $a, b \in V \setminus \{i\}$ such that i as an intermediate vertex and an m -noncollider on π , we define an edge e_π according to Tab. 4.1. Let $A^{\{i\}}$ the set of all such edges e_π . Furthermore, let $E^{\{i\}}$ be the subset of edges in E that have both endpoints in $V \setminus \{i\}$. Then we define $G^{\{i\}} = (V \setminus \{i\}, E^{\{i\}} \cup A^{\{i\}})$ as the graph obtained by marginalizing over $\{i\}$. Furthermore, for $L = \{i_1, \dots, i_n\}$ we set $G^L = ((G^{\{i_1\}})^{\{i_2\}} \dots)^{\{i_n\}}$, that is, we proceed iteratively by marginalizing over i_j , for $j = 1, \dots, n$. Similarly as in Koster (1999), it can be shown that the order of the vertices does not matter and that the graph G^L is indeed well defined.

We note that the graph G^L obtained by marginalizing over the set L in general contains self-loops. Simple considerations, however, show that G^L is Markov-equivalent to a graph \tilde{G}^L with all subpaths of the form $a \cdashrightarrow b \dashrightarrow b$ and $a \leftarrow b \dashrightarrow b$ replaced by $a \cdashrightarrow b$ and $a \leftrightarrow b$, respectively, and all self-loops deleted, that is, the graphs G^L and \tilde{G}^L encode exactly the same Granger-causal and contemporaneous independence relations. It therefore suffices to consider mixed graphs without self-loops. We omit the details.

Now suppose that, for some subsets $A, B, C \subseteq V \setminus L$, π is an m -connecting path between A and B given S . Then all intermediate vertices on π that are in L must be m -noncolliders. Removing these vertices according to Table 4.1, we obtain a path π' in G^L that is still m -connecting. Since the converse is also true, we obtain the following result.

Proposition 4.1. *Let G be a mixed graph, and $L \subseteq V$. Then it holds that, for all distinct $a, b \in V \setminus L$ and all $C \subseteq V \setminus L$, every path between a and b in G is m -blocked given C if and only if the same is true for the paths in G^L . Furthermore, the same equivalence holds for all pointing paths and for all bi-pointing paths.*

It follows that, if a process \mathbf{X}_V is Markov for a graph G , the subprocess $\mathbf{X}_{V \setminus L}$ is Markov for the smaller graph G^L , which encodes all relationships about $\mathbf{X}_{V \setminus L}$ that are also encoded in G .

We note that insertion of edges according to Tab. 4.1 is sufficient but not always necessary for representing the relations in the subprocess $\mathbf{X}_{V \setminus L}$. This applies in particular to the last two cases in Tab. 4.1. For an example, we consider again the process (4.1) with associated path diagram in Fig. 4.1(a).

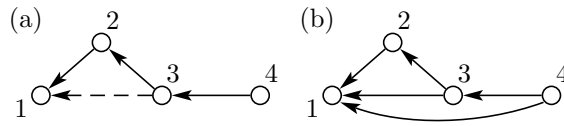


FIGURE 4.2. Two Markov equivalent graphs: (a) non-ancestral graph and (b) corresponding ancestral graph.

By Tab. 4.1, the subpath $1 \leftarrow 4 \rightarrow 2$ should be replaced by $1 \leftrightarrow 2$, which suggests that \mathbf{X}_1 Granger-causes \mathbf{X}_2 (as does the path $1 \leftarrow 4 \rightarrow 2$ in the original path diagram), while in fact the structure can be represented by the graph in Fig. 4.1(c).

4.2. Markov equivalence and dynamic ancestral graphs

The operation of marginalization is of theoretical importance as it shows that the class of general path diagrams is rich enough to represent causal structures of systems with latent variables. It is of much less use for the identification of such structures as the relations determining the complete system are not available. Here, we face the problem that the set of Granger-causal relationships and contemporaneous independences that hold for the observed process does not uniquely determine a graphical representation within the class of general path diagrams. As an example, Fig. 4.2 displays two graphs that are Markov equivalent, that is, they encode the same set of Granger-causal and contemporaneous independence relations among the variables. Therefore, the corresponding graphical models—models that obey the conditional independence constraints imposed by the graph—are statistically indistinguishable.

In order to simplify the identification problem, the search for a suitable graphical representation can be restricted to a smaller class of graphs provided that this class contains at least one representative from every Markov equivalence class. Following Richardson and Spirtes (2002), we consider graphs that satisfy an ancestry property. More precisely, we say that a vertex a is an ancestor of another vertex b if $a = b$ or there exists a directed path $a \rightarrow \dots \rightarrow b$ in G . The set of ancestors of b is denoted by $\text{an}(b)$. Then a graph G is called a dynamic ancestral graph if it satisfies the condition

$$a \in \text{an}(b) \text{ then } a \dashrightarrow b \text{ not in } G \quad (4.2)$$

for all distinct $a, b \in V$. We note that, in contrast to the ancestral graphs introduced by Richardson and Spirtes (2002), we do not require acyclicity (which is hidden in the time ordering) nor that edges are joined by at most one edge. The above condition (4.2) does imply, however, that two vertices a and b can be connected by the two edges $a \rightarrow b$ and $a \dashrightarrow b$ at the same time.

In order to obtain a Markov-equivalent ancestral graph for a general path diagram we have to substitute a dashed directed edge $a \dashrightarrow b$ by a directed edge $a \rightarrow b$ whenever a is an ancestor of b . Furthermore, if G contains the path $c \rightarrow a \dashrightarrow b$, an additional edge $c \rightarrow b$ needs to be inserted; similarly, the edges $c \dashrightarrow a$, $c \leftarrow a$, and $c \dashleftarrow a$ lead to the insertion of edges $c \dashrightarrow b$, $c \leftrightarrow b$, and $c \dashrightarrow b$, respectively. Iterating over all edges, we finally obtain

a graph that satisfies condition (4.2) and is Markov-equivalent to the original graph.

Proposition 4.2. *Let G be a general path diagram. Then there exists a dynamic ancestral graph \tilde{G} that is Markov-equivalent to G .*

From Tab. 4.1 and the way we have constructed dynamic ancestral graphs, it is clear that two vertices a and b are connected by a directed path $a \rightarrow \dots \rightarrow b$ in a dynamic ancestral graph if and only if the same holds true in the path diagram associated with the underlying complete system. This leads us to the following general definition of causal effects in multivariate time series.

Definition 4.3. A series X_a is said to have a causal effect—direct or indirect—on another series X_b if there exists some multivariate process X_V with $a, b \in V$ such that every graph in the Markov equivalence class of dynamic ancestral graphs for X_V contains a directed path $a \rightarrow \dots \rightarrow b$.

A further distinction between different causal patterns such as direct, indirect, or spurious causality is only possible with respect to a given information set \mathcal{I}_V and requires to consider all graphs in the Markov equivalence class of general path diagrams for X_V . We omit further details.

5. LEARNING LATENT VARIABLE STRUCTURES

There are two major approaches for learning causal structures: One approach utilizes constraint-based search algorithms such as the Fast Causal Inference (FCI) algorithm (Spirtes et al. 2001) while the other consists of score-based model selection. In the following, we briefly discuss the former approach for the identification of dynamic ancestral graphs. The latter approach has been investigated in Eichler (2006b).

The constraint-based search tries to find a graph that matches the empirically determined conditional independences. It usually consists of two steps:

1. identification of adjacencies of the graph;
2. identification of the type and the orientation of edges whenever possible.

In the case of ancestral graphs, the first step makes use of the fact that for every ancestral graph there exists a unique Markov-equivalent maximal ancestral graph (MAG), in which every missing edge corresponds to a conditional independence relation among the variables. Here an ancestral graph G is said to be maximal if addition of further edges would change the Markov equivalence class. MAGs are closely related to the concept of inducing paths; in fact, Richardson and Spirtes (2002) used inducing paths to define MAGs and then showed the maximality property.

Definition 5.1. In a (dynamic) ancestral graph, a path π between two vertices a and b is called an inducing path if every intermediate vertex on π is, firstly, a collider on π and, secondly, an ancestor of a or b .

Figures 5.1(a,b) give two examples of dynamic ancestral graphs, in which $2 \dashrightarrow 3 \dashrightarrow 4$ resp. $2 \rightarrow 3 \dashrightarrow 4$ are inducing 4-pointing paths. The graph in (b) shows that—unlike in the case of ordinary ancestral graphs—inducing paths

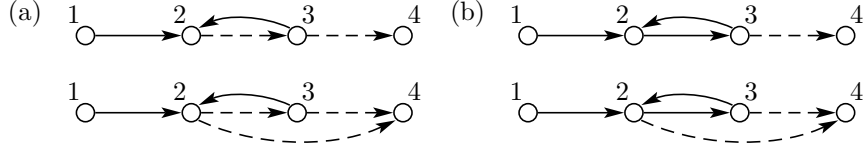


FIGURE 5.1. Inducing paths: (a) Dynamic ancestral graph with an inducing 4-pointing path between 2 and 4 and its corresponding Markov equivalent maximal dynamic ancestral graph. (b) Maximal dynamic ancestral graph with an inducing 4-pointing path between 2 and 4; the graph with additional edge $2 \dashrightarrow 4$ is not Markov equivalent.

may start with a tail at one of two vertices. As a consequence, insertion of an edge $2 \dashrightarrow 4$ or $2 \rightarrow 4$ changes the encoded Granger-causal relationships: while the upper graph implies that X_1 is Granger-noncausal for X_4 with respect to $\mathcal{I}_{\{1,2,3,4\}}$, this is not true for the lower graph. It follows that the method used for identifying adjacencies in ordinary MAGs do not apply to dynamic ancestral graphs.

The problem can be solved by observing that m -connecting pointing paths not only encode Granger-causal relationships but, depending on whether they start with an $a \rightarrow c$ or $a \dashrightarrow c$, also a related type of conditional independences. More precisely, we have the following result.

Proposition 5.2. *Suppose that a and b are not connected by a b -pointing inducing path starting with $a \leftarrow c$, $a \leftarrow\leftarrow c$, or $a \dashrightarrow c$. Then there exist disjoint subsets S_1, S_2 with $b \in S_1$ and $a \notin S_1 \cup S_2$ such that*

$$X_a(t-k) \perp\!\!\!\perp X_b(t+1) \mid \mathcal{I}_{S_1}(t) \vee \mathcal{I}_{S_2}(t-k) \vee \mathcal{I}_a(t-k-1)$$

for all $k \in \mathbb{N}$ and all $t \in \mathbb{Z}$.

The proof is based on the fact that inducing paths starting with an edge $a \rightarrow c$ or $a \dashrightarrow c$ only induce an association between the $X_a(t-k)$ and $X_b(t+1)$ if one conditions on $X_c(t-k+1), \dots, X_c(t)$. To block any other paths, we set S_2 to be the set of all intermediate vertices on all b -pointing inducing paths connecting a and b , and S_1 to be the set of all ancestors of a and b except a and b .

This leads us to the following algorithm for the identification of the Markov equivalence classes of dynamic ancestral graphs. Here, we use dotted directed edges \dashrightarrow to indicate that the tail of the directed edge is (yet) undetermined.

Identification of adjacencies:

1. insert $a \dashrightarrow b$ whenever X_a and X_b are not contemporaneously independent with respect to \mathcal{I}_V ;
2. insert $a \dashrightarrow\!\!\!\rightarrow b$ whenever
 - X_b Granger-causes X_a with respect to \mathcal{I}_S for all $S \subseteq V$ with $a, b \in S$ and
 - $X_a(t-k) \perp\!\!\!\perp X_b(t+1) \mid \mathcal{I}_{S_1}(t) \vee \mathcal{I}_{S_2}(t-k) \vee \mathcal{I}_a(t-k-1)$ for some $k \in \mathbb{N}$, all $t \in \mathbb{Z}$, and all disjoint $S_1, S_2 \subseteq V$ with $b \in S_1$ and $a \notin S_1 \cup S_2$

Identification of tails:

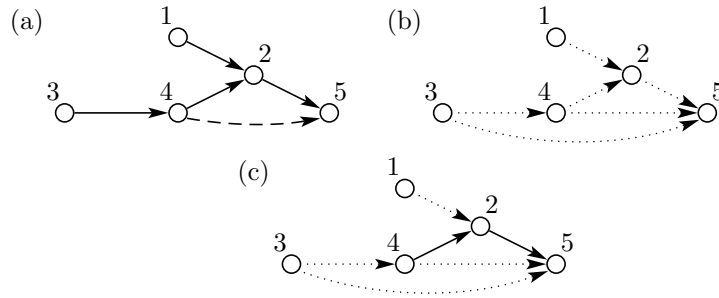


FIGURE 5.2. Identification of dynamic ancestral graphs: (a) underlying structure; (b) adjacencies; (c) identification of tails.

1. *Colliders:*

Suppose that G does not contain $a \cdots\cdots b$, $a \rightarrow b$, or $a \dashrightarrow b$. If $a \cdots\cdots c \cdots\cdots b$ and X_a is Granger-noncausal for X_b with respect to \mathcal{I}_S for some set S with $c \notin S$, replace $c \cdots\cdots b$ by $c \dashrightarrow b$.

2. *Non-colliders:*

Suppose that G does not contain $a \cdots\cdots b$, $a \rightarrow b$, or $a \dashrightarrow b$. If $a \cdots\cdots c \cdots\cdots b$ and X_a is Granger-noncausal for X_b with respect to \mathcal{I}_S for some set S with $c \in S$, replace $c \cdots\cdots b$ by $c \rightarrow b$.

3. *Ancestors:*

if $a \in \text{an}(b)$ replace $a \cdots\cdots b$ by $a \rightarrow b$;

4. *Discriminating paths:*

A fourth rule is based on the concept of discriminating paths. For details, we refer to Ali et al. (2004).

We note that in contrast to the case of ordinary ancestral graphs only the tails of the dotted directed edges need to be identified. The positions of the arrow heads are determined by the time ordering of the Granger-causal relationships. The above algorithm probably can be complemented by further rules, see also Zhang and Spirtes (2005).

To illustrate the identification algorithm, we consider the graphs in Fig. 5.2. The original general path diagram is depicted in (a). Since 4 is an ancestor of 5, this graph is not ancestral. The adjacencies determined by the algorithm are shown in (b). Since X_1 does not Granger-cause X_5 with respect to \mathcal{I}_V , we find that 2 and 5 are connected by $2 \rightarrow 5$. Similarly, X_3 is Granger-noncausal for X_2 with respect to $\mathcal{I}_{\{2,3,4\}}$, which implies that the graph contains also the edge $4 \rightarrow 2$. No further tails can be identified; the final graph is given in (c).

6. CONCLUSION

The concept of Granger causality is widely used for inference about causal relationships from time series observations. One of the main problems in its application, however, is the possible presence of latent variables that affect the measured variables and thus can lead to so-called spurious causalities.

In this paper, we have presented a graphical approach for analysing cause-effect relationship in multivariate time series based on general path diagrams

and dynamic ancestral graphs. In particular, we provided (a sketch of) an constraint-based search algorithm that allows identifying Markov equivalence classes of a dynamic ancestral graphs. The causal interpretation of the resulting graph is based on two fundamental. Firstly, the *causal Markov assumption* requires that all observed dependences are due to causal influences; this is a key assumption underlying all approaches to causal inference from observation data. Secondly, the faithfulness assumption states that the independences observed are structural and not to chance cancellation of several causal influences. This assumption allows to detect spurious causes of type II when observing spurious causes of type I. Under these assumptions, directed edges can be interpreted as causal links although not as direct causes in the original sense of Granger. Due to identifiability up to Markov equivalence only edges that are invariant in a equivalence class—and have their tail identified by the search algorithm—allow an definite interpretation as causes if it is a directed edge and a spurious cause is the edge is a dashed directed edge.

REFERENCES

- Ali, R. A., Richardson, T. S. and Spirtes, P. (2004). Markov equivalence for ancestral graphs. *Technical Report No. 466*, Department of Statistics, University of Washington.
- Boyan, X., Friedman, N. and Koller, D. (1999). Discovering the hidden structure of complex dynamic systems. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 91–100.
- Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, London.
- Eichler, M. (2001). Graphical modelling of multivariate time series. *Technical report*, Universität Heidelberg. ([arXiv:math.ST/0610654](https://arxiv.org/abs/math/0610654)).
- Eichler, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of The Royal Society B* **360**, 953–967.
- Eichler, M. (2006a). Graphical modelling of dynamic relationships in multivariate time series. In M. Winterhalder, B. Schelter and J. Timmer (eds), *Handbook of Time Series Analysis*, Wiley-VCH, pp. 335–372.
- Eichler, M. (2006b). Graphical modelling of multivariate time series with latent variables. *Technical report*, University of Heidelberg.
- Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* **137**, 334–353.
- Florens, J. P. and Mouchart, M. (1985). A linear theory for noncausality. *Econometrica* **53**, 157–175.
- Goldberger, A. S. (1972). Structural equation models in the social sciences. *Econometrica* **40**, 979–1001.
- Good, I. J. (1961). A causal calculus (I). *British Journal for the Philosophy of Science* **11**, 305–318.
- Good, I. J. (1962). A causal calculus (II). *British Journal of the Philosophy of Science* **12**, 43–51.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.

- Granger, C. W. J. (1980). Testing for causality, a personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329–352.
- Granger, C. W. J. (1988). Some recent developments in a concept of causality. *Journal of Econometrics* **39**, 199–211.
- Hosoya, Y. (1977). On the Granger condition for non-causality. *Econometrica* **45**, 1735–1736.
- Hsiao, C. (1982). Autoregressive modeling and causal ordering of econometric variables. *Journal of Economic Dynamics and Control* **4**, 243–259.
- Kamiński, M., Ding, M., Truccolo, W. A. and Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics* **85**, 145–157.
- Koster, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics* **26**, 413–431.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Lauritzen, S. L. (2001). Causal inference from graphical models. In O. E. Barndorff-Nielsen, D. R. Cox and C. Klüppelberg (eds), *Complex stochastic systems*, CRC Press, London, pp. 63–107.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer, New York.
- Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika* **82**, 669–710.
- Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge, UK.
- Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics* **30**, 145–157.
- Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Annals of Statistics* **30**, 962–1030.
- Spirtes, P., Glymour, C. and Scheines, R. (2001). *Causation, Prediction, and Search*. 2nd edn, MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson.
- Spirtes, P., Richardson, T. S., Meek, C., Scheines, R. and Glymour, C. (1998). Using path diagrams as a structural equation modelling tool. *Soc. Methods Res.* **27**, 182–225.
- Suppes, P. (1970). *A probabilistic theory of causality*. North-Holland, Amsterdam.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* **5**, 161–215.
- Zhang, J. and Spirtes, P. (2005). A characterization of Markov equivalence classes for ancestral graphical models. *Technical Report 168*, Department of Philosophy, Carnegie Mellon University.