Granger-causality graphs for multivariate time series

Michael Eichler

Universität Heidelberg

Abstract

In this paper, we discuss the properties of mixed graphs which visualize causal relationships between the components of multivariate time series. In these Granger-causality graphs, the vertices, representing the components of the time series, are connected by arrows according to the Granger-causality relations between the variables whereas lines correspond to contemporaneous conditional association. We show that the concept of Granger-causality graphs provides a framework for the derivation of general noncausality relations relative to reduced information sets by performing sequences of simple operations on the graphs. We briefly discuss the implications for the identification of causal relationships. Finally we provide an extension of the linear concept to strong Granger-causality.

JEL classification: C320

Keywords: Granger-causality, graphical models, spurious causality, multivariate time series

1 Introduction

One of the central concepts in the discussion of economic laws and econometric models is that of causality. There exist various formal definitions of causality in the econometric literature and for a critical survey we refer to Zellner (1979). In this paper we are concerned with the concept of Granger-causality which has been introduced by Granger (1969). This concept is defined in terms of predictability and exploits the direction of the flow of time to achieve a causal ordering of associated variables. Since it does not rely on the specification of an econometric model it is particularly suited for empirical model building strategies as such suggested by Sims (1980). A comprehensive survey of the literature on Granger-causality has been provided by Geweke (1984).

In the original definition of Granger it is supposed that all relevant information is available and included in the analysis. In practice, only a subset of this information may have been observed and omission of important variables could lead to spurious causalities between the variables. Hsiao (1982) addressed this issue formally by introducing concepts for indirect and spurious causality in a trivariate model. In particular, it has been shown that a certain type of spurious causality vanishes if the information set is reduced. This observation led to a strengthened definition of (direct) causality by requiring an improvement in prediction irrespective of the used information set. The work of Hsiao has made clear that for a better understanding of the causal structure of a multivariate time series it is important to study not only noncausality relative to the full information set, but also more general noncausality relations. However, for models with more than three variables the number of possible causal patterns soon becomes too large for a similar, complete characterization in terms of such general noncausality relations.

The objective of the present paper is to introduce a new graphical approach for the modelling, identification and visualization of the causal relationships between the components of a multivariate time series. This approach has been motivated by the idea of graphical models in multivariate statistics. Graphical models have been used sucessfully as a general framework for modelling conditional independence relations between variables. For an introduction to the theory of graphical models we refer to the monographs of Whittaker (1990), Cox and Wermuth (1996), and Lauritzen (1996). More recently, directed acyclic graphs which correspond to factorizations of the joint probability distribution have been associated with concepts for the inference of cause-effect relationships (Pearl, 1995, 2000; Lauritzen, 2000). However, these concepts, which formalize the notion of controlled experiments, often rely on an a priori knowledge of the direction of a possible cause.

The essential feature of the proposed graphical modelling approach is to merge the notion of Granger-causality with graphs. For this we define a new class of mixed graphs for time series in which vertices representing the components of the process are connected by directed edges according to the Granger-causality relations between the variables. Likewise the contemporaneous conditional association structure is given by undirected edges between the vertices. Although such graphs have previously been used to visualize the pairwise causal relations of a multivariate time series, their properties have not yet been discussed in the literature. We can show that these graphs are related to general noncausality relations of a time series. Allowing latent variables to be represented by additional vertices in the graph, the graphical modelling approach can be used for the investigation of spurious causality and thus enables us to gain a better understanding of the causal structure of the time series. In particular, it leads to sufficient conditions for the identification of causal effects. We note that similar graphs with multiple edges allowed between two vertices have been considered by Koster (1996, 1999) in the discussion of the Markov properties of path diagrams of linear structural equation systems.

The paper is organized as follows. In Section 2 we give the definition of Granger-causality graphs. We further consider vector autoregressive processes constrained to a given graph as an important class of graphical time series models. In Section 3 the properties of Granger-causality graphs are discussed. In particular, we develop a method for the derivation of general causal relationships (relative to reduced information sets) which are implied by the graph. This method is based on a concept of separation of vertices and can be executed by a sequence of simple operations on the graph. In Section 4 the results are used to characterize noncausality at all horizons (Dufour and Renault, 1998). Furthermore, we briefly discuss the problem of identification of causal effects. Section 5 provides a generalization of the introduced concept to strong Granger-causality which allows the investigation of nonlinear causal relationships between the studied variables. As an example of a nonlinear graphical time series model we consider a special multivariate ARCH model. The final section gives some concluding remarks. In the appendix we summarize the properties of conditional orthogonality in a Hilbert space. In the second part of the appendix the proofs for the results presented in this paper are collected.

2 Granger-causality graphs

The concepts introduced in this paper are based on the notion of causality in multivariate stochastic processes which has been introduced by Granger (1969). While the original definition has been formulated in terms of mean square prediction we adapt a linear framework previously considered by Hosoya (1977) and Florens and Mourchart (1985). Here noncausality is defined in terms of conditional orthogonality of subspaces in a Hilbert space of square integrable random variables with inner product $\langle X, Y\rangle = \mathbb{E}(XY)$. This Hilbert space of real random variables on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is denoted by $\mathscr{L}^2 = \mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P})$.

Let $X = \{X(t), t \in \mathbb{Z}, t > \tau\}$ be a vector-valued process in $\mathscr{L}^2$ with $X(t) = (X_1(t), \ldots, X_d(t))'$. We are interested in the causal relations between the components of X relative to some information set given by a nondecreasing sequence $I = \{I(t), t \in \mathbb{Z}, t > \tau\}$ of linear subspaces in $\mathscr{L}^2$. Here $I(t)$ represents the information available at time t. We assume that I is conformable with X in the sense that the past and present of X at time t are included in the information set $I(t)$. Denoting by $X(\tau, t]$ the closed linear subspace spanned by $\{X(s), \tau < s \leq t\}$ we thus have $X(\tau, t] \subseteq I(t)$. For simplicity we consider information sets of the form $I_Y(t) = U + Y(\tau, t]$ where U is some closed subspace of $\mathscr{L}^2$ which contains all information available at any time $t > \tau$ such as constants, deterministic variables, or initial conditions (on X or other variables). Further $Y = (X, Z)$ is a multivariate process such that the variables in Z are exogenous for X. An example would be a model for a small open economy where Z are the foreign country variables.

We assume that the conditional variance $\operatorname{var}((X(\tau+1)', \ldots, X(t)')' | I_{Y\setminus X}(t))$ is positive definite for all $t > \tau$ where conditional variance is taken to be the variance about the linear projection. By this deterministic linear relations between the variables are excluded since otherwise the causal effects from different variables might not be identifiable. In the case where X is a weakly stationary process with starting time $\tau = -\infty$ we assume that also the joint process Y is weakly stationary with spectral matrix $f_Y(\lambda)$ and that there exists $c > 0$ such that $c^{-1} I_n \leq f_Y(\lambda) \leq c I_n$ for all $\lambda \in [-\pi, \pi]$.

Within this framework the definition of (linear) Granger-noncausality can be rewritten as follows:

Definition 2.1 (Granger-noncausality) The process X_a is noncausal for the process X_b relative to the information set I_Y, denoted by $X_a \nrightarrow X_b \ [I_Y]$, if

$$X_b(t+1) \perp X_a(\tau, t] \,|\, I_{Y\setminus X_a}(t) \qquad \text{for all } t > \tau.$$

This definition can be retained for vector processes X_A and X_B since for $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$ we have the following composition and decom-

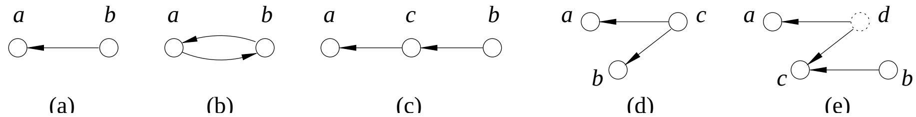

Figure 2.1: Causality patterns: (a) direct causality, (b) direct feedback, (c) indirect causality, (d) spurious causality of type II, and (e) spurious causality of type I.

position property (e.g. Boudjellaba et al., 1992, Corollary 2)

$$X_A \nrightarrow X_B \ [I_Y] \Leftrightarrow X_{a_j} \nrightarrow X_{b_k} \ [I_Y] \quad \forall\, j = 1, \ldots, m \quad \forall\, k = 1, \ldots, n.$$

For processes X with more than two variables the pairwise causality structure of X can be visualized by linking the variables by arrows representing the causal relations. Although this graphical representation seems very natural and allows an intuitive interpretation of the causal structure in terms of feedback, indirect causality, or spurious causality, the exact properties of the graph thus obtained have not yet been discussed in the literature.

For a formal definition of this graph we consider graphs G given by an ordered pair $G = (V, E)$ where V is a finite set of elements called vertices and E is a set of directed or undirected edges which belong to the classes $\{a \longrightarrow b | a, b \in V, a \neq b\}$ and $\{a \text{ --- } b | a, b \in V, a \neq b\}$, respectively. Here we make no distinction between $a \text{ --- } b$ and $b \text{ --- } a$. Since multiple edges of the same type and orientation are not permitted two vertices in the graph may be connected by up to three edges. A graph is called a mixed graph if it contains both types of edges, otherwise the graph is either directed or undirected.

In our context the vertex set will be $V = \{1, \ldots, d\}$, i.e. each vertex a in the graph represents one component X_a of the process. Although our primary interest are the causal relations between the variables it will be crucial for the analysis in the next section to model also the contemporaneous (conditional) association between the variables. We say that X_a and X_b are contemporaneously conditionally orthogonal relative to the information set I_Y, denoted by $X_a \nsim X_b \ [I_Y]$, if

$$X_a(t+1) \perp X_b(t+1) \,|\, I_Y(t) + I_{Y \setminus X_{\{a,b\}}}(t+1).$$

With this definition the dependence structure of X can now be described by the following graph.

Definition 2.2 (Granger-causality graph) The causality graph of a process X relative to the information set I_Y is given by the mixed graph $G = (V, E)$ with vertices $V = \{1, \ldots, d\}$ and edges E such that for all $a, b \in V$ with $a \neq b$

(i) $a \longrightarrow b \notin E \Leftrightarrow X_a \nrightarrow X_b \ [I_Y]$,

(ii) $a \text{ --- } b \notin E \Leftrightarrow X_a \nsim X_b \ [I_Y]$.

For simplicity we will only speak of causality and causality graphs instead of Granger-causality resp. Granger-causality graphs. The directed edges in the causality graph correspond to direct causal relations between the components of X (relative to the chosen information set) whereas causality patterns like spurious or indirect causal relations lead to more complex configurations. The simplest examples of such configurations involving only three variables are depicted in Figure 2.1 (c) and (d). We note that in a bivariate analysis of $X_{\{a,b\}}$ these two configurations may lead to a directed egde $a \longrightarrow b$ and thus would be indistinguishable from configuration (a). Figure 2.1 (e) depicts a configuration where X_b becomes causal for X_a only after including X_c in the information set (assuming that d represents a latent variable). This spurious causality can be detected only by examination of general noncausality relations relative to reduced information sets.

Example 2.3 (VAR-processes) Let X be a weakly stationary vector autoregressive process of order p,

$$X(t) = A(1)X(t-1) + \ldots + A(p)X(t-p) + \varepsilon(t),$$

where $A(j)$ are $d \times d$ matrices and the $\varepsilon(t)$ are independent and identically distributed innovations with mean zero and nonsingular covariance matrix Σ. Setting $I_X(t) = X(-\infty, t\,]$ it is well known (cf. Tjøstheim, 1981; Hsiao, 1982) that X_a is noncausal for X_b if and only if the corresponding entries $A_{ba}(j)$ vanish in all matrices $A(j)$, i.e.

$$X_a \nrightarrow X_b \;\; [I_X] \;\Leftrightarrow\; A_{ba}(j) = 0 \quad \forall j \in \{1, \ldots, p\}. \tag{2.1}$$

Further, X_a and X_b are contemporaneously conditionally orthogonal if and only if the corresponding error components $\varepsilon_a(t)$ and $\varepsilon_b(t)$ are conditionally orthogonal given all remaining components $\varepsilon_{V\setminus\{a,b\}}(t)$. It then follows from the inverse variance lemma (e.g. Whittaker, 1990, Prop. 5.7.3) that contemporaneous conditional orthogonality between the components of X is given by zeros in the inverse covariance matrix $K = \Sigma^{-1}$. More precisely, we have

$$X_a \nsim X_b \;\; [I_X] \;\Leftrightarrow\; \varepsilon_a(t) \perp \varepsilon_b(t) \,|\, \varepsilon_{V\setminus\{a,b\}}(t) \;\Leftrightarrow\; k_{ab} = k_{ba} = 0. \tag{2.2}$$

In the case of normally distributed innovations $\varepsilon(t)$ these conditions correspond to a covariance selection model (Dempster, 1972) for the innovations.

As an example, we consider a five-dimensional VAR(1)-process with parameters

$$A(1) = \begin{pmatrix} a_{11} & 0 & a_{13} & 0 & 0 \\ 0 & a_{22} & 0 & a_{24} & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & a_{45} \\ 0 & 0 & a_{53} & 0 & a_{55} \end{pmatrix}, \qquad K = \begin{pmatrix} k_{11} & k_{12} & k_{13} & 0 & 0 \\ k_{21} & k_{22} & k_{23} & 0 & 0 \\ k_{31} & k_{32} & k_{33} & 0 & 0 \\ 0 & 0 & 0 & k_{44} & 0 \\ 0 & 0 & 0 & 0 & k_{55} \end{pmatrix}.$$

According to conditions (2.1) and (2.2) the zeros in these matrices now correspond to missing edges in the causality graph. The resulting graph is shown in Figure 2.2. From this graph, we immediately can see that for example X_1 is noncausal for X_4 relative to the full information set I_X. The graph further contains a directed path from vertex 1 to 4 (via 3 or via 3 and 5) which indicates that X_1 causes X_4 indirectly. However, since every directed path intersects vertex 3 this suggests that

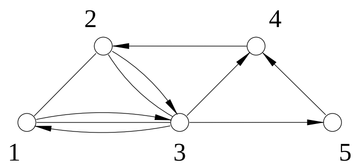

Figure 2.2: Causality graph for the vector autoregressive process in Example 2.3.

X_1 provides no additional information for the prediction of X_4 if X_3 is included in the information set, i.e. $X_1 \nrightarrow X_4 \quad [I_{X_{\{1,3,4\}}}]$. Such general noncausality relations can indeed be derived from the causality graph as will be shown in the next section.

Now suppose that a detailed analysis of the time series showed that $X_{\{1,3\}}$ were noncausal for X_4 after removing X_5 from the information set. Since this general noncausality relation cannot be derived from the graph in Figure 2.2 we would like to know whether this additional knowledge can be used for a modification of the graph such that the noncausality relation holds in the modified graph.

3 Properties of Granger-causality graphs

Granger-causality graphs visualize the pairwise causal relationships between the components of a process X. However, for a better understanding of the causal structure implied by the graph we are interested in the general causality relations relative to reduced information sets. For this we will introduce a concept of separation for mixed graphs which allows to state whether or not in a graph two given subsets of vertices are separated by a third subset of vertices. We then show that all valid separation statements can be linked to general noncausality relations. In the literature on graphical models there are two main approaches for introducing separation in graphs which contain directed edges. The first approach is based on path-oriented criteria as the d-separation for directed acyclic graphs (Pearl, 1988), while the second approach, defined in Frydenberg (1990) for the class of chain graphs, utilizes graph separation in undirected graphs by applying the operation of moralization to subgraphs induced by certain subsets of vertices. In the present paper we follow the latter approach. We start by giving the necessary definitions from graph theory.

3.1 Graph-theoretic definitions

Let $G = (V, E)$ be a mixed graph with vertices V and edges E. Two vertices a and b which are joined by an edge are said to be adjacent. If a and b are connected by an undirected edge $a — b$ they are said to be neighbours. The set of all neighbours of a in G is denoted by $\mathrm{ne}_G(a)$. If there is a directed edge $a \longrightarrow b$ in G then a is a parent of b and b is a child of a. The sets of all parents and of all children are denoted by $\mathrm{pa}_G(a)$ and $\mathrm{ch}_G(a)$, respectively. Finally, a vertex b is said to be an ancestor of a if there exists a directed path $b \longrightarrow \cdots \longrightarrow a$ in G, and $\mathrm{an}_G(a) = \{v \in V | v \longrightarrow \cdots \longrightarrow a \text{ in } G \text{ or } v = a\}$ denotes the set of all ancestors of a. If it is clear which graph G is meant we simple use $\mathrm{ne}(a)$, $\mathrm{pa}(a)$, $\mathrm{ch}(a)$, and $\mathrm{an}(a)$.

Let S be a subset of V. The expressions $\mathrm{ch}(S)$, $\mathrm{pa}(S)$, and $\mathrm{an}(S)$ denote the collection of children, parents, families, and ancestors, respectively, of vertices in S, that is e.g. $\mathrm{ch}(S) = \cup_{s \in S}\mathrm{ch}(s)$. A subset S is called an ancestral set if it contains all its ancestors, i.e. $\mathrm{an}(S) = S$. Further a subset S is complete if all pairs of vertices in S are adjacent.

From a mixed graph G, the undirected subgraph $G^{\mathrm{u}} = (V, E^{\mathrm{u}})$ is obtained by removing all directed edges, i.e. $E^{\mathrm{u}} = \{e \in E | e \text{ is undirected}\}$. Further, if S is a subset of V it induces a subgraph $G_S = (S, E_S)$ where E_S is obtained from E by keeping edges with both endpoints in S.

Finally, we need the concept of separation in undirected graphs. Let $G = (V, E)$ be an undirected graph and A, B, and S disjoint subsets of V. Then the set S separates the sets A and B, denoted by $A \bowtie B \,|\, S$, if every path a — $\cdots$ — b from an element a in A to an element b in B intersects S. We note that the separation in undirected graphs formally satisfies the properties listed in Proposition A.1 (e.g. Lauritzen, 1996).

3.2 Moralization in causality graphs

The essential feature of the graphical modelling approach is to relate the conditional association structure of a multivariate random variable to a graph. One particularly simple, but important class of graphs are the undirected independence graphs $G = (V, E)$ where E consists of undirected edges between all pairs of variables that are not conditionally independent given all other variables. In a linear framework conditional independence may be replaced by conditional orthogonality. In this case we say that the random variable satisfies the pairwise linear Markov property with respect to the graph G. For the interpretation of such graphs one is more interested in a stronger property which holds e.g. under the assumptions of Lemma A.2. Let G be the independence graph of a random vector η. Then the separation properties in G can directly be translated into conditional orthogonality relations between subsets of η. More precisely, if $A \bowtie B | C \;\; [G]$ holds for disjoint subsets A, B, and C of V then $\eta_A \perp \eta_B \,|\, \eta_C$ (e.g. Lauritzen, 1996). This property is known as the global (linear) Markov property for undirected graphs.

In order to exploit the global Markov property for undirected graphs we now introduce a concept of moralization of mixed graphs which yields an undirected graph which in the case of causality graphs reflects certain conditional orthogonality relations of the process X. The key to this concept of moralization is the following theorem.

THEOREM 3.1 *Let $G = (V, E)$ be the causality graph of X relative to I_Y. If for $a, b \in V$ the causality graph G satisfies the following conditions*

(*a*) $a \notin \mathrm{ne}(b)$;

(*b*) $a \notin \mathrm{ch}(b)$ *and* $b \notin \mathrm{ch}(a)$;

(*c*) $\mathrm{ne}(a) \cap \mathrm{ch}(b) = \emptyset$ *and* $\mathrm{ch}(a) \cap \mathrm{ne}(b) = \emptyset$;

(*d*) $\mathrm{ch}(a) \cap \mathrm{ch}(b) = \emptyset$;

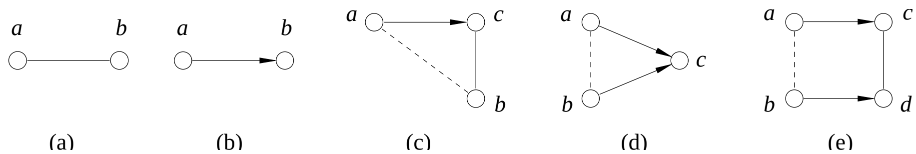

Figure 3.1: Configurations in the mixed graph which lead to an edge between vertices a and b in the moral graph: (a) neighbours, (b) parent and child, (c) flag, (d) immorality, and (e) 2-biflag.

$(e)\quad \mathrm{ne}(\mathrm{ch}(a)) \cap \mathrm{ch}(b) = \emptyset$

then we have $X_a(\tau, t] \perp X_b(\tau, t] \,|\, I_{Y\setminus X_{\{a,b\}}}(t)$.

Let a, b, c, and d be distinct vertices in V. As in Andersson et al. (2001) the subgraph induced by $\{a, b, c\}$ is called a flag (or an immorality) in G if in this graph a and b are not adjacent and therefore satisfy conditions (a) and (b) whereas condition (c) (or condition (d) in the case of an immorality) is violated. Further the subgraph induced by $\{a, b, c, d\}$ forms a 2-biflag if again a and b are not adjacent in the induced subgraph but condition (e) does not hold. These subgraphs are depicted in Figure 3.1. We note that except for an edge between vertices a and b all vertices may be joined by further edges.

Theorem 3.1 now suggests the following definition of a moral graph with edges inserted whenever one of the conditions (c) to (e) is violated. We note that a similar definition is given in Andersson et al. (2001) in the context of chain graphs satisfying the so-called *Alternative Markov property*, whereas the definition differs from the concept of moralization commonly used for chain graphs (Frydenberg, 1990).

Definition 3.2 Let $G = (V, E)$ be a mixed graph. The moral graph $G^{\mathrm{m}} = (V, E^{\mathrm{m}})$ derived from G is defined as the undirected graph obtained by completing all immoralities, flags, and 2-biflags in G and then converting all directed edges in G to undirected edges.

With this definition Theorem 3.1 is equivalent to the statement that the random vector $\eta = (\eta_a)_{a\in V}$ with components $\eta_a = X_a(\tau, t]$ satisfies the pairwise linear Markov property with respect to the moral graph G^{m}.

We note that for weakly stationary time series X undirected graphs which describe the dependence structure of the series have already been considered by Dahlhaus (2000). More precisely an edge a — b is absent in the partial correlation graph G^{pc} of X if and only if $X_a(t)$ and $X_b(s)$ are uncorrelated for all $t, s \in \mathbb{Z}$ after removing the linear effects of all other components $X_{V\setminus\{a,b\}}$. Theorem 3.1 now shows that the partial corrleation graph and the causality graph of a time series are linked by the operation of moralization.

Corollary 3.3 *Let* $G = (V, E)$ *be the causality graph and* $G^{\mathrm{pc}} = (V, E^{\mathrm{pc}})$ *be the partial correlation graph of a weakly stationary process* X. *Then* G^{pc} *is a subgraph of* G^{m}, $G^{\mathrm{pc}} \subseteq G^{\mathrm{m}}$.

Although the two graphs in Corollary 3.3 are identical for most processes this is not always true as shown in the next example.

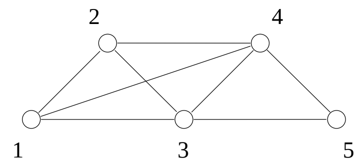

Figure 3.2: Moral graph G^{m} for the vector autoregressive process in Example 2.3.

EXAMPLE 3.4 We consider again the vector autoregressive process in Example 2.3. In the corresponding causality graph shown in Figure 2.2 the subgraph induced by the vertices $\{1,2,4\}$ forms a flag since vertices 1 and 4 are not adjacent. Similarly the subgraph induced by $\{1,2,3,4\}$ is a 2-biflag. Therefore the moral graph G^{m} in Figure 3.2 is obtained by completion of these two subgraphs which results in the insertion of one additional edge joining vertices 1 and 4.

In the partial correlation graph G^{pc}, on the other hand, an edge $a — b$ is absent if and only if there is a zero at the corresponding position in the inverse spectral matrix $f(\lambda)^{-1}$ (cf. Dahlhaus, 2000) which is equivalent to the following parameter constraints (with $K = \Sigma^{-1}$ and $A(1) = (a_{ij})$)

$$\Big(k_{ab} + \sum_{j,k=1}^{5} k_{jk} a_{ja} a_{kb}\Big) = 0, \quad \sum_{k=1}^{5} k_{ak} a_{kb} = 0, \quad \text{and} \quad \sum_{k=1}^{5} k_{bk} a_{ka} = 0.$$

Obviously these constraints are satisfied for all pairs a, b for which conditions (a) to (e) in Theorem 3.1 hold or equivalently for which $a — b \notin G^{\mathrm{m}}$. Further conditional orthogonalities are possible only under additional restriction on the parameters. For example, the edge $1 — 2$ is absent in G^{pc} if $k_{12} = -a_{31}k_{33}a_{32}$, $k_{13} = -k_{12}a_{22}/a_{32}$, and $k_{23} = -k_{12}a_{11}/a_{31}$. In general such constraints only characterize a null set in the parameter space and the two graphs G^{m} and G^{pc} are identical for almost all vector autoregressive processes.

3.3 General noncausality relations

We return now to the original problem of identifying the general noncausality relations relative to reduced information sets which are implied by a causality graph. The identification method presented here is based on an extension of the concept of moralization discussed in the previous section. Although the results in this section also hold with exogenous variables Z included in the information set I_Y, we suppress them for the remainder of this section for notational convenience and set $Y = X$.

We first consider subprocesses X_S where S is an ancestral subset of V. In this case the variables in $X_{V\setminus S}$ are not explanatory for the subprocess X_S and their removal from the analysis does not lead to new indirect or spurious causalities. Therefore the directed edges in the causality graph of X_S can be obtained from G by simply keeping all directed edges $e \in E$ with both endpoints in S. Likewise the reduction of the information set does not lead to additional spurious instantaneous causalities (cf. Granger, 1988) and thus the contemporaneous conditional association structure of X_s can be determined solely from the undirected edges in the full graph G. However two vertices a and b should be joined by an undirected edge $a — b$ in the causality graph of X_S not only if the edge already exists in the full graph G but also if they are connected by an undirected path $a \text{---} c_1 \text{---} \ldots \text{---} c_n \text{---} b$ where the intermediate vertices c_j are not in S. The next proposition states that the mixed graph thus obtained indeed represents the pairwise noncausality relations relative to I_{X_S}.

Proposition 3.5 *Let $G = (V, E)$ be the causality graph of X relative to I_X. For a subset S of V we define $G\langle S\rangle = (\mathrm{an}(S), E\langle S\rangle)$ as the mixed graph derived from G such that for all $a, b \in \mathrm{an}(S)$*

(i) $\quad a \longrightarrow b \notin E\langle S\rangle \;\Leftrightarrow\; a \longrightarrow b \notin E,$

(ii) $\quad a \text{---} b \notin E\langle S\rangle \;\Leftrightarrow\; \{a\} \bowtie \{b\} \,|\, S\backslash\{a, b\} \;\; [G^{\mathrm{u}}].$

Then $G\langle S\rangle$ contains the causality graph of $X_{\mathrm{an}(S)}$.

For general subsets S of V the removal of intermediate or explanatory variables may lead to indirect resp. spurious causality. In this case it is not immediately clear how the causality graph for the subprocess can be derived from the full graph. Instead we consider the larger subprocess $X_{\mathrm{an}(S)}$ which includes all explanatory variables and then use moralization and the global Markov property for undirected graphs for the derivation of noncausality relations between the subsets of interest. However, moral graphs as defined in the previous section have an intrinsically symmetric interpretation and therefore do not allow the identification of unidirectional causal relationships (i.e. X_A noncausal for X_B but not vice versa). More precisely, we have as a consequence of Theorem 3.1 and Proposition 3.5 that for disjoint subsets A, B, and C of $\mathrm{an}(S)$

$$A \bowtie B \,|\, C \;\; [G\langle S\rangle^{\mathrm{m}}] \;\Rightarrow\; X_A(\tau, t\,] \perp X_B(\tau, t\,] \,|\, I_{X_C}(t). \tag{3.1}$$

Therefore we introduce an extended concept of moralization for causality graphs which is based on the idea of splitting the past and the present (at time $t+1$) of certain variables and considering them together in one graph. Suppose we want to know whether $X_A \nrightarrow X_B \;\; [I_{X_S}]$ holds for disjoint subsets A and B of S. The corresponding orthogonality relation

$$X_B(t+1) \perp X_A(\tau, t\,] \,|\, I_{X_{S\backslash A}}(t) \tag{3.2}$$

involves besides the history of X_S up to time t also the value of X_B one step ahead. We therefore seek to modify the moral graph $G\langle S\rangle^{\mathrm{m}}$ such that it includes also the process X_B at time $t+1$. Then the orthogonality relation (3.2) can be verified by means of graph separation in this modified graph.

For any subset B of S the splitting of past and present of variables in X_B can be accomplished by augmenting the moral graph $G\langle S\rangle^{\mathrm{m}}$ with new vertices b^* for all $b \in B$. These represent the variables at time $t+1$, i.e. b^* corresponds to $X_b(t+1)$, whereas the vertices in $\mathrm{an}(A)$ stand for the history of the process up to time t as in (3.2). We then join each new vertex b^* with the corresponding vertex b and its parents of b by a directed edge pointing towards b^*. Furthermore, two vertices b_1^* and b_2^* in B^* are joined by an undirected edge whenever the two corresponding vertices b_1 and b_2 are connected by an undirected edge or an undirected path $b_1 — c_1 — \ldots — c_n — b_2$ such that all intermediate vertices are in $V\backslash B$. This construction leads to a chain graph with two chain components S and B^* (i.e. edges between the two components are all directed and point from S to B^* whereas edges within a component are undirected).

DEFINITION 3.6 (Augmentation chain graphs) Let $G = (V, E)$ be a mixed graph and $B \subseteq V$. Then the augmentation chain graph $G^{\mathrm{aug}}_{B^*} = (V \cup B^*, E^{\mathrm{aug}}_{B^*})$ is given by a chain graph with chain components V and B^* such that for all $v_1, v_2 \in V$ and $b_1, b_2 \in B$

(i) $v_1 — v_2 \notin E^{\mathrm{aug}}_{B^*} \;\Leftrightarrow\; v_1 — v_2 \notin E^{\mathrm{m}}$,

(ii) $v_1 \longrightarrow b_1^* \notin E^{\mathrm{aug}}_{B^*} \;\Leftrightarrow\; v_1 \longrightarrow b_1 \notin E$,

(iii) $b_1^* — b_2^* \notin E^{\mathrm{aug}}_{B^*} \;\Leftrightarrow\; b_1 \bowtie b_2 \,|\, B\backslash\{b_1, b_2\} \;\; [G^{\mathrm{u}}]$.

After moralization of the augmentation chain graph $G\langle S\rangle^{\mathrm{aug}}_{B^*}$ the global Markov property can be applied to check whether or not the noncausality relation (3.2) holds for X. It is important to note that we have introduced moralization only for causality graphs, but it can be shown that the augmentation chain graphs satisfies the AMP Markov property, to which this concept of moralization is also applicable (cf. Andersson et al., 2001). This heuristic argument is made rigorous in the following proposition.

PROPOSITION 3.7 *Let $G = (V, E)$ be the causality graph of X relative to I_X. Further let S be a subset of V with $B \subseteq S$ and define $\eta = (\eta_s)_{s\in S}$ and $\eta^* = (\eta_b^*)_{b\in B}$ as random vectors with components $\eta_s = X_s(\tau, t]$ and $\eta_b^* = X_b(t{+}1)$, respectively. Then the joint vector (η, η^*) satisfies the pairwise linear Markov property with respect to $(G\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}$, i.e. for all $a_1, a_2 \in S$ and $b_1, b_2 \in B$ we have*

(i) $a_1 — a_2 \notin (E\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}} \;\Rightarrow\; X_{a_1}(\tau, t] \perp X_{a_2}(\tau, t] \,|\, I_{X_{S\backslash\{a_1,a_2\}}}(t) + X_B(t+1)$,

(ii) $a_1 — b_1^* \notin (E\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}} \;\Rightarrow\; X_{b_1}(t+1) \perp X_{a_1}(\tau, t] \,|\, I_{X_{S\backslash\{a_1\}}}(t) + X_{B\backslash\{b_1\}}(t+1)$,

(iii) $b_1^* — b_2^* \notin (E\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}} \;\Rightarrow\; X_{b_1}(t+1) \perp X_{b_2}(t+1) \,|\, I_{X_S}(t) + X_{B\backslash\{b_1,b_2\}}(t+1)$.

Since under the assumptions on X the pairwise and global Markov property are equivalent we can not identify the causal relationships between components X_A and X_B relative to the reduced information set I_{X_S} by means of graph separation applied to appropriately chosen augmentation chain graphs. We call this property the global causal Markov property.

THEOREM 3.8 *Let $G = (V, E)$ be the causality graph of X relative to I_X. Further let $S \subseteq V$ be partitioned into disjoint subsets A, B, and C. Then X satisfies the following global causal Markov property with respect to G:*

(i) $A \bowtie B^* \,|\, S\backslash A \;\; [(G\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}] \;\Rightarrow\; X_A \nrightarrow X_B \;\; [I_{X_S}]$;

(ii) $A^* \bowtie B^* \,|\, S \cup C^* \;\; [(G\langle S\rangle^{\mathrm{aug}}_{S^*})^{\mathrm{m}}] \;\Rightarrow\; X_A \nsim X_B \;\; [I_{X_S}]$.

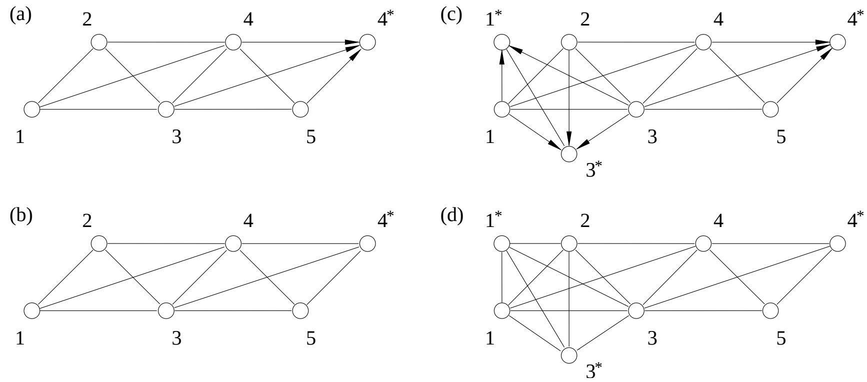

Figure 3.3: Augmentation chain graphs and their moral graphs for the vector autoregressive process in Example 2.3: (a) $G^{\text{aug}}_{\{4^*\}}$, (b) $(G^{\text{aug}}_{\{4^*\}})^{\text{m}}$, (c) $G^{\text{aug}}_{\{1^*,3^*,4^*\}}$, (d) $(G^{\text{aug}}_{\{1^*,3^*,4^*\}})^{\text{m}}$.

EXAMPLE 3.9 We continue with the discussion of Example 2.3. As already mentioned an intuitive interpretation of the causality graph in Figure 2.2 suggests that X_1 has only an indirect effect on X_4 mediated by X_3. That this interpretation is indeed correct can now be shown by deriving the relation $X_1 \nrightarrow X_4 \;\; [I_{X_{\{1,3,4\}}}]$ from the corresponding augmentation chain graph $G\langle \text{an}(1,3,4)\rangle^{\text{aug}}_{\{4^*\}}$.

Since the ancestral set generated by $\{1,3,4\}$ is equal to the full set V we start from the moral graph G^{m} in Figure 3.2. Augmenting the graph with a new vertex 4^* and joining this with vertex 4 and its parents 3 and 5 by arrows pointing towards 4^* we obtain the augmentation chain graph $G^{\text{aug}}_{\{4^*\}}$ in Figure 3.3 (a). As the graph does not contain any flag or 2-biflag the removal of directions yields the corresponding moral graph $(G^{\text{aug}}_{\{4^*\}})^{\text{m}}$ in Figure 3.3 (b). In this graph the vertices 1 and 4^* are separated by the set $\{3,4\}$ and hence the desired noncausality relation follows from Theorem 3.8.

Similarly we find that X_1 and X_4 are contemporaneously partially uncorrelated relative to the same information set. The corresponding augmentation chain graph $G^{\text{aug}}_{\{1^*,3^*,4^*\}}$ and its moral graph are displayed in Figure 3.3 (c) and (d). Since in the augmentation chain graph the subset $\{2,1^*,3^*\}$ forms a flag, the moral graph contains one additional edge between vertices 1^* and 2. In this graph the vertices 1^* and 4^* are separated by the set $\{1,3,4,3^*\}$ and thus $X_1 \nsim X_2 \;\; [I_{X_{\{1,3,4\}}}]$.

4 Identification of causal effects

For a more general discussion of causal effects in multivariate models Lütkepohl (1993) and Dufour and Renault (1998) introduced the notion of noncausality at different horizons h. This concept allows to description of indirect causal effects and the distinction between short-run and long-run causality. In the following we show that that the graphical modelling approach presented in this paper also provides a natural framework for dealing with indirect effects and gives sufficient conditions for noncausality at all horizons.

DEFINITION 4.1 X_a is noncausal for X_b at all horizons relative to I_Y (denoted by

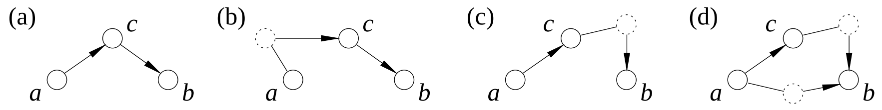

Figure 4.1: Indirect causality $X_a \nrightarrow X_b \;\; [I_{X_{\{a,b,c\}}}]$: (a) and (b) minimal consistent graphs; (c) inconsistent graph; (d) consistent graph.

$X_a \overset{(\infty)}{\nrightarrow} X_b \;\; [I_Y]$) if $X_b(t+h) \perp X_a(\tau, t] \,|\, I_{Y\backslash X_a}(t)$ for all $h \in \mathbb{N}$.

Noncausality from X_a to X_b at all horizons intuitively corresponds to the absence of any direct or indirect effects from X_a to X_b and thus to the nonexistence of a directed path from a to b in the causality graph of X. As in the case of ordinary Granger-noncausality, this can be generalized to arbitrary information sets by relating certain separation properties of the graph to noncausality at all horizons.

Theorem 4.2 *Let $G = (V, E)$ be the causality graph of X relative to I_X. Further let $S \subseteq V$ be partitioned into disjoint subsets A, B, and C. Then we have*

$$A \bowtie \mathrm{an}_G(B) | B \cup C \;\; [G\langle S\rangle^{\mathrm{m}}] \;\Rightarrow\; X_A \overset{(\infty)}{\nrightarrow} X_B \;\; [I_{X_S}]. \tag{4.1}$$

We note that in the case of unreduced information I_X the sufficient condition in (4.1) is fulfilled if and only if the causality graph does not contain any directed paths from A to B, or equivalently $A \subseteq V\backslash\mathrm{an}(B)$. Since further $X_{V\backslash\mathrm{an}(B)} \nrightarrow X_{\mathrm{an}(B)} \;\; [I_X]$ by the composition property, we obtain as a special case of Theorem 4.2 the separation condition for noncausality at all horizons given by Dufour and Renault (1998).

Conversely, the condition of Theorem 4.2 is violated irrespective of the information set whenever for a pair $a \in A$ and $b \in B$ there exists a directed path from a to b. This suggests that one might conclude to a causal effect of X_a on X_b. For this, however, we have to exclude the possibility of a spurious causality. To illustrate the problem let us consider the graph in Figure 4.1 (a) which depicts an indirect causal effect of X_a on X_b, characterized by the noncausality relation $X_a \nrightarrow X_b \;\; [I_{X_{\{a,b,c\}}}]$. This relation does not characterize the configuration uniquely since the edge $a \longrightarrow c$ can be replaced by a spurious causality (Fig. 4.1 (b)) without violating any causality relation which can be derived from the first graph. Thus the noncausality relations of X are not sufficient for the identification of a causal effect of X_a on X_b. On the other hand, substituting a latent structure for the directed edge $c \longrightarrow b$ leads to a violation of $X_a \nrightarrow X_b \;\; [I_{X_{\{a,b,c\}}}]$ as can be seen from the graphs in Figure 4.1 (c) and (d). Rejecting both graphs as a possible model for the causal structure of X we conclude that X_c indeed causes X_b.

This approach for the identification of causal effects can be formally described using the concept of minimal causality graphs consistent with X. It has been motivated by the theory of inferred causation described in Pearl (2000) who addressed the problem of inferring causal effects from multivariate distributions satisfying certain conditional independence relations. In contrast to the graphs described here the causal structures discussed by Pearl need to be directed acyclic graphs.

In the following we consider mixed graphs $G = (\bar{V}, \bar{E})$ with $V \subseteq \bar{V}$. Here, the vertices in $\bar{V}\backslash V$ represent latent variables and are used for modelling noncausality

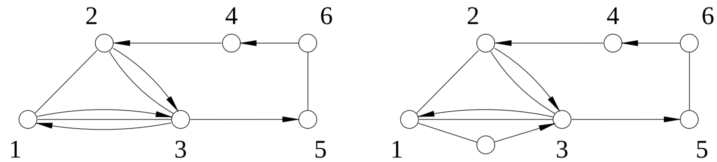

Figure 4.2: Minimal consistent causality graphs for the process in Example 2.3 with additional restriction $X_{\{1,3\}} \nrightarrow X_4 \ [I_{X_{\{1,2,3,4\}}}]$.

relations which are due to spurious causality. A necessary condition for a graph G to serve as a model for the causal structure of a process X is that it does not imply false noncausality relations which do not hold for X.

Definition 4.3 A graph $\bar{G}$ is consistent with X if the process X satisfies the global causal Markov property with respect to $\bar{G}$.

Obviously, this condition is not sufficient for the identification of causal effects since the Markov property trivially holds for a saturated graph with all possible edges included. We therefore need to impose further conditions. For this let $\mathscr{S}(\bar{G})$ be the set of all separation statements of the form $A \bowtie B^* \,|\, B \cup C \ \ [(\bar{G}\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}]$ or $A^* \bowtie B^* \,|\, S \cup C^* \ \ [(\bar{G}\langle S\rangle^{\mathrm{aug}}_{S^*})^{\mathrm{m}}]$, where $S \subseteq V$ with partition $S = A \cup B \cup C$, which are valid for the graph G.

Definition 4.4 A graph $\bar{G}$ is minimal in the class of all graphs consistent with X if for any other consistent graph $\bar{G}'$ we have $\mathscr{S}(\bar{G}) = \mathscr{S}(\bar{G}')$ whenever $\mathscr{S}(\bar{G}) \subseteq \mathscr{S}(\bar{G}')$.

In other words any graph $\bar{G}'$ which implies further noncausality relations additional to those implied by a minimal graph G is not consistent with X. To illustrate these conditions let us consider again the graphs in Figure 4.1. Since the first two graphs both imply the same separation statements they are both minimal consistent with X unless X satisfies further noncausality relations. On the other hand, graph (d) is not minimal since it does not account for the noncausality from X_a to X_b relative to I_X whereas graph (c) is not even consistent with X since it falsely imposes a marginal noncausality from X_a to X_b which does not hold for X.

The minimal graphs consistent with a process X describe all causal models which can be used for an explanation of the observed noncausality relations of X. Therefore if a directed edge $a \longrightarrow b$ is contained in all these graphs the predictability of X_b by X_a cannot be attributed to the influence of a common explanatory variable only and thus necessarily implies the existence of a causal influence of X_a on X_b.

Definition 4.5 (Inferred causation) X_a has an causal effect on X_b if there exists a directed path from a to b in every minimal causality graph $\bar{G}$ consistent with X.

We conclude this section with a final discussion of Example 2.3.

Example 4.6 Suppose that a detailed analysis of the time series showed that $X_{\{1,3\}}$ were noncausal for X_4 after removing X_5 from the information set. Definition 4.5 suggests that there exists a minimal graph consistent with X without a directed path from 3 to 4. Replacing the edge $3 \longrightarrow 5$ by a path $3 \longleftarrow \bar{v} \longrightarrow 5$ with a latent variable $\bar{v}$, we obtain an additional noncausality relation $X_1 \nrightarrow X_5 \ [I_{X_{\{1,5\}}}]$ which does not hold for X. Similarly, the edge $3 \longrightarrow 4$ cannot be replaced by a latent structure either, which leaves us with the graphs depicted in Figure 4.2. It can be easily shown that both graphs are minimal and consistent with X and that there exist no further minimal and consistent graphs. Thus we can conclude e.g. that X_4 has a causal effect on X_5 whereas it remains unclear whether also X_1 has a causal effect on X_5.

5 Extension to strong causality

Since causality graphs as defined so far capture only the linear causal relationships between the components of a multivariate time series they are inappropriate when describing e.g. financial time series which exhibit strong conditional heteroscedasticity. For such time series the definition of causality graphs needs to be generalized to include also nonlinear causal relations between the variables.

Florens and Mouchart (1982) have defined a stronger version of Granger-causality in terms of conditional independence. Denoting the conditional independence of random variables X and Y given Z by $X \perp\!\!\!\perp Y \,|\, Z$ we can modify the definition of Granger causality by substituting $\perp\!\!\!\perp$ for $\perp$. Thus X_a is said to be strongly noncausal for X_b relative to the information set I_Y if

$$X_b(t+1) \perp\!\!\!\perp X_a(\tau, t] \,|\, I_{Y \setminus X_a}(t) \quad \text{for all } t > \tau.$$

Similarly we replace contemporaneous conditional orthogonality by contemporaneous conditional independence. With these definitions we obtain causality graphs which take into account the full causal structure of a process.

The results in Section 3 have been derived from the properties of the conditional orthogonality summarized in Proposition A.1. While (i), (iii), (iv), and the decomposition property also hold when substituting conditional independence for conditional orthogonality, additional assumptions are needed to guarantee the composition and the intersection property. The latter holds under the assumption of measurable separability of the variables (Florens et al., 1990) which for finite-dimensional random vectors is satisfied if the joint probability has a positive and continuous density.

The composition property is only required to establish the equivalence of the noncausality for single component processes X_a, $a \in A$ and the noncausality for the joint vector process X_A, i.e.

$$X_a(t+1) \perp\!\!\!\perp X_B(\tau, t] \,|\, I_{Y \setminus X_B}(t) \, \forall \, a \in A \ \Leftrightarrow \ X_A(t+1) \perp\!\!\!\perp X_B(\tau, t] \,|\, I_{Y \setminus X_B}(t). \quad (5.1)$$

We give three examples of classes of processes for which this equivalence still holds in the case of strong causality.

(a) X is a Gaussian process. Then conditional independence corresponds to conditional orthogonality for which the composition property holds.

(b) The components of X are contemporaneously conditionally independent. Then we have trivially (setting $Y = X$ for convenience) $X_a(t+1) \perp\!\!\!\perp X_{A\setminus\{a\}}(t+1) \,|\, I_X(t)$ for all $a \in A \subseteq V$. Together with the left hand side in (5.1) this implies

$$X_a(t+1) \perp\!\!\!\perp \bar{X}_B(t+1) \,|\, I_X(t) + X_{A\setminus\{a\}}(t+1) \quad \forall a \in A,$$

from which the right hand side in (5.1) now follows by the intersection property.

(c) X is a nonlinear autoregressive process of the form

$$X_a(t+1) = f_a(X(\tau, t\,], \varepsilon_a(t+1)) \quad a = 1, \dots, d \tag{5.2}$$

where f_a are measurable functions strictly monotone in $\varepsilon_a(t+1)$ for fixed $X(\tau, t\,]$ and the innovations $\varepsilon(t)$ are independent of $X(\tau, t\,]$ and have a positive density on $\mathbb{R}$. Then if X_b is noncausal for X_a the function $f_a(X(\tau, t\,], \varepsilon_a(t))$ is constant in $X_b(\tau, t\,]$ almost surely. To show this we write $f_{x_b}(y) = f_a(x, y)$ for any $X(\tau, t\,] = x$ to denote that we leave all components of $X(\tau, t\,]$ except $X_b(\tau, t\,]$ fixed. Since $\varepsilon_a(t+1)$ is independent from $X(\tau, , t\,]$ we have

$$\mathbb{P}\big(X_a(t+1) \leq y \,|\, X(\tau, t\,] = x\big) = \mathbb{P}\big(f_{x_b}(\varepsilon_a(t+1)) \leq y\big).$$

The strong noncausality of X_a for X_b then implies that the left hand side does not depend on x_b. Because of the strict monotonicity of the function f_{x_b} for each x_b this is equivalent to $f_{x_b} = f_{x_b'}$ almost surely.

Since the vector $(f_a, a \in A)$ does not depend on $X_b(\tau, t\,]$ if and only if this holds for each component f_a separately this implies the equivalence in (5.1).

The last class of processes is fairly general and includes in particular nonlinear autoregressive models with additive non-Gaussian errors. As an example for a model with nonadditive errors we consider a multivariate ARCH process.

Example 5.1 (ARCH(q) process) Let X be a stationary process with conditional normal distribution

$$\mathscr{L}(X(t) \,|\, X(t-1), \dots, X(t-q)) \sim \mathcal{N}(0, H(t)^{1/2} \Sigma H(t)^{1/2}),$$

where Σ is a symmetric and positive definite matrix and $H(t)$ is a diagonal matrix with elements

$$h_a(t) = \sigma_a^2 + \sum_{u=1}^{q} X(t-u)' B^{(a)}(u) X(t-u), \quad a = 1, \dots, d$$

for nonnegative definite matrices $B^{(a)}(u)$. This process is a special case of a multivariate ARCH(q) process (e.g. Gouriéroux, 1997). We can rewrite $X(t)$ as a nonlinear autoregressive process of the form (5.2)

$$X_a(t) = f_a(X(t-1), \dots, X(t-q), \varepsilon_a(t)) = \sqrt{h_a(t)}\varepsilon_a(t), \quad a = 1, \dots, d,$$

where $\varepsilon(t)$ are independent and identically distributed with mean zero and covariance matrix Σ. Since the conditional variances $h_a(t)$ are positive the function f_a is monotonically increasing in $\varepsilon_a(t)$ and thus satisfies the conditions of the example above.

Noncausality between the components can now be expressed in terms of the parameters $B^{(a)}(u)$. If $B^{(a)}_{bk}(u) = B^{(a)}_{kb}(u) = 0$ for all $k = 1, \ldots, d$ and $u = 1, \ldots, q$ the conditional variance $h_a(t)$ does not depend on past values of X_b and consequently X_b is noncausal for X_a. Further zeros in the inverse of the covariance matrix Σ correspond to contemporaneous conditional independence as in the case of linear autoregressive processes.

6 Concluding remarks

In this paper, we have introduced a graphical modelling approach for time series based on the concept of Granger-causality graphs. Let us summarize some of the advantages of this approach.

First, graphs can be easily visualized and thus provide a concise way to communicate the pairwise noncausality relations of a time series. Second, under mild assumptions on the time series the global causal Markov property holds. This property enables one to conclude from the pairwise noncausality relations reflected by the causality-graph to general noncausality relations relative to information subsets. Such general noncausality relations are central for the discussion of direct, indirect, and spurious causality. Third, spurious causality can be modelled explicitly by inclusion of latent variables, which are represented by additional vertices in the graph. Thus causality graphs can be obtained which imply the same general noncausality relations which have been empirically found for the time series. Although the causality graph may not be uniquely determined, the identification of all such graphical causal models leads to sufficient conditions for the identification of causal effects. Finally, graphs are simple objects which can be easily implemented on the computer. This can be exploited when investigating causal structures of high-dimensional time series.

The results presented here mainly focus on theoretical aspects of the graphical modelling approach. In particular, we have not been concerned with statistical methods for the estimation of general noncausality relations from time series data. Furthermore, search algorithms need to be developed which allow the identification of all minimal graphs consistent with these relations.

Appendix

A.1 Conditional orthogonality

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. The usual orthogonality with respect to this inner product is denoted by $x \perp y$, i.e. $\langle x, y \rangle = 0$. For a closed linear subspace L of H let $L^{\perp}$ be the orthogonal complement of L, i.e. $L^{\perp} = \{x \in H | \langle x, L \rangle = 0\}$, and Lx be the orthogonal projection of the vector $x \in H$ on L, i.e. the unique vector $y \in L$ such that $x - y \perp L$. Finally, the closed sum of two linear subspaces L_1 and L_2 is denoted by $L_1 + L_2$.

Let L_1, L_2, and L_3 be closed linear subspaces of H. Then L_1 and L_2 are defined to be conditional orthogonal given L_3 if

$$x - L_3 x \perp y - L_3 y \quad \forall\, x \in L_1\, \forall\, y \in L_2$$

or equivalently $L_3^{\perp} L_1 \perp L_3^{\perp} L_2$. The conditional orthogonality is denoted by $L_1 \perp L_2 \,|\, L_3$. The basic properties of this relation are summarized in the following proposition.

Proposition A.1 *Let L_i, $i = 1, \ldots, 4$ be closed linear subspaces of H. Then the conditional orthogonality in H has the following properties:*

(i) *Symmetry.*
If $L_1 \perp L_2 \,|\, L_3$ then $L_2 \perp L_1 \,|\, L_3$.

(ii) *Composition/decomposition:*
$L_1 \perp L_2 \,|\, L_4$ and $L_1 \perp L_3 \,|\, L_4$ if and only if $L_1 \perp L_2 + L_3 \,|\, L_4$.

(iii) *Weak union:*
If $L_1 \perp L_2 \,|\, L_3$ then $L_1 \perp L_2 \,|\, L_3 + U$ for all closed linear subspaces $U \subseteq L_2$.

(iv) *Contraction:*
If $L_1 \perp L_2 \,|\, L_4$ and $L_1 \perp L_3 \,|\, L_2 + L_4$ then $L_1 \perp L_2 + L_3 \,|\, L_4$.

If further $L_2 + L_3$ is separable and $(L_2 + L_4) \cap (L_3 + L_4) = L_4$ then we have additionally

(v) *Intersection:*
If $L_1 \perp L_2 \,|\, L_3 + L_4$ and $L_1 \perp L_3 \,|\, L_2 + L_4$ then $L_1 \perp L_2 + L_3 \,|\, L_4$.

Proof. The first four properties can be proved easily using the properties of ordinary orthogonality in H (e.g. Florens and Mouchart, 1985). For the proof of the last statement we first consider the case where L_2 is finite-dimensional and $L_4 = \{0\}$. By the definition of conditional orthogonality we get $L_1 \perp L_3^{\perp} L_2 + L_2^{\perp} L_3$. Since $L_2 \cap L_3 = \{0\}$ we get $L_2 L_3^{\perp} L_2 = L_2$ and hence $L_2 + L_3 = L_2 + L_2^{\perp} L_3 = L_2 L_3^{\perp} L_2 + L_2^{\perp} L_3$. On the other hand $L_2^{\perp} L_3 = L_2^{\perp} \cup (L_2 + L_3)$ and consequently $L_3^{\perp} L_2 + L_2^{\perp} L_3 = L_2 L_3^{\perp} L_2 + L_2^{\perp} L_3$. Thus $L_1 \perp L_2 + L_3$.

For general separable L_2 we consider orthogonal decompositions $L_2 = U_1 \oplus U_2$ where U_1 is finite-dimensional and apply the previous case to $U_2 L_1$, U_1, and $U_2 L_3$, thus obtaining $L_1 \perp U_1 + L_3 \,|\, U_2$. By the orthogonality of U_1 and U_2 and the decomposition and contraction property this yields $L_1 \perp U_1$ and for $U_1 \nearrow L_2$ $L_1 \perp L_2$. The desired conditional orthogonality now follows from $L_1 \perp L_3 \,|\, L_2$ and the contraction property.

Finally, the general case of $L_4 \neq \{0\}$ can now be derived from this by substituting $L_4^{\perp} L_i$ for L_i. □

For countable subsets X of $\mathscr{L}^2$ let $L_X = \overline{\mathrm{sp}}(X)$ denote the closed linear subspace generated by X. Then X and Y are conditional orthogonal given Z denoted by $X \perp Y \,|\, Z$ if $L_X \perp L_Y \,|\, L_Z$. Further if $f(X, Z) = g(Y, Z)$ a.s. for linear functions f and g implies that $f(X, Z) = h(Z)$ a.s. for some linear function h then the linear subspaces L_X, L_Y, and L_Z satisfy the condition in Proposition A.1 (v), i.e. $(L_X + L_Z) \cap (L_Y + L_Z) = L_Z$. In this case we say that X and Y are linearly separated conditionally on Z.

Lemma A.2 *For $I \subseteq \mathbb{N}$ let $X_I = (X_i)_{i \in I}$ be a random vector in $\mathscr{L}^2$ with positive definite covariance matrix Σ such that $\inf_{x \in \mathbb{R}^{|I|}} \|\Sigma x\| / \|x\| \geq c > 0$. Then for disjoint subsets A, B, and C of I the vectors X_A and X_B are linearly separated conditionally on X_C.*

Proof. Let $f(X_A, X_C) = \alpha' X_A + \gamma' X_C$ and $g(X_B, X_C) = \beta' X_B + \tilde{\gamma}' X_C$. If $f(X_A, X_C) = g(X_B, X_C)$ a.s. it follows from the assumption on Σ that

$$0 = \mathrm{var}(f(X_A, X_C) - g(X_B, X_C)) \geq c(\|\alpha\|^2 + \|\beta\|^2 + \|\gamma - \tilde{\gamma}\|^2).$$

The last term vanishes only if $\alpha = 0$ and $\beta = 0$. Hence $f(X_A, X_C) = \gamma' X_C$ a.s. □

The lower bound on the eigenvalues of Σ is in particular satisfied by any stochastic process under the assumptions in Section 2.

A.2 Proofs

For notational convenience we give the proofs only for $I_Y(t) = X(\tau, t\,]$.

Proof of Theorem 3.1. We show that

$$X_a(s) \perp X_b(s') \,|\, X(\tau, t\,] \backslash \{X_a(s), X_b(s')\}$$

for all $\tau < s, s' \leq t$, which is equivalent to the asserted conditional orthogonality relation. Let $s \leq s'$, otherwise we swap the indices a and b. First, we consider the case $s < s'$. By conditions (b) and (c) X_a is noncausal for $X_{\mathrm{ne}(b)\cup\{b\}}$ which yields

$$X_b(s') \perp X_a(s) \,|\, X(\tau, s'-1\,] \backslash \{X_a(s)\}, X_{\mathrm{ne}(b)}(s').$$

Since further X_b and $X_{V\backslash(\mathrm{ne}(b)\cup\{b\})}$ are contemporaneously conditionally orthogonal by the definition of neighbours it follows by the contraction property that

$$X_b(s') \perp X_a(s) \,|\, X(\tau, s'] \backslash \{X_a(s), X_b(s')\},$$

which in the case $s = s'$ directly follows from (a).

Now let us assume

$$X_b(s') \perp X_a(s) \,|\, X(\tau, t\,] \backslash \{X_a(s), X_b(s')\}$$

for some $t \geq s'$. Since by condition (c) X_b is noncausal for $X_{V\backslash(\mathrm{ch}(b)\cup\{b\})}$ this can be extended to

$$X_b(s') \perp X_a(s) \,|\, X(\tau, t\,] \backslash \{X_a(s), X_b(s')\}, X_{V\backslash(\mathrm{ch}(b)\cup\{b\})}(t+1). \tag{A.1}$$

Similarly X_a is noncausal for $X_{V\backslash(\mathrm{ch}(a)\cup\{a\})}$ and further by conditions (b) and (d) $\mathrm{ch}(b) \cup \{b\} \subseteq V\backslash(\mathrm{ch}(a) \cup \{a\})$, which together leads to

$$X_a(s) \perp X_{\mathrm{ch}(b)\cup\{b\}}(t+1) \,|\, X(\tau, t] \backslash \{X_a(s)\}, X_{V\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1). \tag{A.2}$$

Conditions (a), (c), and (e) now imply

$$X_{\mathrm{ch}(a)\cup\{a\}}(t+1) \perp X_{\mathrm{ch}(b)\cup\{b\}}(t+1) \,|\, X(\tau, t\,], X_{V\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1)$$

which yields together with (A.2)

$$X_a(s) \perp X_{\mathrm{ch}(b)\cup\{b\}}(t+1) \,|\, X(\tau, t\,] \backslash \{X_a(s)\}, X_{V\backslash(\mathrm{ch}(b)\cup\{b\})}(t+1).$$

With (A.1) using the contraction and the weak union property we then obtain

$$X_a(s) \perp X_b(s') \,|\, X(\tau, t+1\,] \backslash \{X_a(s), X_b(s')\}.$$

Application of the intersection property for all s, s' now proves the theorem. □

Proof of Proposition 3.5. By the definition of an ancestral set, S has no predecessors in $V \backslash S$, hence

$$X_A(t+1) \perp X_{V\backslash S}(\tau, t] \,|\, X_S(\tau, t] \tag{A.3}$$

for all $A \subseteq S$. Suppose that $a, b \in S$ are not connected by a directed edge $a \longrightarrow b$ in $G\langle S\rangle$. By definition of $G\langle S\rangle$ the edge $a \longrightarrow b$ is also missing in G, which yields

$$X_a(\tau, t] \perp X_b(t+1) \,|\, X_{V\backslash\{a\}}(\tau, t]$$

and because of (A.3) and the weak union property

$$X_a(\tau, t] \perp X_b(t+1) \,|\, X_S(\tau, t].$$

Next, if a and b are not connected by an undirected edge in $G\langle S\rangle$ then

$$X_a(t+1) \perp X_b(t+1) \,|\, X(\tau, t], X_{S\backslash\{a,b\}}(t+1),$$

which implies again with (A.3)

$$X_a(t+1) \perp X_b(t+1) \,|\, X_S(\tau, t], X_{S\backslash\{a,b\}}(t+1).$$

and thus the contemporaneous conditional orthogonality of X_a and X_b relative to X_S. □

Proof of Proposition 3.7. First we consider vertices a and b in S which are unconnected in $(G\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}$. Then the children of a and b are separated by $B\backslash(\mathrm{ch}(a,b) \cup \{a,b\})$ in G^{u} and hence

$$X_{(\mathrm{ch}(a)\cup\{a\})\cap B}(t+1) \perp X_{(\mathrm{ch}(b)\cup\{b\})\cap B}(t+1) \,|\, X_S(\tau, t], X_{B\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1). \tag{A.4}$$

Further X_a is noncausal for $X_{S\backslash(\mathrm{ch}(a)\cup\{a\})}$ with respect to X_S and thus

$$X_a(\tau, t] \perp X_{(\mathrm{ch}(b)\cup\{b\})\cap B}(t+1) \,|\, X_{S\backslash\{a\}}(\tau, t], X_{B\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1). \tag{A.5}$$

Together with (A.4) this yields

$$X_a(\tau, t], X_{(\mathrm{ch}(a)\cup\{a\})\cap B}(t+1) \perp X_{(\mathrm{ch}(b)\cup\{b\})\cap B}(t+1) \,|\, X_{S\backslash\{a\}}(\tau, t], X_{B\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1). \tag{A.6}$$

Noting that $X_a(\tau, t] \perp X_b(\tau, t] \,|\, X_{S\backslash\{a,b\}}(\tau, t]$ we obtain from (A.5)

$$X_a(\tau, t], X_{(\mathrm{ch}(a)\cup\{a\})\cap B}(t+1) \perp X_b(\tau, t] \,|\, X_{S\backslash\{a,b\}}(\tau, t], X_{B\backslash(\mathrm{ch}(a,b)\cup\{a,b\})}(t+1)$$

and further by the intersection and weak union property

$$X_a(\tau, t] \perp X_b(\tau, t] \,|\, X_{S\backslash\{a,b\}}(\tau, t], X_B(t+1).$$

Next, let $a \in S$ and $b \in B^*$ be unconnected in $(G\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}$. Similarly as for (A.6) in the case before we get

$$X_a(\tau, t], X_{(\mathrm{ch}(a)\cup\{a\})\cap B}(t+1) \perp X_b(t+1) \,|\, X_{S\backslash\{a\}}(\tau, t], X_{B\backslash(\mathrm{ch}(a)\cup\{a,b\})}(t+1),$$

which leads to the desired conditional independence

$$X_a(\tau, t] \perp X_b(t+1) \,|\, X_{S\backslash\{a\}}(\tau, t], X_{B\backslash\{b\}}(t+1).$$

Finally, for $a, b \in B^*$ unconnected in $(G\langle S\rangle^{\mathrm{aug}}_{B^*})^{\mathrm{m}}$ we immediately get

$$X_a(t+1) \perp X_b(t+1) \,|\, X_S(\tau, t], X_{B\backslash\{a,b\}}(t+1)$$

by the definition of edges between vertices in B^*. □

Proof of Theorem 3.8. By Lemma 3.7 the random variables $X_a(\tau,t]$, $a \in \mathrm{an}(S)$ and $X_b(t)$, $b \in B$ satisfy the pairwise Markov property for the the moral graph $(G\langle S\rangle_{B^*}^{\mathrm{aug}})^{\mathrm{m}}$. Using the equivalence of pairwise and global Markov property the first part of the lemma follows. The second part can be derived in the same way. □

Proof of Theorem 4.2. Since the set $\mathrm{an}(B)$ is ancestral there is no directed edge from $V\backslash\mathrm{an}(B)$ to $\mathrm{an}(B)$ in G. Consequently $X_{V\backslash\mathrm{an}(B)} \nrightarrow X_{\mathrm{an}(B)}$ $[I_X]$ which implies for $h \in \mathbb{N}$

$$X_{\mathrm{an}(B)}(t+j) \perp X_{V\backslash\mathrm{an}(B)}(\tau,t] \,|\, X_{\mathrm{an}(B)}(\tau,t+j-1], \quad j=1,\ldots,h.$$

Iterative contraction yields $X_{\mathrm{an}(B)}(t,t+h] \perp X_{V\backslash\mathrm{an}(B)}(\tau,t] \,|\, X_{\mathrm{an}(B)}(\tau,t]$, which finally can be reduced to

$$X_{\mathrm{an}(B)}(t,t+h] \perp X_A(\tau,t] \,|\, X_{\mathrm{an}(B)\cup C}(\tau,t].$$

Further noting that by (3.1) the assumption $A \bowtie \mathrm{an}(B) \,|\, B\cup C$ $[G\langle S\rangle]$ implies $X_A(\tau,t] \perp X_{\mathrm{an}(B)}(\tau,t] \,|\, X_{B\cup C}(\tau,t]$ we obtain

$$X_{\mathrm{an}(B)}(\tau,t+h] \perp X_A(\tau,t] \,|\, X_{B\cup C}(\tau,t],$$

from which the assertion of the lemma follows by application of the decomposition property. □

Acknowledgement

Part of the results presented here have been taken from the author's doctoral thesis (Eichler, 1999). The author is indepted to his supervisor R. Dahlhaus for his support and many helpful comments on an earlier version of the paper.

References

Andersson, S.A., D. Madigan, and M.D. Perlman, 2001, Alternative Markov properties for chain graphs, Scandinavian Journal of Statistics 28, 33-85.

Boudjellaba, H., J.M. Dufour, and R. Roy, 1992, Testing causality between two vectors in multivariate autoregressive moving average models, Journal of the American Statistical Association 87, 1082-1090.

Brillinger, D.R., 1996, Remarks concerning graphical models for time series and point processes, Revista de Econometria 16, 1-23.

Cox, D.R., and N. Wermuth, 1996, Multivariate Dependencies (Chapman and Hall, London).

Dahlhaus, R., 2000, Graphical interaction models for multivariate time series, Metrika 51, 157-172.

Dempster, A.P., 1972, Covariance selection, Biometrics 28, 157-175.

Dufour, J.M., and E. Renault, 1998, Short run and long run causality in time series: theory, Econometrica 66, 1099-1125.

Eichler, M., 1999, Graphical Models in Time Series Analysis, Doctoral thesis, University of Heidelberg, Germany.

Florens, J.P., and M. Mouchart, 1982, A note on noncausality, Econometrica 50, 583-591.

Florens, J.P., and M. Mouchart, 1985, A linear theory for noncausality, Econometrica 53, 157-175.

Florens, J.P., M. Mouchart, and J.M. Rolin, 1990, Elements of Bayesian statistics (Marcel Dekker, New York).

Frydenberg, M., 1990, The chain graph Markov property, Scandinavian Journal of Statistics 17, 333-353.

Geweke, J., 1984, Inference and causality in economic time series, in: Z. Griliches and M.D. Intriligator, eds., Handbook of Econometrics, Vol. 2 (North-Holland, Amsterdam) 1101-1144.

Gouriéroux, C., 1997, ARCH Models and Financial Applications (Springer, New York).

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424-438.

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.

Koster, J.T.A., 1996, Markov properties of nonrecursive causal models, Annals of Statistics 24, 2148-2177.

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., 2000, Causal inference from graphical models, in: E. Barndorff-Nielsen, D.R. Cox, and C. Klüppelberg, eds., Complex Stochastic Systems (CRC Press, London).

Lütkepohl, H., 1993, Testing for causation between two variables in higher dimensional VAR models, in: H. Schneeweiss and K.F. Zimmerman, eds., Studies in Applied Econometrics (Physica-Verlag, Heidelberg) 75-91.

Pearl, J., 1988, Probabilistic Inference in Intelligent Systems (Morgan Kaufmann, San Mateo, California).

Pearl, J., 1995, Causal diagrams for empirical research (with discussion), Biometrika 82, 669-710.

Pearl, J., 2000, Causality (Cambridge University Press, Cambridge, UK).

Sims, C.A., 1980, Macroeconomics and reality, Econometrica 48, 1-48.

Tjøstheim, D., 1981, Granger-causality in multiple time series, Journal of Econometrics 17, 157-176.

Whittaker, J., 1990, Graphical Models in Applied Multivariate Statistics (John Wiley, Chichester).

Zellner, A., 1979, Causality and econometrics, in: K. Brunner and A.H. Meltzer, eds., Three Aspects of Policy and Policymaking (North-Holland, Amsterdam), 9-54.