Markov properties for graphical time series models

Michael Eichler

Universität Heidelberg

Abstract

This paper deals with the Markov properties of a new class of graphical time series models which focus on the dynamic interrelationships between the components of multivariate time series. The modelling approach is based on the concept of strong Granger-causality and thus can also be applied to nonlinear models. The constraints defining the models are encoded by mixed graphs in which each component series is represented by one vertex and directed edges indicates possible Granger-causal relationships between the variables while undirected edges are used to map the contemporaneous dependence structure of the time series. We discuss various Markov properties for time series models and how these are related to each other. In particular we introduce the notion of causal Markov property. We show that global Markov properties can be formulated in terms of a moralization criterion based on separation in undirected graphs or alternatively using a pathwise separation criterion called $p$-separation. As an example of graphical time series models we consider multivariate ARCH models which satisfy the causal Markov property with respect to a given graph.

Keywords: Graphical models, multivariate time series, Granger-causality, Markov property

1 Introduction

Graphical models provide a suitable framework for the statistical analysis of multivariate data sets. The key feature of the graphical modelling approach is to merge the probabilistic concept of conditional independence with graph theory by representing possible dependencies among the variables of a multivariate distribution in a graph. This leads to graphical methods for the identification of all conditional independence relations which are implied by the models associated with a given graph. Furthermore, graphs can be easily visualized and allow an intuitive understanding of a complex dependence structure, thus facilitating the communication of statistical results. For an introduction to graphical models we refer to the monographs by Whittaker (1990), Edwards (1995), and Cox and Wermuth (1996). A mathematical rigorous treatment can be found in Lauritzen (1996).

Recently the graphical modelling approach has also been applied in multivariate time series analysis. Some general remarks concerning graphical models and time series can be found in Brillinger (1996). Lynggaard and Walther (1993) considered dynamic chain graph models based on conditional Gaussian distributions. In these models each variable at a specific time is represented by a separate vertex in the graph. Edges within the chain components, which consist of the variables at a fixed
time, are undirected while all other edges are directed and point in the direction of
time. A similar modelling approach is taken by Queen and Smith (1993) who base
their definition of multiregression dynamic models on directed acyclic graphs reflect-
ing a causal driving mechanism which includes also contemporaneous relationships
between the variables. These models are closely related to structural vector autore-
gressions whose recursive structure can also be represented by a directed acyclic
graph. Reale and Tumnicliffe Wilson (2000) discuss the identification of such models
using undirected conditional independence graphs.

In all of these models, variables at different times are represented by separate
vertices in the graph. A different approach which leads to much simpler graphs has
been suggested by Dahlhaus (2000) who considered undirected graphs in which each
vertex is associated with a complete time series. These partial correlation graphs
generalize the concept of concentration graphs to multivariate stationary time series
and indicate which variables are linearly related after removing the effect of all other
variables.

The same approach could also be used to define e.g. directed acyclic graphs for
time series. This, however, does not seem appropriate since in the time series case
the direction of edges should reflect the dynamics of the time series which requires
a different type of conditional independence relations. Eichler (2000) introduced
Granger-causality graphs which consist of directed and undirected edges and are
defined in terms of Granger-noncausality. This notion of noncausality exploits the
fact that a cause must precede its effect in time. Although it does not satisfy all
requirements for a proper measure of noncausality it is a commonly used notion for
studying dynamic relationships between time series.

While partial correlation graphs and Granger-causality graphs have been intro-
duced as exploratory tools for the analysis of multivariate time series, we focus in
this paper on the graphical modelling approach. For this we introduce the notion
of causal Markov property which is used to define graphical time series models as-
associated with mixed graphs. The constraints defining these models are formulated
in terms of strong Granger-causality (e.g. Florens and Mouchart, 1982) and thus
allow arbitrary nonlinear relationships between the variables to be included in the
models.

In Section 2 we introduce definitions from graph theory and give some prelimi-
ary results on conditional independence for stochastic processes. The concept
of causal Markov properties and the definition of graphical time series models is
given in Section 3. The interpretation of mixed graphs associated with these models
is enhanced by so-called global Markov properties which relate certain separation
properties of the graph to conditional independence or Granger-noncausality rela-
tions. In Section 4 we formulate these global Markov properties using the concept
of moralization and discuss the relationships between them. In Section 5 we provide
an alternative pathwise separation criterion related to $p$-separation in chain graphs
with the AMP Markov property (Levitz et al., 2001). As an example of nonlinear
graphical time series models we briefly discuss in Section 6 vector autoregressive
conditional heteroscedasticity models. Finally Section 7 offers some concluding re-
marks.
2 Preliminaries

In this section we define the class of mixed graphs discussed in this paper and introduce basic graphical notation and definitions. The second part is concerned with conditional independence and some of its properties within the framework of time series analysis. In particular we state our main assumptions on stochastic processes.

2.1 Graphical concepts

The graphs that are used in this paper are mixed graphs with possibly two kind of edges, namely directed and undirected edges. Suppose that $V$ is a finite and nonempty set. Then a graph $G$ over $V$ is given by an ordered pair $(V, E)$ where the elements in $V$ represent the vertices or nodes of the graph and $E$ is a collection of edges $e$ denoted as $a \rightarrow b$, $a \leftarrow b$, or $a \leftrightarrow b$ for distinct nodes $a, b$ in $V$. The edges $a \rightarrow b$ and $a \leftarrow b$ are called directed edge or arrow while $a \leftrightarrow b$ is called undirected edge or line. The nodes $a$ and $b$ are called the endpoints of the edge $e$ and we write $[e] = \{a, b\}$. If $e = a \rightarrow b$, $b$ is the head of $e$ and $a$ is its tail. Sometimes we will write $a \cdots b$ to indicate that there is some edge connecting $a$ and $b$. In this case the two nodes $a$ and $b$ are said to be adjacent.

Here we will distinguish between $a \rightarrow b$ and $b \leftarrow a$ and also between $a \leftrightarrow b$ and $b \leftrightarrow a$ as oriented versions of the same edge. If $e = a \rightarrow b$ we define the reversely oriented edge $\bar{e}$ by $\bar{e} = b \leftarrow a$. Similarly if $e = a \leftrightarrow b$ we set $\bar{e} = b \rightarrow a$. We call two edges $e$ and $e'$ equivalent if they differ only in orientation, that is, if $e' \in \{e, \bar{e}\}$.

Let $E(V)$ be the set of equivalence classes in the class of all oriented edges between distinct vertices of $V$. Then the edge set $E$ of a graph $G = (V, E)$ is formally given by a subset of $E(V)$. However, for notational convenience we do not distinguish between oriented edges and their equivalence classes in $E(V)$.

Let $a$ and $b$ be vertices of a graph $G = (V, E)$. Then a path from $a$ to $b$ in $G$ is a sequence $\pi = \langle e_1, \ldots, e_n \rangle$ of oriented edges $e_k \in E$ such that $e_k = v_{k-1} \cdots v_k$ for $k = 1, \ldots, n$ and some vertices $v_0, \ldots, v_n$ with $v_0 = a$ and $v_n = b$. The vertices $a$ and $b$ are the endpoints of $\pi$ while $v_1, \ldots, v_{n-1}$ are called intermediate points of $\pi$. The reverse path $\bar{\pi}$ of $\pi$ is given by $\bar{\pi} = \langle \bar{e}_n, \ldots, \bar{e}_1 \rangle$. Further $\pi$ is called a directed path if $\pi$ or $\bar{\pi}$ is of the form $a \rightarrow \ldots \rightarrow b$. On the other hand if $\pi$ consists only of undirected edges the path is called undirected.

Suppose that $\pi$ is a path in a mixed graph $G$ over $V$ and $S$ is a subset of $V$. Then we say that $\pi$ is hit by $S$ if one of its points $v_k$ belongs to $S$. Otherwise if $\pi$ is not hit by $S$ it is called $S$-bypassing.

Two nodes $a$ and $b$ which are connected by an undirected edge in $G$ are said to be neighbours. If $a \leftrightarrow b \in E$ then $a$ is a parent of $b$ and $b$ is a child of $a$. The sets of all neighbours, parents, and children of $a$ are denoted by $\text{ne}_G(a)$, $\text{pa}_G(a)$, and $\text{ch}_G(a)$, respectively. If it is clear which graph $G$ is meant we omit the index $G$. Further for a subset $A$ of $V$ the expressions $\text{ne}(A)$, $\text{pa}(A)$, and $\text{ch}(A)$ denote the collection of neighbours, parents, and children, respectively, of vertices in $A$ that are not themselves elements of $A$, that is, $\text{pa}(A) = \cup_{a \in A} \text{pa}(a) \setminus A$ etc.

As in Frydenberg (1990), a node $b$ is said to be an ancestor of $a$ if either $b = a$ or
there exists a directed path $b \rightarrow \cdots \rightarrow a$ in $G$. The set of all ancestors of elements in $A$ is denoted by $\text{an}_G(A)$ or short $\text{an}(A)$. A subset $A$ is called an ancestral set if it contains all its ancestors, that is, $\text{an}(A) = A$.

Let $G = (V, E)$ and $G' = (V', E')$ be mixed graphs. Then $G'$ is a subgraph of $G$ if $V' \subseteq V$ and $E' \subseteq E$. If $A$ is a subset of $V$ it induces the subgraph $G_A = (A, E_A)$ where $E_A = \{e \in E | e \subseteq A \}$. A subset $A$ such that all vertices in the induced subgraph $G_A$ are adjacent is called complete.

Finally, a graph is called undirected if it contains only undirected edges. For a mixed graph $G = (V, E)$ the undirected subgraph $G^u = (V, E^u)$ is obtained by removing all directed edges. Thus we have $E^u = \{e \in E | e = a \rightarrow b \}$. Undirected graphs have a particularly simple concept of separation. Let $A, B,$ and $S$ be disjoint subsets of $V$. Then the set $S$ separates the sets $A$ and $B$, denoted by $A \nparallel B \mid S \ [G]$, if every path in $G$ from $A$ to $B$ necessarily is hit by $S$.

### 2.2 Conditional independence and stochastic processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_1, \mathcal{F}_2,$ and $\mathcal{F}_3$ sub-$\sigma$-algebras of $\mathcal{F}$. Here and elsewhere we assume for simplicity that all $\sigma$-algebras are completed. The smallest $\sigma$-algebra generated by $\mathcal{F}_i \cup \mathcal{F}_j$ is denoted as $\mathcal{F}_i \vee \mathcal{F}_j$. Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are said to be independent conditionally on $\mathcal{F}_3$ if $\mathbb{E}(X|\mathcal{F}_2 \vee \mathcal{F}_3) = \mathbb{E}(X|\mathcal{F}_3)$ a.s. for all real-valued, bounded, $\mathcal{F}_1$-measurable random variables $X$. Using the notation of Dawid (1979) we write $\mathcal{F}_1 \perp \perp \mathcal{F}_2 \mid \mathcal{F}_3 \ [\mathbb{P}]$. If the $\sigma$-algebras are generated by random elements $\eta_i$, we also write $\eta_1 \perp \perp \eta_2 \mid \eta_3 \ [\mathbb{P}]$ or $\eta_1 \perp \perp \eta_2 \mid \eta_3$ if the reference to $\mathbb{P}$ is clear.

For some purposes it will be necessary to assume that $\mathbb{P}$ satisfies condition (C5) of Lauritzen (1996) which has also been called intersection property by Pearl (1988). More precisely we require that

$$\mathcal{F}_1 \perp \perp \mathcal{F}_2 \mid \mathcal{F}_3 \vee \mathcal{F}_4 \text{ and } \mathcal{F}_1 \perp \perp \mathcal{F}_3 \mid \mathcal{F}_2 \vee \mathcal{F}_4 \Leftrightarrow \mathcal{F}_1 \perp \perp \mathcal{F}_2 \vee \mathcal{F}_3 \mid \mathcal{F}_4 \quad (2.1)$$

holds for certain sub-$\sigma$-algebras $\mathcal{F}_i, i = 1, \ldots, 4,$ of $\mathcal{F}$. A sufficient and necessary condition for (2.1) is given by

$$(\mathcal{F}_2 \vee \mathcal{F}_4) \cap (\mathcal{F}_3 \vee \mathcal{F}_4) = \mathcal{F}_4. \quad (2.2)$$

In that case $\mathcal{F}_2$ and $\mathcal{F}_3$ are said to be measurably separated conditionally on $\mathcal{F}_4$, denoted by $\mathcal{F}_2 \parallel \mathcal{F}_3 \mid \mathcal{F}_4$ (Florens et al., 1990). We note that conditional measurable separability preserves most, but not all, of the properties of conditional independence. For instance, if $\mathcal{F}_1 \subseteq \mathcal{F}_2$ and $\mathcal{F}_1 \parallel \mathcal{F}_2 \mid \mathcal{F}_3$ then $\mathcal{F}_1 \parallel \mathcal{F}_4 \mid \mathcal{F}_3$. For details we refer to Chapter 5.2 of Florens et al. (1990).

We now consider stationary $d$-variate stochastic processes $X = \{X(t), t \in \mathbb{Z}\}$, $X(t) = (X_1(t), \ldots, X_d(t))'$, on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_X = (\mathcal{F}_X(t))$ be the filtration induced by $X$, that is, $\mathcal{F}_X(t)$ is the $\sigma$-algebra generated by all the components of $X$ up to time $t$. The generating set $\{X(s), s \leq t\}$ will be abbreviated by $\hat{X}(t)$. Further let $V$ denote the index set $\{1, \ldots, d\}$. For any $A \subseteq V$ we define $X_A = \{X_A(t)\}$ as the multivariate subprocess with components $X_a, a \in A$. Throughout this paper we assume that the process $X$ satisfies the following conditions.
**Assumption 2.1**

(i) \( X = \{X(t), t \in \mathbb{Z}\} \) is a stationary, mixing, purely nondeterministic stochastic process on some probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

(ii) For all \( t \in \mathbb{Z} \) and \( k \in \mathbb{N} \) there exists a probability measure \( \mathbb{P}' \) on \((\Omega, \mathcal{F}_X(t))\) such that \( \mathbb{P}|_{\mathcal{F}_X(t)} \) and \( \mathbb{P}' \) are equivalent (i.e. have the same null sets) and the random variables \( X_a(s), a \in V, s \leq t \) are mutually independent conditionally on \( \mathcal{F}_X(t - k) \) with respect to \( \mathbb{P}' \).

For disjoint finite subsets \( Y_1, Y_2 \) of \( \bar{X}(t) \) the second assumption implies that \( Y_1 \perp Y_2 \mid \bar{X}(t) \backslash (Y_1 \cup Y_2) \) \([\mathbb{P}']\) for some probability measure \( \mathbb{P}' \) on \((\Omega, \mathcal{F}_X(t))\). It then follows from Corollary 5.2.11 of Florens et al. (1990) that \( Y_1 \) and \( Y_2 \) are measurably separated conditionally on \( \bar{X}(t) \backslash (Y_1 \cup Y_2) \) with respect to the original measure \( \mathbb{P} \).

We note that the assumption is satisfied if for all \( t \in \mathbb{Z} \) the conditional distribution \( \mathbb{P}^X(t+1) \mid X(t) \) is almost surely absolutely continuous with respect to some product measure on \( \mathbb{R}^d \) and has almost surely a positive and continuous density. The next lemma shows that for mixing processes the measurable separability can also be extended to \( \sigma \)-algebras with infinite generating sets.

**Lemma 2.2** Suppose that \( X \) satisfies Assumption 2.1. Then \( \mathcal{F}_{X_A}(t) \) and \( \mathcal{F}_{X_B}(t) \) are measurably separated conditionally on \( \mathcal{F}_{X_V \backslash (A \cup B)}(t) \) for all disjoint subsets \( A \) and \( B \) of \( V \).

In particular, we note that for every \( \mathcal{F} \)-measurable random variable \( Y \) the lemma yields the following intersection property

\[
Y \perp \bar{X}_{\varnothing}(t) \mid \mathcal{F}_{X_V \backslash \{a\}}(t) \iff Y \perp \bar{X}_{\varnothing}(t) \mid \mathcal{F}_{X_V \backslash \{a\}}(t) \quad \forall a \in A. \tag{2.3}
\]

For the proof of the lemma we first show that every stationary, mixing process is also conditionally mixing. Let \( A \subseteq V \). For any set \( S' = \{\omega \in \Omega \mid \bar{X}(t - k - 1) \in T\} \) in \( \mathcal{F}_{X_A}(t - k - 1) \) we define the shifted set \( S'_j \in \mathcal{F}_{X_A}(t - k - j) \) by \( S'_j = \{\omega \in \Omega \mid \bar{X}(t - k - j) \in T\} \). Then the process \( X \) is said to be conditionally mixing (Veijanen, 1990) if for all disjoint subsets \( A \) and \( B \) of \( V \), \( t \in \mathbb{Z} \), \( S \in \sigma \{X_A(t), \ldots, X_A(t - k)\} \), and \( S' \in \mathcal{F}_{X_A}(t - k - 1) \) we have

\[
\mathbb{E}\left| \mathbb{P}(S \cap S'_j \mid \mathcal{F}_{X_B}(t)) - \mathbb{P}(S \mid \mathcal{F}_{X_B}(t))\mathbb{P}(S'_j \mid \mathcal{F}_{X_B}(t)) \right| \to 0
\]
as \( j \to \infty \).

**Lemma 2.3** Every stochastic process \( X \) satisfying Assumption 2.1(i) is also conditionally mixing.

**Proof.** Let \( F \in \sigma \{X_B(t_1), \ldots, X_B(t_m)\} \) with \( m \in \mathbb{N} \) and \( t_j < t, 1 \leq j \leq m \). Then we have for \( S \in \sigma \{X_A(t), \ldots, X_A(t - k)\} \) and \( S' \in \mathcal{F}_{X_A}(t - k - 1) \),

\[
\mathbb{E}\left| \mathbb{E}\left[\left(\mathbb{E}(1_{S \cap S'_j} \mid \mathcal{F}_{X_B}(t)) - \mathbb{E}(1_S \mid \mathcal{F}_{X_B}(t))\mathbb{E}(1_{S'_j} \mid \mathcal{F}_{X_B}(t))\right)1_F\right]\right|
\leq\mathbb{E}\left|\mathbb{E}(1_{S \cap S'_j \cap F}) - \mathbb{E}(1_{S \cap F})\mathbb{E}(1_{S'_j})\right|
\]

\[
+ \mathbb{E}\left[\left(\mathbb{E}(1_{S \cap F} \mid \mathcal{F}_{X_B}^{(n)}(t))\right)1_{S'_j}\right]
- \mathbb{E}\left[\left(\mathbb{E}(1_{S \cap F} \mid \mathcal{F}_{X_B}^{(n)}(t))\right)1_{S'_j}\right]
\]

\[
+ \mathbb{E}\left[\left(\mathbb{E}(1_{S \cap F} \mid \mathcal{F}_{X_B}(t)) - \mathbb{E}(1_{S \cap F} \mid \mathcal{F}_{X_B}(t))\right)1_{S'_j}\right]
\]

5
where \( \mathcal{F}_{X_B}^{(n)}(t) = \sigma\{X_B(t), \ldots , X_B(t - n)\} \). Each of the three terms on the right side converges to zero as \( j - n \) and \( n \) tend to infinity. Since the union of all \( \sigma\{X_B(t_1), \ldots , X_B(t_m)\} \) with \( m \in \mathbb{N} \) and \( t_j < t \) constitutes a \( \cap \)-stable generator of \( \mathcal{F}_{X_B}(t) \) the assertion of the lemma follows. \( \square \)

**Proof of Lemma 2.2.** As noted before Assumption 2.1(ii) implies that \( \mathcal{F}_{X_A}(t) \parallel \mathcal{F}_{X_B}(t) | \mathcal{F}_{X_{A^cB}}(t) \lor \mathcal{F}_X(t - k) \) for all \( t \in \mathbb{Z} \) and \( k \in \mathbb{N} \). Accordingly, we have

\[
(\mathcal{F}_{X_A}(t) \lor \mathcal{F}_X(t - k)) \cap (\mathcal{F}_{X_B}(t) \lor \mathcal{F}_X(t - k)) = (\mathcal{F}_{X_{A^cB}}(t) \lor \mathcal{F}_X(t - k)).
\]

It therefore suffices to establish the identity

\[
\bigcap_{k > 0} \left[ \mathcal{F}_{X_A}(t - k) \lor \mathcal{F}_{X_B}(t) \right] = \mathcal{F}_{X_B}(t) \tag{2.4}
\]

for any subsets \( A \) and \( B \) of \( V \). Let \( \mathcal{T} \) denote the \( \sigma \)-algebra on the left hand side. Since by the previous lemma \( X \) is conditionally mixing the \( \sigma \)-algebras \( \mathcal{T} \) and \( \mathcal{F}_{X_A}(t) \) are conditionally independent given \( \mathcal{F}_{X_B}(t) \). Thus we have for all \( T \in \mathcal{T} \)

\[
\mathbb{E}(1_T | \mathcal{F}_{X_B}(t)) = \mathbb{E}(1_T | \mathcal{F}_{X_A}(t) \lor \mathcal{F}_{X_B}(t)) \rightarrow 1_T
\]

as \( k \rightarrow \infty \). Hence \( \mathcal{T} \subseteq \mathcal{F}_{X_B}(t) \) whereas the converse relation is obvious. \( \square \)

### 3 Graphical time series models

In graphical modelling the focus is on multivariate statistical models of which the possible dependencies between the studied variables can be represented by a graph. More precisely, let \( \xi_v, v \in V \) be random variables on a common probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with values in \( \mathcal{X} \). Then a graphical model for the random vector \( \xi = (\xi_v)_{v \in V} \) is given by a family of probability distributions \( \{P_{\theta} | \theta \in \Theta\} \) on \( \mathcal{X}^V \) and a graph \( G = (V, E) \) in which the vertices represent the random variables and the absence of edges indicates conditional independences which hold for all distributions \( P_{\theta} \) in the model. In the case of real-valued random variables these conditional independence constraints are typically of the form that \( \xi_a \perp \perp \xi_b | \xi_{R_{ab}} [P_{\theta}] \) whenever \( a \) and \( b \) are not adjacent in \( G \) and the set \( R_{ab} \) depends on the type of graph. For undirected graphs we have \( R_{ab} = V \setminus \{a, b\} \), that is, if two variables are not adjacent in the graph then they are conditionally independent given all other variables. This is known as the pairwise Markov property for undirected graphs.

In this paper, we consider graphical models where the variables \( \xi_v \) are the components \( X_v \) of a multivariate stationary time series \( X \). Although such models could be specified by the same type of conditional independence constraints as used for multivariate data in \( \mathbb{R}^d \), namely \( \xi_a \perp \perp \xi_b | \xi_{R_{ab}} \), this approach is in general not appropriate for modelling the dynamics of time series as it e.g. cannot indicate feedback between two variables. Instead we choose a new type of conditional independence constraints which are more suitable for time series modelling.

One concept which is particularly useful for studying dynamic relationships between time series is that of Granger-(non)causality introduced by Granger (1969). This notion of causality exploits the fact that an effect cannot precede its cause in
time. More precisely, one time series $X_a$ is said to be Granger-causal for another series $X_b$ if at time $t$ we are better able to predict the next value of $X_b$ using the entire information up to time $t$ than if the same information apart from the former series $X_a$ had been used. In practice not all relevant variables may be available and thus the notion of Granger-causality clearly depends on the used information set which limits its appropriateness as a formal description of causality. Nevertheless this does not weaken its usefulness in the construction, estimation, and application of multivariate time series models.

While the original definition of Granger-noncausality has been formulated in terms of mean square prediction, we base our definition of graphical time series models on the notion of strong Granger-noncausality (e.g. Florens and Mouchart, 1982) which is defined in terms of conditional independence and $\sigma$-algebras. In contrast to the linear approach presented in Eichler (2000) this allows also the treatment of models with nonlinear dependencies between the components of the process.

In this context, the information available at times $t$ can be represented by $\sigma$-algebras $\mathcal{F}(t)$ which, under the assumption that information is not lost as $t$ increases, form a filtration $\mathcal{F} = \{ \mathcal{F}(t) \}$ on $\mathbb{Z}$. We further assume that the past and present of the process of interest, say $X$, belongs to the information set at time $t$. For notational convenience we suppress any further environmental variables from the discussion and thus set $\mathcal{F} = \mathcal{F}_X$ although most of the results in this paper also hold in the more general case. We then obtain the following definition of strong Granger-noncausality.

**Definition 3.1** (Granger-noncausality) Let $A$ and $B$ disjoint subsets of $V$. Then $X_A$ is strongly Granger-noncausal for $X_B$ relative to the filtration $\mathcal{F}_X$ if

$$X_B(t+1) \perp \perp X_A(t) \mid \mathcal{F}_{X \setminus A}(t)$$

for all $t \in \mathbb{Z}$. This will be denoted by $X_A \not\rightarrow X_B \ [\mathcal{F}_X]$.

For a complete description of the dependence structure of a process $X$ we additionally need to model dependencies between variables at the same time. We do this by taking into account all information including that of the other variables at the same time which leads to the following definition.

**Definition 3.2** Let $A$ and $B$ disjoint subsets of $V$. Then $X_A$ and $X_B$ are contemporaneously conditionally independent relative to the filtration $\mathcal{F}_X$ if

$$X_A(t+1) \perp \perp X_B(t+1) \mid \mathcal{F}_X(t) \vee \mathcal{F}_{X \setminus (A \cup B)}(t+1)$$

for all $t \in \mathbb{Z}$. This will be denoted by $X_A \not\sim X_B \ [\mathcal{F}_X]$.

For a stationary process $X$ the pairwise Granger-noncausality and contemporaneous conditional independence relations can be visualized by a mixed graph $G = (V, E)$ as shown in Eichler (2000). In these Granger-causality graphs the vertices represent the components of $X$ while the absence or presence of directed edges
a \rightarrow b$ and undirected edges $a \leftarrow b$ indicates whether the corresponding relations $X_a \rightarrow X_b [\mathcal{F}_X]$ and $X_a \sim X_b [\mathcal{F}_X]$, respectively, hold for $X$ or not.

Conversely, the absence of directed and undirected edges in a mixed graph $G = (V,E)$ immediately implies a set of pairwise Granger-noncausality and contemporaneous conditional independence constraints. Such a set of conditional independence relations encoded by a graph is generally known as the Markov property associated with the graph. In the context of time series, graphs may encode different types of conditional independence relations and we therefore speak of causal Markov properties when dealing with Granger-noncausality and contemporaneous conditional independence relations.

**Definition 3.3 (Pairwise causal Markov property)** Let $X$ be a stationary process and $G = (V,E)$ a mixed graph. Then $X$ satisfies the pairwise causal Markov property (PC) with respect to $G$ if for all $a, b \in V$ with $a \neq b$

(i) $a \rightarrow b \notin E \Rightarrow X_a \rightarrow X_b [\mathcal{F}_X],$
(ii) $a \leftarrow b \notin E \Rightarrow X_a \sim X_b [\mathcal{F}_X].$

Under Assumption 2.1 the equivalence in (2.3) leads to a composition and decomposition property for Granger-noncausality relations,

$$X_A \rightarrow X_B [\mathcal{F}_X] \Leftrightarrow X_a \rightarrow X_B [\mathcal{F}_X] \quad \forall a \in A.$$ \hfill (3.1)

Similarly we have for the contemporaneous conditional independence

$$X_A \sim X_B [\mathcal{F}_X] \Leftrightarrow X_a \sim X_B [\mathcal{F}_X] \quad \forall a \in A, \forall b \in B.$$ \hfill (3.2)

Collecting for each vertex $v \in V$ all vertices which are not parents (resp. neighbours) of $v$, it thus follows from the pairwise causal Markov property that

$$X_{V \setminus (pa(v) \cup \{v\})} \rightarrow X_v [\mathcal{F}_X] \quad \text{and} \quad X_{V \setminus (ne(v) \cup \{v\})} \sim X_v [\mathcal{F}_X]$$ \hfill (3.3)

for all $v \in V$. If (3.3) holds for a process $X$ and a graph $G$ we say that $X$ satisfies the local causal Markov property (LC) with respect to $G$.

Conversely if a process $X$ satisfies the local causal Markov property with respect to a mixed graph $G = (V,E)$ it follows from the properties of conditional independence that for arbitrary subsets $A$ of $V$ the graph implies $X_A \rightarrow X_b [\mathcal{F}_X]$ if $A$ contains neither $b$ nor parents of $b$. In particular we find that $X$ also satisfies the pairwise Markov property with respect to $G$. On the other hand the local causal Markov property provides no means to decide whether the process $X_A$ is Granger-noncausal for another vector process $X_B$ (relative to $\mathcal{F}_X$) since in contrast to the composition and decomposition property in (3.1) the required equivalence

$$X_A \rightarrow X_B [\mathcal{F}_X] \Leftrightarrow X_A \sim X_b [\mathcal{F}_X] \quad \forall b \in B$$ \hfill (3.4)

does not hold in general but only in special cases (e.g. Eichler 2000, Section 5). For the treatment of more general time series models we therefore strengthen the local causal Markov property by extending it to arbitrary subsets of $V$. 

8
**Definition 3.4** (Block-recursive causal Markov property) Let $G = (V, E)$ be a mixed graph. Then $X$ satisfies the block-recursive causal Markov property (BC) with respect to $G$ if for all subsets $A$ of $V$

$$X_{V \setminus (pa(A) \cup A)} \not\rightarrow X_A \ [F_X] \quad \text{and} \quad X_{V \setminus (ne(A) \cup A)} \not\leftarrow X_A \ [F_X].$$

In the next section we show that the block-recursive causal Markov property with respect to a graph $G$ yields a complete model for the dependence structure of a process $X$ in terms of Granger-noncausality and contemporaneous conditional independence relations. In other words, the edges in the graph visualize all possible dependencies between the components of the process. We therefore define a graphical time series model given by a mixed graph $G$ as a family of stochastic processes $X(\theta) = \{X(\theta)(t)\}$, $\theta \in \Theta$, which satisfy the block-recursive causal Markov property with respect to $G$.

We note that all causal Markov properties are basically properties of the probability distribution $P = P^X$ of the process $X$. Therefore we could alternatively define graphical time series models in terms of probability distributions $P_\theta$ on the sample space $\mathcal{X}^d$ where $\mathcal{X} = \mathbb{R}^2$. Although this probabilistic approach is commonly used in the theory of graphical models it rarely can be found in the literature on time series analysis and therefore will not be used in this paper.

**Example 3.5** (Vector autoregressive processes) Let $X$ be a stationary vector autoregressive process of order $p$,

$$X(t) = A(1)X(t-1) + \ldots + A(p)X(t-p) + \varepsilon(t), \quad \varepsilon(t) \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma), \quad (3.5)$$

where $A(j)$ are $d \times d$ matrices and the variance matrix $\Sigma$ is nonsingular. Suppose that $X$ satisfies the pairwise causal Markov property with respect to a graph $G = (V, E)$. Since $X_a$ is Granger-noncausal for $X_b$ if and only if the corresponding entries $A_{ba}(j)$ vanish in all matrices $A(j)$ (e.g. Tjøstheim, 1981) it follows that

$$a \rightarrow b \notin E \Rightarrow A_{ba}(j) = 0 \quad \forall j = 1, \ldots , p. \quad (3.6)$$

Further, $X_a$ and $X_b$ are contemporaneously conditionally independent if and only if the corresponding error components $\varepsilon_a(t)$ and $\varepsilon_b(t)$ are conditionally independent given all remaining components $\varepsilon_{V \setminus \{a,b\}}(t)$. Since for Gaussian variables conditional independences are given by zeros in the inverse covariance matrix $K = \Sigma^{-1}$ (e.g. Lauritzen, 1996) the pairwise causal Markov property implies the condition

$$a \rightarrow b \notin E \Rightarrow K_{ab} = K_{ba} = 0. \quad (3.7)$$

Since the conditional distribution of $X(t)$ given $\mathcal{F}_X(t-1)$ depends on the past only in its mean, the process apparently satisfies condition (3.4) and hence the pairwise and block-recursive causal Markov property are equivalent for such processes. Thus the graphical VAR($p$) model associated with the graph $G$ is given by the set of all stationary VAR($p$) processes whose parameters are constrained to zero according to the conditions (3.6) and (3.7).
Example 3.6 (Nonlinear autoregressive processes) More generally, we may consider nonlinear time series models \( \{ X^{(\theta)}, \theta \in \Theta \} \) consisting of stationary processes with components

\[
X_a^{(\theta)}(t) = f_a^{(\theta)}(X^{(\theta)}(t-1), \ldots, X^{(\theta)}(t-p)) + \sigma_a^{(\theta)}(X^{(\theta)}(t-1), \ldots, X^{(\theta)}(t-p)) \varepsilon_a^{(\theta)}(t),
\]

where \( \varepsilon^{(\theta)}(t) \stackrel{\text{iid}}{\sim} P_{\theta} \). We further assume that \( P_{\theta} \) has a positive and continuous density with respect to Lebesgue-measure and that the functions \( \sigma_a^{(\theta)} \) are positive and bounded away from zero. For a mixed graph \( G = (V, E) \) we then define \( \Theta_G \) as the subset of all \( \theta \in \Theta \) such that first the functions \( f_a^{(\theta)} \) and \( \sigma_a^{(\theta)} \) are both constant in their arguments \( X_b^{(\theta)}(t-1), \ldots, X_b^{(\theta)}(t-p) \) whenever \( b \rightarrow a \notin E \) and second \( P_{\theta} \) satisfies the pairwise Markov property with respect to the undirected subgraph \( G^a \). It then can be shown that \( X_b^{(\theta)} \) is Granger-noncausal for \( X_a^{(\theta)} \) if the edge \( b \rightarrow a \) is absent in \( G \) (Eichler, 2000) and hence that \( X^{(\theta)} \) satisfies the pairwise causal Markov property with respect to \( G \). Furthermore \( X^{(\theta)} \) fulfills the conditions for the equivalence of the pairwise and the block-recursive causal Markov properties, namely (3.1), (3.2), and (3.4), which follows by similar arguments as in the previous example. Therefore \( \{ X^{(\theta)}, \theta \in \Theta_G \} \) is the graphical submodel associated with the graph \( G \).

Further examples of graphical time series models are given in Section 6 where we discuss various multivariate ARCH models satisfying the block-recursive causal Markov property with respect to a given graph. In particular we will give an example where condition (3.4) does not hold and hence the block-recursive causal Markov property is strictly stronger than the pairwise.

4 Global Markov properties

The interpretation of graphs describing the dependence structure of graphical models in general is enhanced by global Markov properties which merge the notion of conditional independence with a purely graph theoretical concept of separation allowing one to state whether two subsets of vertices are separated by a third subset of vertices. For undirected graphs the notion of separation is straightforward, namely two sets are separated by a third one if each path connecting the first two sets intersects the third set. For graphs which contain also directed edges the concept of separation depends on the kind of conditional independence relations to be encoded by the graph, but there exist two main approaches for defining such a concept of separation: The first approach utilises separation in undirected graphs by applying the operation of ‘moralization’ to appropriate subgraphs (e.g. Frydenberg, 1990; Lauritzen et al., 1990) while the other approach is based on path-oriented criteria like \( d \)-separation (Pearl, 1988) for directed acyclic graphs. In this section we will follow the first approach to define a global causal Markov property for graphical time series models. The concept of moralization which we present for this is closely
related to other classical global Markov properties which do not refer to noncausal-
ity relations but to conditional independence between complete components of the
time series.

4.1 Global Markov property for undirected graphs

Graphical time series models associated with undirected graphs have been first con-
sidered by Dahlhaus (2000) who introduced partial correlation graphs to describe
the linear dependence structure of multivariate time series. This notion of graphs
can be generalized to include arbitrary nonlinear dependencies between the com-
ponents of the time series and leads to the following version of the global Markov
property for undirected graphs.

**Definition 4.1** (Global Markov property) Let \( G' = (V, E') \) be an undirected
graph. Then a stationary process \( X \) satisfies the global Markov property (G) with
respect to \( G' \) if for all disjoint subsets \( A, B, S \) of \( V \)

\[
A \ni B \mid S \ [G'] \Rightarrow X_A \perp \perp X_B \mid X_S.
\]

Now suppose that a stationary process satisfies the block-recursive causal Markov
property with respect to the mixed graph \( G = (V, E) \). The basic idea of moralization
is to map the mixed graph \( G \) to an undirected graph such that \( X \) satisfies the global
Markov property with respect to this undirected graph. We illustrate this operation
by the processes in Example 3.5.

**Example 4.2** Let \( X \) be a VAR\((p)\) process of the form (3.5). Further let \( f(\lambda) =
(2\pi)^{-1}A^{-1}(e^{-i\lambda})\Sigma A^{-1}(e^{i\lambda})' \) be the spectral matrix of \( X \) and \( g(\lambda) = f(\lambda)^{-1} \) its
inverse. Then it has been shown by Dahlhaus (2000) that two components \( X_a \) and \( X_b \)
are conditionally independent given all other components if and only if \( g_{ab}(\lambda) = 0 \)
for all \( \lambda \in [-\pi, \pi] \). This leads to the following constraints on the parameters

\[
\sum_{u=0}^{p+h} \sum_{j,k=1}^{d} K_{jk} A_{ja}(u) A_{kb}(u + h) = 0, \quad \forall h = -p, \ldots, p \quad (4.1)
\]

with \( A(0) = I \) and \( K = \Sigma^{-1} \).

Now suppose that \( X \) satisfies the block-recursive causal Markov property with
respect to a mixed graph \( G = (V, E) \). Then it follows from (3.6) and (3.7) that (4.1)
holds if the two nodes \( a \) and \( b \) are neither adjacent nor linked by paths of the form
\( a \rightarrow c \leftarrow b \), \( a \rightarrow c \rightarrow b \), or \( a \rightarrow c \leftarrow d \leftarrow b \). Thus \( X \) satisfies the pairwise and hence
the global Markov property with respect to the undirected graph which is obtained
from \( G \) by inserting undirected edges \( a \rightarrow b \) whenever one of the above paths exists
in \( G \) and then converting all directed edges to undirected edges.

Following Andersson et al. (2000) we call the subgraph induced by vertices \( a, b, c \)
an immorality (a flag) if it contains \( a \rightarrow c \leftarrow b \) (\( a \rightarrow c \rightarrow b \)) but \( a \) and \( b \) are not
adjacent in \( G \). Further the subgraph induced by \( a, b, c, d \) is called a 2-biflag if again
\( a \) and \( b \) are not adjacent and linked by the path \( a \rightarrow c \rightarrow d \leftarrow b \). The possible forms
of immoralities, flags, and 2-biflags in mixed graphs are depicted in Figure 4.1. We note that in the present context a subgraph can be both an immorality and a flag and further that a 2-biflag does not necessarily consist of two flags.

In order to construct the undirected graph in Example 4.2 we have completed all immoralities, flags, and 2-biflags in $G$ by inserting undirected edges such that all vertices in these subgraphs are adjacent. This leads to the following definition of moralization.

**Definition 4.3** Let $G = (V, E)$ be a mixed graph. The moral graph $G^m = (V, E^m)$ derived from $G$ is defined to be 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.

We note that in Andersson et al. (2000) this operation of moralization and the resulting graph have been called augmentation and augmentation graph, respectively, in order to distinguish it from moralization for chain graphs as defined in Frydenberg (1990). In this paper we will use the term augmentation only when referring to the augmentation of additional vertices.

In the next theorem we show that the global Markov property holds for graphical time series models satisfying the block-recursive causal Markov property.

**Theorem 4.4** Suppose $X$ satisfies the block-recursive causal Markov property with respect to $G$. Then $X$ satisfies the global Markov property with respect to $G^m$.

We first prove a simpler version of the theorem which will be of use for the definition of the global causal Markov property.

**Lemma 4.5** Suppose $X$ satisfies the block-recursive causal Markov property with respect to $G$. Let $\xi = (\xi_1, \ldots, \xi_d)'$ with $\xi_a = \bar{X}_a(t)$ for some $t \in \mathbb{Z}$. Then $\xi$ satisfies the global Markov property with respect to $G^m$.

**Proof.** Because of Lemma 2.2 the global Markov property is equivalent to the pairwise Markov property with respect to $G^m$ and it therefore suffices to show that

$$a \not\rightarrow b \notin E^m \Rightarrow \xi_a \perp \perp \xi_b \mid \xi_{V \setminus \{a,b\}}.$$

First we note that the block-recursive causal Markov property allows us to conclude from the absence of $a \rightarrow b$ for all $b \in B$ that $X_a$ is noncausal for $X_B$ without referring
to condition (3.4). Thus we can proceed as in the proof of Theorem 3.1 in Eichler (2000) to show that

\[ X_a(s) \perp \! \! \! \perp X_b(s') \mid \bar{X}(t) \setminus \{X_a(s), X_b(s')\} \]

for all \( s, s' \leq t \). Because of Assumption 2.1(ii) we can use the intersection property repeatedly which yields \( \bar{X}_a(t) \perp \! \! \! \perp \bar{X}_b(t) \mid \bar{X}_{V \setminus \{a, b\}}(t) \cup \bar{X}(t - k) \) for all \( k < t \). For \( k \to \infty \) the asserted pairwise conditional independence follows from (2.4).

**Proof of Theorem 4.4.** Suppose that \( A, B, \) and \( S \) are disjoint subsets of \( V \) such that \( A \nabla B \mid S \mid [G^m] \). Let \( \xi \) be any \( \mathcal{F}_X(t) \) measurable random variable with \( \mathbb{E}[\xi] < \infty \), where \( \mathcal{F}_X(t) = \bigvee_{t \in \mathbb{Z}} \mathcal{F}_X(t) \) denotes the \( \sigma \)-algebra generated by \( X_A \).

Setting \( \xi(t) = \mathbb{E}(\xi | \mathcal{F}_X(t)) \) we get, as \( t \to \infty \),

\[ \mathbb{E}(\xi(t) | \mathcal{F}_{S \cup B}(t)) \to \mathbb{E}(\xi | \mathcal{F}_{S \cup B}) \text{ in } L^1. \tag{4.2} \]

On the other hand, the previous lemma implies \( \xi(t) \perp \! \! \! \perp \mathcal{F}_B(t) | \mathcal{F}_S(t) \). Hence we obtain, as \( t \to \infty \),

\[ \mathbb{E}(\xi(t) | \mathcal{F}_{S \cup B}(t)) = \mathbb{E}(\xi(t) | \mathcal{F}_S(t)) \to \mathbb{E}(\xi | \mathcal{F}_S) \text{ in } L^1. \tag{4.3} \]

Since the limits in (4.2) and (4.3) must be equal in \( L^1 \) and thus almost surely, this proves that \( \mathcal{F}_X(t) \perp \! \! \! \perp \mathcal{F}_B | \mathcal{F}_S \).

\[ \square \]

### 4.2 Global AMP Markov property

The number of conditional independence statements derived from the global Markov property for undirected graphs can be increased by applying the operation of moralization only to appropriate subgraphs of \( G \). This leads to the notion of global AMP Markov property which has been introduced by Andersson et al. (2000) in the case of chain graphs. Here we give a slightly different definition of the global AMP Markov property which is based on marginal ancestral graphs instead of extended subgraphs.

Let \( G = (V, E) \) be a mixed graph. We then define the marginal ancestral graph induced by \( A \), denoted by \( G(A) = (an(A), E(A)) \), as the graph which is obtained from the induced subgraph \( G_{an(A)} \) by insertion of additional undirected edges \( a \rightarrow b \) whenever \( a \) and \( b \) are not separated by \( an(A) \setminus \{a, b\} \) in \( G^a \). Due to these additional undirected edges the subprocess \( X_{an(A)} \) apparently satisfies the pairwise causal Markov property with respect to the marginal ancestral graph \( G(A) \) if \( X \) did so with respect to \( G \). As the following lemma shows the same inheritance property also holds for the block-recursive causal Markov property.

**Lemma 4.6** Suppose that \( X \) satisfies the block-recursive causal Markov property with respect to the mixed graph \( G \) and let \( U \subseteq V \). Then the subprocess \( X_{an(U)} \) satisfies the block-recursive causal Markov property with respect to \( G(U) \).

**Proof.** First, since \( an(U) \) is an ancestral set we have \( pa_{G}(A) = pa_{G(U)}(A) \) for each subset \( A \subseteq an(U) \). Therefore \( X_{V \setminus \{pa(A) \cup A\}} \) is noncausal for \( X_A \) relative to \( \mathcal{F}_X \) if and only if \( X_{an(A) \setminus \{pa(A) \cup A\}} \) is noncausal for \( X_A \) relative to the smaller filtration \( \mathcal{F}_{X_{an(A)}} \).

13
Second, it is clear from the construction of $G(U)$ that for any $A \subseteq \text{an}(U)$ the set of neighbours in $G(U)$, $\text{ne}(G(U))(A)$, separates the sets $A$ and $\text{an}(U) \setminus (\text{ne}(G(U))(A) \cup A)$ in the undirected subgraph $G^m$. Hence there exists a subset $A^* \subseteq \text{an}(U)$ such that $A \subseteq A^*$ and $\text{ne}(G(A^*)) = \text{ne}(G(U))(A)$. The block-recursive causal Markov property then implies

$$X_A(t+1) \perp X_{\text{an}(U) \setminus (\text{ne}(G(U))(A) \cup A)}(t+1) \mid X_{\text{ne}(G(U))(A)}(t+1), X(t),$$

from which we obtain the desired relation $X_A \sim X_{\text{an}(U) \setminus (\text{ne}(G(U))(A) \cup A)} \perp \perp [\mathcal{F}_{\text{an}(U)}]$ by noting that $X_{\text{an}(U)}(t+1) \perp \perp X_{V \setminus \text{an}(U)}(t) \mid X_{\text{an}(U)}(t)$.

We are now able to strengthen the global Markov property with respect to the moral graph $G^m$ by requiring that the subprocess $X_{\text{an}(U)}$ satisfies the global Markov property with respect to the moral marginal ancestral subgraph $G(U)^m$ for all possible subsets $U \subseteq V$.

**Definition 4.7** (Global AMP Markov property) Let $G = (V,E)$ be a mixed graph. Then $X$ satisfies the global AMP Markov property (GA) with respect to $G$ if for all disjoint subsets $A,B,S$ of $V$

$$A \bowtie B \mid S \ [G(ABS)^m] \Rightarrow X_A \perp \perp X_B \mid X_S,$$

where $ABS = A \cup B \cup S$.

In particular, if for disjoint subsets $A$ and $B$ of $V$ we set $S = V \setminus (A \cup B)$ the corresponding marginal ancestral graph is the moral graph $G^m$ of $G$ itself irrespectively of the choice of $A$ and $B$. Thus the global AMP Markov property with respect to $G$ implies the global Markov property with respect to $G^m$. That it can be strictly stronger can be illustrated by the following example. Consider the mixed graph $1 \rightarrow 3 \leftarrow 2$. With $A = \{1\}$, $B = \{2\}$ and $S = \emptyset$ the global AMP Markov property yields $X_1 \perp \perp X_2$. On the other hand the moral graph $G^m$ is complete and consequently no conditional independence relation can be read off the graph.

**Theorem 4.8** Suppose $X$ satisfies the block-recursive Markov property with respect to $G$. Then $X$ also satisfies the global AMP Markov property with respect to $G$.

**Proof.** The global AMP Markov property is an immediate consequence of Lemma 4.6 and Theorem 4.4.

Due to Lemma 4.5 the AMP Markov property can also be interpreted in terms of conditional independences between processes up to some time $t$, that is, if $X$ satisfies the block-recursive Markov property with respect to $G$ then

$$A \bowtie B \mid S \ [G(ABS)^m] \Rightarrow X_A(t) \perp \perp X_B(t) \mid X_S(t).$$

(4.4)

### 4.3 Global causal Markov property

In the mixed graphs associated with a graphical time series models each component $X_a$ of the process is represented only by a single vertex $a$. Thus for the derivation of the noncausality relations encoded by a graph each vertex $a$ has to be interpreted sometimes as $X_a(t)$ and sometimes as $\bar{X}_a(t)$ depending on the type and the direction
of the adjacent edges. Since this information about the ordering in time in the relationships is removed by moralization the global Markov properties discussed so far do not allow to retrieve noncausality relations between the variables.

In order to preserve the information about the causal ordering of the variables when moralising mixed graphs, Eichler (2000) introduced a special separation concept for mixed graphs which is based on the idea of splitting the past and the present of certain variables and considering them together in a chain graph. For any subset $B$ of $V$ this splitting of the past and the present for the variables in $B$ can be accomplished by augmenting the moral graph $G^m$ with new vertices $b^*$ for all $b \in B$ which then represent the present of these variables, i.e. $b^*$ corresponds to $X_b(t)$, whereas all other vertices stand for the past at time $t$, i.e. $a$ for $X_a(t - 1)$. The new vertices are joined by edges such that we obtain a chain graph with two chain components $V$ and $B^*$ which reflects the pairwise dependence structure between the present and the past values of the variables.

More precisely, we define the augmentation chain graph $G_{B^*}^{aug}$ of $G$ as the chain graph with chain components $V$ and $B^*$ and edge set $E_{B^*}^{aug}$ such that for all $a_1, a_2 \in U$ and $b_1, b_2 \in B$

\[
\begin{align*}
    a_1 \rightarrow a_2 &\notin E_{B^*}^{aug} \iff a_1 \rightarrow a_2 \notin E^m, \\
    a_1 \rightarrow b_1^* &\notin E_{B^*}^{aug} \iff a_1 \rightarrow b_1 \notin E, \\
    b_1^* \rightarrow b_2^* &\notin E_{B^*}^{aug} \iff b_1 \ni b_2 \mid B \backslash \{b_1, b_2\} [G^m].
\end{align*}
\]

Each vertex in this chain graph represents one of the components of the random vector $\xi = (\xi_a, a \in V \cup B^*)$ where $\xi_a = X_a(t - 1)$ and $\xi_{b^*} = X_b(t)$. Due to the way in which the chain graph is obtained from the original graph $G$ the random variable $\xi$ apparently satisfies the pairwise AMP Markov property with respect to $G_{B^*}^{aug}$ (Andersson et al., 2000) if $X$ satisfies the pairwise causal Markov property with respect to $G$. This suggests to define the global causal Markov property in mixed graphs by the global AMP Markov property in augmentation chain graphs.

**Definition 4.9 (Global causal Markov property)** Let $G = (V, E)$ be a mixed graph. Then $X$ satisfies the global causal Markov property (GC) with respect to $G$ if for all disjoint subsets $A, B, S$ of $V$

\[
\begin{align*}
    A \ni B^* \mid S \cup B &\ [(G(U\backslash B^*)^{aug})^m] \Rightarrow X_A \rightarrow X_B [\mathcal{F}_{X_U}], \\
    A^* \ni B^* \mid U \cup S^* &\ [(G(U\backslash U^*)^{aug})^m] \Rightarrow X_A \sim X_B [\mathcal{F}_{X_U}],
\end{align*}
\]

where $U = A \cup B \cup S$.

We begin our study of the global causal Markov property by an alternative description in terms of separation in marginal ancestral graphs. For this we generalize the definition of the neighbours of a set. Let $A$ and $U$ be subsets of $V$ with $A \subseteq U$. Then we define the set of neighbours of $A$ relative to $U$, denoted by $\text{ne}(A, U)$, as the set of all $u \in U \setminus A$ such that there exists an undirected path in $G$ between $u$ and $A$ which is not hit by $U \setminus (A \cup \{u\})$. We note that $\text{ne}(A) = \text{ne}(A, V)$. 

15
PROPOSITION 4.10 Let $G = (V, E)$ be a mixed graph. Then we have for disjoint subsets $A$, $B$, and $S$ of $V$ and $U = A \cup B \cup S$

$$A \Join B^* \mid S \cup B \quad [(G(U)_{B^*})^m] \Leftrightarrow A \Join \pa(B) \mid S \cup B \quad [(G(U))^m].$$

Further let $S_A = \ne(A, A \cup S)$ and $S_B = \ne(B, B \cup S)$. Then

$$A^* \Join B^* \mid U \cup S^* \quad [(G(U)_{U^*})^m] \Leftrightarrow \pa(A \cup S_A) \Join \pa(B \cup S_B) \mid U \quad [G(U)^m] \land A \Join B \mid S \quad [G(U)^u].$$

PROOF. Since the parents of $B$ together with $B$ separate $B^*$ from the remaining vertices in $H^m = (G(U)_{B^*})^m$ we obtain

$$A \Join B^* \mid S \cup B \quad [H^m] \Leftrightarrow A \Join \pa(B) \mid S \cup B \quad [H^m].$$

The moralisation of $H$ affects only edges between vertices in $\pa(B)$ and thus any $S \cup B$-bypassing path in $H^m$ must be also a $S \cup B$-bypassing path in $G(U)^m$ and vice versa, which proves the first equivalence.

Next, let $H = G(U)_{U^*}$. Then the neighbours of $A^*$ in $H^m$ are $\pa(A \cup S_A) \cup S_A$. Hence

$$A^* \Join B^* \mid U \cup S^* \quad [H^m] \Leftrightarrow A^* \cup \pa(A \cup S_A) \Join B^* \cup \pa(B \cup S_B) \mid U \cup S^* \quad [H^m].$$

This holds if and only if first there is no edge between $A^*$ and $B^*$ in $H$ and thus $A \Join B \mid S \quad [G(U)^u]$ and second

$$\pa(A \cup S_A) \Join \pa(B \cup S_B) \mid U \cup S^* \quad [H^m].$$

Since any $U$-bypassing path in $G(U)^m$ is also a $U \cup U^*$-bypassing path in $H^m$ and vice versa this shows the second part of the proposition. \hfill \Box

We are now in the position to state the main result of this section in which we establish the equivalence of the block-recursive and the global causal Markov property. The result justifies our choice of the block-recursive causal Markov property for the definition of graphical time series models since it is simpler and thus better suited for the formulation of the constraints imposed by a graph.

THEOREM 4.11 Suppose that Assumption 2.1 holds. Then the process $X$ satisfies the block-recursive causal Markov property with respect to the mixed graph $G = (V, E)$ if and only if it satisfies the global causal Markov property with respect to $G$.

PROOF. Setting $S = \pa(B)$ and $A = V \setminus S$ for arbitrary $B \subseteq V$, the global causal Markov property directly yields the first relation in (BC). The second relation can be derived similarly.

Conversely, let $H = G(U)_{B^*}$ where $U = A \cup B \cup S$. Suppose that $A$ and $B^*$ are separated by $S \cup B$ in $H^m$. By Proposition 4.10 $S \cup B$ separates $A$ and $\pa(B)$ in $G(U)^m$ which by (4.4) implies

$$\hat{X}_{\pa(B)}(t) \perp \hat{X}_A(t) \mid \hat{X}_{S \cup B}(t).$$
Further we obtain from the block-recursive causal Markov property that

\[ X_B(t + 1) \perp \bar{X}_A(t) \mid \bar{X}_{pa(B) \cup S_B}(t) \]

since \( A \) and \( pa(B) \) are disjoint. Together with the previous relation this proves the noncausality of \( X_A \) for \( X_B \) relative to \( \mathcal{F}_U \).

For the second part, let \( H = G(U)^{aug} \) and assume that \( A^* \not\perp \not\perp B^* \mid U \cup S^* \mid [H^m] \). Let \( S_0 = \{ s \in S \mid pa(s) \subseteq U \} \), which in particular includes all \( s \in S \) which have no parents. For any \( s \in S \) the set \( pa(s) \) is complete due to moralization of immoralities. Therefore the set \( pa(s) \setminus U \), if nonempty, must be separated from \( A^* \) or \( B^* \) by \( U \cup S^* \).

Thus we define \( S_A = \{ s \in S \setminus S_0 \mid pa(s) \setminus U \not\perp \not\perp B^* \mid U \cup S^* \mid [H^m] \} \) and \( S_B = S \setminus (S_0 \cup S_A) \). It is immediately clear that \( pa(S_B) \) is separated from \( A^* \) by \( U \cup S^* \).

Further it follows that \( pa(A \cup S_A) \) and \( pa(B \cup S_B) \) are separated by \( U \cup U^* \) in \( H^m \). Since nonadjacent vertices in \( an(U) \setminus U \) are not joined by moralisation of \( H \), the two sets are also separated by \( U \) in \( G(U)^m \). Hence we get with (4.4) that

\[ \bar{X}_{pa(A \cup S_A) \setminus U}(t) \perp \bar{X}_{pa(B \cup S_B) \setminus U}(t) \mid \bar{X}_U(t). \]  \hspace{1cm} (4.5)

Noting that \( pa(S_0) \subseteq U \), the block-recursive causal Markov property yields

\[ X_{S_0}(t + 1) \perp \bar{X}_{U \setminus U}(t) \mid \bar{X}_U(t) \]  \hspace{1cm} (4.6)

and further with (4.5)

\[ \bar{X}_{pa(A \cup S_A) \setminus U}(t) \perp \bar{X}_{pa(B \cup S_B) \setminus U}(t) \mid \bar{X}_U(t), X_{S_0}(t + 1). \]  \hspace{1cm} (4.7)

Since \( pa(A \cup S_A) \subseteq V \setminus pa(B \cup S_B) \) the block-recursive causal Markov property combined with (4.6) implies

\[ X_{B \cup S_B}(t + 1) \perp \bar{X}_{pa(A \cup S_A) \setminus U}(t) \mid \bar{X}_{U \cup pa(B \cup S_B)}(t), X_{S_0}(t + 1) \]

and further with (4.7)

\[ X_{B \cup S_B}(t + 1) \perp \bar{X}_{pa(A \cup S_A) \setminus U}(t) \mid \bar{X}_U(t), X_{S_0}(t + 1). \]  \hspace{1cm} (4.8)

Next, we note that \( S_0 \) separates \( S_A \) and \( S_B \) in the undirected subgraph \( G(U)^u \) since otherwise there existed adjacent vertices \( s_1^* \in S_A^* \) and \( s_2^* \in S_B^* \). Then for any vertices \( v_1 \in pa(S_A) \setminus U \) and \( v_2 \in pa(S_B) \setminus U \) the 2-biflag \( v_1 \rightarrow s_1^* \leftarrow s_2^* \rightarrow v_2 \) is in \( H \). Hence \( v_1 \) and \( v_2 \) are adjacent in \( H^m \). This contradicts the fact that \( pa(S_A) \setminus U \) and \( B^* \) are separated by \( U \cup S^* \) since by definition of \( S_B \) \( v_2 \) is not separated from \( B^* \) by \( U \cup S^* \). It then follows from the the block-recursive causal Markov property and (4.6) that

\[ X_{A \cup S_A}(t + 1) \perp X_{B \cup S_B}(t + 1) \mid \bar{X}_{an(U)}(t), X_{S_0}(t + 1) \]

Since on the other hand we also have

\[ X_{A \cup S_A}(t + 1) \perp \bar{X}_{V \setminus (U \cup pa(A \cup S_A))}(t) \mid \bar{X}_{U \cup pa(A \cup S_A)}(t), X_{S_0}(t + 1), \]
Figure 4.2: Illustration of separation by moralisation: (a) Mixed graph $G$, (b) moralized marginal ancestral graph $G(U)^m$, and (c) augmentation chain graph $G(U)^{aug}$.

we get

$$X_{A \cup S_A} (t+1) \perp \perp X_{B \cup S_B} (t+1) \mid X_{U \cup \text{pa}(A \cup S_A)} (t), X_{S_0} (t+1).$$

Contraction with (4.8) finally gives

$$X_{A \cup S_A} (t+1) \perp \perp X_{B \cup S_B} (t+1) \mid X_U (t), X_{S_0} (t+1),$$

from which the desired relation follows by decomposition.

The next result summarizes the relations between the various Markov properties discussed in this paper.

**Corollary 4.12** Suppose that Assumption 2.1 holds. Then the causal Markov properties are related as follows:

$$(GC) \leftrightarrow (BC) \Rightarrow (LC) \leftrightarrow (PC),$$

$$(BC) \Rightarrow (GA) \Rightarrow (G)$$

If additionally condition (3.4) holds then the four causal Markov properties are all equivalent.

**Proof.** The equivalence of (PC) and (LC) has already been shown in Section 3. Further (BC) obviously implies (LC). Conversely, for $A \subseteq V$ it follows from (LC) that $X_{V \setminus (\text{pa}(A) \cup A)} \not\rightarrow X_a [\mathcal{F}_X]$ for all $a \in A$. If condition (3.4) holds this implies $X_{V \setminus (\text{pa}(A) \cup A)} \not\rightarrow X_A [\mathcal{F}_X]$. All other implications have been proved before.

**Example 4.13** We illustrate the presented separation concepts and global Markov properties by application to the mixed graph $G$ in Figure 4.2 (a). Suppose that $X$ is a stationary time series which satisfies the block-recursive Markov property with respect to $G$ and that we want to know about the relationship between the components $X_a$ and $X_b$ taking into account only information given by the subprocess $X_U$ where $U = \{a, b, x, y, z\}$.

For the construction of the marginal ancestral graph we note that the set of ancestors of $U$ is $V \setminus \{d\}$. Removing this node and all adjacent edges from the graph, we need to insert one additional edge $c \rightarrow y$ to obtain the marginal ancestral graph $G(U)$. The corresponding moral graph $G(U)^m$ is depicted in Figure 4.2 (b). In this graph the two vertices $a$ and $b$ clearly are not separated by $\{x, y, z\}$ and therefore...
the graph suggests that $X_a$ and $X_b$ are not conditionally independent given $X_{\{x,y,z\}}$. However, it should be noted that the global AMP Markov property does not provide any sufficient criterion for this interpretation of nonseparation in graphs.

Next, in order to check whether $X_a$ is noncausal for $X_b$ relative to the chosen filtration $\mathcal{F}_{X_U}$ the graph $G(U)^m$ in Figure 4.2 (b) has to be augmented by one additional vertex $b^*$. Connecting the new vertex with the parents of $b$ we obtain the augmentation chain graph $G(U)^{aug}$ shown in Figure 4.2. Since moralization of this graph does not lead to any additional edges we find that $a$ and $b^*$ are separated by the set $\{b, x, y, z\}$. Hence it follows from the the global causal Markov property that $X_a$ is indeed noncausal for $X_b$ relative to $\mathcal{F}_{X_U}$.

Similarly it can be shown that $X_a$ and $X_b$ are contemporaneously conditional independent relative to $\mathcal{F}_{X_A}$ whereas the global causal Markov property with respect to $G$ does not imply $X_b \rightarrow X_a \ [\mathcal{F}_{X_A}]$. Intuitively this is clear because in the graph the directed path from $b$ to $a$ does not intersect the set $A$ and thus suggests that $X_b$ has a direct effect on $X_a$ which is not mediated by any of the variables in $X_{\{x,y,z\}}$.

### 5 Pathwise separation criterion

The concept of separation via moralization does not provide a separation criterion in the mixed graph $G$ itself since the marginal ancestral graphs and the augmentation chain graphs which appear in the definition of the global AMP and the global Markov property vary with the subsets $A$, $B$, and $S$. Furthermore the resulting moral graphs, having all directions removed, do not allow any conclusions about the possible nature of a causal relationship (in the sense of Granger) between two variables. As an example, consider the two mixed graphs $G$ and $G'$ in Figure 5.1 and suppose we want to know whether $X_a$ is noncausal for $X_b$ in a bivariate setting. Augmentation of an additional node $b^*$ leads for both graphs to the same augmentation chain graph. The corresponding moral graph is depicted in Figure 5.1 (c). In this graph $a$ and $b^*$ are not separated by $b$ and hence $X_a$ possibly causes $X_b$. However, the first graph $G$ indicates an indirect effect of $X_a$ on $X_b$ mediated by $X_c$ whereas in the second graph $G'$ the dependence between $X_a$ and $X_b$ is due to a common influence from $X_c$. This information cannot be retrieved from the moralized augmentation chain graph in Figure 5.1 (c).

In this section, we discuss an alternative, path-oriented approach for defining separation in graphs which relates conditional independence to the blocking of certain pathways in the graph. This approach is due to Pearl (1988), who introduced the notion of $d$-separation in directed acyclic graphs, and has been applied to various types of graphs (e.g. Spirtes, 1995; Studený and Bouckaert, 1998; Koster, 1999, 2000). In the case of the AMP Markov property for chain graph Markov models a
pathwise separation criterion, called \( p \)-separation, has been presented by Levitz et al. (2001). In the following we will show that the concept of \( p \)-separation can also applied more generally to the class of mixed graphs discussed in this paper.

Let \( G = (V, E) \) be a mixed graph and \( \pi = \langle e_1, \ldots, e_n \rangle \) a path in \( G \) with edges \( e_i = v_{i-1} \cdots v_i \). Recall that according to our definition of paths the vertices \( v_i \) do not need to be different, that is, the path may be self-intersecting. We then call an intermediate point \( v_i \), \( 1 \leq i \leq n-1 \) a collider in \( \pi \) if the adjacent edges in \( \pi \) meet either head-to-head or head-to-line at \( v_i \), that is if \( \pi \) contains either \( v_{i-1} \leftarrow v_i \rightarrow v_{i+1} \), \( v_{i-1} \rightarrow v_i \rightarrow v_{i+1} \), or \( v_{i-1} \rightarrow v_i \leftarrow v_{i+1} \) as a subpath. Otherwise the vertex \( v_i \) is called a noncollider. We further define the sets \( A \subseteq V \) subsets of \( V \) intermediate point \( v \) need not to be different, that is, the path may be self-intersecting. We then call an \( e \) applied more generally to the class of mixed graphs discussed in this paper.

Let \( G = (V, E) \) be a mixed graph and \( \pi = \langle e_1, \ldots, e_n \rangle \) a path in \( G \) with edges \( e_i = v_{i-1} \cdots v_i \). Recall that according to our definition of paths the vertices \( v_i \) do not need to be different, that is, the path may be self-intersecting. We then call an intermediate point \( v_i \), \( 1 \leq i \leq n-1 \) a collider in \( \pi \) if the adjacent edges in \( \pi \) meet either head-to-head or head-to-line at \( v_i \), that is if \( \pi \) contains either \( v_{i-1} \leftarrow v_i \rightarrow v_{i+1} \), \( v_{i-1} \rightarrow v_i \rightarrow v_{i+1} \), or \( v_{i-1} \rightarrow v_i \leftarrow v_{i+1} \) as a subpath. Otherwise the vertex \( v_i \) is called a noncollider. We further define the sets \( C_\pi = \{ v \in V | v \) is a collider in \( \pi \} \) and \( N_\pi = \{ v \in V | v \) is a noncollider in \( \pi \} \).

**Definition 5.1** \((p\text{-separation})\) Let \( S \) be a subset of \( V \). Then a path \( \pi \) in \( G \) is \( S \)-open if \( C_\pi \subseteq S \) and \( N_\pi \subseteq V \setminus S \), otherwise the path is \( S \)-blocked. Furthermore, let \( A, B, \) and \( S \) be disjoint subsets of \( V \). Then \( S \) \( p \)-separates \( A \) and \( B \) in \( G \), denoted by \( A \nabla_p B \mid S \ [G] \), if every path in \( G \) between \( A \) and \( B \) is \( S \)-blocked.

The next theorem generalizes Theorem 4.1 of Levitz et al. (2001) to the class of mixed graphs.

**Theorem 5.2** Let \( G = (V, E) \) be a mixed graph and let \( A, B, \) and \( S \) be disjoint subsets of \( V \). Then \( S \) \( p \)-separates \( A \) and \( B \) in the mixed graph \( G \) if and only if \( S \) separates \( A \) and \( B \) in \( G(\text{ABS})^m \),

\[
A \nabla_p B \mid S \ [G] \iff A \nabla B \mid S \ [G(\text{ABS})^m],
\]

where \( \text{ABS} = A \cup B \cup S \).

We first show that in order to check \( p \)-separation in a graph it is sufficient to consider only paths in the set of ancestors of the studied sets \( A, B, \) and \( S \) since all other paths are always \( S \)-blocked.

**Lemma 5.3** Let \( G = (V, E) \) be a mixed graph and \( A, B, S \) disjoint subsets of \( V \). Then the set \( S \) \( p \)-separates \( A \) and \( B \) in \( G \) if and only if \( S \) \( p \)-separates \( A \) and \( B \) in \( G(\text{ABS}) \).

**Proof.** Let \( \pi = \langle e_1, \ldots, e_n \rangle \) be a \( S \)-open path between \( A \) and \( B \) in \( G(\text{ABS}) \). If all \( e_j \) are edges in \( G \) the path \( \pi \) is also \( S \)-open in \( G \), so we assume that the edges \( e_j, \ldots, e_m \) in \( \pi \) do not occur in \( G \). These edges \( e_j \) are necessarily undirected since all directed edges in \( G(\text{ABS}) \) also occur in \( G \). Let \( v_{j_k} = v_{j_k} \rightarrow v_{j_k+1} \). Then by definition of the marginal ancestral graph there exists an undirected path \( \phi_{j_k} \) between \( v_{j_{k-1}} \) and \( v_{j_k} \) which bypasses an(\text{ABS})\setminus\{v_{j_{k-1}}, v_{j_k}\} \) and therefore is \( S \)-open. Replacing all edges \( e_{j_k} \) in \( \pi \) by the corresponding paths \( \phi_{j_k} \) we obtain a new path \( \pi' \) which connects \( A \) and \( B \) in \( G \). This path \( \pi' \) is also \( S \)-open since the replacement of \( e_{j_k} \) by the undirected and \( S \)-open path \( \phi_{j_k} \) does not change the collider resp. noncollider status of the nodes \( v_{j_{k-1}} \) and \( v_{j_k} \).

Conversely let \( \pi = \langle e_1, \ldots, e_n \rangle \) be a \( S \)-open path between \( A \) and \( B \) in \( G \). Then all edges in \( \pi \) with both endpoints in an(\text{ABS}) also occur in \( G(\text{ABS}) \) since \( G_{\text{an}(\text{ABS})} \).
is a subgraph of $G(\text{ABS})$. We first show that the endpoints of any directed edge $e_j$ in $\pi$ are in $\text{an}(\text{ABS})$. Let $e_j = v_j \rightarrow v_{j+1}$ (the case $e_j = v_j \leftarrow v_{j+1}$ is treated similarly). Then there exists a directed subpath $\langle e_j, \ldots, e_{j+r} \rangle$ of maximal length such that either $v_{j+r}$ is an endpoint of $\pi$ and thus in $A \cup B$ or $e_{j+r+1} \in \{v_{j+r} \leftarrow v_{j+r+1}, v_{j+r} \rightarrow v_{j+r+1}\}$. In the latter case $v_{j+r}$ is a collider and thus in $S$ since $\pi$ is $S$-open. It follows that $v_j, v_{j+1} \in \text{an}(\text{ABS})$.

Next if $e_j$ is an edge in $\pi$ which does not occur in $G(\text{ABS})$ at least one of its endpoints $v_{j-1}$ and $v_j$ is not in $\text{an}(\text{ABS})$. Thus there exists an undirected subpath $\psi_{i,k} = \langle e_i, \ldots, e_k \rangle$ with $i \leq j \leq k$ such that $v_{i-1}, v_k \in \text{an}(\text{ABS})$ but all intermediate points are not in $\text{an}(\text{ABS})$. In other words, $v_{i-1}$ and $v_k$ are not separated by $\text{an}(\text{ABS}) \backslash \{v_{j-1}, v_k\}$ in $G$ which implies the presence of the undirected edge $f_{i,k} = v_{i-1} - v_k$ in $G(\text{ABS})$. Replacing all undirected subpaths $\phi_{i,k}$ with intermediate points not in $\text{an}(\text{ABS})$ by the corresponding edge $f_{i,k}$, we obtain a path between $A$ and $B$ in $G(\text{ABS})$ which still has all its colliders in $S$ and all its noncolliders outside $S$ and therefore is $S$-open. 

\[ \square \]

**Lemma 5.4** Let $\pi$ be a path in the mixed graph $G = (V,E)$. Then $\pi$ does not contain sequences of more than two consecutive colliders.

**Proof.** Since every tail of a directed edge in $\pi$ is a noncollider any two consecutive colliders $v_k$ and $v_{k+1}$ in $\pi$ must be connected by an undirected edge in $\pi$. On the other hand if $v_k \leftarrow v_{k+1} \leftarrow v_{k+2}$ the middle vertex $v_{k+1}$ is a noncollider which excludes sequences of more than two consecutive colliders. 

\[ \square \]

**Proof of Theorem 5.2.** By Lemma 5.3 it suffices to show $S$ separates $A$ and $B$ in $G(\text{ABS})^m$ if and only if $S$ $p$-separates $A$ and $B$ in $G(\text{ABS})$.

Let $\pi = \langle e_1, \ldots, e_n \rangle$ be an $S$-bypassing path between $A$ and $B$ in $G(\text{ABS})^m$. We first show that then there exists a path $\pi'$ which connects $A$ and $B$ in $G(\text{ABS})$ such that all intermediate points in $S$ are colliders. Let $e_k = v_{k-1} \leftarrow v_k$ be an edge in $\pi$. If in $G(\text{ABS})$ the vertices $v_{k-1}$ and $v_k$ are joined by an edge $e'_k$ we set $\phi_k = \langle e'_k \rangle$. Otherwise $e_k$ is due to the moralisation of $G(\text{ABS})$ and there exists a path $\phi_k$ in $G(\text{ABS})$ which is of the form $v_{k-1} \rightarrow u \leftarrow v_k$, $v_{k-1} \leftarrow u \rightarrow v_k$, or $v_{k-1} \leftarrow u \rightarrow w \leftarrow v_k$ with $u, w \in \text{an}(\text{ABS})$. Concatenating the paths $\phi_k$ corresponding to the edges $e_k$ in $\pi$ we obtain a path $\pi' = \langle \phi_1, \ldots, \phi_n \rangle$ in $G(\text{ABS})$ connecting $A$ and $B$. Since $v_k \notin S$ by assumption on $\pi$ and all intermediate points of the subpaths $\phi_k$ are colliders the path $\pi'$ has the required property.

Next we show the existence of a $S$-open path $\bar{\pi}$ in $G(\text{ABS})$ connecting $A$ and $B$. Let $\pi' = \langle e_1, \ldots, e_n \rangle$ be the path constructed above and assume that $v_0 \in A$. Further let $v_{j_1}, \ldots, v_{j_m}$ denote the colliders in $\pi'$ and set $v_{j_{m+1}} = v_n$. We proceed by induction over $v_{j_k}$. First $\pi_1 = \langle e_1, \ldots, e_{j_1} \rangle$ is an $S$-open path between $A$ and $v_{j_1}$ since all intermediate points $v_1, \ldots, v_{j_1-1}$ are noncolliders and not in $S$. Assume that $\pi_k$ is an $S$-open path between $A$ and $v_{j_k}$.

(i) If there exists a directed path $\phi = v_{j_k} \rightarrow \ldots \rightarrow b \in B$ in $G(\text{ABS})$ with all intermediate points not in $S$ then the path $\pi^* = \langle \pi_k, \phi \rangle$ is already an $S$-open path between $A$ and $B$. 

21
(ii) If \( v_{jk} \in \text{an}(S) \) there exists a directed path \( \phi = v_{jk} \rightarrow \ldots \rightarrow s \in S \) with all intermediate points not in \( S \). Then \( \pi_{k+1} = \langle \pi_k, \phi, \tilde{e}_{jk+1}, \ldots, e_{jk+1} \rangle \) is an \( S \)-open path between \( A \) and \( v_{jk+1} \).

(iii) Otherwise since \( v_{jk} \in \text{an}(ABS) \) there exists a directed path \( \phi = a \leftarrow \ldots \leftarrow v_{jk} \) for some \( a \in A \) with all intermediate points outside \( S \). Then the path \( \pi_{k+1} = \langle \phi, e_{jk+1}, \ldots, e_{jk+1} \rangle \) is an \( S \)-open path between \( A \) and \( v_{jk+1} \).

Continuing this construction we eventually obtain a path \( \pi^* \) which connects \( A \) and \( B \) and is \( S \)-open.

Conversely, let \( \pi \) be an \( S \)-open path between \( A \) and \( B \) in \( G(ABS) \) with colliders \( v_1, \ldots, v_m \). If \( v_{jk} \) is an isolated collider (i.e. \( v_j \) and \( v_{jk}+1 \) are noncolliders) then the subgraph induced by \( v_{jk-1}, v_j, \) and \( v_{jk+1} \) forms either an immorality or a flag and thus \( v_{jk-1} \) and \( v_{jk+1} \) are joined by an edge in the moral graph \( G(ABS)^m \). Next if \( v_{jk} \) and \( v_{jk+1} \) is a pair of consecutive colliders then by Lemma 5.4 the adjacent vertices \( v_{jk-1} \) and \( v_{jk+2} \) are noncolliders and the subgraph induced by \( v_{jk-1}, \ldots, v_{jk+2} \) forms a 2-biflag. Therefore \( v_{jk-1} \) and \( v_{jk+2} \) are connected in \( G(ABS)^m \). It follows that the sequence of noncolliders in \( \pi \) forms an \( S \)-bypassing undirected path connecting \( A \) and \( B \) in \( G(ABS)^m \).

Like separation in undirected graphs the notion of \( p \)-separation is clearly symmetric since it only refers to paths between two sets \( A \) and \( B \). In order to describe noncausality relations graphically in terms of \( p \)-separation it is therefore necessary to apply the concept only to those paths which can be associated with a causal relationship between the variables.

Let \( G = (V, E) \) be a mixed graph. A path \( \pi = \langle e_1, \ldots, e_n \rangle \) in \( G \) between vertices \( a \) and \( b \) is said to be \( b \)-pointing if it has a head at node \( b \), that is \( e_n = v_{n-1} \rightarrow b \). More generally, we call \( \pi \) a \( B \)-pointing path if \( \pi \) is \( b \)-pointing for some \( b \in B \). Furthermore following Studený and Bouckaert (1998), we define a slide in \( G \) from a vertex \( a \) to a vertex \( b \) as a path \( \langle e_1, \ldots, e_n \rangle \) with \( n \geq 1 \) such that \( e_1 = a = v_{1} \rightarrow v_{1} \) and \( e_j = v_{j-1} \rightarrow v_{j} \) for all \( j = 2, \ldots, n \) with \( v_{n} = b \). The vertex \( a \) is then called the top node of the slide. With this definition we call a path \( \pi \) in \( G \) a slide-to-slide path if \( \pi = \langle \sigma, \tilde{\pi}, \tau \rangle \) where \( \sigma \) and \( \tau \) are slides with top nodes \( t_{\sigma} \) and \( t_{\tau} \), respectively, and \( \tilde{\pi} \) is a path between \( t_{\sigma} \) and \( t_{\tau} \).

In the next theorem we show that the global causal Markov property can be formulated alternatively in terms of blocked path of certain types.

**Theorem 5.5** \( p \)-separation criterion Let \( G = (V, E) \) be a mixed graph and \( A, B, S \) disjoint subsets of \( V \). Then with \( \text{ABS} = A \cup B \cup S \)

(i) \( A \) and \( B^* \) are separated by \( S \cup B \) in \( (G(ABS)^{\text{aug}}) \) if and only if all \( B \)-pointing paths in \( G \) between \( A \) and \( B \) are blocked by \( S \cup B \);

(ii) \( A^* \) and \( B^* \) are separated by \( S \cup S^* \) in \( (G(ABS)^{\text{aug}}) \) if and only if all undirected and all slide-to-slide paths in \( G \) between \( A \) and \( B \) are blocked by \( \text{ABS} \).

**Proof.** We first note that every \( B \)-pointing path \( \pi \) between \( A \) and \( B \) is of the form \( \pi = \langle \tilde{\pi}, e \rangle \) where \( e \) is a directed edge \( u \rightarrow b \) for some \( b \in B \). Thus \( \pi \) is \( S \cup B \)-open
if and only if $u \notin S \cup B$ and $\tilde{\pi}$ is a $S \cup B$-open path between $A$ and $u \in \text{pa}(B)$. In other words every $B$-pointing path between $A$ and $B$ is $S \cup B$-blocked if and only if $A$ and $\text{pa}(B)$ are $p$-separated by $S \cup B$. The first part of the theorem then follows from Lemma 4.10 and Theorem 5.2.

For the second part, we observe that undirected paths between the vertices in an($ABS$) are preserved when replacing $G$ by $G(ABS)$. Consequently all undirected paths in $G$ between $A$ and $B$ are $ABS$-blocked if and only if every path between $A$ and $B$ in $G(ABS)^u$ is hit by $S$, that is, $A \bowtie B \mid S \{G(ABS)^u\}$.

Next, let us assume that all undirected paths between $A$ and $B$ are blocked by $ABS$ and let $\pi$ be a slide-to-slide path between $A$ and $B$. Then $\pi$ is of the form $\pi = (\sigma_a, \tilde{\pi}, \sigma_b)$ where $\sigma_a$ is a slide between $t_a$ and $a \in A$ with top node $t_a$, $\sigma_b$ is a slide between $t_b$ and $b \in B$ with top node $t_b$, and $\tilde{\pi}$ is a path between $t_a$ and $t_b$. Both top nodes $t_a$ and $t_b$ are noncolliders as they have at least one adjacent tail. Noting that $A \bowtie B \mid S \{G(ABS)^u\}$ by the above assumption, the path $\pi$ is $ABS$-open if and only if $\sigma_a$ and $\sigma_b$ are $S$-open, $t_a, t_b \notin ABS$, and $\tilde{\pi}$ is $ABS$-open. Let $\sigma_a = (t_a \rightarrow c_a, \tau_a)$ where $\tau_a$ is an undirected path between $c_a$ and $a$. Then $\sigma_a$ is $S$-open if and only if $c_a \in S$ and $\tau_a$ is not hit by $S$. In other words, $A$ and $c_a$ are not separated by $S\{c_a\}$ and hence $c_a \in S_A$ and $t_a \in \text{pa}(S_A)$. Similarly we have $t_b \in \text{pa}(S_B)$. It follows that $\tilde{\pi}$ is $ABS$-open if and only if there exists an $ABS$-open path, namely $\tilde{\pi}$, between $\text{pa}(S_A \cup A)$ and $\text{pa}(S_B \cup B)$ which are thus not $p$-separated by $ABS$. The second part now follows again from Lemma 4.10 and Theorem 5.2.

Example 5.6 We consider again the situation in Example 4.13. In order to find out whether $X_a$ is noncausal for $X_b$ relative to $\mathcal{F}_{X_U}$ where $U = \{a, b, x, y, z\}$, we have to show that all $b$-pointing paths between $a$ and $b$ are blocked by $S = \{b, x, y, z\}$. Three such paths are given in Figure 5.2. All three paths are blocked by noncolliders in $S$ which are marked as black dots. Note that the last path $b$ is blocked by $b$ itself which also occurs as an intermediate point.

By similar arguments it can be shown that any slide-to-slide path between $a$ and $b$ is $S$-blocked. Since there is no undirected path between $a$ and $b$ this implies the contemporaneous conditional independence of $X_a$ and $X_b$ relative to $\mathcal{F}_{X_U}$. Finally since the directed path from $b$ to $a$ is $S$-open we cannot conclude that $X_b$ is noncausal for $X_a$ relative to $\mathcal{F}_{X_U}$.
6 Graphical ARCH models

In the examples of graphical time series models given in Section 3 we concentrated mainly on modelling the conditional mean of a stochastic process. In the analysis of financial time series which exhibit strong volatility the investigation of the conditional variance becomes more interesting which led to the development of the autoregressive conditional heteroscedasticity (ARCH) model and its various sub-sidiaries. Here a stationary stochastic process $X$ is said to follow a multivariate ARCH process if the conditional mean $\mathbb{E}(X(t) \mid \mathcal{F}(t - 1))$ is zero and the conditional covariance matrix

$$\mathbb{E}(X(t)X(t)' \mid \mathcal{F}(t - 1)) = \Sigma(t)$$

is $\mathcal{F}(t - 1)$-measurable and depends non-trivially on the past of the process. For an overview of multivariate ARCH models we refer to Bollerslev et al. (1994) and Gouriéroux (1997).

One key issue in the specification of multivariate ARCH models is the restriction of the number of parameters involved, which in a general setting can be very large. Various parametrization which allow different levels of complexity have been suggested. Here the graphical modelling approach can help to achieve a further reduction of the number of parameters.

In the following we formulate the constraints defining a graphical ARCH models associated with a graph $G = (V,E)$ for four different parametrizations of $\Sigma(t)$. For this we consider the class of all stationary process $X$ with conditional distribution $\mathcal{N}(0, \Sigma(t))$. Then each of the following sets of constraints for $\Sigma(t)$ defines a graphical ARCH($q$) model.

(i) **Constant conditional correlations:** The constant conditional correlation model of Bollerslev (1990) provides the most parsimonious parametrization of $\Sigma(t)$. The conditional variances are given by

$$\sigma_{ii}(t) = \sigma_{ii}^0 + \sum_{u=1}^{q} \sum_{k \in \text{pa}(i)} \alpha_i^k(u)X_k(t - u)^2,$$

whereas the conditional covariances are determined by the set of equations

$$\sigma_{ij}(t) = \sigma_{ij}^{1/2} \sigma_{jj}^{1/2} \rho_{ij} \quad \text{if } i \rightarrow j \in E,$$

$$K_{ij}(t) = 0 \quad \text{if } i \rightarrow j \notin E.$$

Here $K(t) = \Sigma(t)^{-1}$ is the inverse conditional covariance matrix.

(ii) **Constant conditional correlations with interaction:** In this parametrization the conditional variances $\sigma_{ii}(t)$ additionally depend on the interaction terms $X_k(t - u)X_l(t - u)$ if $k$ and $l$ are both parents of $i$. Thus the conditional variance can be written as

$$\sigma_{ii}(t) = \sigma_{ii}^0 + \sum_{u=1}^{q} \sum_{k,l \in \text{pa}(i): k < l} \alpha_{ki}(u)X_k(t - u)X_l(t - u).$$

The entries $\sigma_{ij}(t)$ have the same form as in (i).
(iii) **Vector ARCH model without interaction:** The vector ARCH model due to Kraft and Engle (1982) the correlation between the components of $X(t)$ may also depend on the past values of $X$. This leads to conditional covariances $\sigma_{ij}(t)$, $i \leq j$, of the form

$$
\sigma_{ij}(t) = \sigma_{ij}^0 + \sum_{u=1}^{q} \sum_{k \in \text{pa}(i) \cap \text{pa}(j)} \alpha_{ik}^{ij}(u)X_k(t-u)^2 \quad \text{if } i = j \text{ or } i \rightarrow j \in E,
$$

$$
K_{ij}(t) = 0 \quad \text{if } i \rightarrow j \notin E.
$$

(iv) **Vector ARCH model:** Again this model can be generalised by inclusion of interaction terms, that is, the conditional covariances $\sigma_{ij}(t)$ can be written as

$$
\sigma_{ij}(t) = \sigma_{ij}^0 + \sum_{u=1}^{q} \sum_{k,l \in \text{pa}(i) \cap \text{pa}(j): k < l} \alpha_{kl}^{ij}(u)X_k(t-u)X_l(t-u)
$$

if $i = j$ or $i \rightarrow j \in E$ while the conditions $K_{ij}(t) = 0$ for $i \rightarrow j \notin E$ remain unchanged.

For the constant conditional correlation models it is easy to derive conditions to ensure that the conditional covariances are positive definite almost surely for all $t$. In contrast, such conditions are difficult to impose and verify for the vector ARCH model. Therefore Engle and Kroner (1995) suggested an alternative representation for the multivariate ARCH($q$) model in which $\Sigma(t)$ is guaranteed to be positive definite almost surely for all $t$. In this BEKK-representation the conditional covariances of a graphical ARCH model are parametrized by

$$
\sigma_{ij}(t) = \sigma_{ij}^0 + \sum_{n=1}^{N} \sum_{u=1}^{q} \sum_{k,l \in \text{pa}(i) \cap \text{pa}(j): k < l} \alpha_{ki}^{ij}(u)\alpha_{lj}^{ij}(u)X_k(t-u)X_l(t-u).
$$

In this form it is immediately clear that if $\sigma_{ij}(t)$ depends on the past of $X_k$ then at least one of the conditional variances $\sigma_{ii}(t)$ and $\sigma_{jj}(t)$ must also depend on $X_k$. Although less obvious the same can be shown for the vector ARCH model in the original parametrization noting that the conditional covariance matrix $\Sigma(t)$ must be positive definite. Hence the graphical vector ARCH models fulfills condition (3.4). For the constant conditional correlation model condition (3.4) is trivially fulfilled.

Although condition (3.4) is satisfied by a wide variety of time series models it does not hold generally. As an example we consider a multivariate generalization of the Qualitative Threshold ARCH($q$) model of Gouriéroux and Monfort (1992). For this, we divide the sample space into $K$ intervals and let $I_k(X_l(t-1)) = 1$ if $X_l(t-1)$ is in the $k$th interval and zero elsewhere. Then the conditional covariance matrix of a QTARCH($q$) model associated with a graph $G$ is given by

$$
\sigma_{ij}(t) = \sigma_{ij}^0 + \sum_{k=1}^{K} \sum_{l \in \text{pa}(i) \cap \text{pa}(j)} \sum_{u=1}^{q} \alpha_{kl}^{ij}(u)I_k(X_l(t-u))
$$

if $i = j$ or $i \rightarrow j \in E$ while $K_{ij}(t) = 0$ for $i \rightarrow j \notin E$. Unlike in the graphical ARCH models above the effect of $X_l(t-1)$ is bounded. For suitable choices of $(\sigma_{ij}^0)$ it is therefore possible to have $\alpha_{ii}^{ii} = \alpha_{kk}^{jj} = 0$ even though some parameters $\alpha_{kl}^{ij}$ are nonzero which violates condition (3.4).
7 Concluding remarks

In this paper, a new class of graphical time series models has been introduced which is particularly useful for modelling the dynamics of a multivariate time series. The models are formulated in terms of strong Granger-noncausality and thus allow the inclusion of arbitrary nonlinear dependencies. The graphical modelling approach can help to reduce the number of parameters involved in modelling high-dimensional nonlinear time series while encoding the constraints on the parameters in a simple graph which is easy to communicate. In these graphs the vertices represent the components of the time series while the edges visualize the dynamic relationships between the variables.

We have shown that the interpretation of these graphs, which for many models are built only from pairwise noncausality relations, is greatly enhanced by global Markov properties by which the separation properties of the graph can be related to conditional independence or noncausality statements about the process. The interpretation is particularly facilitated by the path-oriented concept of $p$-separation by which Granger-causal relations between the variables can be attributed to certain pathways in the graphs.

Two important issues have not been addressed in this paper. First, in many applications there is little prior knowledge about the causal relationships between the variables and empirical methods have to be used to find an appropriate graphical model. This step of model selection is hampered by the large number of possible models by which an exhaustive search becomes infeasible even for moderate dimensions. Therefore model search strategies are needed to lessen the computational burden.

A second issue, which is related to the problem of model selection, is the identification of causal effects. It is clear from the definition of Granger-causality that only if all relevant variables are included in a study we can conclude from Granger-causality to the existence of a causal effect whereas the omission of important variables can lead to spurious causalities. However, Hsiao (1982) noted that such spurious causalities may vanish if the information set is reduced. In other words two processes which both satisfy the pairwise causal Markov property with respect to a graph $G$ will in general exhibit different noncausality relation relative to information subsets due to the presence or absence of spurious causalities. Following the ideas in the monograph by Pearl (2000) it is possible to construct an extended graph in which additional vertices represent latent variables and which encodes all noncausality relations satisfied by the process under study. Although the graph does not need to be uniquely determined the class of all possible graphs allows to detect edges which cannot be due to spurious causality and thus correspond to a causal effect between the variables. For more details we refer to Eichler (2000) who discussed this approach in a linear framework.

Finally, we note that the graphical modelling approach presented here can be generalized to processes with exogenous variables or to nonstationary time series on an interval $[0, T]$. This more general setting has been considered by Eichler (2000) again in a linear framework.
References


