

---

# Graphical Gaussian Modelling of Multivariate Time Series with Latent Variables

---

Michael Eichler

Department of Quantitative Economics  
Maastricht University, NL  
m.eichler@maastrichtuniversity.nl

## Abstract

In time series analysis, inference about cause-effect relationships among multiple time series is commonly based on the concept of Granger causality, which exploits temporal structure to achieve causal ordering of dependent variables. One major problem in the application of Granger causality for the identification of causal relationships is the possible presence of latent variables that affect the measured components and thus lead to so-called spurious causalities. In this paper, we describe a new graphical approach for modelling the dependence structure of multivariate stationary time series that are affected by latent variables. To this end, we introduce dynamic maximal ancestral graphs (dMAGs), in which each time series is represented by a single vertex. For Gaussian processes, this approach leads to vector autoregressive models with errors that are not independent but correlated according to the dashed edges in the graph. We discuss identifiability of the parameters and show that these models can be viewed as graphical ARMA models that satisfy the Granger causality restrictions encoded by the associated dynamic maximal ancestral graph.

## 1 INTRODUCTION

The notion of causality and the identification of new causal relationships play a central role in scientific research. In time series analysis, inference about cause-effect relationships is commonly based on the concept

---

Appearing in Proceedings of the 13<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna Resort, Sardinia, Italy. Volume 9 of JMLR: W&CP 9. Copyright 2010 by the authors.

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

Recent advances in the understanding of such latent variable structures were based on graphical models, which provide a general framework for describing and inferring causal relations (e.g., Pearl 2000, Lauritzen 2001). For time series, this graphical approach for the discussion of causal relationships in systems that are affected by latent variables has been considered first in Eichler (2005). Based on previously developed graphical representations of Granger-causal relationships in multivariate time series (Eichler 2001, 2007), a new class of path diagrams for the representation of the interrelationships in multivariate time series with latent variables was introduced. These general path diagrams allow a better encoding of the conditional independencies if the system is affected by latent variables. In Eichler (2005), a multi-step procedure for the identification of such general path diagrams was proposed, where each step requires the fitting of a new autoregressive model. As a consequence, this approach prohibits the application of most model selection strategies as well as the statistical comparison of two graphical representations of the dependence structure of a process.

In this paper, we consider an alternative approach and

present a new class of graphical time series models that are associated with general path diagrams and, in particular, dynamic maximal ancestral graphs (dMAGs). In section 2, we review path diagrams for vector autoregression processes and their relation to the notion of Granger causality as well as their Markov properties. In section 3, we discuss graphical representations of multivariate time series affected by latent variables and introduce dMAGs. The parametric model for Gaussian processes is presented in Section 4; Section 5 contains some concluding remarks.

## 2 VECTOR AUTOREGRESSIONS AND PATH DIAGRAMS

Let  $\mathbf{X}_V = (\mathbf{X}_V(t))_{t \in \mathbb{Z}}$  with  $\mathbf{X}_V(t) = (X_v(t))'_{v \in V}$  be a stationary Gaussian process with mean zero and covariances  $\mathbf{\Gamma}(u) = \mathbb{E}\mathbf{X}_V(t)\mathbf{X}_V(t-u)'$ . Throughout the paper, we assume that  $\mathbf{X}_V$  has a mean-square convergent autoregressive representation

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

where  $\mathbf{\Phi}(u)$  is a square summable sequence of  $V \times V$  matrices and  $(\boldsymbol{\varepsilon}_V(t))$  is a Gaussian white noise process with non-singular covariance matrix  $\mathbf{\Sigma}$ . The autoregressive structure of the process  $\mathbf{X}_V$  can be visualized by a path diagram in which the vertices correspond to the variables of  $\mathbf{X}_V$  while the edges—arrows and lines—between vertices indicate non-zero coefficients in the autoregressive representation of  $\mathbf{X}_V$ .

**Definition 2.1.** Let  $\mathbf{X}_V$  be a stationary Gaussian process with autoregressive representation (1). Then the *path diagram* associated with  $\mathbf{X}_V$  is the graph  $G = (V, E)$  with vertex set  $V$  and edge set  $E$  such that for distinct  $a, b \in V$

- (i)  $a \rightarrow b \notin E \Leftrightarrow \Phi_{ba}(u) = 0 \quad \forall u \in \mathbb{N}$ ;
- (ii)  $a \text{ --- } b \notin E \Leftrightarrow \Sigma_{ab} = \Sigma_{ba} = 0$ .

Since path diagrams of this form may contain two types of edges, they will be referred to as mixed graphs. Furthermore, we note that two vertices  $a$  and  $b$  may be connected by up to three edges, namely  $a \rightarrow b$ ,  $a \leftarrow b$ , and  $a \text{ --- } b$ . Similar path diagrams have been used to represent linear structural equation models (e.g., Wright 1934, Goldberger 1972, Koster 1999, Drton et al. 2009)<sup>1</sup>.

<sup>1</sup>In path diagrams for structural equation systems, correlated errors commonly are represented by bi-directed edges ( $\leftrightarrow$ ) instead of dashed lines (---). Since in our approach directions are associated with temporal ordering, we prefer (dashed) undirected edges to indicate correlation between the error variables. Dashed edges with a similar connotation are used for covariance graphs (e.g. Cox and

### 2.1 GRANGER CAUSALITY AND PATH DIAGRAMS

One fundamental tool for the investigation of the dynamic dependencies and the causal relationships among the variables of a multivariate time series is the concept of Granger causality (Granger 1969). This concept of causality is based on the common sense conception that causes always precede their effects. This temporal ordering implies that the past and present values of a series  $\mathbf{X}$  that influences another series  $\mathbf{Y}$  should help to predict future values of this latter series  $\mathbf{Y}$ . Furthermore, the improvement in the prediction should persist after any other relevant information for the prediction has been exploited. To make this idea more precise, let  $\overline{\mathbf{X}}_A(t) = (\mathbf{X}_A(s))_{s \leq t}$  for any  $A \subseteq V$  denote the past and present values of  $\mathbf{X}_A$  at time  $t$ .

**Definition 2.2.** Let  $A$  and  $B$  be disjoint subsets of  $S \subseteq V$ . Then  $\mathbf{X}_A$  is *Granger-noncausal* for  $\mathbf{X}_B$  with respect to  $\mathbf{X}_S$ , denoted as  $\mathbf{X}_A \nrightarrow \mathbf{X}_B [\mathbf{X}_S]$ , if

$$\mathbb{E}((\mathbf{X}_B(t+1) | \overline{\mathbf{X}}_S(t)) = \mathbb{E}((\mathbf{X}_B(t+1) | \overline{\mathbf{X}}_{S \setminus A}(t))$$

for all  $t \in \mathbb{Z}$ . Furthermore,  $\mathbf{X}_A$  and  $\mathbf{X}_B$  are said to be *contemporaneously uncorrelated* with respect to  $\mathbf{X}_S$ , denoted as  $\mathbf{X}_A \approx \mathbf{X}_B [\mathbf{X}_S]$ , if

$$\text{corr}(\mathbf{X}_A(t+1), \mathbf{X}_B(t+1) | \overline{\mathbf{X}}_S(t)) = 0$$

for all  $t \in \mathbb{Z}$ .

It is well known that the pairwise Granger-causal relationships among the variables of a Gaussian process  $\mathbf{X}_V$  are reflected in the autoregressive coefficients of the process and, thus, in the presence and absence of edges in the associated path diagram. More precisely, we have the following result (see Eichler 2007).

**Lemma 2.3.** Let  $G = (V, E)$  be the path diagram associated with a stationary Gaussian process  $\mathbf{X}_V$  of the form (1). Then

- (i)  $a \rightarrow b \notin E \Leftrightarrow \mathbf{X}_a \nrightarrow \mathbf{X}_b [\mathbf{X}_V]$
- (ii)  $a \text{ --- } b \notin E \Leftrightarrow \mathbf{X}_a \approx \mathbf{X}_b [\mathbf{X}_V]$

Because of this result, the path diagram associated with a process  $\mathbf{X}_V$  has also been called the *Granger causality graph* of  $\mathbf{X}_V$  (e.g., Eichler 2006).

### 2.2 MARKOV PROPERTIES

Under the assumptions imposed on  $\mathbf{X}_V$ , more general Granger-causal relationships than those in Lemma 2.3 can be derived from the path diagram associated with  $\mathbf{X}_V$ . This global Markov interpretation is based on

Wermuth 1996), whereas undirected edges (—) are commonly associated with nonzero entries in the inverse of the variance matrix (e.g. Lauritzen 1996).

a path-oriented concept of separating subsets of vertices in a mixed graph, which has been used previously to represent the Markov properties of linear structural equation systems (e.g. Spirtes et al. 1998, Koster 1999). Following Richardson (2003) we will call this notion of separation in mixed graphs *m*-separation.

Let  $G = (V, E)$  be a mixed graph and  $a, b \in V$ . A *path*  $\pi$  in  $G$  is a sequence  $\pi = \langle e_1, \dots, e_n \rangle$  of edges  $e_i \in E$  with an associated sequence of nodes  $v_0, \dots, v_n$  such that  $e_i$  is an edge between  $v_{i-1}$  and  $v_i$ . The vertices  $v_0$  and  $v_n$  are the *endpoints* while  $v_1, \dots, v_{n-1}$  are the *intermediate vertices* of the path. Notice that paths may be self-intersecting since we do not require that the vertices  $v_i$  are distinct. An intermediate vertex  $c$  on a path  $\pi$  is said to be an *m-collider* on  $\pi$  if the edges preceding and succeeding  $c$  both have an arrowhead or a dashed tail at  $c$  (i.e.  $\rightarrow c \leftarrow$ ,  $\rightarrow c \dashrightarrow$ ,  $\dashrightarrow c \leftarrow$ ,  $\dashrightarrow c \dashrightarrow$ ); otherwise  $c$  is said to be an *m-noncollider* on  $\pi$ . A path  $\pi$  between  $a$  and  $b$  is said to be *m-connecting* given a set  $C$  if

- (i) every *m-noncollider* on  $\pi$  is not in  $C$  and
- (ii) every *m-collider* on  $\pi$  is in  $C$ ;

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

With this notion of separation, it can be shown that path diagrams for multivariate time series have a similar Markov interpretation as path diagrams for linear structural equation systems (cf Koster 1999), that is, if two sets  $A$  and  $B$  are *m-separated* given a third set  $C$ , then the corresponding series  $\mathbf{X}_A$  and  $\mathbf{X}_B$  are conditionally independent given  $\mathbf{X}_C$ ; for details we refer to Eichler (2007). Derivation of such conditional independence statements requires that all paths between two sets are *m-blocked*. In contrast, for the derivation of Granger-causal relationships, it suffices to consider only a subset of these paths, namely those having an arrowhead at one endpoint. For a formal definition, we say that a path  $\pi$  between  $a$  and  $b$  is *b-pointing* if it has an arrowhead at the endpoint  $b$ ; similarly, a path between sets  $A$  and  $B$  is said to be *B-pointing* if it is *b-pointing* for some  $b \in B$ . Then, to establish Granger noncausality from  $\mathbf{X}_A$  to  $\mathbf{X}_B$ , it suffices to consider only all *B-pointing* paths between  $A$  and  $B$ . Similarly, a graphical condition for contemporaneous correlation can be obtained based on *bi-pointing* path, which have an arrowhead at both endpoints.

**Definition 2.4.** A stationary Gaussian process  $\mathbf{X}_V$  is Granger-Markov for a graph  $G$  if, for all disjoint subsets  $A, B, C \subseteq V$ , the following two conditions hold:

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

With this definition, it can be shown that path diagrams for vector autoregressions can be interpreted in terms of such global Granger-causal relationships.

**Theorem 2.5.** Let  $\mathbf{X}_V$  be a stationary Gaussian process given by (1), and let  $G$  be the associated path diagram. Then  $\mathbf{X}_V$  is Granger-Markov for  $G$ .

### 3 SYSTEMS WITH LATENT VARIABLES

The notion of Granger causality is based on the assumption that all relevant information is included in the analysis (Granger 1969, 1980). It is well known, that the omission of important variables can lead to temporal correlations among the observed components that are falsely detected as causal relationships. The detection of such so-called spurious causalities (Hsiao 1982) becomes a major problem when identifying the structure of systems that may be affected by latent variables.

Of particular interest will be spurious causalities of type I, where a Granger-causal relationship with respect to the complete process vanishes when only a subprocess is considered. Since the path diagrams from the previous section are defined, by Lemma 2.3, in terms of the pairwise Granger-causal relationships with respect to the complete process, they provide no means to distinguish such spurious causalities of type I from true causal relationships. To illustrate this remark, we consider the four-dimensional vector autoregressive process  $\mathbf{X}_V$  with components

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

where  $\varepsilon_i(t)$ ,  $i = 1, \dots, 4$  are uncorrelated white noise processes with mean zero and variance one. The true dynamic structure of the process is shown in Fig. 1(a). In this graph, the 1-pointing path  $3 \rightarrow 2 \leftarrow 4 \rightarrow 1$  is *m-connecting* given  $S = \{2\}$ , but not given the empty set. By Theorem 2.5, we conclude that  $\mathbf{X}_3$  is Granger-noncausal for  $\mathbf{X}_1$  in a bivariate analysis, but not necessarily in an analysis based on  $\mathbf{X}_{\{1,2,3\}}$ .

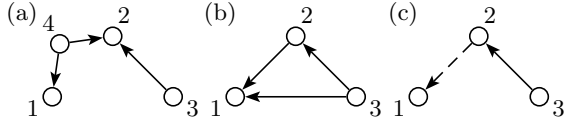


Figure 1: Graphical representations of the four-dimensional VAR(2) process in (2): (a) path diagram associated with  $X_{\{1,2,3,4\}}$ ; (b) path diagram associated with  $X_{\{1,2,3\}}$ ; (c) general path diagram for  $X_{\{1,2,3\}}$ .

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

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

$$X_2(t) = \gamma X_3(t-1) + \tilde{\varepsilon}_2(t),$$

$$X_3(t) = \varepsilon_3(t),$$

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

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

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

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

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

### 3.1 GENERAL PATH DIAGRAMS

We consider mixed graphs that may contain three types of edges, namely undirected edges ( $---$ ), directed edges ( $\rightarrow$ ), and dashed directed edges ( $\dashrightarrow$ ). For the

Table 1: Creation of edges by marginalizing over  $i$ .

Subpath $\pi$ in $G$	Associated edge $e_\pi$ in $G^{\{i\}}$
$a \rightarrow i \rightarrow b$	$a \rightarrow b$
$a \dashrightarrow i \rightarrow b$	$a \dashrightarrow b$
$a --- i \rightarrow b$	$a \dashrightarrow b$
$a \leftarrow i \rightarrow b$	$a \leftrightarrow b$
$a \leftarrow i \dashrightarrow b$	$a \leftrightarrow b$

sake of simplicity, we also use  $a \leftrightarrow b$  as an abbreviation for the triple edge  $a \dashrightarrow b$ . Unlike path diagrams for autoregressions, these graphs in general are not defined in terms of pairwise Granger-causal relationships, but only through the global Markov interpretation according to Definition 2.4. To this end, we simply extend the concept of  $m$ -separation introduced in the previous section by adapting the definition of  $m$ -noncolliders and  $m$ -colliders. Let  $\pi$  be a path in a mixed graph  $G$ . Then an intermediate vertex  $n$  is called an  $m$ -noncollider on  $\pi$  if at least one of the edges preceding and succeeding  $n$  on the path is a directed edge ( $\rightarrow$ ) and has its tail at  $n$ . Otherwise,  $n$  is called an  $m$ -collider on  $\pi$ . With this extension, we leave all other definitions such as  $m$ -separation or pointing paths unchanged.

The main difference between the class of mixed graphs with directed ( $\rightarrow$ ) and undirected ( $---$ ) edges and the more general class of mixed graphs that has been just introduced is that the latter class is closed under marginalization. This property makes it suitable for representing systems with latent variables.

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

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

self-loops. We omit the details.

Now suppose that, for some subsets  $A, B, C \subseteq V \setminus L$ ,  $\pi$  is an  $m$ -connecting path between  $A$  and  $B$  given  $S$ . Then all intermediate vertices on  $\pi$  that are in  $L$  must be  $m$ -noncolliders. Removing these vertices according to Table 1, we obtain a path  $\pi'$  in  $G^L$  that is still  $m$ -connecting. Since the converse is also true, we obtain the following result.

**Proposition 3.1.** *Let  $G = (V, E)$  be a mixed graph, and  $L \subseteq V$ . Then it holds that, for all distinct  $a, b \in V \setminus L$  and all  $C \subseteq V \setminus L$ , every path between  $a$  and  $b$  in  $G$  is  $m$ -blocked given  $C$  if and only if the same is true for the paths in  $G^L$ . Furthermore, the same equivalence holds for all pointing path and for all bi-pointing paths.*

It follows that, if a process  $\mathbf{X}_V$  is Markov for a graph  $G$ , the subprocess  $\mathbf{X}_{V \setminus L}$  is Markov for  $G^L$  and  $G^L$  encodes all relationships about  $\mathbf{X}_{V \setminus L}$  that are also encoded in  $G$ . Fig. 2 illustrates this marginalization property. Fig. 2(a) shows the complete path diagram including two latent variables  $L_1$  and  $L_2$  while Fig. 2(b) depicts the corresponding generalized path diagram after marginalizing over  $L_1$  and  $L_2$ . The only independence relationship encoded by both graphs is that  $X_1$  and  $X_3$  are contemporaneously uncorrelated in a bivariate analysis.

We note that insertion of edges according to Tab. 1 is sufficient but not always necessary for representing the relations in the subprocess  $\mathbf{X}_{V \setminus L}$ . This applies in particular to the last two cases in Tab. 1. For an example, we consider again the process (2) with associated path diagram in Fig. 1(a). By Tab. 1, the subpath  $1 \leftarrow 4 \rightarrow 2$  should be replaced by  $1 \leftrightarrow 2$ , which suggests that  $X_1$  Granger-causes  $X_2$  (as does the path  $1 \leftarrow 4 \rightarrow 2$  in the original path diagram), while in fact the structure can be represented by the graph in Fig. 1(c) (see Section 4.3).

### 3.2 DYNAMIC MAXIMAL ANCESTRAL GRAPHS

For systems with latent variables, the set of Granger-causal relationships and contemporaneous independencies that hold for the observed process does not uniquely determine a graphical representation within the class of general path diagrams. As an example, Fig. 3 displays two graphs that are Markov equivalent, that is, they encode the same set of Granger-causal and contemporaneous independence relations among the variables. Therefore, the corresponding graphical models—models that obey the conditional independence constraints imposed by the graph—are statistically indistinguishable. This suggests to choose one unique representative for each Markov equivalence class and to restrict model selection to these.

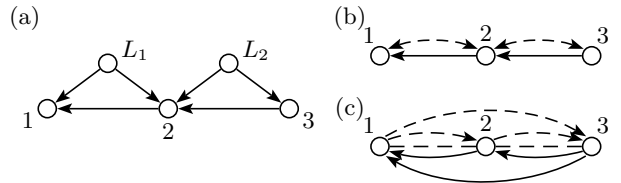


Figure 2: Graphical representations: (a) Full graph with two latent variables; (b) general path diagram; (c) dynamic maximal ancestral graph.

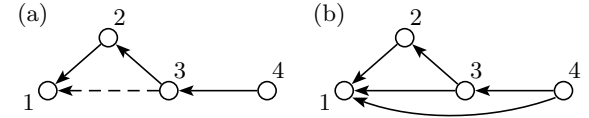


Figure 3: Two Markov equivalent graphs: (a) non-ancestral graph and (b) corresponding ancestral graph.

Following Richardson and Spirtes (2002), one suitable choice are maximal ancestral graphs. For vertices  $a, b \in V$ , we say that  $a$  is an ancestor of  $b$  if  $a = b$  or there exists a directed path  $a \rightarrow \dots \rightarrow b$  in  $G$ . The set of ancestors of  $b$  is denoted by  $\text{an}(b)$ . Then  $G = (V, E)$  is an ancestral graph if

$$a \in \text{an}(b) \Rightarrow a \dashrightarrow b \notin E. \quad (3)$$

for all distinct  $a, b \in V$ . We note that, in contrast to Richardson and Spirtes (2002), we do not require acyclicity (which is hidden in the time ordering). Furthermore, an ancestral graph  $G$  is maximal if addition of further edges changes the independence models; for details, we refer to Richardson and Spirtes (2002). We call these graphs dynamic maximal ancestral graphs (dMAG). For the example in Fig. 2, a Markov equivalent dMAG is shown in Fig. 2(c).

## 4 PARAMETRIC MODELLING

In this section, we present a parametric model that is associated with general path diagrams. In particular, the absence of an edge in a general path diagram will correspond to zero constraints on one or more model parameters.

The undirected edges in path diagrams associated with vector autoregressions are defined by the dependencies of the error process  $\varepsilon$ . To allow dashed directed edges in path diagrams, we weaken the restriction that the error process  $\varepsilon$  is white noise.

More precisely, we consider multivariate stationary Gaussian processes  $\mathbf{X}_V$  that are given by

$$\mathbf{X}_V(t) = \sum_{u=1}^p \Phi(u) \mathbf{X}_V(t-u) + \varepsilon_V(t), \quad (4)$$

where  $\varepsilon_V$  is a stationary Gaussian process with mean

zero and covariances

$$\text{cov}(\varepsilon_V(t), \varepsilon_V(t-u)) = \begin{cases} \mathbf{\Omega}(u) & \text{if } |u| \leq q \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

for some  $q \in \mathbb{N}$ . The distribution of these processes can be parametrized by the vector  $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\omega})$  with

$$\begin{aligned} \boldsymbol{\phi} &= \text{vec}(\boldsymbol{\Phi}(1), \dots, \boldsymbol{\Phi}(p)) \\ \boldsymbol{\omega} &= (\text{vech}(\boldsymbol{\Omega}(0))', \text{vec}(\boldsymbol{\Omega}(1), \dots, \boldsymbol{\Omega}(q))')', \end{aligned}$$

where as usual the  $\text{vec}$  operator stacks the columns of a matrix and the  $\text{vech}$  operator stacks only the elements contained in the lower triangular submatrix. The parameter  $\boldsymbol{\phi}$  corresponds to the autoregressive structure of the process  $\mathbf{X}_V$  while  $\boldsymbol{\omega}$  parametrizes the dependence structure of the error process  $\varepsilon_V$ .

**Theorem 4.1.** *Let  $\mathbf{X}_V$  be a stationary Gaussian process of the form (4) and (5). Furthermore, let  $G = (V, E)$  be a mixed graph such that:*

- (i) if  $a \rightarrow b \notin E$  then  $\Phi_{ba}(u) = 0$  for all  $u > 0$ ;
- (ii) if  $a \dashrightarrow b \notin E$  then  $\Omega_{ba}(u) = 0$  for all  $u > 0$ ;
- (iii) if  $a \dashdash b \notin E$  then  $\Omega_{ba}(0) = \Omega_{ab}(0) = 0$ .

Then  $\mathbf{X}_V$  is Granger-Markov for  $G$ .

*Proof.* For a sketch of the proof, we regard  $\mathbf{X}_V$  as a linear equation system with correlated errors. Similar to Koster (1999), it can be shown that this system is Markov for the path diagram  $G'$  by representing each variable  $X_v(t)$  as a separate node  $v_t$  and inserting edges  $a_{t-u} \rightarrow b_t$  and  $a_{t-u} \leftrightarrow b_t$  whenever the corresponding parameters  $\Phi_{ba}(u)$  and  $\Omega_{ba}(u)$ , respectively, are not constrained to zero. The two path diagrams  $G$  and  $G'$  are related by

$$\begin{aligned} a \rightarrow b \notin E &\Leftrightarrow a_{t-u} \rightarrow b_t \notin E' \quad \forall u \in \mathbb{N} \forall t \in \mathbb{Z}, \\ a \dashrightarrow b \notin E &\Leftrightarrow a_{t-u} \leftrightarrow b_t \notin E' \quad \forall u \in \mathbb{N} \forall t \in \mathbb{Z}, \\ a \dashdash b \notin E &\Leftrightarrow a_t \leftrightarrow b_t \notin E' \quad \forall t \in \mathbb{Z}. \end{aligned}$$

Noting that every  $b$ -pointing path from  $a$  to  $b$  in the general path diagram  $G$  corresponds to a path from  $\{a_{t-u} : u \in \mathbb{N}\}$  to  $b_t$  in the path diagram  $G'$  (and similar for bi-pointing paths), the result follows from the Markov property for  $G'$ .  $\square$

#### 4.1 IDENTIFIABILITY

In order to prove that the parameters of the model are identifiable if the associated graph is a dMAG, we first derive equations that are of similar form as the Yule-Walker equations for vector autoregressions (e.g., Brockwell and Davis 1991).

Multiplying the model equation (4) by  $\mathbf{X}_V(t-h)'$ , we obtain

$$\mathbf{X}_V(t)\mathbf{X}_V(t-h)' = \sum_{k=1}^p \boldsymbol{\Phi}(k)\mathbf{X}_V(t-k)\mathbf{X}_V(t-h)' + \varepsilon_V(t)\mathbf{X}_V(t-h)' \quad (6)$$

Since the assumptions imply that the characteristic polynomial  $\boldsymbol{\Phi}(z)$  is invertible,  $\mathbf{X}_V$  has a moving average representation

$$\mathbf{X}_V(t-h) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}(k)\varepsilon_V(t-h-k),$$

where  $\boldsymbol{\Psi}(0) = \mathbf{I}$ . Substituting into (6) and taking expectations, we obtain

$$\boldsymbol{\Gamma}(h) = \sum_{k=1}^p \boldsymbol{\Phi}(k)\boldsymbol{\Gamma}(h-k) + \sum_{k=0}^{q-h} \boldsymbol{\Omega}(h+k)\boldsymbol{\Psi}(k)' \quad (7)$$

for  $u = 0, \dots, p$ , where we have used that  $\boldsymbol{\Omega}(h+k) = 0$  if  $k > p-h$ . We note that these equations are similar to the Yule-Walker equations for autoregressive processes.

By condition (3), it follows that for any dashed directed edge  $a \dashrightarrow b$  in  $G$  the vertex  $a$  cannot be an ancestor of  $b$ ,  $a \notin \text{an}(b)$ . Furthermore, the child-parent relations among the vertices correspond to the nonzero non-diagonal entries in the matrices  $\boldsymbol{\Phi}(k)$ . Since the matrix  $\boldsymbol{\Psi}(z)$  is the inverse of  $\boldsymbol{\Phi}(z) = \mathbf{I} - \boldsymbol{\Phi}(1)z - \dots - \boldsymbol{\Phi}(p)z^p$ , it follows by a geometric series expansion that  $a \notin \text{an}(b)$  implies  $\Psi_{ba}(u) = 0$  for all  $u \in \mathbb{N}$ . Summarizing we find that

$$\Omega_{ba}(k) \neq 0 \text{ for some } k \in \mathbb{N} \Rightarrow \Psi_{ba}(u) = 0 \quad \forall u \in \mathbb{N}.$$

Let  $\text{ps}(b) = \{v \in V | v \dashrightarrow b\}$  by the set of past spouses of  $b$ . Then if  $a \rightarrow b \in E$  we have for  $h = 1, \dots, p$

$$\begin{aligned} \Gamma_{ba}(h) &= \sum_{u=1}^p \boldsymbol{\Phi}_{b\text{pa}(b)}(u)\boldsymbol{\Gamma}_{\text{pa}(b)a}(h-u) \\ &\quad + \sum_{u=0}^{q-h} \boldsymbol{\Omega}_{b\text{ps}(b)}(h+u)\boldsymbol{\Psi}_{a\text{ps}(b)}(u) \\ &= \sum_{u=1}^p \boldsymbol{\Phi}_{b\text{pa}(b)}(u)\boldsymbol{\Gamma}_{\text{pa}(b)a}(h-u). \end{aligned}$$

Hence, we get the equation system for  $\boldsymbol{\phi}$

$$\boldsymbol{\gamma}_\phi = \mathbf{P}_\phi(\boldsymbol{\Gamma} \otimes \mathbf{I})\mathbf{P}'_\phi \cdot \mathbf{P}_\phi\boldsymbol{\phi}$$

where  $\mathbf{P}_\phi$  is the projection onto the unconstrained parameters in  $\boldsymbol{\phi}$ ,  $\boldsymbol{\gamma}_\phi = \mathbf{P}_\phi \text{vec}(\boldsymbol{\Gamma}(1), \dots, \boldsymbol{\Gamma}(p))$  and  $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}(u-v))_{u,v=1,\dots,p}$ . The equation system uniquely determines the unconstrained parameters in  $\boldsymbol{\phi}$ . Next, noting that  $\boldsymbol{\Psi}(0) = \mathbf{I}$ , the parameters  $\boldsymbol{\Omega}(1), \dots, \boldsymbol{\Omega}(q)$  can be determined from the equations by solving (7) iteratively for  $\boldsymbol{\Omega}(k)$ ,  $k = q, \dots, 1$ :

$$\begin{aligned} \boldsymbol{\Omega}_{b\text{ps}(b)}(h) &= \boldsymbol{\Gamma}_{b\text{sp}(b)}(h) \\ &\quad - \sum_{k=1}^p \boldsymbol{\Phi}_{b\text{pa}(b)}(k)\boldsymbol{\Gamma}_{\text{pa}(b)\text{ps}(b)}(h-k) \\ &\quad - \sum_{k=h+1}^q \boldsymbol{\Omega}_{b\text{ps}(b)}(k)\boldsymbol{\Psi}_{\text{ps}(b)\text{ps}(b)}(k-h)'. \end{aligned}$$

Finally,  $\boldsymbol{\Omega}(0)$  can be determined similarly.

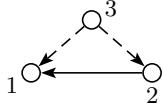


Figure 4: General path diagram for the trivariate process defined in (9) and (10).

It follows that under condition (3) the model parameters  $\phi$  and  $\omega$  are uniquely determined by the covariance function  $\Gamma(u)$  of the process, which implies that the parameters are identifiable.

#### 4.2 RELATION TO MULTIVARIATE ARMA MODELS

In the following, we briefly show that the models introduced in the previous section can be viewed as graphical multivariate ARMA models that satisfy the conditional independence constraints encoded by general path diagrams. We note that under the assumptions the error process  $\varepsilon$  has a moving average representation. Since the covariance function  $\Omega(u)$  vanishes for  $|u| > q$ , it follows that the moving average representation is of order  $q$  and we have

$$\varepsilon_V(t) = \sum_{u=0}^q \Theta(u) \eta_V(t-u)$$

where  $\Theta(0) = \mathbf{I}$  and  $\eta_V$  is a Gaussian white noise process with mean zero and covariance matrix  $\Sigma$ . The coefficients of this moving average representation are uniquely determined by the equation system

$$\Omega(v) = \sum_{u=0}^{q-v} \Theta(u)' \Sigma \Theta(u+v), \quad (8)$$

which can be iteratively solved for  $\Theta(u)$  and  $\Sigma$  (Tunnicliffe Wilson 1972). It follows that the process  $\mathbf{X}_V$  can be represented as a multivariate ARMA( $p, q$ ) process

$$\begin{aligned} \mathbf{X}_V(t) &= \sum_{u=1}^p \Phi(u) \mathbf{X}_V(t-u) \\ &\quad + \sum_{u=1}^q \Theta(u) \eta_V(t-u) + \eta_V(t). \end{aligned}$$

We note that, because of (8), the zero constraints on the matrices  $\Omega(u)$  do not translate into equally simple constraints on the parameters  $\Theta(1), \dots, \Theta(q)$  and  $\Sigma$ .

#### 4.3 EXAMPLE

As an illustration, we consider the trivariate process  $\mathbf{X}_V$  given by

$$\begin{aligned} X_1(t) &= \phi X_2(t-1) + \bar{\varepsilon}_1(t), \\ X_2(t) &= \bar{\varepsilon}_2(t), \\ X_3(t) &= \bar{\varepsilon}_3(t) \end{aligned} \quad (9)$$

with error processes

$$\begin{aligned} \bar{\varepsilon}_1(t) &= \alpha \varepsilon_4(t-2) + \varepsilon_1(t), \\ \bar{\varepsilon}_2(t) &= \delta \varepsilon_5(t-2) + \varepsilon_2(t), \\ \bar{\varepsilon}_3(t) &= \beta \varepsilon_4(t-1) + \gamma \varepsilon_5(t-1) + \varepsilon_1(t), \end{aligned} \quad (10)$$

where  $\varepsilon_1(t), \dots, \varepsilon_5(t)$  are independent and identically distributed with mean 0 and variance  $\sigma^2$ . The processes  $\varepsilon_4$  and  $\varepsilon_5$  can be viewed as latent variables acting as common confounders for  $X_3$  and  $X_1$  respectively for  $X_3$  and  $X_2$ . The process  $\mathbf{X}_{\{1,2,3\}}$  belongs to the graphical model associated with the general path diagram  $G$  in Fig. 4. The model has orders  $p = 1$  and  $q = 1$  and parameters given by

$$\begin{aligned} \Phi(1) &= \begin{pmatrix} 0 & \phi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \Omega(0) &= \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{33} \end{pmatrix}, \\ \Omega(1) &= \begin{pmatrix} 0 & 0 & \omega_{13} \\ 0 & 0 & \omega_{23} \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

In particular, we have for the above process  $\omega_{11} = (\alpha^2 + 1)\sigma^2$ ,  $\omega_{22} = (\beta^2 + \gamma^2 + 1)\sigma^2$ ,  $\omega_{33} = (\delta^2 + 1)\sigma^2$ ,  $\omega_{13} = \alpha\beta\sigma^2$ , and  $\omega_{23} = \gamma\delta\sigma^2$ .

As mentioned in the previous subsection, the process can also be expressed in the usual ARMA parametrization,

$$\mathbf{X}_V(t) = \Phi \mathbf{X}_V(t-1) + \Theta \varepsilon_V(t-1) + \varepsilon_V(t)$$

with  $\text{var}(\varepsilon_V(t)) = \Sigma$ . Simple calculations show that the ARMA parameters of  $\mathbf{X}_V$  are given by

$$\begin{aligned} \Phi &= \begin{pmatrix} 0 & \phi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \Theta &= \begin{pmatrix} 0 & 0 & \theta_{13} \\ 0 & 0 & \theta_{23} \\ 0 & 0 & 0 \end{pmatrix}, \\ \Sigma &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \end{aligned}$$

with the additional restriction that

$$\sigma_{12} = -\theta_{13}\theta_{23}\sigma_{33}.$$

Unlike the dMAG parametrization, the ARMA model has seven parameters which are restricted by one nonlinear constraint.

## 5 CONCLUDING REMARKS

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

In this paper, we have proposed a new graphical modelling approach that allows to fit models under conditional independence constraints typical for systems affected by latent variables. These models can be regarded as time series versions of linear equation systems with correlated errors and are equivalent to multivariate ARMA models under the conditional independence relations imposed by the associated path diagram. The aim of this approach is to learn the dependence structure of a process by fitting such models; the graphical representations associated with the best fitting models then allow conclusions about the possible causal structures and the presence of latent variables.

We have shown that the model parameters are identifiable if the associated path diagram is an ancestral graph. Although every general path diagram is Markov equivalent to an ancestral graph, such path diagrams in general impose additional non-conditional-independence constraints. Therefore, it is of interest to extend the identifiability result to more general classes such as, for example, bow-free path diagrams (e.g., Brito and Pearl 2002, Drton et al. 2009).

We have not yet addressed the problem of parameter estimation and model selection. Having a parametric model associated with a particular causal structure, identification can be accomplished by minimization of a model selection criterion such as AIC or BIC. For estimation of the parameters, we could substitute empirical covariances into the Yule-Walker type equations that have been used for the proof of identifiability. The resulting method of moment estimators could be easily implemented. However, it is well known (e.g. Dahlhaus 1988, Drton et al. 2009) that such method of moment often are inefficient and may even violate model restrictions such as positive definiteness of covariance matrices or stability of a process. In a current project, maximum likelihood estimation for the parameters is investigated. As the number of possible models for the examination of causal structures is large even for moderate dimensions, fast convergence is crucial for successful application of the method; one possible candidate is the iterative conditional fitting for fitting ancestral graph models (Drton and Richardson 2004).

## References

- Boyen, X., Friedman, N. and Koller, D. (1999). Discovering the hidden structure of complex dynamic systems. *Proceedings of the 15<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 91–100.
- Bruto, C. and Pearl, J. (2002). A new identification condition for recursive models with correlated errors. *Structural Equation Modeling* **9**, 59–474.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd edn, Springer, New York.
- Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies - Models, Analysis and Interpretation*. Chapman & Hall, London.
- Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Annals of Statistics* **18**, 808–841.
- Drton, M., Eichler, M. and Richardson, T. S. (2009). Computing maximum likelihood estimates in recursive linear models with correlated errors. *Journal of Machine Learning Research* **10**, 2329–2348.
- Drton, M. and Richardson, T. S. (2004). Iterative conditional fitting for Gaussian ancestral graph models. In M. Chickering and J. Halpern (eds), *Proceedings of the 20<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, pp. 130–137.
- Eichler, M. (2001). Graphical modelling of multivariate time series. *Technical report*, Universität Heidelberg. (arXiv:math.ST/0610654).
- Eichler, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of The Royal Society B* **360**, 953–967.
- Eichler, M. (2006). Graphical modelling of dynamic relationships in multivariate time series. In M. Winterhalder, B. Schelter and J. Timmer (eds), *Handbook of Time Series Analysis*, Wiley-VCH, pp. 335–372.
- Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* **137**, 334–353.
- Goldberger, A. S. (1972). Structural equation models in the social sciences. *Econometrica* **40**, 979–1001.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.
- Granger, C. W. J. (1980). Testing for causality, a personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329–352.
- Hsiao, C. (1982). Autoregressive modeling and causal ordering of econometric variables. *Journal of Economic Dynamics and Control* **4**, 243–259.
- Koster, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics* **26**, 413–431.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Lauritzen, S. L. (2001). Causal inference from graphical models. In O. E. Barndorff-Nielsen, D. R. Cox and C. Klüppelberg (eds), *Complex stochastic systems*, CRC Press, London, pp. 63–107.
- Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika* **82**, 669–710.
- Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge, UK.
- Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics* **30**, 145–157.
- Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Annals of Statistics* **30**, 962–1030.
- Spirtes, P., Richardson, T. S., Meek, C., Scheines, R. and Glymour, C. (1998). Using path diagrams as a structural equation modelling tool. *Soc. Methods Res.* **27**, 182–225.
- Tunnicliffe Wilson, G. (1972). The factorization of matrix spectral densities. *SIAM J. Appl. Math.* **23**, 420–426.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* **5**, 161–215.