Journal of Multivariate Analysis 107 (2012) 90–103

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Graphical models for multivariate Markov chains

R. Colombi [a], S. Giordano [b,*]

[a] Dipartimento di Ingegneria dell'Informazione e Metodi Matematici, Università di Bergamo, Viale G. Marconi, 5, 24044 Dalmine - Bergamo, Italy

[b] Dipartimento di Economia e Statistica, Università della Calabria, cubo 0C, 87036 Arcavacata di Rende - Cosenza, Italy

ARTICLE INFO

Article history:
Received 17 June 2009
Available online 10 January 2012

AMS subject classifications:
68R10
60J10
62H99

Keywords:
Granger noncausality
Conditional independence
Generalized marginal interactions

ABSTRACT

The aim of this paper is to provide a graphical representation of the dynamic relations among the marginal processes of a first order multivariate Markov chain. We show how to read Granger-noncausal and contemporaneous independence relations off a particular type of mixed graph, when directed and bi-directed edges are missing. Insights are also provided into the Markov properties with respect to a graph that are retained under marginalization of a multivariate chain. Multivariate logistic models for transition probabilities are associated with the mixed graphs encoding the relevant independencies. Finally, an application on real data illustrates the methodology.

1. Introduction

The identification of the existence of proper dynamic relations among variables, simultaneously observed over time, is a revealing task in many areas.

Basically, two types of dependence relations in multivariate time series models are relevant: the dependence of the present of a subset of variables on the past of all the variables, and the contemporaneous association among variables at any time point that cannot be ruled out by conditioning on the past.

In this paper, we address the use of graphical models for the analysis of the dynamic relations among the marginal processes of a time-homogeneous first order multivariate Markov chain. We employ a mixed graph, whose nodes represent the univariate marginal processes of the Markov chain and whose directed and bi-directed edges describe the dependence structure among them. The approach adopted here enables us to interpret the lack of directed edges as Granger-noncausal relationships while the missing bi-directed edges are used to visualize the contemporaneous independence relations between the marginal processes of the chain. The transition probabilities of the multivariate Markov chain are required to obey the set of Markov properties implied by such a graph and a multivariate logistic parameterization for the transition probabilities that satisfy these Markov properties is provided. We also present the conditions that ensure Granger noncausality, contemporaneous independence and Markovian features to be preserved by the marginal processes of a multivariate chain.

A similar graphical approach was used by Eichler [12,15] to describe the dynamic structure of multivariate time series with autoregressive representation, while we basically restrict ourselves to the case of multivariate Markov chain models. We believe that graphical models for multivariate Markov chains are worth examination for different reasons: first because Markov chain models are basic tools in modeling categorical multidimensional time series; moreover, under the assumption of Markovianity, Granger noncausal and contemporaneous independence relationships are remarkably simplified and satisfy relevant properties; and, finally, providing a suitable parameterization of the transition probabilities which meet the Markov properties of noncausal and contemporaneous independence is a nontrivial task and useful in applications.

* Corresponding author.
E-mail addresses: roberto.colombi@unibg.it (R. Colombi), sabrina.giordano@unical.it (S. Giordano).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.01.010

The paper is organized as follows. In Section 2, we introduce some basic notation concerning multivariate Markov chains, while the key definitions of Granger noncausality and contemporaneous independence are set out in Section 3.

In Section 4, the features of mixed graphs are described; how they can be used to represent independence relations among the components of a multivariate Markov chain is shown and the definition of multivariate Markov chains, satisfying the conditions of Granger noncausality and contemporaneous independence with respect to a mixed graph is given. In Section 5, we point out some results on causal and contemporaneous independence properties in the framework of Markov chains. Section 6 provides a parameterization for the transition probabilities that meet the independencies described in Section 4. In Section 7, we discuss when a marginal process of a multivariate Markov chain is Markov with respect to a mixed subgraph. Different approaches through which certain alternative conditional independencies can be read off the graph are briefly described in Section 8. Finally, an example on real data concludes the work and the Appendix contains technical proofs.

A notation used throughout this paper is that of conditional independence [9], that is, we write $X \perp\!\!\!\perp Y|W$ when the random variables X and Y are independent once the value of a third variable W is given. We mention two basic properties [24] of the conditional independence relation which will be used in later proofs: *contraction* and *intersection* properties. The *contraction* property states that $X \perp\!\!\!\perp Z|W$ and $X \perp\!\!\!\perp Y|ZW$ are equivalent to $X \perp\!\!\!\perp YZ|W$. Under the assumption of a strictly positive probability function, the intersection property asserts that $X \perp\!\!\!\perp Y|ZW$ and $X \perp\!\!\!\perp Z|YW$ hold if and only if $X \perp\!\!\!\perp YZ|W$ is valid.

Most of the later results are proved by applying the intersection property under the strong restriction of strictly positive transition probabilities even if such property holds under more general conditions [30] as well. However, our restrictive choice is convenient for simplifying the proofs and it is essential to use the parameterization introduced in Section 6.

2. Basic notation for multivariate Markov chains

Given a set of integers $\mathcal{V} = \{1, \ldots, q\}$, let $\mathbf{A}_{\mathcal{V}} = \{A_{\mathcal{V}}(t) : t \in \mathbb{N}\} = \{A_j(t) : t \in \mathbb{N}, j \in \mathcal{V}\}$ be a time-homogeneous first order multivariate Markov chain (MMC), in a discrete time interval $\mathbb{N} = \{0, 1, 2, \ldots\}$.

For all $t \in \mathbb{N}$, $A_{\mathcal{V}}(t) = \{A_j(t) : j \in \mathcal{V}\}$ is a discrete random vector with each element $A_j(t)$ taking on values in a finite set $\mathcal{A}_j = \{a_{j1}, \ldots, a_{js_j}\},\ j \in \mathcal{V}$.

For every $\mathcal{S} \subset \mathcal{V}$, a marginal process of the chain is represented by $\mathbf{A}_{\mathcal{S}} = \{A_{\mathcal{S}}(t) : t \in \mathbb{N}\}$ where $A_{\mathcal{S}}(t) = \{A_j(t) : j \in \mathcal{S}\}$. When $\mathcal{S} = \{j\}$, the univariate marginal process is indicated as $\mathbf{A}_j,\ j \in \mathcal{V}$.

In order to simplify the notation, the history up to time $t-1$ of the MMC is denoted by $\bar{A}_{\mathcal{V}}(t-1) = \{A_{\mathcal{V}}(r) : r \leq t-1\}$.

The first order multivariate Markov chain property

$$A_{\mathcal{V}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{V}}(t-2)|A_{\mathcal{V}}(t-1), \quad t \in \mathbb{N} \setminus \{0, 1\} \tag{1}$$

asserts that $A_{\mathcal{V}}(t)$ is conditionally independent of the remote past $\bar{A}_{\mathcal{V}}(t-2) = \{A_{\mathcal{V}}(r) : r \leq t-2\}$, given the knowledge of the most recent past $A_{\mathcal{V}}(t-1)$.

3. Granger noncausality and contemporaneous independence

A deeper understanding of the joint behavior of the component processes of an MMC requires investigation of both the effect of the past of one marginal process on the present of another and the relation among marginal processes at the same time, given the past of the chain.

Motivated by these considerations, we present the definitions of Granger noncausality, Granger [20], (also G-noncausality hereafter), and contemporaneous independence which play a central role in our work.

Definition 1 (*Granger Noncausality*)**.** Given two disjoint marginal processes $\mathbf{A}_{\mathcal{T}}$ and $\mathbf{A}_{\mathcal{S}}$ of a time series $\mathbf{A}_{\mathcal{V}}$, $\mathbf{A}_{\mathcal{T}}$ is not Granger caused by $\mathbf{A}_{\mathcal{S}}$ with respect to $\mathbf{A}_{\mathcal{V}}$ if and only if the following condition holds for every $t \in \mathbb{N} \setminus \{0\}$

$$A_{\mathcal{T}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{S}}(t-1)|\bar{A}_{\mathcal{V}\setminus\mathcal{S}}(t-1). \tag{2}$$

This condition states that the past of $\mathbf{A}_{\mathcal{S}}$ does not contain additional information on the present of $\mathbf{A}_{\mathcal{T}}$, given the past of the marginal process $\mathbf{A}_{\mathcal{V}\setminus\mathcal{S}}$.

The above definition of Granger noncausality in terms of conditional independence is due to Chamberlain [6]. An in-depth discussion of the Granger causality in the context of time series has been proposed by Eichler [13], who addressed the problem of distinguishing direct causal relationships from spurious causalities due to the presence of latent variables. Furthermore, the concept of G-noncausality in a Markov chain framework has been discussed in the econometric literature by Bouissou et al. [3], Chamberlain [6], Florens et al. [17] and Gouriéroux et al. [19]. For the special case of bivariate binary Markov chains, Mosconi and Seri [27] dealt with causality by allowing the transition probabilities to depend on covariates.

Definition 2 (*Contemporaneous Independence*)**.** Two disjoint marginal processes $\mathbf{A}_{\mathcal{T}}$ and $\mathbf{A}_{\mathcal{S}}$ of a time series $\mathbf{A}_{\mathcal{V}}$ are contemporaneously independent with respect to $\mathbf{A}_{\mathcal{V}}$ if and only if the following restriction holds for every $t \in \mathbb{N} \setminus \{0\}$

$$A_{\mathcal{T}}(t) \perp\!\!\!\perp A_{\mathcal{S}}(t)|\bar{A}_{\mathcal{V}}(t-1). \tag{3}$$

In other words, this definition means that two marginal processes are independent at each time point, given all available past information.

Statements (2) and (3) refer to a general multivariate time series $\mathbf{A}_{\mathcal{V}}$, our aim, however, is to investigate Granger noncausal relations and contemporaneous independencies among marginal processes of a first order multivariate Markov chain.

As we shall discuss in-depth in Section 5, the Markovian assumption simplifies independence restrictions (2) and (3) by reducing the conditioning set to the information at time $t-1$.

A convenient way for dealing with such dynamic relations is to represent them with a graph whose nodes correspond to the univariate marginal processes of the MMC and the edges describe the dependence structure.

In the central part of this work, we will discuss the noncausality and contemporaneous independence restrictions encoded by a particular kind of graph, and a suitable parameterization for the associated models that assess those restrictions on the transition probabilities of the MMC.

4. Mixed graphs

4.1. Basic concepts of mixed graphs

A graph G is defined by a pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the set of edges, connecting pairs of nodes.

Graphical Markov models associate missing edges of a graph with some conditional independence restrictions imposed on a multivariate probability distribution. The rules for reading such conditional independence relations off the graph are known as Markov properties.

Key references for the extensive literature on graphical models are Whittaker [31], Edwards [11], Cox and Wermuth [8] and Lauritzen [24].

The basic problem here lies in finding a graph that represents the G-noncausal and contemporaneous independence relations among the component processes of an MMC.

This is achieved by a particular graphical structure allowing different types of edges, called *mixed graph*. We use the term mixed graph in keeping with the usual terminology of the literature on graphical models. Other mixed graphs featuring different Markov properties are discussed by many authors, i.e. Andersson et al. [1], Cox and Wermuth [8] and Richardson [29], among others. Mixed graphs have been proposed by Eichler [12] in the context of multivariate time series. As in [12], in the mixed graph $G = (\mathcal{V}, \mathcal{E})$ considered here, there exists a one-to-one correspondence between the nodes $j \in \mathcal{V}$ and the univariate marginal processes $\mathbf{A}_j,\ j \in \mathcal{V}$, of the MMC $\mathbf{A}_{\mathcal{V}}$.

A pair of nodes $i, k \in \mathcal{V}$ of the considered mixed graph may be joined by the directed edges $i \rightarrow k$, $i \leftarrow k$, and by the bi-directed edge $i \leftrightarrow k$.

Each pair of distinct nodes $i, k \in \mathcal{V}$ can be connected by up to all the three types of edges. For each single node $i \in \mathcal{V}$, the bi-directed edge $i \leftrightarrow i$ is implicitly introduced and the directed edge $i \rightarrow i$ may or may not be present. The introduction of the edges $i \leftrightarrow i$ is a matter of convenience and simplifies the next definition of *district* of a node. The fact that the self-loop $i \rightarrow i$ may be inserted or not will allow us to take into account the conditional independencies (7) discussed in Section 4.3. In all the examples throughout the paper, these edges will be assumed implicitly inserted even if they are not shown in the graphs.

In the next section, we shall clarify how to interpret the lack of each type of edge by means of the Markov properties associated with the mixed graph.

We now briefly review basic graphical concepts applied to mixed graphs which are needed later in the paper.

If $i \rightarrow k \in \mathcal{E}$, then i is a *parent* of k and k is a *child* of i. The sets $\mathrm{pa}(i) = \{j \in \mathcal{V} : j \rightarrow i \in \mathcal{E}\}$ and $\mathrm{ch}(i) = \{j \in \mathcal{V} : i \rightarrow j \in \mathcal{E}\}$ are the sets of parents and children of $i, i \in \mathcal{V}$, respectively. When $i \leftrightarrow k \in \mathcal{E}$ the nodes i, k are *spouses*. The set of spouses of i is denoted by $\mathrm{sp}(i) = \{j \in \mathcal{V} : i \leftrightarrow j \in \mathcal{E}\}$. Note that i is a spouse of itself ($i \in \mathrm{sp}(i)$) as $i \leftrightarrow i$ is always implicitly present.

A *path* τ is a sequence of edges $e_i, e_i \in \mathcal{E},\ i = 1, \ldots, r$ between the nodes j_{i-1}, j_i, of an ordered set $\{j_0, j_1, \ldots, j_r\}$ of not necessarily distinct nodes of $\mathcal{V}$. The endpoint nodes j_0, j_r must be distinct. A path of only bi-directed edges is a bi-directed path.

The *district* of i, say $\mathrm{dis}(i)$, is the set of nodes connected to i by a bi-directed path, i.e. $\mathrm{dis}(i) = \{j \in \mathcal{V} : j \leftrightarrow \ldots \leftrightarrow i\}$.

Let $\mathcal{S} \subset \mathcal{V}$ be a non-empty subset of nodes, $\mathrm{pa}(\mathcal{S}) = \bigcup_{i \in S} \mathrm{pa}(i)$, $\mathrm{ch}(\mathcal{S}) = \bigcup_{i \in S} \mathrm{ch}(i)$, and $\mathrm{sp}(\mathcal{S}) = \bigcup_{i \in S} \mathrm{sp}(i)$ are the collection of parents, children and spouses of nodes in $\mathcal{S}$.

A set $\mathcal{S} \subseteq \mathcal{V}$ is an ancestral set if $\mathrm{pa}(\mathcal{S}) \subseteq \mathcal{S}$. For any subset $\mathcal{S} \subseteq \mathcal{V}$, $\mathrm{an}(\mathcal{S})$ indicates the smallest ancestral set containing $\mathcal{S}$.

G^b will denote the graph obtained from the mixed graph G by retaining the bi-directed edges and removing all directed edges.

Given a mixed graph $G = (\mathcal{V}, \mathcal{E})$, the induced subgraph $G_{\mathcal{M}}$ has the node set $\mathcal{M} \subset \mathcal{V}$ and its edge set contains every edge of $\mathcal{E}$ connecting nodes of $\mathcal{M}$. A subset $\mathcal{M}$ of $\mathcal{V}$ is *bi-connected* if every pair of its nodes is linked by a bi-directed path in $G_{\mathcal{M}}$ and it is *bi-complete* if every pair of its nodes is linked by a bi-directed edge. The symbol $\mathcal{B}(G)$ will be used to denote the family of bi-connected subsets of $\mathcal{V}$. If $\mathcal{M}$ is bi-connected (bi-complete), $G_{\mathcal{M}}$ is a bi-connected (bi-complete) subgraph of G.

The symbols $pa_{\mathcal{M}}(\mathcal{S})$, $\mathrm{sp}_{\mathcal{M}}(\mathcal{S})$, $\mathrm{ch}_{\mathcal{M}}(\mathcal{S})$, $\mathrm{dis}_{\mathcal{M}}(\mathcal{S})$ will refer to all parents, spouses, children, and the district of $\mathcal{S}$ in $G_{\mathcal{M}}$, $\mathcal{S} \subset \mathcal{M} \subset \mathcal{V}$. Finally, $\mathcal{P}(\mathcal{V})$ will denote the family of all non-empty subsets of the node set $\mathcal{V}$.

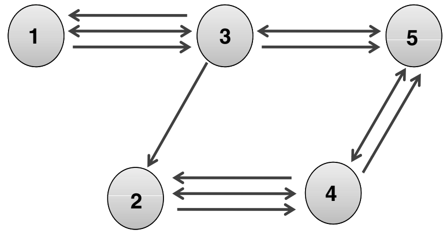

Fig. 1. In this graph, it holds that $\{1\} \bowtie_m \{2\}|\{3\}$ and $\{1\} \bowtie_m \{2\}|\{3, 4\}$ by Lemma 1.

4.2. The m-separation criterion

The Markov properties encoded by the mixed graph in this paper can be described in terms of the well-known notion of m-separation [29,12]. In particular, the next Lemma 1 gives an equivalent formulation of the m-separation which most of our results provided in Section 8 rely on.

We remind the reader that the edges $s \leftrightarrow t$ or $s \rightarrow t$ have an arrowhead at t. Let $G = (\mathcal{V}, \mathcal{E})$ be a mixed graph and τ a path in G with endpoints j_1 and j_r. An intermediate node j_i, $i = 2, \ldots, r-1$, of τ is said to be an m-collider if the edges e_i, e_{i+1} belonging to $\mathcal{E}$ end with an arrowhead at j_i, otherwise it is considered an m-noncollider. A path τ is an m-connecting path, given a set of nodes $\mathcal{C}$, if $\mathcal{C}_\tau \cap \mathcal{C} = \mathcal{C}_\tau$ and $\bar{\mathcal{C}}_\tau \cap \mathcal{C} = \emptyset$, where the set $\mathcal{C}_\tau$ contains the m-colliders and $\bar{\mathcal{C}}_\tau$ denotes the set of the m-noncolliders of τ. Two disjoint subsets of nodes $\mathcal{S}, \mathcal{T}$ of $\mathcal{V}$ are m-separated, given a set $\mathcal{C}$, commonly indicated in short form $\mathcal{S} \bowtie_m \mathcal{T}|\mathcal{C}$, if there are no m-connecting paths, given $\mathcal{C}$, with endpoints $s \in \mathcal{S}$ and $t \in \mathcal{T}$. Hereafter, the prefix m- will be omitted for the sake of simplicity.

The following lemma will play a central role in the proofs in Section 7.

Lemma 1. *If $\mathcal{S}, \mathcal{T}$ are disjoint subsets of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$, then $\mathcal{S}$ and $\mathcal{T}$ are m-separated, given $\mathcal{C}$, i.e. $\mathcal{S} \bowtie_m \mathcal{T}|\mathcal{C}$, if and only if there exist two disjoint subsets $\mathcal{S}'$ and $\mathcal{T}'$ of $\mathcal{V}$ such that $\mathcal{S} \subseteq \mathcal{S}'$, $\mathcal{T} \subseteq \mathcal{T}'$, $\mathcal{M} = \mathcal{S}' \cup \mathcal{T}' \cup \mathcal{C} = \mathrm{an}(\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ and $\mathrm{dis}_{\mathcal{M}}(\mathcal{S}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{S}')) \cap \mathrm{dis}_{\mathcal{M}}(\mathcal{T}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{T}')) = \emptyset$.*

The proof follows from Proposition 2 by Koster [21], Lemma B.1 by Eichler [12] and its converse [14].

Example 1. Let $\mathcal{S} = \{1\}$, $\mathcal{T} = \{2\}$ be two disjoint subsets of the node set of the mixed graph in Fig. 1. They are m-separated, given different sets $\mathcal{C}$. Choosing $\mathcal{S}' = \{1\}$ and $\mathcal{T}' = \{2, 4\}$, the set $\mathcal{C} = \{3\}$, which contains the common parent of the nodes in $\mathcal{S}$ and $\mathcal{T}$, allows $\mathcal{M} = \mathcal{S}' \cup \mathcal{T}' \cup \mathcal{C} = \{1, 2, 3, 4\}$ to be ancestral. Moreover, since the intersection set between $\mathrm{dis}_{\mathcal{M}}(\mathcal{S}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{S}')) = \{1, 3\}$ and $\mathrm{dis}_{\mathcal{M}}(\mathcal{T}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{T}')) = \{2, 4\}$ is empty, it holds that $\{1\} \bowtie_m \{2\}|\{3\}$ by Lemma 1. Alternatively, if we consider $\mathcal{S}' = \mathcal{S} = \{1\}$ and $\mathcal{T}' = \mathcal{T} = \{2\}$, the set $\mathcal{C}$ that leads to the previous ancestral marginal set $\mathcal{M} = \mathcal{S}' \cup \mathcal{T}' \cup \mathcal{C} = \{1, 2, 3, 4\}$ is $\mathcal{C} = \{3, 4\}$, while $\mathrm{dis}_{\mathcal{M}}(\mathcal{S}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{S}')) = \{1, 3\}$ and $\mathrm{dis}_{\mathcal{M}}(\mathcal{T}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{T}')) = \{2, 4\}$ do not change and have an empty intersection. In this case, however, it holds that $\{1\} \bowtie_m \{2\}|\{3, 4\}$.

4.3. Markov properties of mixed graphs

Markov properties link sets of G-noncausality and contemporaneous independence restrictions with missing directed and bi-directed edges of mixed graphs, respectively.

In particular, missing bi-directed edges lead to independencies concerning marginal processes at the same point of time; missing directed edges, instead, refer to independencies which involve marginal processes at two consecutive instants.

This is expressed formally in the definition below that supplies a key idea of our work.

Definition 3 (*MMC is Markov w.r.t. a Graph*)**.** A multivariate Markov chain is Markov with respect to a mixed graph G if and only if its transition probabilities satisfy the following conditional independencies for all $t \in \mathbb{N} \setminus \{0\}$

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1) | A_{\mathrm{pa}(\mathcal{S})}(t-1) \quad \mathcal{S} \in \mathcal{P}(\mathcal{V}) \tag{4}$$

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) | A_{\mathcal{V}}(t-1) \quad \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{5}$$

Condition (4) is equivalent, for all $t \in \mathbb{N} \setminus \{0\}$, to the two statements

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S}) \cup \mathcal{S}}(t-1) | A_{\mathrm{pa}(\mathcal{S}) \cup \mathcal{S}}(t-1) \quad \mathcal{S} \in \mathcal{P}(\mathcal{V}) \tag{6}$$

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{S}\setminus \mathrm{pa}(\mathcal{S})}(t-1) | A_{\mathrm{pa}(\mathcal{S})}(t-1) \quad \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{7}$$

In the next section, we will see that conditions (4) and (5) are coherent with the general definitions of G-noncausality and contemporaneous independence provided in Section 3, although they include only the variables observed at the preceding time point in the conditioning set.

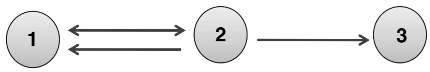

Fig. 2. The graph encodes the independence relations $\mathbf{A}_{\{1,2\}} \nleftrightarrow \mathbf{A}_3$; $\mathbf{A}_1 \nrightarrow \mathbf{A}_3$; $\mathbf{A}_{\{1,3\}} \nrightarrow \mathbf{A}_2$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_1$; $\mathbf{A}_1 \nrightarrow \mathbf{A}_{\{2,3\}}$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}$.

For all $\mathcal{S}' \in \mathcal{P}(\mathcal{V})$, where $\mathcal{S}' \cap (\mathrm{pa}(\mathcal{S}) \cup \mathcal{S}) = \emptyset$, (6) implies $A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{S}'}(t-1) | A_{\mathcal{V}\setminus\mathcal{S}'}(t-1)$ which means that the most recent past of $\mathbf{A}_{\mathcal{S}'}$ is irrelevant for predicting $\mathbf{A}_{\mathcal{S}}$ once the most recent past of $\mathbf{A}_{\mathcal{V}\setminus\mathcal{S}'}$ is known. In this case, it is usual to say that $\mathbf{A}_{\mathcal{S}'}$ does not G-cause $\mathbf{A}_{\mathcal{S}}$ with respect to $\mathbf{A}_{\mathcal{V}}$. Thus, $\mathrm{pa}(\mathcal{S})$ identifies the maximal marginal process $\mathbf{A}_{\mathcal{V}\setminus\mathrm{pa}(\mathcal{S})\cup\mathcal{S}}$ which does not G-cause $\mathbf{A}_{\mathcal{S}}$ with respect to $\mathbf{A}_{\mathcal{V}}$.

Whenever $\mathcal{S} \setminus \mathrm{pa}(\mathcal{S}) \neq \emptyset$, which implies that a number of self-loops $i \to i$ ($i \in \mathcal{S}$) are missing, statement (7) concerns the case of variables at the current time-point which are not affected by their immediate past.

Henceforth, we will refer to (4) with the term *Granger noncausality* condition for MMCs saying that $\mathbf{A}_{\mathcal{S}}$ is not G-caused by $\mathbf{A}_{\mathcal{V}\setminus\mathrm{pa}(\mathcal{S})}$ with respect to $\mathbf{A}_{\mathcal{V}}$, and use the shorthand notation $\mathbf{A}_{\mathcal{V}\setminus\mathrm{pa}(\mathcal{S})} \nrightarrow \mathbf{A}_{\mathcal{S}}$.

On the other hand, condition (5) is a restriction on marginal transition probabilities because it does not involve the marginal processes $\mathbf{A}_j : j \in \mathrm{sp}(\mathcal{S}) \setminus \mathcal{S}$, at time t and, more precisely, it states that the transition probabilities must satisfy the bi-directed Markov property [29] with respect to the graph obtained by removing the directed edges from G. Here, we will refer to (5) with the term *contemporaneous independence* condition for MMCs using a shorthand notation $\mathbf{A}_{\mathcal{S}} \nleftrightarrow \mathbf{A}_{\mathcal{V}\setminus\mathrm{sp}(\mathcal{S})}$, and say that $\mathbf{A}_{\mathcal{S}}$ and $\mathbf{A}_{\mathcal{V}\setminus\mathrm{sp}(\mathcal{S})}$ are contemporaneously independent with respect to $\mathbf{A}_{\mathcal{V}}$.

From the above definition it follows that, if an MMC is Markov with respect to a mixed graph G, the lack in G of a directed edge from node i to k, ($i, k \in \mathcal{V}$), implies the independence of the present of the univariate marginal process $\mathbf{A}_k$ from the immediate past of $\mathbf{A}_i$ given the most recent past of $\mathbf{A}_{\mathcal{V}\setminus\{i\}}$, that is, for all $t \in \mathbb{N} \setminus \{0\}$

$$i \to k \notin \mathcal{E} \Longrightarrow A_k(t) \perp\!\!\!\perp A_i(t-1) | A_{\mathcal{V}\setminus\{i\}}(t-1). \tag{8}$$

From Definition 3, moreover, if an MMC is Markov with respect to a mixed graph G, a missing bi-directed edge between i and k implies that the corresponding marginal processes are contemporaneously independent, given the recent past of the MMC, that is, for all $t \in \mathbb{N} \setminus \{0\}$

$$i \leftrightarrow k \notin \mathcal{E} \Longrightarrow A_i(t) \perp\!\!\!\perp A_k(t) | A_{\mathcal{V}}(t-1). \tag{9}$$

The conditional independencies (8) and (9) are interpretable in terms of pairwise Granger noncausality and contemporaneous independence conditions, respectively. The more general noncausal and contemporaneous independence statements of Definition 3 are needed because the pairwise restrictions, associated with missing edges, are not, in general, sufficiently strong for the encoding of all the Granger noncausal relations and contemporaneous independence properties among the components of an MMC. This follows from the fact that the composition property [12] does not hold in the context of multivariate Markov chains. For example, the condition

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_i(t-1) | A_{\mathcal{V}\setminus\{i\}}(t-1), \quad t \in \mathbb{N} \setminus \{0\}$$

is not equivalent to

$$A_k(t) \perp\!\!\!\perp A_i(t-1) | A_{\mathcal{V}\setminus\{i\}}(t-1), \quad t \in \mathbb{N} \setminus \{0\},\ k \in \mathcal{S},$$

which means that the G-noncausality for the joint process $\mathbf{A}_{\mathcal{S}}$ is not equivalent to the G-noncausality for all the univariate processes $\mathbf{A}_k,\ k \in \mathcal{S}$.

Note that in [12, Lemma 2.3] concerning stationary Gaussian processes the composition property holds and the implications in (8) and (9) can be replaced by equivalences (see also the comments in Section 5 of Eichler [15]).

Example 2. The graph in Fig. 2 displays the contemporaneous independence relation $\mathbf{A}_{\{1,2\}} \nleftrightarrow \mathbf{A}_3$ and the G-noncausal restrictions: $\mathbf{A}_1 \nrightarrow \mathbf{A}_3$; $\mathbf{A}_{\{1,3\}} \nrightarrow \mathbf{A}_2$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_1$; $\mathbf{A}_1 \nrightarrow \mathbf{A}_{\{2,3\}}$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}$. Besides, the pairwise G-noncausality conditions associated to the missing directed edges in the graph are: $\mathbf{A}_1 \nrightarrow \mathbf{A}_2$; $\mathbf{A}_1 \nrightarrow \mathbf{A}_3$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_1$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_2$. If we consider, for example, that relations $\mathbf{A}_3 \nrightarrow \mathbf{A}_1$; $\mathbf{A}_3 \nrightarrow \mathbf{A}_2$ are not equivalent to $\mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}$, it immediately becomes evident that the pairwise conditions do not imply the more general causal restriction (4). Similarly, the pairwise contemporaneous independence conditions $\mathbf{A}_1 \nleftrightarrow \mathbf{A}_3$, $\mathbf{A}_2 \nleftrightarrow \mathbf{A}_3$ do not imply $\mathbf{A}_{\{1,2\}} \nleftrightarrow \mathbf{A}_3$ given by (5).

The previous example helps us to clarify that an MMC can exhibit more independencies than those encoded by a single mixed graph and that an MMC can be Markov with respect to more than one mixed graph. In fact, if the pairwise G-noncausal relationships $\mathbf{A}_3 \nrightarrow \mathbf{A}_1$, $\mathbf{A}_3 \nrightarrow \mathbf{A}_2$ hold but $\mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}$ is not true, then the MMC is not Markov with respect to the graph illustrated in Fig. 2, while it is Markov with respect to the graphs obtained by adding to Fig. 2 the edge $3 \to 1$ or $3 \to 2$. Clearly, these graphs do not encode all the independencies satisfied by the MMC. This simply happens since the pairwise conditions (8), (9) and the conditions of Definition 3 are not equivalent.

5. Main results on G-noncausality and contemporaneous independence in MMC

The lemma and the proposition of this section are useful to clarify why the Markov properties (4) and (5), associated with a mixed graph, are to be considered as Granger noncausality and contemporaneous independence restrictions in the context of multivariate Markov chains.

Lemma 2. *Under the assumption (1) of first order MMC, the conditional independence statements (4) and (5) are equivalent to the restrictions*

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t), A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1), \bar{A}_{\mathcal{V}}(t-2) | A_{\mathrm{pa}(\mathcal{S})}(t-1) \quad t \in \mathbb{N} \setminus \{0, 1\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}) \tag{10}$$

$$A_{\mathcal{S}}(1) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(1), A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(0) | A_{\mathrm{pa}(\mathcal{S})}(0). \tag{11}$$

Proof. According to the contraction property, conditions (4) and (5) may be equivalently expressed by

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t), A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1) | A_{\mathrm{pa}(\mathcal{S})}(t-1) \quad t \in \mathbb{N} \setminus \{0\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}) \tag{12}$$

which, for $t = 1$, gives (11). Moreover, from the Markovianity assumption (1), it is easy to deduce

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{V}}(t-2) | A_{\mathrm{pa}(\mathcal{S})}(t-1), A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1), A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) \quad t \in \mathbb{N} \setminus \{0, 1\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{13}$$

The proof is now complete as the equivalence between (10) and (12, 13) follows for $t \in \mathbb{N} \setminus \{0, 1\}$ by applying the contraction property, and the converse immediately holds since (10, 11) imply both (4) and (5). □

It is important to observe that conditions (10, 11) imply the statements of G-noncausality and contemporaneous independence which involve all the past history of the process

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1) | \bar{A}_{\mathrm{pa}(\mathcal{S})}(t-1) \quad t \in \mathbb{N} \setminus \{0\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}), \tag{14}$$

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) | \bar{A}_{\mathcal{V}}(t-1) \quad t \in \mathbb{N} \setminus \{0\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{15}$$

The principal advantage of the Markov assumption on $\mathbf{A}_{\mathcal{V}}$ is that it allows conditions (14) and (15) to be equivalent to restrictions (4) and (5) which involve only a finite set of conditioning variables. Proposition 1 will prove this general equivalence.

Proposition 1. *Under assumption (1) of first order Markovianity and the assumption of strictly positive transition probabilities, conditions (4) and (5) are equivalent to (14) and (15).*

Proof. The results of Lemma 1 entail that conditions (4) and (5) are sufficient to yield (14) and (15). To prove the converse of the previous lemma, we need to recall that the restriction

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{V}}(t-2) | A_{\mathrm{pa}(\mathcal{S})}(t-1), A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1), \quad t \in \mathbb{N} \setminus \{0, 1\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}) \tag{16}$$

holds by the Markovianity assumption, whereas (14) implies

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1) | A_{\mathrm{pa}(\mathcal{S})}(t-1), \bar{A}_{\mathcal{V}}(t-2) \quad t \in \mathbb{N} \setminus \{0, 1\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{17}$$

Since the transition probabilities are strictly positive, the intersection property enables us to write (16) and (17) in an equivalent expression

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t-1), \bar{A}_{\mathcal{V}}(t-2) | A_{\mathrm{pa}(\mathcal{S})}(t-1) \quad t \in \mathbb{N} \setminus \{0, 1\},\ \mathcal{S} \in \mathcal{P}(\mathcal{V}). \tag{18}$$

If $t > 1$, the contraction property ensures that (15) and (18) are equivalent to (10), hence (4) and (5) hold. It also becomes clear that, for $t = 1$, (4) and (5) are identical to (14) and (15), respectively. This completes the proof. □

Proposition 1 is related to Theorems 3.1bis and 3.2bis provided by Florens et al. [17], who thoroughly examined, in the framework of Markov chains, the relations between the Granger causality conditions defined by conditioning either on all the past history of the involved variables or on their immediate past only. Nevertheless, Proposition 1 differs from the results by Florens et al., since it also considers conditions of contemporaneous independence and does not take into account the marginal Markovianity of the noncaused process. This aspect will be further discussed in Propositions 3 and 4.

6. A multivariate logistic model for transition probabilities

We remind the reader that $\mathcal{I} = \times_{j \in \mathcal{V}} \mathcal{A}_j$ is the joint state space. For a pair of states $\boldsymbol{i} \in \mathcal{I}, \boldsymbol{i}' \in \mathcal{I}$, the time-homogeneous joint transition probabilities are denoted by $p(\boldsymbol{i}|\boldsymbol{i}')$. Given a vector $\boldsymbol{i} = (i_1, i_2, \ldots, i_q)' \in \mathcal{I}$, if $\mathcal{M} \subset \mathcal{V}$ then $\boldsymbol{i}_{\mathcal{M}}$ denotes the vector with components $i_j : j \in \mathcal{M}$. If $\boldsymbol{i}_{\mathcal{M}\cup\mathcal{N}}$ is a vector such that $\boldsymbol{i}_{\mathcal{M}} = \boldsymbol{h}_{\mathcal{M}}, \boldsymbol{i}_{\mathcal{N}} = \boldsymbol{k}_{\mathcal{N}}$, with disjoint sets $\mathcal{M}, \mathcal{N}$, we also write $\boldsymbol{i}_{\mathcal{M}\cup\mathcal{N}} = (\boldsymbol{h}_{\mathcal{M}}, \boldsymbol{k}_{\mathcal{N}})$. For every marginal process $\mathbf{A}_j$, the first a_{j1} ($a_{j1} \in \mathcal{A}_j$) is called the baseline category. Any state which includes categories $a_{ji} \in \mathcal{A}_j$, for $j \notin S, S \subset \mathcal{V}$, at the baseline value is denoted by $(\boldsymbol{i}_S, \boldsymbol{i}^*_{\mathcal{V}\setminus S})$. Given a state $\boldsymbol{i}' \in \mathcal{I}$, for the transition probabilities $p(\boldsymbol{i}|\boldsymbol{i}'), \boldsymbol{i} \in \mathcal{I}$, we adopt a Glonek–McCullagh [18] multivariate logistic model whose marginal interaction parameters are denoted by $\eta^P(\boldsymbol{i}_P|\boldsymbol{i}')$, for every non empty subset P of $\mathcal{V}$ and for every $\boldsymbol{i}_P \in \times_{j\in P}\mathcal{A}_j$. The Glonek–McCullagh baseline interactions $\eta^P(\boldsymbol{i}_P|\boldsymbol{i}')$ are expressed as contrasts of logarithms of marginal transition probabilities $p(\boldsymbol{i}_P|\boldsymbol{i}')$ from the state $\boldsymbol{i}'$ to one of the states in $\times_{j\in P}\mathcal{A}_j$

$$\eta^P(\boldsymbol{i}_P|\boldsymbol{i}') = \sum_{\mathcal{K}\subseteq P}(-1)^{|P\setminus\mathcal{K}|}\log p((\boldsymbol{i}_{\mathcal{K}}, \boldsymbol{i}^*_{P\setminus\mathcal{K}})|\boldsymbol{i}'). \tag{19}$$

Note that the Glonek–McCullagh interactions are not log-linear parameters because they are not contrasts of logarithms of the joint transition probabilities $p(\boldsymbol{i}|\boldsymbol{i}')$.

In order to model the dependence of the transition probabilities on the conditioning states $\boldsymbol{i}' \in \mathcal{I}$, we adopt the usual factorial expansion of the Glonek–McCullagh marginal interactions

$$\eta^P(\boldsymbol{i}_P|\boldsymbol{i}') = \sum_{Q\subseteq\mathcal{V}}\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q). \tag{20}$$

The Möbius inversion theorem [24] ensures that

$$\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q) = \sum_{\mathcal{H}\subseteq Q}(-1)^{|Q\setminus\mathcal{H}|}\eta^P(\boldsymbol{i}_P|(\boldsymbol{i}'_{\mathcal{H}}, \boldsymbol{i}^{*\prime}_{\mathcal{V}\setminus\mathcal{H}})). \tag{21}$$

Eqs. (20) and (21) provide that the transition probabilities are parameterized by the interaction parameters $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q)$, $P \subseteq \mathcal{V}, P \neq \emptyset, Q \subseteq \mathcal{V}, \boldsymbol{i}_P \in \times_{j\in P}\mathcal{A}_j, \boldsymbol{i}'_Q \in \times_{j\in Q}\mathcal{A}_j$.

The next proposition shows that (4) and (5) for an MMC being Markov with respect to a graph correspond to simple linear constraints on the $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q)$ parameters and thus testing the hypotheses (4, 5) is a standard parametric problem.

Proposition 2. *For an MMC with strictly positive time-homogeneous transition probabilities, it holds that:* (i) *the Granger noncausality condition* (4) *is equivalent to* $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q) = 0$ *for all* $Q \not\subseteq \mathrm{pa}(P), \boldsymbol{i}_P \in \times_{j\in P}\mathcal{A}_j, \boldsymbol{i}'_Q \in \times_{j\in Q}\mathcal{A}_j$, *and* (ii) *the contemporaneous independence condition* (5) *is equivalent to* $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q) = 0$ *for all* $P \notin \mathcal{B}(G), \boldsymbol{i}_P \in \times_{j\in P}\mathcal{A}_j, \boldsymbol{i}'_Q \in \times_{j\in Q}\mathcal{A}_j$.

Proof. Classical results on the logistic regression ensure equivalence (i), while a result due to Lupparelli et al. [25] implies (ii). □

The expression (19) of the Glonek–McCullagh interactions in terms of baseline log-linear contrasts of marginal transition probabilities is not necessarily the most convenient. When the interpretation of the non null parameters $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q)$, for all $Q \subseteq \mathrm{pa}(P)$ and $P \in \mathcal{B}(G)$ is of interest and the $\mathcal{A}_j$, $j \in \mathcal{V}$, are ordered sets, more general types of Glonek–McCullagh interactions can be used, as shown by Bartolucci et al. [2].

Similar to [2], it can be proved that the set of zero restrictions imposed on parameters $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q)$ can be written in the form $\boldsymbol{C}\ln(\boldsymbol{M}\boldsymbol{\pi}) = \boldsymbol{0}$, where $\boldsymbol{\pi}$ is the vector of all the transition probabilities and $\boldsymbol{C}$ and $\boldsymbol{M}$ are matrices of known constants. The procedures for the maximum likelihood estimation and hypothesis testing, developed by Lang [22] and Cazzaro and Colombi [5] under the assumption of Poisson-multinomial sampling and constraints $\boldsymbol{C}\ln(\boldsymbol{M}\boldsymbol{\pi}) = \boldsymbol{0}$, can be easily adapted to the MMC context. These procedures are implemented in the R function *Mphfit* [23] and in the R-package *hmmm* [4].

7. G-noncausality and contemporaneous independence in marginal processes of MMC

While the earlier sections deal explicitly with all processes which jointly compose an MMC, in this section, we focus on the marginal processes of a multivariate Markov chain. The results illustrated here, that enhance our ability to read independencies off a mixed graph, concern properties of Markovianity, contemporaneous independence and Granger noncausality relations that are preserved after marginalization.

In general, it is not automatically the case that dynamic relations which characterize a process still hold for its marginal subprocesses. For this, it is essential to invoke additional assumptions. More precisely, a marginal process of a multivariate Markov chain is not necessarily a Markov chain. Nevertheless, the Granger noncausality relation gives some insight into the Markovianity of a marginal process and this can be expressed formally in the proposition below.

Proposition 3. *A marginal process* $\mathbf{A}_{\mathcal{M}}$ *of the multivariate Markov chain* $\mathbf{A}_{\mathcal{V}}$, *with positive transition probabilities, is marginally a Markov chain if it is not* G-*caused by* $\mathbf{A}_{\mathcal{V}\setminus\mathcal{M}}$, *that is*

$$A_{\mathcal{M}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{V}\setminus\mathcal{M}}(t-1)|\bar{A}_{\mathcal{M}}(t-1), \quad t \in \mathbb{N}\setminus\{0,1\}. \tag{22}$$

This result was first obtained by Florens et al. [17, Theorem 3.2bis] under the weaker condition of measurable separability. An alternative proof is in [7]. Note that under the stronger condition

$$A_{\mathcal{M}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus\mathcal{M}}(t-1)|A_{\mathcal{M}}(t-1), \quad t \in \mathbb{N}\setminus\{0,1\}, \tag{23}$$

the above proposition holds without the assumption of strict positivity on the transition probabilities or any other assumption that assures the validity of the intersection property [17, Theorem 3.1bis].

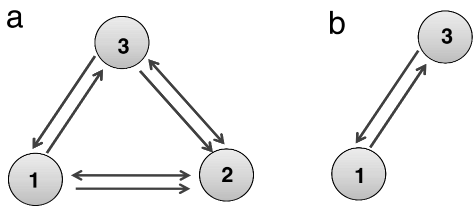

Fig. 3. (a) Mixed graph G; (b) subgraph $G_{\mathcal{M}}$ of G induced by the ancestral set $\mathcal{M} = \{1, 3\}$.

The previous proposition states that a marginal process $\mathbf{A}_{\mathcal{M}}$ retains the Markovian feature if an appropriate Granger noncausality condition is satisfied, whereas the following proposition allows us to identify the mixed graph with respect to which the marginal MMC $\mathbf{A}_{\mathcal{M}}$ is Markov.

Proposition 4. *If the MMC $\mathbf{A}_{\mathcal{V}}$ is Markov with respect to a mixed graph G and if the subset of nodes $\mathcal{M}$, $\mathcal{M} \subset \mathcal{V}$, is ancestral, then $\mathbf{A}_{\mathcal{M}}$ is a multivariate Markov chain which is Markov with respect to the mixed subgraph $G_{\mathcal{M}}$ induced by $\mathcal{M}$.*

Proof. See Appendix. □

Example 3. In the mixed graph G of Fig. 3(a), the set of nodes $\mathcal{M} = \{1, 3\}$ is ancestral. Hence, according to Proposition 4, the marginal process $\mathbf{A}_{\{1,3\}}$ is a bivariate MC which is Markov with respect to the subgraph $G_{\mathcal{M}}$, in Fig. 3(b) induced by $\mathcal{M} = \{1, 3\}$, as $\mathbf{A}_{\{1,2,3\}}$ is Markov with respect to G. Note that the Markov properties of the bivariate process are proved directly by the following simple considerations that illustrate how Proposition 4 works. Since the MC $\mathbf{A}_{\{1,2,3\}}$ is Markov with respect to the initial graph (a), the G-noncausality property $A_{\{1,3\}}(t) \perp\!\!\!\perp A_2(t-1)|A_{\{1,3\}}(t-1)$ and the contemporaneous independence $A_1(t) \perp\!\!\!\perp A_3(t)|A_{\{1,3\}}(t-1)$ hold. This last statement together with $A_1(t) \perp\!\!\!\perp A_2(t-1)|A_{\{1,3\}}(t-1)$, true by the previous G-noncausality condition, leads to $A_1(t) \perp\!\!\!\perp A_3(t), A_2(t-1)|A_{\{1,3\}}(t-1)$. Then, the contemporaneous independence associated to $\mathbf{A}_1 \not\leftrightarrow \mathbf{A}_3$ encoded by the subgraph in Fig. 3(b) immediately follows.

The problem of investigating G-noncausality and contemporaneous independence relations for marginal processes of an MMC is now addressed explicitly, but first, we need to introduce a useful definition for later results.

Definition 4 (*Full Independence*)**.** Two marginal processes $\mathbf{A}_{\mathcal{S}}, \mathbf{A}_{\mathcal{T}}$ of the multivariate Markov chain $\mathbf{A}_{\mathcal{V}}$ are fully independent with respect to $\mathbf{A}_{\mathcal{V}}$ if and only if

$$\overline{A}_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{T}}(t)|\overline{A}_{\mathcal{V}\setminus(\mathcal{S}\cup\mathcal{T})}(t), \quad t \in \mathbb{N} \setminus \{0\}. \tag{24}$$

Later results are analogous to those of Eichler [12] concerning multivariate autoregressive processes. However, the main difference with Eichler's findings is that we rely on statements (4) and (5) and not on the pairwise definitions of Granger noncausality and contemporaneous independence.

In all the following lemmas and propositions, it is implicitly assumed that the MMC $\mathbf{A}_{\mathcal{V}}$ is Markov with respect to a given mixed graph $G = (\mathcal{V}, \mathcal{E})$ (see Definition 3).

The next two lemmas specify when two marginal processes are fully independent both with respect to the whole chain $\mathbf{A}_{\mathcal{V}}$ or to a subprocess $\mathbf{A}_{\mathcal{M}}$, $\mathcal{M} \subset \mathcal{V}$.

Lemma 3. *If the disjoint subsets $\mathcal{S}, \mathcal{T}$ of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$ are m-separated, given $\mathcal{V} \setminus (\mathcal{S} \cup \mathcal{T})$ then the marginal processes $\mathbf{A}_{\mathcal{S}}, \mathbf{A}_{\mathcal{T}}$ of the multivariate Markov chain $\mathbf{A}_{\mathcal{V}}$, with positive transition probabilities, are fully independent with respect to $\mathbf{A}_{\mathcal{V}}$.*

Proof. This lemma is analogous to Lemma B.2 in [12] and a more detailed proof is reported in [7]. □

Lemma 4. *If the disjoint subsets $\mathcal{S}, \mathcal{T}$ of $\mathcal{M}$, $\mathcal{M} \subset \mathcal{V}$, are m-separated, given $\mathcal{M} \setminus (\mathcal{S} \cup \mathcal{T})$, then the marginal processes $\mathbf{A}_{\mathcal{S}}, \mathbf{A}_{\mathcal{T}}$ of the multivariate Markov chain $\mathbf{A}_{\mathcal{V}}$, with positive transition probabilities, are fully independent with respect to the marginal process $\mathbf{A}_{\mathcal{M}}$ of $\mathbf{A}_{\mathcal{V}}$, that is, it holds*

$$\overline{A}_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{T}}(t)|\overline{A}_{\mathcal{M}\setminus(\mathcal{S}\cup\mathcal{T})}(t), \quad t \in \mathbb{N} \setminus \{0\}. \tag{25}$$

Proof. It easily follows by Lemma 3 and Proposition 4. □

Example 4. Let G be the mixed graph in Fig. 4. Consider the marginal set $\mathcal{M} = \{1, 2, 3\}$, and the disjoint subsets $\mathcal{S} = \{1\}$, $\mathcal{T} = \{2\}$, then $\{1\} \bowtie_m \{2\}|\{3\}$ by Lemma 1. Hence, Lemma 4 implies that $\mathbf{A}_1$ and $\mathbf{A}_2$ are fully independent with respect to the marginal process $\mathbf{A}_{\{1,2,3\}}$.

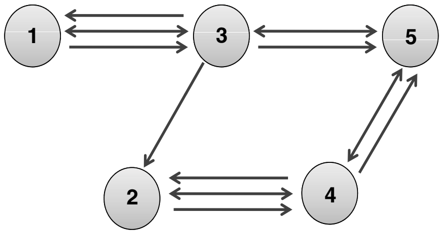

Fig. 4. Lemma 4 implies the full independence between $\mathbf{A}_1$ and $\mathbf{A}_2$ with respect to $\mathbf{A}_{\{1,2,3\}}$.

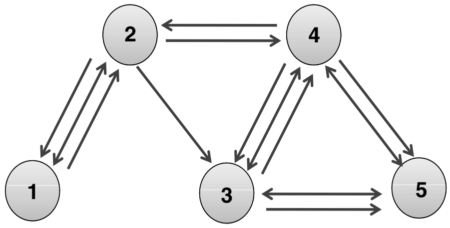

Fig. 5. Proposition 5 ensures that $\mathbf{A}_3 \nrightarrow \mathbf{A}_2$ with respect to $\mathbf{A}_{\{2,3,4\}}$.

The meaning of the m-separation properties involved in the next Propositions 5 and 6 are clarified in detail in Lemmas 5 and 6 reported in the Appendix and it is worthwhile observing that these m-separation properties are not affected by self-loops, since the definition of the sets $\mathrm{pa}(\mathcal{S}) \setminus \mathcal{S}$ and $\mathcal{S} \cup \mathrm{ch}(\mathcal{S})$ does not change with or without the self-loops $i \to i,\ i \in \mathcal{S}$, in the mixed graph.

To understand the significance of the next proposition, we can rewrite the Granger noncausality condition (2) as follows, involving disjoint sets of nodes $\mathcal{S}, \mathcal{T}, \mathcal{C}$,

$$A_{\mathcal{T}}(t) \perp\!\!\!\perp \bar{A}_{\mathcal{S}}(t-1) | \bar{A}_{\mathcal{V}\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t-1), \bar{A}_{\mathcal{T}\cup\mathcal{C}}(t-1) \quad t \in \mathbb{N} \setminus \{0, 1\}. \tag{26}$$

The proposition below highlights when $\bar{A}_{\mathcal{V}\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t-1)$ can be omitted in the conditioning set of (26) without destroying the G-noncausality property. Proposition 5, indeed, clarifies when a marginal process $\mathbf{A}_{\mathcal{S}}$ is not Granger causal for another one $\mathbf{A}_{\mathcal{T}}$ with respect to a marginal process $\mathbf{A}_{\mathcal{M}}$, $\mathcal{M} \subset \mathcal{V}$.

Similar remarks also apply to the contemporaneous independence relation as illustrated in Proposition 6.

Proposition 5. *If $\mathcal{S}, \mathcal{T}, \mathcal{C}$ are disjoint subsets of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$ and if $\mathcal{S}$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$ are m-separated, given $\mathcal{T} \cup \mathcal{C}$, then the marginal process $\mathbf{A}_{\mathcal{S}}$ does not Granger cause $\mathbf{A}_{\mathcal{T}}$ with respect to the marginal process $\mathbf{A}_{\mathcal{M}}$, $\mathcal{M} = \mathcal{S} \cup \mathcal{T} \cup \mathcal{C}$, of the multivariate Markov chain $\mathbf{A}_{\mathcal{V}}$, with positive transition probabilities. More precisely, the following condition holds, for all $t \in \mathbb{N}$*

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \bar{A}_{\mathcal{S}}(t) | \bar{A}_{\mathcal{T}\cup\mathcal{C}}(t). \tag{27}$$

Proof. See Appendix. $\square$

Remark. Note that if $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$ is an empty set, the thesis of Proposition 5 follows directly by (14).

Example 5. Consider the disjoint sets of nodes $\mathcal{S} = \{3\}, \mathcal{T} = \{2\}$, and $\mathcal{C} = \{4\}$ in the mixed graph of Fig. 5. Setting $\mathcal{S}' = \mathcal{S}$ and $\mathcal{T}' = \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C}) = \{1\}$, the marginal set $\mathcal{M} = \mathcal{S}' \cup \mathcal{T}' \cup (\mathcal{T} \cup \mathcal{C}) = \{1, 2, 3, 4\}$ is ancestral, the sets $\mathrm{dis}_{\mathcal{M}}(\mathcal{T}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{T}')) = \{1, 2\}$ and $\mathrm{dis}_{\mathcal{M}}(\mathcal{S}' \cup \mathrm{ch}_{\mathcal{M}}(\mathcal{S}')) = \{3, 4\}$ have no common nodes, and consequently it holds that $\{3\} \bowtie_m \{1\} | \{2, 4\}$. Proposition 5, then ensures that $\mathbf{A}_3 \nrightarrow \mathbf{A}_2$ with respect to the marginal process $\mathbf{A}_{\{2,3,4\}}$.

Proposition 6. *Let $\mathcal{S}, \mathcal{T}$ be disjoint subsets of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$. If $\mathcal{T} \subseteq \mathcal{V} \setminus \mathrm{sp}(\mathcal{S})$ and if the subsets $\mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ are m-separated, given $\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}$, then the marginal processes $\mathbf{A}_{\mathcal{S}}$ and $\mathbf{A}_{\mathcal{T}}$ of the multivariate Markov chain $\mathbf{A}_{\mathcal{V}}$, with positive transition probabilities, are contemporaneously independent with respect to $\mathbf{A}_{\mathcal{M}}$, $\mathcal{M} = \mathcal{S} \cup \mathcal{T} \cup \mathcal{C}$, $\mathcal{M} \subset \mathcal{V}$. More precisely, for all $t \in \mathbb{N}$, it holds that*

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp A_{\mathcal{T}}(t+1) | \bar{A}_{\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t). \tag{28}$$

Proof. See Appendix. $\square$

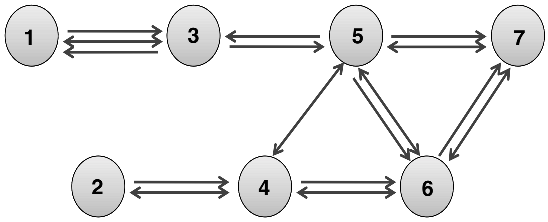

Fig. 6. Proposition 6 states that $\mathbf{A}_3 \not\leftrightarrow \mathbf{A}_{\{4,6\}}$ with respect to $\mathbf{A}_{\{3,4,5,6\}}$.

Example 6. Consider the disjoint sets of nodes $\mathcal{S} = \{3\}, \mathcal{T} = \{4, 6\}, \mathcal{C} = \{5\}$ of the mixed graph displayed in Fig. 6. The sets $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}) = \{2\}$ and $\mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}) = \{1\}$ are m-separated, given $\{3, 4, 5, 6\}$. This can be easily shown by assuming, for example, $\mathcal{S}' = \mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ and $\mathcal{T}' = \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$, so that $\mathcal{M}' = \mathcal{S}' \cup \mathcal{T}' \cup (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}) = \{1, 2, 3, 4, 5, 6\}$ is ancestral and the sets $\mathrm{dis}_{\mathcal{M}'}(\mathcal{T}' \cup \mathrm{ch}_{\mathcal{M}'}(\mathcal{T}')) = \{2, 4, 5, 6\}$ and $\mathrm{dis}_{\mathcal{M}'}(\mathcal{S}' \cup \mathrm{ch}_{\mathcal{M}'}(\mathcal{S}')) = \{1, 3\}$ are disjoint. Hence, Proposition 6 enables us to deduce that $\mathbf{A}_3 \not\leftrightarrow \mathbf{A}_{\{4,6\}}$ with respect to the marginal process $\mathbf{A}_{\{3,4,5,6\}}$.

Remark. It is clear that the marginal process $\mathbf{A}_{\mathcal{M}}$, involved in Propositions 5 and 6, is not in general an MMC unless it is not G-caused by $\mathbf{A}_{\mathcal{V}\setminus\mathcal{M}}$ as required in Proposition 3, or equivalently unless the set $\mathcal{M}$ is ancestral. In the case of a marginal MMC, all the past history of the processes involved in (27) and (28) is not necessary and it can be replaced by the most immediate past.

The foregoing Propositions 5 and 6, together with Lemmas 5 and 6 in the Appendix, highlight the fact that the Markov properties of Definition 3 of an MMC imply the global Markov properties (27) and (28) which are equivalent to those of Eichler and Didelez [16, Def. 4.6].

It is easy to deduce the following result, which is the partial converse of Propositions 5 and 6 since in the previous two propositions, the MMC $\mathbf{A}_{\mathcal{V}}$ is assumed to satisfy (5, 6, 7) with respect to a given mixed graph $G = (\mathcal{V}, \mathcal{E})$.

Proposition 7. *If condition (27) is true, for every disjoint subsets $\mathcal{S}, \mathcal{T}, \mathcal{C}$ of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$ such that $\mathcal{S}$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$ are m-separated, given $\mathcal{T} \cup \mathcal{C}$, then (6) holds. Moreover, (5) is true if (28) is satisfied for all disjoint subsets $\mathcal{S}, \mathcal{T}$ such that:* (i) $\mathcal{T} \subseteq \mathcal{V} \setminus \mathrm{sp}(\mathcal{S})$ *and* (ii) *the sets $\mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})$ are m-separated, given $\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}$.*

In our opinion, the Markov properties of Definition 3 are to be preferred to the global ones as they are simpler to use in technical proofs, they encode statement (7) which refers to the independence of variables from their own recent past and, moreover, lead directly to a parameterization of the transition probabilities.

8. Alternative approaches

The following contemporaneous independence relations can be used instead of (5)

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) | A_{\mathrm{sp}(\mathcal{S})\setminus\mathcal{S}}(t), A_{\mathcal{V}}(t-1), \quad t \in \mathbb{N} \setminus \{0\}. \tag{29}$$

This alternative specification associates different sets of conditional independence restrictions with missing bi-directed edges of the mixed graph G. Conditions (4, 29) were used by Eichler [15] to present a block recursive Granger-causal Markov property clearly different from that presented in Definition 3.

It is interesting to read the Markov properties (4, 5) and (4, 29) under another perspective based on a two-component chain graph, hereafter referred to as G^* (for a review on chain graphs see [24]). This other point of view helps to understand better the meaning of the parametric constraints of Proposition 2 and makes clear that our hypotheses are basically restrictions on transition probabilities.

The nodes of the chain graph G^* belonging to chain components τ_0 and τ_1 correspond to the scalar random variables $A_j(0),\ j = 1, 2, \ldots, q$, and $A_j(1),\ j = 1, 2, \ldots, q$, respectively. Here, the choice of the time points $t = 0$ and $t = 1$ is arbitrary, indeed, any pair of contiguous time points can be considered. All the edges in the subgraph induced by a chain component are bi-directed. Moreover, the graph induced by the chain component τ_0 is bi-complete and the subgraph $G^*_{\tau_1}$, induced by τ_1, has the same edges as G^b. This means that the nodes related to the scalar random variables $A_j(1), A_k(1)$ are connected in $G^*_{\tau_1}$ by a bi-directed edge if and only if the nodes corresponding to the marginal processes $\mathbf{A}_j, \mathbf{A}_k$ are linked by a bi-directed edge in the mixed graph G. Furthermore, the directed edges in graph G^* point from τ_0 toward τ_1. In fact, a directed edge joins the nodes of the random variable $A_j(0)$ to $A_k(1)$ in the chain graph G^* if and only if the nodes representing the marginal processes $\mathbf{A}_j$ and $\mathbf{A}_k$ are linked in the mixed graph G by a directed edge pointing to $\mathbf{A}_k$.

Drton [10] has described four types of Markov properties, called *block recursive Markov properties*, which identify four different classes of models associated with a chain graph. When $t = 1$, conditions (4, 5) encoded by the mixed graph G coincide with Drton's type IV block recursive Markov properties of the above mentioned two-component chain graph G^*

Table 1
Hypothesis tests.

Hyp	Missing edges	*LRT*	*df*	*p*-value	Relations
1	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 2 \leftrightarrow 3$	80.37	32	0.00	$\mathbf{A}_1 \not\leftrightarrow \mathbf{A}_2 \not\leftrightarrow \mathbf{A}_3$
2	$1 \leftrightarrow 2, 1 \leftrightarrow 3$	22.32	24	0.56	$\mathbf{A}_1 \not\leftrightarrow \mathbf{A}_{\{2,3\}}$
3	$1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 1$	23.81	18	0.16	$\mathbf{A}_1 \nrightarrow \mathbf{A}_{\{2,3\}}, \mathbf{A}_{\{2,3\}} \nrightarrow \mathbf{A}_1$
4	$2 \rightarrow 1, 2 \rightarrow 3, 1 \rightarrow 2, 3 \rightarrow 2$	32.47	18	0.02	$\mathbf{A}_2 \nrightarrow \mathbf{A}_{\{1,3\}}, \mathbf{A}_{\{1,3\}} \nrightarrow \mathbf{A}_2$
5	$3 \rightarrow 1, 3 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3$	24.95	18	0.13	$\mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}, \mathbf{A}_{\{1,2\}} \nrightarrow \mathbf{A}_3$
6	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 1$	45.09	42	0.34	$\mathbf{A}_1 \not\leftrightarrow \mathbf{A}_{\{2,3\}}, \mathbf{A}_1 \nrightarrow \mathbf{A}_{\{2,3\}}, \mathbf{A}_{\{2,3\}} \nrightarrow \mathbf{A}_1$
7	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 3 \rightarrow 1, 3 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 3$	45.30	38	0.19	$\mathbf{A}_1 \not\leftrightarrow \mathbf{A}_{\{2,3\}}, \mathbf{A}_3 \nrightarrow \mathbf{A}_{\{1,2\}}, \mathbf{A}_{\{1,2\}} \nrightarrow \mathbf{A}_3$
8	$1 \leftrightarrow 2, 1 \leftrightarrow 3, 2 \leftrightarrow 3, 1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 1,$ $3 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 2$	110.2	50	0.00	$\mathbf{A}_i \not\leftrightarrow \mathbf{A}_{\{j,k\}}, \mathbf{A}_i \nrightarrow \mathbf{A}_{\{j,k\}}, \mathbf{A}_{\{j,k\}} \nrightarrow \mathbf{A}_i\ i \neq j \neq k,\ i, j, k = \{1, 2, 3\}$

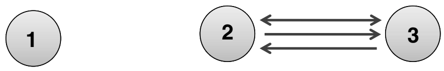

Fig. 7. Mixed graph for web data. It encodes the G-noncausality and contemporaneous independence relations: $\mathbf{A}_1 \nrightarrow \mathbf{A}_{\{2,3\}}, \mathbf{A}_{\{2,3\}} \nrightarrow \mathbf{A}_1; \mathbf{A}_1 \not\leftrightarrow \mathbf{A}_{\{2,3\}}$.

and the parameterization of the transition probabilities introduced in Section 6 is a special case of that proposed for the multivariate regression chain graph models [26]. Thus, the zero restrictions on parameters imposed in Proposition 2 for an MMC to be Markov with respect to a mixed graph G state also that the transition probabilities satisfy the type IV block recursive Markov properties encoded by G^*.

Moreover, type II block recursive Markov properties of G^* are known as the Andersson–Madigan–Perlman (AMP) Markov properties [1] of a chain graph and when $t = 1$ coincide with the conditions (4) and (29) of the mixed graph G.

Unlike types II and IV, Types I and III block recursive Markov properties of G^* does not seem to be appropriate when dealing with G-noncausality because they replace the Granger noncausality condition (4) with

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V} \setminus \mathrm{pa}(\mathcal{S})}(t-1) | A_{\mathcal{V} \setminus \mathcal{S}}(t), A_{\mathrm{pa}(\mathcal{S})}(t-1), \quad t \in \mathbb{N} \setminus \{0\} \tag{30}$$

which introduces $A_{\mathcal{V} \setminus \mathcal{S}}(t)$ in the conditioning information set and does not meet the noncausality in the sense of Granger.

Finally, Drton proved that the type I and IV block recursive properties are the only ones which lead to smooth parameterizations for discrete chain graph models (for an in-depth discussion of this aspect, see the aforementioned work). This result further justifies our preference for the block recursive properties (4, 5) of Definition 3 instead of (4, 29) of [15].

9. Example

The proposed methodology is used to analyze the binary data collected every day for 6 months by an Italian mobile telephone company. The data (available from the authors) consist of a 3-dimensional time series of the daily utilization rate level (low and high) of 3 web servers located in Rome (Italy). The joint dynamic behavior of the series is described by a first order 3-variate Markov chain.

As the servers are all simultaneously operational, it is useful to verify whether the status of a server depends on the utilization levels of the others on the same day, given the past use of all servers. Moreover, it is also important to ascertain whether the current working of a server is influenced by the extent that the others worked the previous day.

An answer to these questions can be attained by testing the hypotheses of G-noncausality and contemporaneous independence. This problem can be easily reduced to establishing which Markov properties of type (4) and (5) are satisfied by the transition probabilities of the web server Markov chain and to identifying the mixed graph which represents them.

To this end, various hypotheses associated with missing edges in the mixed graph have been tested. A few results are illustrated in Table 1. All the hypotheses are tested by using the package *hmmm* [4] in the R language (R Development Core Team [28]). Among the hypotheses in Table 1 involving both G-noncausality and contemporaneous independence restrictions, the 6th and 7th are accepted, but we focus on the 6th, as it has a lower value of the likelihood ratio statistic test *LRT*, higher degrees of freedom and is easier to interpret.

Hypothesis 6 refers to noncausal and contemporaneous dependence relations between the servers 1 and 2, 3. This means that yesterday's utilization levels of servers 2 and 3 do not add helpful information when predicting the first one's operation level today and vice versa and, moreover, there is no influence between the contemporaneous working of servers 1 and 2, 3. The conditions under hypothesis 6 are encoded by the mixed graph $G = (\mathcal{V}, \mathcal{E})$ displayed in Fig. 7, with one node for each server. Therefore, the Markov chain of the web data is Markov with respect to this mixed graph. Note that a directed edge from each node to itself belongs to $\mathcal{E}$, even if it is not drawn in Fig. 7.

The missing edges in Fig. 7 are equivalent to a set of zero constraints on Glonek–McCullagh interactions $\theta^{P,Q}(\boldsymbol{i}_P|\boldsymbol{i}'_Q)$ as explained in Proposition 2. In this example, all the marginal processes have only two states, hence for every P and Q there is only one Glonek–McCullagh multivariate logistic interaction which will be denoted by $\theta^{P,Q}$. In each row of Table 2, an interaction $\theta^{P,Q}$ is identified by the sets $P \subseteq \{1, 2, 3\}$ and $Q \subseteq \{1, 2, 3\}$. The short form 123 is used to denote the set $\{1, 2, 3\}$, 12 is used to denote $\{1, 2\}$ and so on. The symbols $*$ and $\times$ in Table 2 indicate the interactions $\theta^{P,Q}$ which have to

Table 2
Estimates (standard error) of the interactions $\theta^{P,Q}$ of the model associated with the mixed graph in Fig. 7. Interactions with the significant Wald Statistic are boldfaced. The symbols $(*)$ and $(\times)$ indicate the interactions which are null for the conditions of G-noncausality and contemporaneous independence, respectively, encoded by the mixed graph in Fig. 7.

P	Q	Estimates	(s.e.)	P	Q	Estimates	(s.e.)	P	Q	Estimates	(s.e.)
1		**−1.124**	(0.250)	12		×		123		×	
	1	**2.365**	(0.351)		1	×			1	×	
	2	$*$			2	×			2	×	
	3	$*$			3	×			3	×	
	12	$*$			12	×			12	×	
	13	$*$			13	×			13	×	
	23	$*$			23	×			23	×	
	123	$*$			123	×			123	×	
2		**−2.085**	(0.352)	13		×					
	1	$*$			1	×					
	2	**2.085**	(0.783)		2	×					
	3	1.693	(0.963)		3	×					
	12	$*$			12	×					
	13	$*$			13	×					
	23	0.707	(1.254)		23	×					
	123	$*$			123	×					
3		**−2.085**	(0.352)	23		**3.823**	(0.915)				
	1	$*$			1	$*$					
	2	0.205	(1.091)		2	−0.281	(5.232)				
	3	**2.478**	(0.963)		3	0.782	(5.271)				
	12	$*$			12	$*$					
	13	$*$			13	$*$					
	23	1.803	(1.466)		23	1.740	(7.511)				
	123	$*$			123	$*$					

be set at zero in order to meet G-noncausality and contemporaneous independence conditions which we can read off the mixed graph in Fig. 7.

The comparison of the model defined by hypothesis 6 in Table 1 with the model in which also the interactions $\theta^{23,2}, \theta^{23,3}, \theta^{23,23}$ are equal to zero provides the value $LRT = 3.88$ ($df = 3$) of the likelihood ratio test statistic. This means that the odds ratio measuring the association between the daily utilization of servers 2 and 3 does not depend on the utilization levels on the day before. This form of constant association is not encoded by missing edges in the graph.

Appendix. Technical proofs

A.1. Some results on the m-separation

The following lemmas help to clarify the assumptions of Propositions 5 and 6 reported in Section 7.

Recall that a path τ in a mixed graph $G = (\mathcal{V}, \mathcal{E})$ with endpoints s and t is pointing to t if its last edge has an arrowhead at the endpoint t; while τ is bi-pointing if it has an arrowhead at both endpoints s and t.

Lemma 5. *If $\mathcal{S}, \mathcal{T}$ are disjoint subsets of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$, then $\mathcal{S} \bowtie_m \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C}) | \mathcal{T} \cup \mathcal{C}$, if and only if there are no connecting paths, with endpoints $s \in \mathcal{S}$ and $t \in \mathcal{T}$ pointing to t.*

Proof. Let us assume that the sets $\mathcal{S}$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$ are not m-separated, given $\mathcal{T} \cup \mathcal{C}$, this means that there exists a connecting path, given $\mathcal{T} \cup \mathcal{C}$, with endpoints $s \in \mathcal{S}$ and $p \in \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$. If we extend this path by adding a directed edge from p to $t \in \mathcal{C}$, we get a connecting path from $s \in \mathcal{S}$ to $t \in \mathcal{T}$ pointing to t. Conversely, let us consider that $\mathcal{S} \bowtie_m \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C}) | \mathcal{T} \cup \mathcal{C}$ and, moreover, that there exists a connecting path τ from $s \in \mathcal{S}$ to $t \in \mathcal{T}$ pointing to t. Let $u \to t$ be the last edge of this path. Since τ is connecting and u is a noncollider node it must be $u \in \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$, while the remaining noncollider nodes of τ must not belong to $\mathcal{T} \cup \mathcal{C}$ and the collider nodes must be into $\mathcal{T} \cup \mathcal{C}$. This contradicts the assumed hypothesis of separation and, consequently, such a path cannot exist. □

Lemma 6. *If $\mathcal{S}, \mathcal{T}$ are disjoint subsets of nodes of the mixed graph $G = (\mathcal{V}, \mathcal{E})$, then $\mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}) \bowtie_m \mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C}) | \mathcal{S} \cup \mathcal{T} \cup \mathcal{C}$ if and only if there are no connecting bi-pointing paths with endpoints $s \in \mathcal{S}$ and $t \in \mathcal{T}$.*

Proof. The proof is analogous to that of Lemma 5 and is omitted. □

A.2. Proofs of the propositions in Section 7

Proof of Proposition 4. First, let us consider that from $\mathrm{pa}(\mathcal{M}) \subseteq \mathcal{M}$ and (4), true for hypothesis, we have

$$A_{\mathcal{M}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus\mathcal{M}}(t-1) | A_{\mathcal{M}}(t-1), \quad t \in \mathbb{N} \setminus \{0, 1\}.$$

Since the joint process is Markov, it also follows that

$$A_{\mathcal{M}}(t) \perp\!\!\!\perp A_{\mathcal{M}}(t-2) | A_{\mathcal{V}}(t-1), \quad t \in \mathbb{N} \setminus \{0, 1, 2\}.$$

According to the composition property, the previous two conditions imply that

$$A_{\mathcal{M}}(t) \perp\!\!\!\perp A_{\mathcal{M}}(t-2), A_{\mathcal{V}\setminus\mathcal{M}}(t-1) | A_{\mathcal{M}}(t-1), \quad t \in \mathbb{N} \setminus \{0, 1, 2\}. \tag{A.1}$$

Hence, the claim that the marginal process $\mathbf{A}_{\mathcal{M}}$ is an MMC follows directly from (A.1).

Now, in order to prove that MMC $\mathbf{A}_{\mathcal{M}}$ is also Markov with respect to the mixed graph $G_{\mathcal{M}}$ induced by $\mathcal{M} \subseteq \mathcal{V}$, we need to show that the G-noncausality and contemporaneous independence conditions of Definition 3 continue to hold with respect to the subset $\mathcal{M}$ of $\mathcal{V}$ as well.

More precisely, it is worth noting that $\mathrm{pa}(\mathcal{M}) \subseteq \mathcal{M}$ and (14) imply

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{M}\setminus pa_{\mathcal{M}}(\mathcal{S})}(t-1) | \overline{A}_{pa_{\mathcal{M}}(\mathcal{S})}(t-1) \quad t \in \mathbb{N} \setminus \{0\}, \ \mathcal{S} \in \mathcal{P}(\mathcal{M}),$$

which evidently is the Granger noncausality condition in $G_{\mathcal{M}}$.

We only have to show the same result for contemporaneous independence. This can be demonstrated by a few arguments. Note that from (15), we obtain

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) | \overline{A}_{\mathcal{V}\setminus\mathcal{M}}(t-1), \overline{A}_{\mathcal{M}}(t-1) \quad t \in \mathbb{N} \setminus \{0\}, \ \mathcal{S} \in \mathcal{P}(\mathcal{V}), \tag{A.2}$$

and if $\mathcal{S} \in \mathcal{P}(\mathcal{M})$, by Lemma 2 the foregoing condition (A.1) gives

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus\mathcal{M}}(t-1) | \overline{A}_{\mathcal{M}}(t-1) \quad t \in \mathbb{N} \setminus \{0\}. \tag{A.3}$$

Without loss of generality, we can continue to assume that $\mathcal{S} \in \mathcal{P}(\mathcal{M})$, hence it is simple to prove that (A.2, A.3) can be equivalently written as follows

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus\mathcal{M}}(t-1), A_{\mathcal{V}\setminus \mathrm{sp}(\mathcal{S})}(t) | \overline{A}_{\mathcal{M}}(t-1) \quad t \in \mathbb{N} \setminus \{0\}. \tag{A.4}$$

Since the nodes belonging to $\mathcal{M}$ which are spouses in the original graph G remain spouses in the subgraph $G_{\mathcal{M}}$ and $\mathcal{S} \in \mathcal{P}(\mathcal{M})$, then the set $\mathcal{V} \setminus \mathrm{sp}(\mathcal{S})$ contains $\mathcal{M} \setminus \mathrm{sp}_{\mathcal{M}}(\mathcal{S})$. Therefore, we can immediately deduce from (A.4) the desired result

$$A_{\mathcal{S}}(t) \perp\!\!\!\perp A_{\mathcal{M}\setminus \mathrm{sp}_{\mathcal{M}}(\mathcal{S})}(t) | \overline{A}_{\mathcal{M}}(t-1) \quad t \in \mathbb{N} \setminus \{0\}, \ \mathcal{S} \in \mathcal{P}(\mathcal{M}) \tag{A.5}$$

which states the contemporaneous independence condition in $G_{\mathcal{M}}$.

Finally, the independencies (A.1, A.5) establish that the multivariate Markov chain $\mathbf{A}_{\mathcal{M}}$ is Markov with respect to the mixed subgraph $G_{\mathcal{M}}$ induced by $\mathcal{M}$ after marginalizing $\mathcal{V}$. $\square$

Proof of Proposition 5. Under the hypotheses of this proposition, Lemma 4 entails

$$\overline{A}_{\mathcal{S}}(t) \perp\!\!\!\perp \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N} \setminus \{0\}. \tag{A.6}$$

Furthermore, the next independence

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus(\mathrm{pa}(\mathcal{T})\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{T})\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}$$

follows by condition (14).

It is worth noting that $\mathcal{S} \subset \mathcal{V} \setminus (\mathrm{pa}(\mathcal{T}) \cup \mathcal{T} \cup \mathcal{C})$ since $\mathcal{S}, \mathcal{T}, \mathcal{C}$ are disjoint and $\mathcal{S}$ and $\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{T} \cup \mathcal{C})$ are m-separated sets given $\mathcal{T} \cup \mathcal{C}$. Consequently, the above statement yields

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{S}}(t) | \overline{A}_{\mathrm{pa}(\mathcal{T})\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}.$$

The thesis of this proposition is immediately obtained as the earlier independence and (A.6) are equivalent to

$$A_{\mathcal{T}}(t+1), \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{T}\cup\mathcal{C})}(t) \perp\!\!\!\perp \overline{A}_{\mathcal{S}}(t) | \overline{A}_{\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}$$

which directly implies

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{S}}(t) | \overline{A}_{\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}. \quad \square$$

Proof of Proposition 6. The hypotheses of the proposition and Lemma 4 lead to

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp A_{\mathcal{T}}(t+1) | \overline{A}_{\mathcal{V}}(t), \quad t \in \mathbb{N}, \tag{A.7}$$

$$\overline{A}_{\mathrm{pa}(\mathcal{S})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) \perp\!\!\!\perp \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}. \tag{A.8}$$

Then, using (14) at time $t+1$,

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus \mathrm{pa}(\mathcal{S})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{S})}(t), \quad t \in \mathbb{N}$$

we find

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus(\mathrm{pa}(\mathcal{S})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{S})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}. \tag{A.9}$$

For our purpose, it is important to note that (A.7, A.9) are equivalent to

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp A_{\mathcal{T}}(t+1), \overline{A}_{\mathcal{V}\setminus(\mathrm{pa}(\mathcal{S})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{S})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}$$

from which it unfolds that

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp A_{\mathcal{T}}(t+1), \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{S})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N} \tag{A.10}$$

since the hypothesis of m-separated sets ensures that $(\mathrm{pa}(\mathcal{T}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})) \cap (\mathrm{pa}(\mathcal{S}) \setminus (\mathcal{S} \cup \mathcal{T} \cup \mathcal{C})) = \emptyset$.

Moreover, condition (14) provides

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathcal{V}\setminus(\mathrm{pa}(\mathcal{T})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{T})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}$$

and, hence, we have

$$A_{\mathcal{T}}(t+1) \perp\!\!\!\perp \overline{A}_{\mathrm{pa}(\mathcal{S})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathrm{pa}(\mathcal{T})\cup\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}. \tag{A.11}$$

Now, applying the contraction property to (A.8, A.11) gives, for all $t \in \mathbb{N}$

$$\overline{A}_{\mathrm{pa}(\mathcal{S})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) \perp\!\!\!\perp A_{\mathcal{T}}(t+1), \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t). \tag{A.12}$$

Finally, we see that conditions (A.10, A.12) are equivalently expressed as

$$A_{\mathcal{S}}(t+1), \overline{A}_{\mathrm{pa}(\mathcal{S})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) \perp\!\!\!\perp A_{\mathcal{T}}(t+1), \overline{A}_{\mathrm{pa}(\mathcal{T})\setminus(\mathcal{S}\cup\mathcal{T}\cup\mathcal{C})}(t) | \overline{A}_{\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t)$$

that demonstrates the thesis of the proposition

$$A_{\mathcal{S}}(t+1) \perp\!\!\!\perp A_{\mathcal{T}}(t+1) | \overline{A}_{\mathcal{S}\cup\mathcal{T}\cup\mathcal{C}}(t), \quad t \in \mathbb{N}. \quad \square$$

References

[1] S. Andersson, D. Madigan, M. Perlman, Alternative Markov properties for chain graphs, Scand. J. Stat. 28 (2001) 33–85.
[2] F. Bartolucci, R. Colombi, A. Forcina, An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints, Statist. Sinica 17 (2007) 691–711.
[3] M.B. Bouissou, J.J. Laffont, Q. Vuong, Tests of non causality under Markov assumptions for qualitative panel data, Econometrica 54 (1986) 395–414.
[4] M. Cazzaro, R. Colombi, Hierarchical multinomial marginal models: the R package hmmm, 2008. www.unibg.it/pers/?colombi.
[5] M. Cazzaro, R. Colombi, Multinomial-Poisson models subject to inequality constraints, Stat. Model. 9 (2009) 215–233.
[6] G. Chamberlain, The general equivalence of Granger and Sims causality, Econometrica 50 (1982) 569–581.
[7] R. Colombi, S. Giordano, Proofs of Proposition 3 and Lemma 3 in graphical models for multivariate Markov chains, Technical Report, 2010. Available on www.ecostat.unical.it/giordano.
[8] D.R. Cox, N. Wermuth, Multivariate Dependencies-Models, Analysis and Interpretation, Chapman and Hall, London, 1996.
[9] A.P. Dawid, Conditional independence in statistical theory (with discussion), J. R. Stat. Soc. Ser. B Stat. Methodol. 41 (1979) 1–13.
[10] M. Drton, Discrete chain graph models, Bernoulli 15 (3) (2009) 753–763.
[11] D. Edwards, Introduction to Graphical Modelling, Springer, New York, 1995.
[12] M. Eichler, Granger causality and path diagrams for multivariate time series, J. Econometrics 137 (2007) 334–353.
[13] M. Eichler, Causal inference from multivariate time series: what can be learned from Granger causality, in: C. Glymour, W. Wang, D. Westerstahl (Eds.), Proceedings from the 13th International Congress of Logic, Methodology and Philosophy of Science, King's College Publications, London, 2009.
[14] M. Eichler, A note on global Markov properties for mixed graphs, 2011. arXiv.org/abs/1107.3036.
[15] M. Eichler, Graphical modelling of multivariate time series, Probab. Theory Related Fields (2011) doi:10.1007/s00440-011-0345-8.
[16] M. Eichler, V. Didelez, Causal reasoning in graphical time series models, in: Proceedings of the 23rd Annual Conference on Uncertainty in Artifical Intelligence, Vancouver, 2007, pp. 109–116.
[17] J.P. Florens, M. Mouchart, J.M. Rolin, Noncausality and marginalization of Markov processes, Econometric Theory 9 (1993) 241–262.
[18] G.J.N. Glonek, P. McCullagh, Multivariate logistic models, J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1995) 533–546.
[19] C. Gouriéroux, A. Monfort, E. Renault, Kullback causality measures, Ann. Économ. Statist. 67 (1987) 369–410.
[20] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (1969) 424–438.
[21] J.T.A. Koster, On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors, Scand. J. Stat. 26 (1999) 413–431.
[22] J.B. Lang, Homogeneous linear predictor models for contingency tables, J. Amer. Statist. Assoc. 100 (2005) 121–134.
[23] J.B. Lang, mph.fit: an R program for maximum likelihood fitting of multinomial-Poisson homogeneous (MPH) models for contingency tables. www.stat.uiowa.edu/~jblang/#software.
[24] S.L. Lauritzen, Graphical Models, Clarendon Press, Oxford, 1996.
[25] M. Lupparelli, G.M. Marchetti, W. Bergsma, Parameterizations and fitting of bi-directed graph models to categorical data, Scand. J. Stat. 36 (2009) 559–576.
[26] G.M. Marchetti, M. Lupparelli, Chain graph models of multivariate regression type for categorical data, Bernoulli 17 (3) (2011) 827–844.
[27] R. Mosconi, R. Seri, Non-causality in bivariate binary time series, J. Econometrics 132 (1996) 379–407.
[28] R development Core Team, R: A Language and Environment for Statistical Computing. R Version 2.8.0, R Foundation for Statistical Computing, Vienna, ISBN: 3-900051-07-0, 2008.
[29] T. Richardson, Markov properties for acyclic directed mixed graphs, Scand. J. Stat. 30 (2003) 145–157.
[30] E. San Martìn, M. Mouchart, J.M. Rolin, Ignorable common information, null sets and Basu's first theorem, Sankhyā 67 (4) (2005) 674–698.
[31] J. Whittaker, Graphical Models in Applied Multivariate Statistics, John Wiley, Chichester, 1990.