



Multiple hidden Markov models for categorical time series



R. Colombi^a, S. Giordano^{b,*}

^a Department of Engineering, University of Bergamo, Italy

^b Department of Economics, Statistics and Finance, University of Calabria, Italy

ARTICLE INFO

Article history:

Received 31 March 2014

Available online 12 April 2015

AMS subject classifications:

68R10

60J10

62H99

Keywords:

Conditional independence

Granger noncausality

Graphical models

Marginal models

Markov properties

Multivariate Markov chains

ABSTRACT

We introduce multiple hidden Markov models (MHMMs) where a multivariate categorical time series depends on a latent multivariate Markov chain. MHMMs provide an elegant framework for specifying various independence relationships between multiple discrete time processes. These independencies are interpreted as Markov properties of a mixed graph and a chain graph associated respectively to the latent and observation components of the MHMM. These Markov properties are also translated into zero restrictions on the parameters of marginal models for the transition probabilities and the distributions of observable variables given the latent states.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

In several applications involving time series, it is interesting to describe how the evolution of variables over time depends on latent characteristics or the focus may be on the dynamics of unobservable characteristics measured by variables observed on consecutive time occasions. These issues are addressed by hidden Markov models and a widespread application of them has occurred in several fields such as speech recognition, signal processing, digital communication, biology, reliability, etc., standard references are [3,13,15,16], among others.

Basically, a hidden Markov model assumes that an observable time series depends on a latent Markov chain in such a way that the joint process is also Markovian. Note that, throughout the paper, the term observable never refers to the observability property of state-space models (see [18] and [21], among others), but it is used with its literal meaning to distinguish the variables that can be directly observed from the latent ones.

In this work, we focus on hidden Markov models in discrete time with a multivariate categorical process depending on a multivariate latent chain. In these models, several variables are observed and their distribution is supposed to be affected by one or more latent variables. For the multidimensionality of latent and observation components, we will refer to these models as multiple hidden Markov models (MHMMs).

The MHMM can be seen as a variant of the usual hidden Markov model (HMM) that allows modeling opportunities not available in the standard approach with the same clarity, interpretability and parsimony.

In particular, MHMMs enable us to formulate meaningful independence structures for the latent component and for the observable variables given the latent states. Such independence hypotheses are restrictions on the transition

* Corresponding author.

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

probabilities of the latent chain and on the state-dependent distributions of the observation process. The same hypotheses can be investigated using the classical hidden Markov models, but formulating these hypotheses as simple conditional independencies and turning them into tractable constraints on parameters are much more demanding than in the MHMM context.

MHMMs can be well suited to applications where all the observable time series are affected by one common unobservable factor (general effect) and each observable variable is also governed by its specific latent variable. MHMMs may also handle time series data where an unobservable aspect influences the marginal dynamics of each observable variable while another latent factor influences the association among them. In the framework of MHMMs, different sets of observable categorical time series are allowed to depend on different sets of unobservable processes and the observable variables are not required to be independent given the latent states (local independence assumption), but the association between them is also modeled. Moreover, in MHMMs, the multivariate latent process can satisfy the hypotheses of Granger noncausality and contemporaneous independence described by Colombi and Giordano [7] for multivariate Markov chains.

So great is the variety of roles (specific, generic, association-affecting, etc.) assigned to the unobserved component in MHMMs, that it is hard to imagine this variety being easily handled by any classical HMM where only one latent process is allowed.

This approach, for example, responds to the shortcomings highlighted in Zucchini and Guttorp [23] who proposed a model for describing the sequence of wet and dry days at 5 sites without taking into account the spatial dependence among sites situated in closed locations and without allowing for a multivariate state process with sites in different regions responding to different components of the latent process. Other examples illustrated in the literature can be enriched by more flexible and realistic hypotheses using MHMMs.

To appreciate the possibility offered by MHMMs of modeling various hypotheses on the temporal evolution of the latent components and on the influence of the latent states on the observable variables, it is crucial to formulate such hypotheses by compact and simple expressions. The methodology of graphical models (Lauritzen [14]) serves this need. In fact, the graphs can visually represent the complex independence structures related to the latent and observation components of the MHMM and all the hypotheses can be described and handled by careful combination of simple elements in graphical models. This is the reason why we take advantage of the graphical models by presenting the independence conditions behind the MHMMs as Markov properties of the associated graphs and testing them as simple linear constraints on parameters.

The paper is organized as follows. MHMMs are presented in Section 2 and Section 3 illustrates that the transition probabilities of the latent model and the distributions of observable variables given the latent states of MHMMs can be required to obey the Markov properties of a mixed and a chain graph, respectively. In Section 4, the conditional independencies, defining the MHMMs and interpreted as Markov properties of graphs in Section 3, are shown to be equivalent to linear constraints on parameters of Gloneck and McCullagh models [12] for the transition probabilities and the state-dependent distributions. This last result is extremely important for fitting and testing MHMMs. Finally, in the last section several MHMMs are used to analyze two data sets.

2. Multiple hidden Markov models

Let $\mathbf{E}_{\mathcal{U}}$ be a r -variate process of categorical variables, $\mathbf{E}_{\mathcal{U}} = \{E_{\mathcal{U}}(t) : t \in \mathbb{N}\} = \{E_i(t) : t \in \mathbb{N}, i \in \mathcal{U}\}$, $\mathcal{U} = \{1, \dots, r\}$, $\mathbb{N} = \{0, 1, 2, \dots\}$ and let $\mathbf{F}_{\mathcal{V}}$ be a s -dimensional process of categorical variables $\mathbf{F}_{\mathcal{V}} = \{F_{\mathcal{V}}(t) : t \in \mathbb{N}\} = \{F_j(t) : t \in \mathbb{N}, j \in \mathcal{V}\}$, $\mathcal{V} = \{1, \dots, s\}$. The random variables $E_i(t), F_j(t)$ take values in finite sets \mathcal{E}_i and \mathcal{F}_j , $i \in \mathcal{U}, j \in \mathcal{V}$.

For every subset $\mathcal{T} \subset \mathcal{U}$ and $\mathcal{R} \subset \mathcal{V}$, marginal processes are represented by $\mathbf{E}_{\mathcal{T}} = \{E_i(t) : i \in \mathcal{T}, t \in \mathbb{N}\}$ and $\mathbf{F}_{\mathcal{R}} = \{F_j(t) : j \in \mathcal{R}, t \in \mathbb{N}\}$. Univariate marginal processes will be denoted by $\mathbf{E}_i, \mathbf{F}_j$, $i \in \mathcal{U}, j \in \mathcal{V}$.

The following definition states when $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is an MHMM.

Definition 1. The joint process $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is an MHMM if

- (a) $\mathbf{E}_{\mathcal{U}}$ is unobservable
- (b) $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is a first order multivariate Markov chain
- (c) $E_{\mathcal{U}}(t) \perp F_{\mathcal{V}}(t-1) | E_{\mathcal{U}}(t-1)$
- (d) $F_{\mathcal{V}}(t) \perp E_{\mathcal{U}}(t-1), F_{\mathcal{V}}(t-1) | E_{\mathcal{U}}(t)$.

In particular, condition (c) implies that $\mathbf{E}_{\mathcal{U}}$ is a first order Markov chain (see [6]).

A marginal process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$, $\mathcal{T} \subset \mathcal{U}$ and $\mathcal{R} \subset \mathcal{V}$, in general, is not a hidden Markov model. The following theorem clarifies when the properties of an MHMM are preserved after marginalizing the latent and observation processes. This helps to specify under which restrictions all the attractive features of the hidden models (forecast distributions, smoothing and filtering algorithms, etc.) are still valid for an MHMM with marginalized components.

Theorem 1. Let $\mathbf{E}_{\mathcal{T}}$ and $\mathbf{F}_{\mathcal{R}}$, $\mathcal{T} \subset \mathcal{U}$ and $\mathcal{R} \subset \mathcal{V}$, be marginal processes of the latent and observation components of an MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$. The marginal process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$ is still an MHMM if the following conditions are fulfilled for all $t \in \mathbb{N} \setminus \{0\}$

$$E_{\mathcal{T}}(t) \perp E_{\mathcal{U} \setminus \mathcal{T}}(t-1) | E_{\mathcal{T}}(t-1) \quad (1)$$

$$F_{\mathcal{R}}(t) \perp E_{\mathcal{U} \setminus \mathcal{T}}(t) | E_{\mathcal{T}}(t). \quad (2)$$

Proof. We will show that under the hypotheses of the theorem, the process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$ satisfies the conditions of [Definition 1](#).

Let $Pr(E_{\mathcal{U}}(t), F_{\mathcal{V}}(t)|E_{\mathcal{U}}(t-1), F_{\mathcal{V}}(t-1))$ denote the transition probability of an MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$. Conditions (b), (c) and (d) of [Definition 1](#) imply the following factorization

$$Pr(E_{\mathcal{T}}(t), F_{\mathcal{R}}(t)|E_{\mathcal{U}}(t-1), F_{\mathcal{V}}(t-1)) = Pr(E_{\mathcal{T}}(t)|E_{\mathcal{U}}(t-1))Pr(F_{\mathcal{R}}(t)|E_{\mathcal{U}}(t)).$$

This formula, together with conditions (1) and (2), gives

$$Pr(E_{\mathcal{T}}(t), F_{\mathcal{R}}(t)|E_{\mathcal{U}}(t-1), F_{\mathcal{V}}(t-1)) = Pr(E_{\mathcal{T}}(t)|E_{\mathcal{T}}(t-1))Pr(F_{\mathcal{R}}(t)|E_{\mathcal{T}}(t))$$

which entails that $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$ is a Markov chain. Moreover, this last equality proves the independencies $E_{\mathcal{T}}(t) \perp F_{\mathcal{R}}(t-1)|E_{\mathcal{T}}(t-1)$ and $F_{\mathcal{R}}(t) \perp E_{\mathcal{T}}(t-1), F_{\mathcal{R}}(t-1)|E_{\mathcal{T}}(t)$. Therefore, $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$ meets the conditions defining an MHMM ([Definition 1](#)) and this result completes the proof. \square

In [Theorem 1](#), condition (1) ensures that the marginal process $\mathbf{E}_{\mathcal{T}}$ of the latent component $\mathbf{E}_{\mathcal{U}}$ is a Markov chain, as shown in Colombi and Giordano [6], and (2) states that the observable variables $F_{\mathcal{R}}(t)$ at time t depends only on the latent variables $E_{\mathcal{T}}(t)$. In [Corollary 1](#) of the next section, conditions (1) and (2) will be presented as special cases of the Markov properties of graphical models associated to MHMMs and in [Section 4](#) they will be tested as zero constraints on parameters.

Some special cases of the previous theorem may be advisable.

In particular, it is easy to verify that the marginal process $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{R}})$, $\mathcal{R} \subset \mathcal{V}$, of an MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is hidden Markov too, whereas when only the latent component is marginalized, testing if the resulting process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{V}})$, $\mathcal{T} \subset \mathcal{U}$ is still an MHMM leads to a non-standard problem due to the presence of unidentifiable parameters. In fact, for the marginal process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{V}})$, $\mathcal{T} \subset \mathcal{U}$ of the MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$, conditions (1) and (2) of [Theorem 1](#) become

$$E_{\mathcal{T}}(t) \perp E_{\mathcal{U} \setminus \mathcal{T}}(t-1)|E_{\mathcal{T}}(t-1), \quad F_{\mathcal{V}}(t) \perp E_{\mathcal{U} \setminus \mathcal{T}}(t)|E_{\mathcal{T}}(t) \quad (3)$$

ensuring that also $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{V}})$ is an MHMM. But, as already discussed by Colombi and Giordano [6], under (3) some parameters of the model $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ are not identifiable, and therefore the problem of testing (3) cannot be solved by standard methods.

3. Graphical models for MHMMs

The advantage of modeling a variety of independence hypotheses for the latent and observation processes offered by the MHMMs is enhanced by exploiting the peculiarities of graphical models. With this aim in mind, in this section, we address the use of graphical models associated to MHMMs in the sense that the conditional independencies defining an MHMM will be presented as Markov properties encoded by certain graphs.

Graphical models for MHMMs associate the missing edges of a graph with conditional independence restrictions imposed on the probabilities of observable variables given the latent states and the transition probabilities of the latent process.

The notation used throughout the paper follows Colombi and Giordano [7]. We would just briefly recall that $G = (V, E)$ is a graph with a finite set of nodes V and a set of edges E ; moreover, for every non-empty subset of nodes \mathcal{S} , $pa_G(\mathcal{S})$, $ch_G(\mathcal{S})$, and $sp_G(\mathcal{S})$ are the collection of parents, children and spouses of nodes in the set \mathcal{S} . Note that every node is spouse of itself. Finally, $\mathfrak{B}(G)$ indicates the family of the bi-connected sets in the graph. For details on the terminology and the general theory of graphical models see Lauritzen [14]. For models associated to chain graphs and to bi-directed or mixed graphs see Drton [10] and Richardson [19], respectively.

3.1. MHMM Markov with respect to a mixed and a chain graph

We consider two types of graphs: a mixed graph for the latent component of the MHMM and a chain graph for the observation component. In particular, the transition probabilities of the multivariate latent Markov chain $\mathbf{E}_{\mathcal{U}}$ in the MHMM are required to obey a set of Markov properties encoded by a mixed graph G whose basic features are discussed in Colombi and Giordano [7], for Markov chains, while the Markov properties satisfied by the distribution of the observable variables conditioned on the latent states are read off a chain graph G^* , see Drton [10] among others.

In a mixed graph G , a node i corresponds to the marginal process \mathbf{E}_i , for every $i \in \mathcal{U}$, and independence restrictions are associated with missing directed and bi-directed edges, respectively. In particular, missing bi-directed edges lead to independencies of marginal processes at the same point in time; missing directed edges, instead, refer to independencies that involve marginal processes at two consecutive instants.

A chain graph G^* with two chain components τ_0 and τ_1 serves the need to encode the independence relations among observable and latent variables of the MHMM at a given point in time. The nodes of the chain graph G^* belonging to τ_0 correspond to the random variables $E_i(t^*)$, $i \in \mathcal{U}$, and the nodes belonging to τ_1 correspond to the random variables $F_j(t^*)$, $j \in \mathcal{V}$, for any arbitrary $t^* \in \mathbb{N}$.

All the edges in the subgraph induced by a chain component are bi-directed and the graph induced by the chain component τ_0 is bi-complete. Furthermore, the directed edges in graph G^* point in the same direction from τ_0 towards τ_1 .

The definition below specifies the properties for an MHMM to be Markov with respect to a mixed and a chain graph.

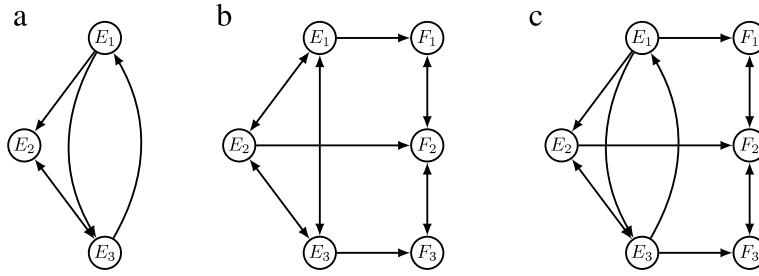


Fig. 1. Mixed (a), chain graph (b) and mixed-chain graph (c) associated to an MHMM.

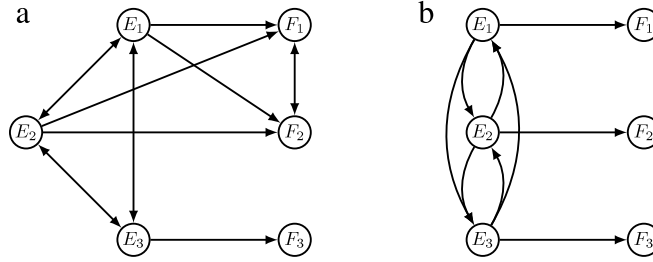


Fig. 2. Mixed-chain graphs associated to MHMMs described in Example 1.

Definition 2 (MHMM Markov wrt a Mixed and a Chain Graph). The latent process $\mathbf{E}_{\mathcal{U}}$ of an MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is Markov wrt a mixed graph G if and only if its transition probabilities fulfill the following conditional independencies for all $t \in \mathbb{N} \setminus \{0\}$ associated to missing directed and bi-directed edges of G , respectively:

$$E_{\mathcal{T}}(t) \perp\!\!\!\perp E_{\mathcal{U} \setminus pa_G(\mathcal{T})}(t-1) | E_{pa_G(\mathcal{T})}(t-1) \quad \forall \mathcal{T} \subset \mathcal{U} \quad (4)$$

$$E_{\mathcal{T}}(t) \perp\!\!\!\perp E_{\mathcal{U} \setminus sp_G(\mathcal{T})}(t) | E_{\mathcal{U}}(t-1) \quad \forall \mathcal{T} \subset \mathcal{U}. \quad (5)$$

The observation process $\mathbf{F}_{\mathcal{V}}$ is Markov wrt a chain graph G^* if and only if the distribution of the observable variables given the latent states fulfills the following conditional independencies for all $t \in \mathbb{N} \setminus \{0\}$ associated with missing bi-directed and directed edges of G^* , respectively:

$$F_{\mathcal{R}}(t) \perp\!\!\!\perp F_{\mathcal{V} \setminus sp_{G^*}(\mathcal{R})}(t) | E_{\mathcal{U}}(t) \quad \forall \mathcal{R} \subset \mathcal{V} \quad (6)$$

$$F_{\mathcal{R}}(t) \perp\!\!\!\perp E_{\mathcal{U} \setminus pa_{G^*}(\mathcal{R})}(t) | E_{pa_{G^*}(\mathcal{R})}(t) \quad \forall \mathcal{R} \subset \mathcal{V}. \quad (7)$$

Therefore, an MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is said to be Markov wrt a mixed and a chain graph when the latent component is Markov wrt a mixed graph, and the observation component given the latent states is Markov wrt a chain graph.

In the context of first order multivariate Markov chains, condition (4) corresponds to a notion of Granger noncausality and, when $pa_G(\mathcal{T}) = \mathcal{T}$, it ensures that the marginal process $\mathbf{E}_{\mathcal{T}}$ is a Markov chain [7,11]. Henceforth, we will refer to (4) with the term Granger noncausality (G-noncausality) condition saying that the latent process $\mathbf{E}_{\mathcal{T}}$ is not G-caused by $\mathbf{E}_{\mathcal{U} \setminus pa_G(\mathcal{T})}$ with respect to $\mathbf{E}_{\mathcal{U}}$. Condition (5) states that the transition probabilities satisfy the bi-directed Markov property [19] with respect to the graph obtained by removing the directed edges from the mixed graph. Here, we will refer to (5) with the term contemporaneous independence condition and say that the latent processes $\mathbf{E}_{\mathcal{T}}$ and $\mathbf{E}_{\mathcal{U} \setminus sp_G(\mathcal{T})}$ are contemporaneously independent.

On the other hand, conditions (6) and (7) encoded by the chain graph G^* refer to the observation component of the MHMM and are equivalent to the type IV Markov properties C2b and C3b discussed by Drton [10]. Note that conditions (6) are bi-directed Markov properties that describe a local independence assumption, the independencies (7) of a set of manifest variables from a set of latent variables, at time t , are conditioned by the remaining latent variables, but not by the remaining observable variables.

As a matter of convenience, the mentioned mixed graph and the two component chain graph can be superimposed as shown in Fig. 1 to form a unique graph that we will refer to as mixed-chain graph for simplicity.

In all the examples throughout the paper, we implicitly assume that each node of the mixed-chain graph, corresponding to a latent variable, is parent of itself, even if the edge \odot is not reported.

Example 1. The mixed-chain graph (a) of Fig. 2 is associated to an MHMM where the latent variables are not contemporaneously independent, and each of them depends only on its own past, i.e. the encoded Granger noncausality conditions (4) are $E_1(t) \perp\!\!\!\perp E_{\{2,3\}}(t-1) | E_1(t-1)$, $E_2(t) \perp\!\!\!\perp E_{\{1,3\}}(t-1) | E_2(t-1)$, $E_3(t) \perp\!\!\!\perp E_{\{1,2\}}(t-1) | E_3(t-1)$, $E_{\{1,2\}}(t) \perp\!\!\!\perp E_3(t-1) | E_{\{1,2\}}(t-1)$,

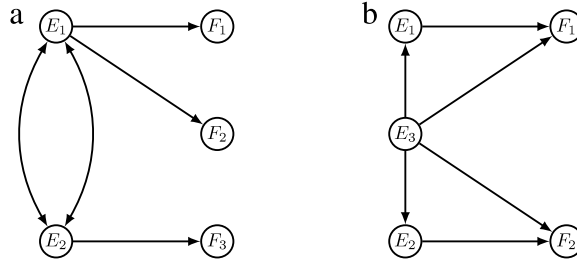


Fig. 3. Mixed-chain graphs associated to MHMMs illustrated in Example 2.

$E_{\{2,3\}}(t) \perp\!\!\!\perp E_1(t-1) | E_{\{2,3\}}(t-1)$, $E_{\{1,3\}}(t) \perp\!\!\!\perp E_2(t-1) | E_{\{1,3\}}(t-1)$; while the independencies (7) of manifest from latent variables are $F_{\{1,2\}}(t) \perp\!\!\!\perp E_3(t) | E_{\{1,2\}}(t)$ and $F_3(t) \perp\!\!\!\perp E_{\{1,2\}}(t) | E_3(t)$; the local independence (6) is $F_1(t) \perp\!\!\!\perp F_{\{2,3\}}(t) | E_{\{1,2,3\}}(t)$. The mixed-chain graph (b) of Fig. 2 encodes that the latent variables of the three dimensional MHMM are Granger-caused reciprocally but are contemporaneously independent, i.e. $E_1(t) \perp\!\!\!\perp E_2(t) \perp\!\!\!\perp E_3(t) | E_{\{1,2,3\}}(t-1)$; moreover, every observable variable depends only on its own latent variable, i.e. $F_1(t) \perp\!\!\!\perp E_{\{2,3\}}(t) | E_1(t)$, $F_2(t) \perp\!\!\!\perp E_{\{1,3\}}(t) | E_2(t)$, $F_3(t) \perp\!\!\!\perp E_{\{1,2\}}(t) | E_3(t)$, $F_{\{1,2\}}(t) \perp\!\!\!\perp E_3(t) | E_{\{1,2\}}(t)$, $F_{\{2,3\}}(t) \perp\!\!\!\perp E_1(t) | E_{\{2,3\}}(t)$, $F_{\{1,3\}}(t) \perp\!\!\!\perp E_2(t) | E_{\{1,3\}}(t)$; and, at every time point, the observable variables given the latent states are independent, i.e. $F_1(t) \perp\!\!\!\perp F_2(t) \perp\!\!\!\perp F_3(t) | E_{\{1,2,3\}}(t)$.

Example 2. The mixed-chain graph (a) in Fig. 3 is associated with an MHMM encoding the dynamic relations of three observable time series that depend on two latent factors. The observable time series can represent, for example, the sales levels of three products that can be interpreted as indicators of the motivational latent states of the customers behaviors. The distributions of the motivational states may be of more interest than those of the product sales. The graph underlies that the first two products belong to the same category so their sales series (F_1, F_2) depend on the latent variable E_1 (i.e. $F_{\{1,2\}}(t) \perp\!\!\!\perp E_2(t) | E_1(t)$), the sales of the third product respond to a different latent variable E_2 (i.e. $F_3(t) \perp\!\!\!\perp E_1(t) | E_2(t)$); the observable sales levels are independent given the latent states (i.e. $F_1(t) \perp\!\!\!\perp F_2(t) \perp\!\!\!\perp F_3(t) | E_{\{1,2\}}(t)$); the two motivational latent processes are G-caused reciprocally, so that the past motivational states influence the actual state of each latent variable, but they are contemporaneously independent (i.e. $E_1(t) \perp\!\!\!\perp E_2(t) | E_{\{1,2\}}(t-1)$).

The mixed-chain graph (b) in Fig. 3 can describe a model for financial series where specific and generic latent effects exist. The series F_1, F_2 can indicate the trading patterns of two financial traded shares in two different financial sectors. The presence or the absence of trading may depend on a specific latent aspect typical of each financial sector where the trading takes place (e.g. the turbulence of the sector), so each observable variable depends on one specific latent variable: F_1 on E_1 , F_2 on E_2 , but all the trading patterns may be influenced by one common unobservable variable (E_3) such as Market turbulence with states corresponding to calm or turbulent Market phases (i.e. $F_1(t) \perp\!\!\!\perp E_2(t) | E_{\{1,3\}}(t)$, $F_2(t) \perp\!\!\!\perp E_1(t) | E_{\{2,3\}}(t)$). There is local independence $F_1(t) \perp\!\!\!\perp F_2(t) | E_{\{1,2,3\}}(t)$. Moreover, the past Market turbulence can affect the current turbulence of each financial sector (G-causality) whereas a specific latent variable of a sector does not Granger cause the other specific latent variable and the turbulence of the Market, (i.e. $E_3(t) \perp\!\!\!\perp E_{\{1,2\}}(t-1) | E_3(t-1)$, $E_{\{1,3\}}(t) \perp\!\!\!\perp E_2(t-1) | E_{\{1,3\}}(t-1)$, $E_{\{2,3\}}(t) \perp\!\!\!\perp E_1(t-1) | E_{\{2,3\}}(t-1)$). Finally there is no contemporaneous relation among the three turbulences (i.e. $E_1(t) \perp\!\!\!\perp E_2(t) \perp\!\!\!\perp E_3(t) | E_{\{1,2,3\}}(t-1)$).

One immediate consequence of Theorem 1 and Definition 2 is the following result.

Corollary 1. If the MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ is Markov wrt a mixed graph G and a chain graph G^* , such that $pa_G(\mathcal{T}) = \mathcal{T}$ and $pa_{G^*}(\mathcal{R}) = \mathcal{T}$, then the marginal process $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$, $\mathcal{T} \subset \mathcal{U}$, $\mathcal{R} \subset \mathcal{V}$, is an MHMM.

Proof. Conditions (4) and (7) for $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ being Markov with respect to the graphs where $pa_G(\mathcal{T}) = \mathcal{T}$ and $pa_{G^*}(\mathcal{R}) = \mathcal{T}$ coincide with the conditions (1) and (2) ensuring that $(\mathbf{E}_{\mathcal{T}}, \mathbf{F}_{\mathcal{R}})$, $\mathcal{T} \subset \mathcal{U}$, $\mathcal{R} \subset \mathcal{V}$, is still a hidden process.

3.2. Special cases of graphical MHMMs: linked and coupled MHMMs

To further illustrate the richness of the proposed models, we discuss some special cases of graphical MHMMs that generalize the linked and coupled hidden Markov models applied in the literature in several contexts such as, for example, speech recognition [22], human gesture [2], voicing and speech detection [1], disease interactions [20], among others.

In our generalized version of linked hidden Markov models, every set of a partition of the latent variables is affected only by its own past, while in the coupled hidden Markov models, these sets are independent conditionally on the past of every latent variable. In all mentioned models, every set of the partition of the latent variables influences one and only one set of a partition of the manifest variables and the sets of the observable variables are independent given the latent states.

A formal definition is presented below.

Definition 3. An MHMM $(\mathbf{E}_{\mathcal{U}}, \mathbf{F}_{\mathcal{V}})$ which is Markov wrt a mixed graph G and a chain graph G^* is a linked MHMM with l components if, for a partition of the latent variables $\mathcal{U} = \bigcup_{i=1}^l \mathcal{T}_i$ and a partition of the observable variables $\mathcal{V} = \bigcup_{i=1}^l \mathcal{R}_i$,

$i = 1, \dots, l$, it holds that: $pa_G(\mathcal{T}_i) = \mathcal{T}_i$, $pa_{G^*}(\mathcal{R}_i) = \mathcal{T}_i$ and $sp_{G^*}(\mathcal{R}_i) = \mathcal{R}_i$. If the condition $pa_G(\mathcal{T}_i) = \mathcal{T}_i$ is replaced by $sp_G(\mathcal{T}_i) = \mathcal{T}_i$, the process $(\mathbf{E}_u, \mathbf{F}_v)$ Markov wrt such graphs is a coupled MHMM.

When the conditions (5), (4), (6) and (7) are specified in the special case of Definition 3, we deduce that a linked MHMM satisfies the following independencies for all the unions of variables $\mathcal{U}_s = \bigcup_{i \in s} \mathcal{T}_i$, $\mathcal{V}_s = \bigcup_{i \in s} \mathcal{R}_i$, $s \subset \{1, 2, \dots, l\}$

$$E_{\mathcal{U}_s}(t) \perp\!\!\!\perp E_{\mathcal{U} \setminus \mathcal{U}_s}(t-1) | E_{\mathcal{U}_s}(t-1) \quad (8)$$

$$F_{\mathcal{V}_s}(t) \perp\!\!\!\perp E_{\mathcal{U} \setminus \mathcal{U}_s}(t) | E_{\mathcal{U}_s}(t) \quad (9)$$

$$F_{\mathcal{R}_1}(t) \perp\!\!\!\perp F_{\mathcal{R}_2}(t) \perp\!\!\!\perp \dots \perp\!\!\!\perp F_{\mathcal{R}_l}(t) | E_{\mathcal{U}}(t). \quad (10)$$

In a coupled MHMM, condition (8) is replaced by

$$E_{\mathcal{T}_1}(t) \perp\!\!\!\perp E_{\mathcal{T}_2}(t) \perp\!\!\!\perp \dots \perp\!\!\!\perp E_{\mathcal{T}_l}(t) | E_{\mathcal{U}}(t-1). \quad (11)$$

The above independence (8) corresponds to the G-noncausality condition (4) specified in the special case of Definition 3, so the conditional independence (9) corresponds to (7), condition (10) to the local independence (6), while (11) to the contemporaneous independence (5).

It is worth noting that according to Theorem 1, the marginal processes $(\mathbf{E}_{\mathcal{U}_s}, \mathbf{F}_{\mathcal{V}_s})$, $s \subset \{1, 2, \dots, l\}$, of a linked MHMM are still MHMMs, but this is not true for a coupled MHMM.

For instance, in Fig. 2, the mixed-chain graph (a) corresponds to a linked MHMM with two components and the mixed-chain graph (b) to a coupled MHMM with three components.

In Section 5, the hypotheses defining linked and coupled MHMMs will be tested on real data.

3.3. Equivalent mixed-chain graphs

Standard hidden Markov models, and generally mixture models, suffer from a problem of non-identifiability due to the invariance with respect to the relabeling of hidden states. For this reason, MHMMs are identifiable only up to switching the latent categories and the labels of the latent variables. Due to the invariance with respect to switching labels of the latent variables, there are equivalent mixed-chain graphs in the sense that they correspond to the same MHMM. Below we further clarify this point.

Two mixed-chain graphs, say G_1 and G_2 , with the same sets of nodes, are equivalent if there is a bijection ν which maps the set of nodes corresponding to the latent variables onto itself ensuring that: (i) the nodes E_i of G_1 and $\nu(E_i)$ of G_2 are associated to latent variables with the same number of states; (ii) if E_i, E_j are connected by an edge in G_1 then $\nu(E_i)$ and $\nu(E_j)$ are connected by the same type of edge in G_2 ; (iii) if E_i, E_j are not connected by any edges in G_1 then $\nu(E_i)$ and $\nu(E_j)$ are not connected in G_2 ; (iv) if E_i is a parent of F_j in G_1 then $\nu(E_i)$ is a parent of F_j in G_2 .

It is evident that two mixed-chain graphs, equivalent according to the previous conditions, encode conditional independencies that differ only for the labels assigned to the latent variables.

An example illustrates the mentioned concepts.

Example 3. The first line (graphs a, b, c) of Fig. 4 shows three equivalent mixed-chain graphs corresponding to an MHMM where a latent variable affects both the observable variables (generic latent effect) and each manifest variable is governed by its specific latent variable (specific latent effect). In the second line (graphs d, e) of Fig. 4, the two equivalent mixed-chain graphs identify the MHMM where two latent variables affect only the first two observable variables and another latent variable is specific for the third observable variable. Moreover, the latent variable specific for F_3 is contemporaneously independent of the other two latent variables, and there is no Granger-causality between this variable and one of the remaining unobservable variables.

3.4. The number of latent variables

A consistent estimation of the number s of latent states in a hidden Markov model is a prerequisite to the parameter estimation (Cappé [3]) and here we will assume that s is known or that it has been estimated, for example, as shown in Celeux and Durand [4]. When s is given and it is not a prime number, what remains is to establish how many latent variables are compatible with the fixed s . For example, $s = 12$ can be provided by two dichotomous variables and one variable with three categories, or two latent variables with four and three categories, respectively, or a couple of latent variables, one with six and another with two categories.

In the proposed approach, the number of latent variables is a way to tune the richness of the family of models and limiting it would reduce the ability of formulating hypotheses corresponding to conditional independencies of interest, as illustrated at the end of Example 4.

Basically, we believe that the number of latent variables involved in the model should be chosen a priori on the grounds of considerations specific for the data at hand or the aims of the analysis. Nevertheless, the necessity of comparing models with the same number of states s but with a different number of latent variables cannot be excluded a priori. In these cases, the models to be compared can be equivalent, nested or non-nested as shown in Example 4.

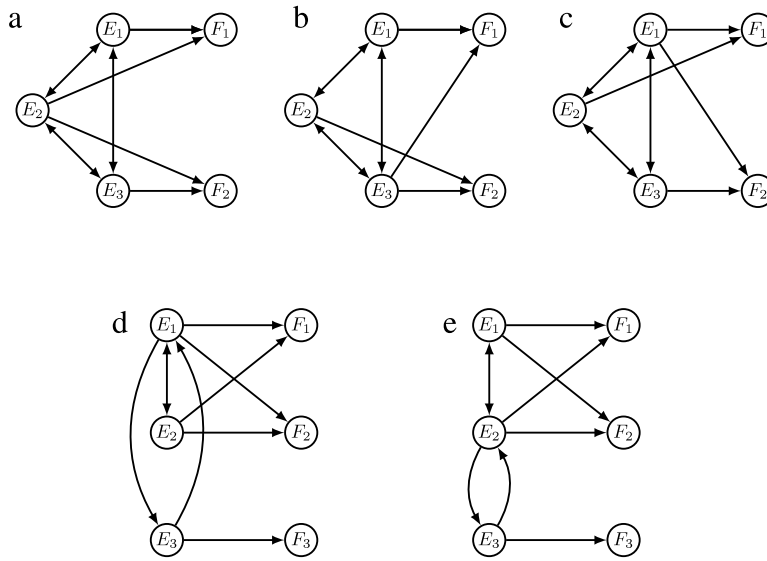


Fig. 4. Equivalent mixed-chain graphs.

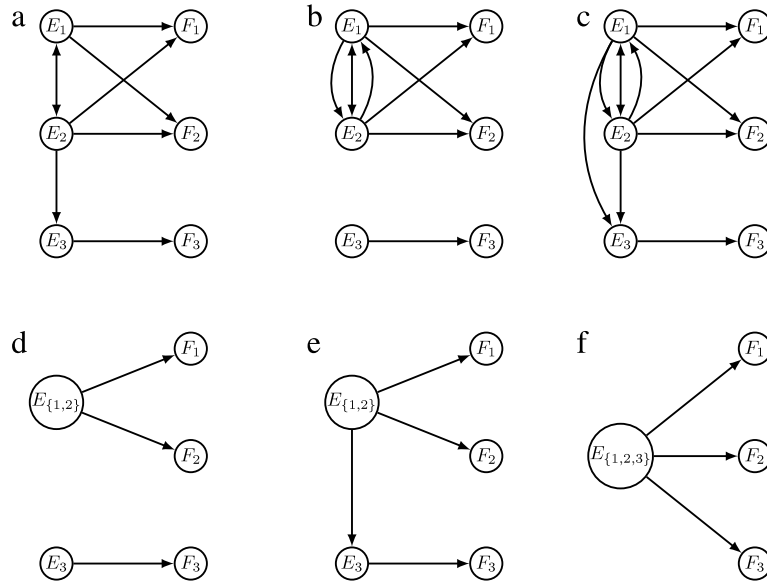


Fig. 5. Mixed-chain graphs associated to equivalent, nested or non-nested MHMMs.

The non equivalent models can be compared through AIC or BIC criteria and if they are nested via standard maximum likelihood based testing methods.

On the other hand, the comparison among models with a different number s of states and a different number of latent variables is more controversial, encounters parameter identifiability problems and deserves an in-depth study.

Example 4. The first three graphs (a), (b), (c) of Fig. 5 are associated to MHMMs with three dichotomous latent variables and graphs (c) and (d) describe models with two latent variables, one binary and another with four states. These last two models depicted in graphs (c) and (d) may be the result of collapsing two latent variables in the first three models associated to graphs (a), (b) and (c). Note that the models represented by graphs (a) and (b) are non-nested (every graph has at least one edge which is missing in the other graph). On the contrary, models corresponding to graphs (a) and (b) are nested in the model associated to graph (c). The MHMM of graph (e) is equivalent to the model of graph (c) because the latent variables, represented by two distinct nodes in graph (c) and collapsed in only one node in graph (e), are connected by all the three types of edges in (c) and have the same connections with the remaining variables in both graphs (c) and (e). It follows that the models of graphs (a) and (b) are nested in the model corresponding to graph (e). The model of graph (d) is equivalent to the model associated to graph (b) for the same reasons, so the MHMMs represented by graphs (a) and (d) are non-nested.

Finally, all the previous models are nested in the conventional HMM with one latent variable with eight states associated to graph (f).

Moreover, consider for instance the need of testing the hypothesis of reciprocal G-noncausality of two latent variables E_1 and E_2 . Such hypothesis is encoded in graph (a) but can also be easily introduced, and tested as shown in the next section, by deleting the directed edges among the nodes corresponding to E_1 and E_2 in graph (b) or (c). On the contrary, the same hypothesis would not easily testable starting from the graph (e) where instead of two distinct latent variables E_1 and E_2 there is only one $E_{\{1,2\}}$. Even more difficult would be testing the same hypotheses if the initial graph is (f) associated to the standard HMM with one latent variable. This example shows a case where reducing the number of latent variables shortens the possibilities of testing the independence hypotheses of interest.

4. A parameterization for MHMMs

Until now we have described the hypotheses underlying MHMMs, now we need to illustrate how to test such hypotheses. In particular, all the conditional independencies so far illustrated are related to the natural parameters of MHMMs: the state-dependent probabilities or the probabilities of the observable variables given the latent states, and the transition probabilities of the latent process.

In this section, we aim to parameterize both the state-dependent distributions and the transition probabilities of the latent process through marginal models and test the above mentioned conditional independence hypotheses by linear constraints on parameters.

Some notation is needed. One realization of the observable process \mathbf{F}_v at a given time is denoted by $\mathbf{f} = (f_1, f_2, \dots, f_s) \in \mathcal{F} = \times_{j \in \mathcal{V}} \mathcal{F}_j$, and the realization of the latent process \mathbf{E}_u is $\mathbf{e} = (e_1, e_2, \dots, e_r) \in \mathcal{E} = \times_{i \in \mathcal{U}} \mathcal{E}_i$. The time-homogeneous joint transition probabilities are denoted by $\phi(\mathbf{e}|\mathbf{e}')$ for every pair of states $\mathbf{e}' \in \mathcal{E}$, $\mathbf{e} \in \mathcal{E}$. Moreover, $\phi_{\mathcal{T}}(\mathbf{e}_{\mathcal{T}}|\mathbf{e}')$ are the marginal transition probabilities from state $\mathbf{e}' \in \mathcal{E}$ to state $\mathbf{e}_{\mathcal{T}}$ with components $e_i : i \in \mathcal{T} \subset \mathcal{U}$. Furthermore, $\varphi(\mathbf{f}|\mathbf{e})$ indicate the state-dependent probabilities that is the conditional probabilities of the observable variables given the latent state \mathbf{e} and $\varphi_{\mathcal{R}}(\mathbf{f}_{\mathcal{R}}|\mathbf{e})$ represent the marginal probabilities of the observable variables in the set \mathcal{R} , $\mathcal{R} \subset \mathcal{V}$, given the latent state \mathbf{e} .

In this work, we adopt a Gloneck–McCullagh multivariate logistic model [12] whose interaction parameters, involving the variables in the set \mathcal{P} , are log-linear interactions defined on the marginal distributions $\phi_{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}')$, $\varphi_{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e})$.

For every observable or latent categorical variable, the first category is called baseline. Any observation \mathbf{f} which includes categories at the baseline level for variables $j \notin \mathcal{J}$, $\mathcal{J} \subset \mathcal{V}$, is denoted by $(\mathbf{f}_{\mathcal{J}}, \mathbf{f}_{\mathcal{V} \setminus \mathcal{J}}^*)$. A similar notation holds for the latent state \mathbf{e} . For every non-empty subset \mathcal{P} of the observable variables \mathcal{V} and for every $\mathbf{f}_{\mathcal{P}} \in \times_{j \in \mathcal{P}} \mathcal{F}_j$, the baseline interactions $\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e})$, $\mathcal{P} \subseteq \mathcal{V}$, of the Gloneck–McCullagh marginal model for the observable variables are contrasts of logarithms of marginal state-dependent probabilities

$$\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}) = \sum_{\mathcal{K} \subseteq \mathcal{P}} (-1)^{|\mathcal{P} \setminus \mathcal{K}|} \log \varphi_{\mathcal{P}}(\mathbf{f}_{\mathcal{K}}, \mathbf{f}_{\mathcal{P} \setminus \mathcal{K}}^*|\mathbf{e}).$$

In order to model the dependence of the distribution of the observable variables on the states \mathbf{e} , we adopt the usual factorial expansion

$$\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}) = \sum_{\mathcal{Q} \subseteq \mathcal{U}} \theta^{\mathcal{P}, \mathcal{Q}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}_{\mathcal{Q}}). \quad (12)$$

Analogously, in the marginal model for the latent component of MHMMs, we define, for every $\mathcal{P} \subseteq \mathcal{U}$, the marginal parameters

$$\lambda^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}') = \sum_{\mathcal{K} \subseteq \mathcal{P}} (-1)^{|\mathcal{P} \setminus \mathcal{K}|} \log \phi_{\mathcal{P}}(\mathbf{e}_{\mathcal{K}}, \mathbf{e}_{\mathcal{P} \setminus \mathcal{K}}^*|\mathbf{e}')$$

and the factorial expansion

$$\lambda^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}') = \sum_{\mathcal{Q} \subseteq \mathcal{U}} \delta^{\mathcal{P}, \mathcal{Q}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}'_{\mathcal{Q}}). \quad (13)$$

4.1. Parametric constraints for conditional independencies

The properties of graphical models for MHMMs (Definition 2) correspond to zero restrictions on the parameters $\theta^{\mathcal{P}, \mathcal{Q}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}_{\mathcal{Q}})$ and $\delta^{\mathcal{P}, \mathcal{Q}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}'_{\mathcal{Q}})$, introduced in (12) and (13), as illustrated in the next theorem.

Theorem 2. For a latent model with strictly positive time-homogeneous transition probabilities, the Granger noncausality condition (4) is equivalent to $\delta^{\mathcal{P}, \mathcal{Q}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}'_{\mathcal{Q}}) = 0$ for all $\mathcal{P} \subseteq \mathcal{T}$, $\mathcal{Q} \not\subseteq pa_G(\mathcal{P})$, while the conditional contemporaneous independence (5) is equivalent to $\delta^{\mathcal{P}, \mathcal{Q}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}'_{\mathcal{Q}}) = 0$ for all $\mathcal{P} \notin \mathcal{B}(G)$, $\mathbf{e}_{\mathcal{P}} \in \times_{i \in \mathcal{P}} \mathcal{E}_i$, $\mathbf{e}'_{\mathcal{Q}} \in \times_{i \in \mathcal{Q}} \mathcal{E}_i$.

Moreover, if the state-dependent probabilities are strictly positive, the independence (6) is equivalent to $\theta^{\mathcal{P}, \mathcal{Q}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}_{\mathcal{Q}}) = 0$ for all $\mathcal{P} \notin \mathcal{B}(G^*)$, while condition (7) corresponds to $\theta^{\mathcal{P}, \mathcal{Q}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}_{\mathcal{Q}}) = 0$ for all $\mathcal{P} \subseteq \mathcal{R}$, $\mathcal{Q} \not\subseteq pa_{G^*}(\mathcal{P})$, $\mathbf{f}_{\mathcal{P}} \in \times_{j \in \mathcal{P}} \mathcal{F}_j$, $\mathbf{e}_{\mathcal{Q}} \in \times_{i \in \mathcal{Q}} \mathcal{E}_i$.



Fig. 6. Mixed-chain graphs whose Markov properties correspond to the constraints on marginal parameters described in Example 5.

Proof. The equivalence between the zero restrictions on δ parameters and conditions (4) and (5) follows from Proposition 2 by Colombi and Giordano [7]. These authors dealt with analogous Granger noncausality conditions and conditional contemporaneous independencies in the case of observed multivariate Markov chain. On the other hand, restrictions (6) and (7) are Markov properties of type IV (Drton [10]) proposed by Marchetti and Lupporelli [17] in the context of multivariate regression chain graph models; the equivalence between conditions (6) and (7) and the nullity of θ parameters follows from their Theorem 2. \square

The theorem allows a simple implementation of standard methods to fit and test MHMMs under the restrictions (4)–(7). The theorem enhances the use of the Gloneck–McCullagh marginal parameterization since all the conditions (4)–(7) are equivalent to linear constraints on the marginal parameters, the same restrictions under the log-linear parameterization would correspond to nonlinear constraints on the parameters.

An example clarifies which marginal parameters are restricted to zero in order to satisfy the conditional independencies that correspond to the Markov properties of the mixed-chain graph.

Example 5. The mixed-chain graph (a) of Fig. 6 encodes the following Markov properties: $E_1(t) \perp E_2(t-1) | E_1(t-1)$ so that E_1 is no Granger caused by E_2 and $F_i(t) \perp E_j(t) | E_i(t)$ for $i, j = 1, 2, i \neq j$ which reveals a specific effect of the first (second) latent variable on the first (second) observable variable. According to Theorem 1, these conditions ensure that the marginal process $(\mathbf{E}_1, \mathbf{F}_1)$ of the MHMM is still hidden Markov, but $(\mathbf{E}_2, \mathbf{F}_2)$ is not.

Moreover, the local independence condition $F_1(t) \perp F_2(t) | E_{1,2}(t)$ can be also read off the same graph. The mentioned conditional independencies are equivalent to the nullity of the following parameters: $\theta^{1,2}(f_1|e_2), \theta^{1,12}(f_1|e_1, e_2), \theta^{2,1}(f_2|e_1), \theta^{2,12}(f_2|e_1, e_2), \theta^{12}(f_1, f_2), \theta^{12,1}(f_1, f_2|e_1), \theta^{12,2}(f_1, f_2|e_2), \theta^{12,12}(f_1, f_2|e_1, e_2), \delta^{1,2}(e_1|e'_2), \delta^{1,12}(e_1|e'_1, e'_2)$.

In the graph (b) of Fig. 6 there is contemporaneous independence between the two latent variables, i.e. $E_1(t) \perp E_2(t) | E_1(t-1)$ that is equivalent to the zero restrictions on the parameters: $\delta^{12}(e_1, e_2), \delta^{12,1}(e_1, e_2|e'_1), \delta^{12,2}(e_1, e_2|e'_2), \delta^{12,12}(e_1, e_2|e'_1, e'_2)$. As for graph (a), the parameters: $\theta^{12}(f_1, f_2), \theta^{12,1}(f_1, f_2|e_1), \theta^{12,2}(f_1, f_2|e_2), \theta^{12,12}(f_1, f_2|e_1, e_2)$ are null according to the local independence condition.

4.2. Additivity hypotheses

In the framework of hidden Markov models with several latent and observable variables, other interesting hypotheses, that do not correspond to any conditional independence, can be easily formulated. These hypotheses reduce the number of parameters needed to parameterize the transition and state-dependent probabilities. A useful restriction that considerably simplifies the observation model is the hypothesis of *additivity* of the effects of the latent variables on the marginal interactions of the observable variables. This marginal additive dependence allows the interactions to be expressed by the factorial expansion

$$\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}) = \theta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}) + \sum_{k \in \mathcal{U}} \theta^{\mathcal{P},k}(\mathbf{f}_{\mathcal{P}}|e_k). \quad (14)$$

Note that under this hypothesis, the parameters $\theta^{\mathcal{P},Q}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}_Q)$ described in (12) are null if $|Q| > 1$. A similar additivity hypothesis can be used for the interactions of the latent model

$$\lambda^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}') = \delta^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}) + \sum_{k \in \mathcal{U}} \delta^{\mathcal{P},k}(\mathbf{e}_{\mathcal{P}}|e'_k). \quad (15)$$

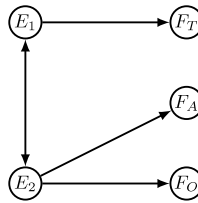
Another hypothesis is that of *invariant association* corresponding to the constraints $\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}}|\mathbf{e}) = \theta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}})$, if $|P| > 1$. According to this hypothesis, the interactions between observable variables do not depend on the states of the latent variables.

Similarly, the constraints for the invariant association can be also imposed to the interactions of the latent model: $\lambda^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}}|\mathbf{e}') = \delta^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}})$, if $|P| > 1$.

Table 1

Constrained latent and observation models for soft-drink data.

Latent model	obs. model	LRT	df	p-value	par	loglike	AIC
noGranger	saturated	4.0705	4	0.3965	112	−694.0009	1612.002
saturated	ci	33.0017	20	0.0337	96	−708.4665	1608.933
noGranger	ci	33.2346	24	0.0992	92	−691.9656	1567.931
saturated	ci + ind	89.6551	92	0.5497	24	−736.7932	1521.586
noGranger	ci + ind	92.6434	96	0.5780	20	−738.2874	1516.575
noGranger + ia	saturated	12.9629	7	0.0730	109	−698.4471	1614.894
saturated	ci + ia	76.2292	80	0.5987	36	−730.0803	1532.161
noGranger + ia	ci	33.5379	27	0.1798	89	−708.7346	1595.469
noGranger	ci + ia	77.4619	84	0.6795	32	730.6966	1525.393
noGranger + ia	ci + ia	77.9029	87	0.7467	29	−730.9171	1519.834
noGranger + ia	ci + ind	94.1538	99	0.6189	17	−739.0426	1512.085

**Fig. 7.** Mixed-chain graph for soft-drinks data.

5. Examples

In this section, we fit different MHMMs on two data sets. The EM algorithm used for estimating the models is described by Colombi and Giordano [6], and implemented in the R-package *hmmm* see [8,9].

The data set of a soft-drink company [5] (available also in the R-package *hmmm*) consists of a one-year time series of daily sales of soft-drinks: lemon tea, orange juice and apple juice, all with categories: low, medium, high level. Changes in sale outcomes over time can depend on time-varying unobservable factors and we consider an MHMM with two dichotomous latent variables to model these data.

The marginal latent processes of the MHMM are denoted by \mathbf{E}_1 , \mathbf{E}_2 , and the marginal observation components by \mathbf{F}_T , \mathbf{F}_O , \mathbf{F}_A .

Table 1 reports the likelihood ratio test statistic (*LRT*) for the models described in the first two columns against the unrestricted model, degrees of freedom (*df*) and *p-value*. The number of parameters (*par*), the log-likelihood value (*loglike*) and the Akaike criterion (*AIC*) are given in the last three columns.

The first five rows refer to MHMMs Markov with respect to mixed-chain graphs. Among these models, the MHMM with the lowest *AIC* value is also the most parsimonious model (5th row in Table 1) which corresponds to the graph illustrated in Fig. 7. This graph displays the hypotheses that the tea sales depend on \mathbf{E}_1 only and the orange and apple juices sales on \mathbf{E}_2 only. Moreover, it encodes condition (4) of reciprocal Granger noncausality (in short *noGranger*) for the latent variables: $E_1(t) \perp E_2(t-1) | E_1(t-1)$ and $E_2(t) \perp E_1(t-1) | E_2(t-1)$ which ensure that \mathbf{E}_1 and \mathbf{E}_2 are univariate Markov chains and the restriction (7) of conditional independencies among the observable and latent variables: $F_T(t) \perp E_2(t) | E_1(t)$ and $F_{\{O,A\}}(t) \perp E_1(t) | E_2(t)$ that we will refer to as *ci*. Note that, according to Theorem 1, these restrictions serve to affirm that the marginal processes $(\mathbf{F}_T, \mathbf{E}_1)$ and $(\mathbf{F}_{\{O,A\}}, \mathbf{E}_2)$ are still hidden Markov models. Another Markov property of the graph 7 is the local independence (6) (in short *ind*) of all the observable variables given the latent chain $F_T(t) \perp F_O(t) \perp F_A(t) | E_{\{1,2\}}(t)$. Finally, we observe that graph 7 is associated to a linked MHMM with two components subject to the additional condition $F_O(t) \perp F_A(t) | E_{\{1,2\}}(t)$.

An additional hypothesis mentioned in Section 4.2 has been also tested: the invariant association (*ia*). For the observation model, it means that the interactions of second and higher order of the observable variables do not depend on the latent states, i.e. $\eta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}} | \mathbf{e}) = \theta^{\mathcal{P}}(\mathbf{f}_{\mathcal{P}})$ for $|\mathcal{P}| > 1$, while for the latent model there is invariant association when second order interactions do not depend on the latent states in past occasions, i.e. $\lambda^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}} | \mathbf{e}') = \delta^{\mathcal{P}}(\mathbf{e}_{\mathcal{P}})$, for $|\mathcal{P}| > 1$. The MHMMs under this hypothesis are tested and the results reported in the last 6 rows of Table 1.

The MHMM with the lowest *AIC* value is the model corresponding to the graph of Fig. 7 with the additional constraints of the invariant association for the latent component.

The second data set, available at <http://archive.ics.uci.edu/ml/>, reports daily measurements (from December 2006 to November 2010) of electric power consumption in one household. The energy (in watt-hour of active energy) from the sub-meter 1 corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave; the energy from the sub-meter 2 is used for the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light; energy from sub-meter 3 corresponds to an electric water-heater and an air-conditioner. Here, the energy consumptions registered

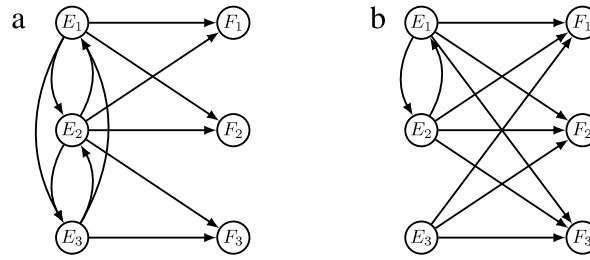


Fig. 8. Mixed-chain graphs for energy data.

over 4 years by the sub-meters 1, 2, and 3, are categorized in low and high levels according to whether the real measurements are under or over an established threshold, and F_1 , F_2 and F_3 indicate the resulting categorical time series.

The categorized consumptions for kitchen, laundry and heater are the manifest variables used as indicators of the true energy demand for eating, cleaning and heating/cooling that is unobservable directly. The latent real demand for energy, in fact, can be lower (higher) than the measured level due, for example, to waste (saving need). Therefore, we model the energy data through an MHMM involving three latent variables with two states and three observable binary variables.

Several hypotheses have been considered to describe relations among the observable series and the latent factors, but below we will focus only on those models that perform better in terms of interpretability, parsimony and fitting.

Let us start by considering the saturated MHMM, i.e. without any restrictions, which attains $AIC = 5493.015$, $loglike = -2634.507$, $par = 112$.

Among the alternatives to the saturated model, a simple and intuitive linked MHMM considers that each of the three observable series of energy consumptions is an indicator of its specific latent variable and the observable variables are locally independent. Regarding the latent model, the three unobservable variables affect each others at the same time but each one depends only on its own past (no Granger causality and no contemporaneous independence). This linked MHMM ($AIC = 5621.251$, $loglike = -2778.626$, $par = 32$) cannot be preferred to the saturated model. Also the coupled MHMM with every observable variable depending on its own latent variable, with contemporaneous independence and Granger causality among the three latent variables ($AIC = 5615.272$, $loglike = -2777.636$, $par = 30$) is outperformed by the saturated model.

Unfortunately, although combined with other different hypotheses on the latent component of the MHMM, the model restricted under the assumption of a specific effect of each latent variable on just one energy consumption series shows an unsatisfactory performance with an higher AIC value than that of the saturated model.

In the following models, the effects of two or more latent variables on the same observable variable will be considered additive as described in the expression (14).

An interesting model with a better fit assumes that: there is a specific effect of one latent variable on the energy consumptions for kitchen and laundry, another latent variable influences the energy consumptions for the heater, and there is a generic effect since a third latent variable affects all the manifest variables. This observation model is combined with the assumption that Granger causality and contemporaneous independence exist among all the latent variables. The model with these restrictions has $AIC = 5470.029$, $loglike = -2702.014$, $par = 33$. Its corresponding mixed-chain graph is reported in Fig. 8(a).

The MHMM which seems the most suitable among several models we fitted for representing the dynamics of the energy consumptions assumes that: the first two latent variables are G-caused reciprocally, the third one depends only on its past, and there is contemporaneous independence among the three latent factors; moreover, there is local independence among energy consumption for eating, cleaning and air-conditioning which seem to depend on all three latent variables ($AIC = 5459.77$, $loglike = -2707.883$, $par = 22$). Its mixed-chain graph is reported in Fig. 8(b). This model is nested in an MHMM which shows a similar fit but admits that there is Granger causality among all the latent variables, the other hypotheses being equal ($AIC = 5467.374$, $loglike = -2697.687$, $par = 36$). When the two models are compared, the LRT ($LRT = 20.4$, $df = 14$, $p-value = 0.118$) confirms that the most parsimonious model can be retained coherently with the AIC results.

Acknowledgments

This research was partially funded by STAR project (University of Naples Federico II-CUP: E68C13000020003, Italy) and by project PRIN-SURWEY (grant 2012F42NS8, Italy).

References

- [1] S. Basu, A linked-HMM model for robust voicing and speech detection, in: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, 2003.
- [2] M. Brand, N. Oliver, A. Pentland, Coupled hidden Markov models for complex action recognition, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, pp. 994–999.

- [3] O. Cappé, E. Moulines, T. Rydén, *Inference in Hidden Markov Models*, Springer, New York, 2005.
- [4] G. Celeux, J.B. Durand, Selecting hidden Markov chain states number with cross-validated likelihood, *Comput. Statist.* 23 (2008) 541–564.
- [5] W.K. Ching, E.S. Fung, M.K. Ng, A multivariate Markov chain for categorical data sequences and its applications in demand predictions, *IMA J. Manag. Math.* 17 (2002) 187–199.
- [6] R. Colombi, S. Giordano, Lumpability for discrete hidden Markov models, *AStA Adv. Stat. Anal.* 95 (2011) 293–311.
- [7] R. Colombi, S. Giordano, Graphical models for multivariate Markov chains, *J. Multivariate Anal.* 107 (2012) 90–103.
- [8] R. Colombi, S. Giordano, M. Cazzaro, R package hmmm: Hierarchical Multinomial Marginal Models, 2012, <http://CRAN.R-project.org/package=hmmm>.
- [9] R. Colombi, S. Giordano, M. Cazzaro, hmmm: An R package for hierarchical multinomial marginal models, *J. Statist. Softw.* 59 (11) (2014) 1–25.
- [10] M. Drton, Discrete chain graph models, *Bernoulli* 15 (3) (2009) 736–753.
- [11] J.P. Florens, M. Mouchart, J.M. Rolin, Noncausality and marginalization of Markov processes, *Econometric Theory* 9 (1993) 241–262.
- [12] G.F.V. Glonek, P. McCullagh, Multivariate logistic models, *J. Roy. Statist. Soc. Ser. B* 57 (1995) 533–546.
- [13] T. Koski, *Hidden Markov Models for Bioinformatics*, Kluwer Academic Publishers, London, 2001.
- [14] S.L. Lauritzen, *Graphical Models*, Clarendon Press, Oxford, 1996.
- [15] I.L. MacDonald, W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman & Hall, London, 1997.
- [16] I.L. MacDonald, W. Zucchini, *Hidden Markov Models for Time Series, An Introduction using R*, Chapman & Hall, London, 2009.
- [17] G.M. Marchetti, M. Lupporelli, Chain graph models of multivariate regression type for categorical data, *Bernoulli* 17 (2011) 827–844.
- [18] G. Petris, S. Petrone, P. Campagnoli, *Dynamic Linear Models with R*, Springer, New York, 2009.
- [19] T. Richardson, Markov properties for acyclic directed mixed graphs, *Scand. J. Stat.* 30 (2003) 145–157.
- [20] C. Sherlock, T. Xifara, S. Telfer, M. Begon, A coupled hidden Markov model for disease interactions, *J. Roy. Statist. Soc. Ser. C* 62 (2013) 609–627.
- [21] R. van Handel, Observability and nonlinear filtering, *Probab. Theory Related Fields* 145 (2009) 35–74.
- [22] P. Wiggers, *Hidden Markov models for automatic speech recognition and their multimodal applications*, Delft University of Technology, Delft, 2001.
- [23] W. Zucchini, P. Guttorp, A hidden Markov model for space–time precipitation, *Water Resources Res.* 27 (1991) 1917–1923.