Bernoulli **23**(2), 2017, 863–883
DOI: 10.3150/15-BEJ762

Type II chain graph models for categorical data: A smooth subclass

FEDERICA NICOLUSSI[1] and ROBERTO COLOMBI[2]

[1]*Dip. Discipline matematiche, finanza matematica e econometria, Catholic University of Milan, largo A. Gemelli, 1-20213 Milano, Italy. E-mail: federica.nicolussi@unimib.it*
[2]*Dip. di Ingegneria gestionale, dell'informazione e della produzione, University of Bergamo, viale Marconi 5-24044 Dalmine (BG), Italy. E-mail: roberto.colombi@unibg.it*

The Probabilistic Graphical Models use graphs in order to represent the joint distribution of q variables. These models are useful for their ability to capture and represent the system of independence relationships among the variables involved, even when complex. This work concerns categorical variables and the possibility to represent symmetric and asymmetric dependences among categorical variables. For this reason we use the Chain Graphical Models proposed by Andersson, Madigan and Perlman (*Scand. J. Stat.* **28** (2001) 33–85), also known as Chain Graphical Models of type II (GMs II). The GMs II allow for symmetric relationships typical of log-linear models and, at the same time, asymmetric dependences typical of Graphical Models for Directed Acyclic Graphs. In general, GMs II are not smooth, however this work provides a subclass of smooth GMs II by parametrizing the probability function through marginal log-linear models. Furthermore, the proposed models are applied to a data-set from the European Value Study for the year 2008 (EVS (2010)).

Keywords: categorical variables; Chain Graph Models; conditional indipendence models; marginal models

1. Introduction

The increasing use of graphical models is due to their ability to represent complex phenomena. The Probabilistic Graphical Models represent the joint probability function of q variables through a graph where each vertex of the graph corresponds to one variable and the arcs are indicators of dependences. Chain Graph Models use both directed and undirected arcs, thus are able to represent symmetrical or directional relationships simultaneously. This topic is largely discussed in literature, see, for instance [10,15].

With the aim to represent simultaneous independence relationships among a collection of categorical variables, we use the Chain Graph Models proposed by [1] also known as Graphical Models of type II (GMs II), see [6]. In a GM II, the variables are partitioned into different sets. Independences typical of log-linear models hold among the variables in the same set, while asymmetrical independences typical of DAG (Directed Acyclic Graph) hold among variables in different sets. It is important to observe that, the dependence between response and explanatory variables is studied marginally compared to other response variables.

A useful way to parametrize the joint distribution of categorical variables is given by the marginal log-linear models [3]; these models have been used to parametrize different types of chain graphical models, see, for instance, [12,14] and [8], but this parametrization, unfortunately, does not exist for the sub-class of GMs II which are non-smooth models [6]. As the parametric marginal models for categorical data have useful properties for the asymptotic theory of the ML estimators, we are interested in investigating the sub-class of GMs II's that can be parametrized by marginal models. A preliminary study of these models was debated in [13].

The work is organized as follows: Section 2 introduces the marginal log-linear models with their properties; Section 3 is dedicated to the GMs II and, after an introduction on basic notation, it proposes new equivalent Markov properties. Section 4 describes the smooth subclass of GMs II which is representable with marginal log-linear models. Finally, Section 5 presents an application on EVS data to illustrate the results obtained in this work.

2. Marginal log-linear models

Here we give only a brief introduction to the marginal log-linear models; for more details, specific to the context of chain graphical models see, for instance, [12] and [8].

Let us consider q categorical variables $V = \{X_1, X_2, \ldots, X_q\}$, with levels $d_1, d_2, \ldots, d_q$, respectively. The *marginal log-linear models* are models where log-linear interaction parameters are defined on marginal distributions. We will refer to these interactions, that are contrasts of logarithms of sum of probabilities, as *marginal parameters*, [3]. Any marginal parameter is distinguished by a pair of sets $(\mathcal{M}; \mathcal{L})$, where $\mathcal{L} \subseteq \mathcal{M} \subseteq V$. The so-called *marginal set* $\mathcal{M}$ specifies the marginal distribution where the parameter is evaluated. The so-called *interaction set* $\mathcal{L}$, is the set of variables involved in the interaction. The collection of interactions defined in $\mathcal{M}$ concerning the interaction set $\mathcal{L}$ are staked in the vector:

$$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} = \mathbf{C}_{\mathcal{L}}^{\mathcal{M}} \log \mathbf{M}_{\mathcal{L}}^{\mathcal{M}} \boldsymbol{\pi}, \tag{2.1}$$

where $\mathbf{C}_{\mathcal{L}}^{\mathcal{M}}$, $\mathbf{M}_{\mathcal{L}}^{\mathcal{M}}$ are a *contrast* and a *marginalization* matrix and $\boldsymbol{\pi}$ is the vector of strictly positive joint probabilities of the q variables [2].

All the marginal parameters are collected in the vector $\boldsymbol{\eta}$ obtained stacking all the previous $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$.

Definition 1. *A class $\mathcal{H} = \{\mathcal{M}_1, \ldots, \mathcal{M}_s\}$ of marginal sets, where $\mathcal{M}_s = V$ and $\mathcal{M}_i \nsubseteq \mathcal{M}_j$ if $j < i, \forall i, j = 1, \ldots, s$, is called hierarchical family of marginal set.*

Definition 2. *Given a hierarchical family of marginal sets, the vector of parameters $\boldsymbol{\eta}$ is complete if there is exactly one $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$ for all $\mathcal{L} \subseteq V$; it is hierarchical if the marginal set $\mathcal{M}$ of $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$ is the first in $\mathcal{H}$ which contains the variables in $\mathcal{L}$.*

A marginal model is characterized by a vector of hierarchical and complete parameters $\boldsymbol{\eta}$ and a remarkable property is that $\boldsymbol{\eta}$ is a smooth parametrization of the probability function of the variables in V (see [3], Theorem 2). On the other hand, marginal models are not able to represent all lists of conditional independences $\{A_i \perp B_i | C_i, i = 1, \ldots, k\}$. Concerning this problem, [14], gave a sufficient condition according to which a list of independences is representable by a marginal model. The contribution of this paper is founded on this result reported by the following

Theorem 2.1. Let us consider, for each independence $A_i \perp B_i | C_i$, the subclass D_i of interaction sets

$$D_i = \mathcal{P}(A_i \cup B_i \cup C_i) \setminus \bigl(\mathcal{P}(A_i \cup C_i) \cup \mathcal{P}(B_i \cup C_i)\bigr), \tag{2.2}$$

where $\mathcal{P}(S)$ is the power set of S. The elements of this class contain at least one element of A_i, at least one element of B_i and possibly elements of C_i. Let $\mathcal{M}(\mathcal{L})$ be the first marginal set in the hierarchical class $\mathcal{H} = \{\mathcal{M}_1, \dots, \mathcal{M}_m\}$ which contains the interaction $\mathcal{L}$, $\forall \mathcal{L} \subseteq V$.

Theorem 2.1. *Let us consider q variables, a hierarchical class of marginal sets $\mathcal{H}$ and the conditional independences list $\{A_i \perp B_i | C_i, i = 1, \dots, k\}$. The class of probability distribution functions of the q variables that satisfies the previous conditional independence system is equivalent to the marginal model where*

$$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})} = 0, \qquad \forall \mathcal{L} \in \bigcup_{i=1}^{k} D_i \quad \textit{and} \quad \mathcal{M}(\mathcal{L}) \in \mathcal{H} \tag{2.3}$$

if the next condition is satisfied

$$C_i \subseteq \mathcal{M}(\mathcal{L}) \subseteq (A_i \cup B_i \cup C_i) \qquad \forall \mathcal{L} \in D_i, i = 1, \dots, k. \tag{2.4}$$

In this work marginal sets, interaction sets and, more generally, sets of nodes will be denoted by enclosing the list of elements in round brackets while the members of a family of sets will be enclosed in curly brackets.

Example 1. Let us consider a collection of four variables V_1, V_2, V_3, V_4 and the independences $V_1 \perp V_3 | V_4$ and $V_4 \perp V_1, V_2 | V_3$. Let us take a hierarchical class of marginal sets, for instance, $\mathcal{H} = \{(V_1, V_3, V_4); (V_1, V_2, V_3, V_4)\}$. In order to verify the condition (2.4), we need the classes D_i, $i = 1, 2$. Thus, from the independence $V_1 \perp V_3 | V_4$, we have $D_1 = \{(V_1, V_3); (V_1, V_3, V_4)\}$, and from independence $V_4 \perp V_1, V_2 | V_3$ we have $D_2 = \{(V_1, V_4); (V_2, V_4); (V_1, V_2, V_4); (V_1, V_3, V_4); (V_2, V_3, V_4); (V_1, V_2, V_3, V_4)\}$, see formula (2.2). In the first case, the condition (2.4) becomes $(V_4) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (V_1, V_3, V_4)$ that always holds since $\mathcal{M}(\mathcal{L}) = (V_1, V_3, V_4)$ for all $\mathcal{L} \in D_1$. Thus, the vectors of parameters to constrain to zero are $\boldsymbol{\eta}_{13}^{134}$, $\boldsymbol{\eta}_{134}^{134}$. According to the second independence the condition (2.4) becomes $(V_3) \subseteq \mathcal{M}(\mathcal{L}) \subseteq (V_1, V_2, V_3, V_4)$. Even in this case, the condition holds for all $\mathcal{L} \in D_2$ thus the second independence is represented by annulling the vectors $\boldsymbol{\eta}_{14}^{134}$, $\boldsymbol{\eta}_{134}^{134}$, $\boldsymbol{\eta}_{24}^{1234}$, $\boldsymbol{\eta}_{234}^{1234}$, $\boldsymbol{\eta}_{124}^{1234}$, $\boldsymbol{\eta}_{1234}^{1234}$. Note that, if we introduce the statement $V_4 \perp V_1 | V_2, V_3$ which is implied by $V_4 \perp V_1, V_2 | V_3$, the interaction referring to $\mathcal{L} = (V_1, V_3, V_4)$ must be set to zero also to satisfy this last independence. But, this time the condition (2.4) does not hold since $(V_2, V_3) \nsubseteq \mathcal{M}(\mathcal{L}) = (V_1, V_3, V_4)$. Thus, if we take into account the third independence statement (i.e., unnecessary) we can not tell if the list of independences is parametrizable by a marginal model.

A closely related result, about the equivalence of conditional independence statements and constraints on marginal log-linear parameters, that does not require the specification of $\mathcal{H}$, is given by [9].

3. Graphical models of type II

Section 3.1 introduces the main notation and definitions on graph theory useful to understand the graphical models of type II described in Section 3.2; for a more general treatment see [10]. Section 3.3 introduces a new equivalent formulation of the Markov properties of the GMs II.

3.1. Graph theory

Graphs are mathematical objects defined by two sets $G = \{V, E\}$, where $V = \{V_1, \dots, V_q\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges or arcs that can be both directed or undirected. Two vertices are *adjacent* if they are joined by an undirected edge. Given a subset A of V, the set of all vertices both not included in A and adjacent to at least one of the vertices in A, is called *set of neighbours* of A, $\mathrm{nb}(A)$. The *neighbourhood* of A is defined by $\mathrm{Nb}(A) = \mathrm{nb}(A) \cup A$.

On the other hand, when there is a directed arc from a vertex V_i to a vertex V_j, the first vertex is called *parent* of V_j and V_j is called *child*. Given a subset A of V, the set $\mathrm{pa}(A)$ of *parents* of A is the collection of all vertices with at least one child in A. We define the set $\mathrm{ch}(A)$ of children of A as the set of all vertices with at least one parent included in A.

A *directed cycle* is an ordered sequence of vertices, all joined by direction preserving directed arcs (directed-path), starting and ending in the same vertex. A *semi-directed cycle* is an ordered sequence of vertices, joined by both direction preserving directed and undirected arcs (semi-directed path), which starts and ends in the same vertex.

A *Chain Graph* (CG) is a graph that can include both directed and undirected arcs without any directed or semi-directed cycle. A CG is decomposable into *Chain Components*, denoted by $T_1, \dots, T_s$. Within these connected chain components, there are only undirected arcs and between two components there are only directed arcs in the same direction.

Given a CG, the associated *Directed Graph* is a directed acyclic graph where the components $T_1, T_2, \dots$ act for the vertices, and there is a direct arc linking T_j to T_h if at least an element of T_h is a child of an element of T_j. The definition of parents and children also applies to the associated Directed Graph. Thus, for instance, for a component T_h, the parent $\mathrm{pa}_D(T_h)$ is the set of the chain components having as children T_h, $h = 1, \dots, s$. The chain components are ordered in such way that if $j < h$ there is not any directed arc from T_h to T_j.

Example 2. An example of chain graphs is represented in Figure 1 where we can recognize three components: $T_1 = (1)$, $T_2 = (2, 3)$ and $T_3 = (4, 5, 6)$. Within the component T_3, the vertices 4 and 5 are adjacent and the set of neighbours of 5 is $\mathrm{nb}(5) = (4, 6)$. The vertex 2 is child of 1 and parent of 4. The set of parents of $(4, 6)$ is $\mathrm{pa}(4, 6) = (2, 3)$ and the set of parents of the component T_3 is $\mathrm{pa}_D(T_3) = T_2 = (2, 3)$.

Any component T_h can be partitioned in two subsets, respectively, the set CH_h of children and the set NC_h of vertices that are not children. In the first set, there are the vertices $V_j \in T_h$ such that $\mathrm{pa}(V_j) \neq \varnothing$ while in NC_h there are the $V_j \in T_h \setminus \mathrm{CH}_h$.

Example 3 (Continuation of Example 2). In the component T_1, the set CH_1 is empty and $\mathrm{NC}_1 = (1)$. In the component T_2 the set of children is $\mathrm{CH}_2 = (2, 3)$ and NC_2 is empty. Finally, in the component T_3 we have $\mathrm{CH}_3 = (4, 6)$ and $\mathrm{NC}_3 = (5)$.

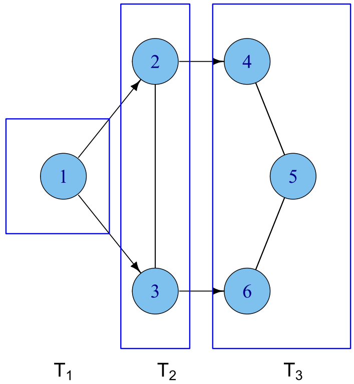

Figure 1. Chain graphs with set of vertex $V = (1, 2, 3, 4, 5, 6)$ and set of edges $E = \{(1, 2); (1, 3); (2, 3); (3, 2); (2, 4); (3, 6); (4, 5); (5, 4); (5, 6); (6, 5)\}$.

A subset A of V is *complete* if every pair of vertices of A is connected by an undirected edge. A complete subset A of V is a *clique* if it is maximal, that is if there are no complete sets containing it. We denote the *family of complete subsets* of a component T_h by $\mathcal{C}_h$ and the *family of cliques* by $\mathcal{C}l_h$.

Example 4 (Continuation of Example 2). As the component T_1 has only one vertex, $\mathcal{C}_1 = \mathcal{C}l_1 = \{(1)\}$. The two vertices of T_2 are adjacent, so $\mathcal{C}_2 = \{(2); (3); (2, 3)\}$ and the family of cliques of T_2 is $\mathcal{C}_2 = \{(2, 3)\}$. For the component T_3, the class of complete sets is $\mathcal{C}_3 = \{(4); (5); (6); (4, 5); (5, 6)\}$ and the family of cliques is $\mathcal{C}l_3 = \{(4, 5); (5, 6)\}$.

3.2. Graphical models of type II

Graphical models (GMs) use graphs to represent multidimensional dependence structures among variables. There are four types of GMs for chain graphs (CG), listed in [6], each of which represents different dependence systems. The Drton's GMs use the CG where the vertices act for the variables and the possible edges act for dependence relationships. When two vertices are connected by an undirected edge, it is possible to assume that the two linked variables are associated. On the other hand, when there is a directed arc, a dependence relationship among the linked variables can be assumed. Since their desirable features, in what follows we will use the Graphical models of type II (GMs II), introduced by [1] as generalization of both GMs for undirected graphs (UG) and GMs for directed acyclic graphs (DAG), for details see [10]. First, the grouping of variables in components allows to split the variables in "*purely explicative*" variables, "*purely response*" variables and "*intervening*" variables. Second, in the GMs II, the relationship among a variable and its explicative variables is considered marginally with respect to the variables in the same component. Finally, the association between the variables, within the same component, is modeled as it is done by log-linear models in contingency table analysis. The rules to read a list of conditional independences from a graph are called *Markov properties* and, for the GM II are the following three:

$$
\begin{aligned}
&\text{(C1)} && T_h \perp \left(\bigcup_{i<h} T_i \setminus \mathrm{pa}_D(T_h)\right)\Big| \mathrm{pa}_D(T_h), \\
&\text{(C2a)} && A \perp T_h \setminus \mathrm{Nb}(A) | \mathrm{pa}_D(T_h) \cup \mathrm{nb}(A) \qquad \forall A \subset T_h, \\
&\text{(C3b)} && A \perp \mathrm{pa}_D(T_h) \setminus \mathrm{pa}(A) | \mathrm{pa}(A) \qquad \forall A \subseteq T_h, \\
& && h = 1, \ldots, s.
\end{aligned}
\tag{3.1}
$$

The first Markov property, (C1), describes the independences between the chain components; the second, (C2a), reads the conditional independences within the components, and the third, (C3b) interprets the lack of directed arcs between variables in different components. The labels (C1), (C2a) and (C3b) are the same used in [6].

Example 5. Applying the Markov properties in formula (3.1) to the graph in Figure 1, we get, respectively, from the (C1) $4, 5, 6 \perp 1|2, 3$, from the (C2a) $4 \perp 6|2, 3, 5$ and from the (C3b) $5 \perp 2, 3$; $4 \perp 3|2$; $6 \perp 2|3$; $4, 5 \perp 3|2$ and $5, 6 \perp 2|3$. Note that, the first three (C3b) independences are implied by the last two and so $5 \perp 2, 3$; $4 \perp 3|2$; $6 \perp 2|3$; can be deleted from the list without loss of information.

Unfortunately, when the variables involved are categorical, these models are not necessarily smooth [6]. This means that the probability function of the q variables under the constraints given by the removed arcs does not always belong to a curved exponential family. Since the parametric marginal models introduced in Section 2 are smooth, in Section 4 we will propose a subclass of GMs II that can be parametrized as marginal models. In order to do this, it is convenient to use the equivalent lists of independences introduced in Section 3.3 in order to remove the statements that can lead to a "wrong" inconclusive declaration, as it is shown by Example 1 and more generally by [9]. Remind that two lists of independences are *equivalent* if all independences of one list can be obtained from the other list.

3.3. Alternative Markov properties for GMs II

In this section, we introduce two alternative Markov properties useful to introduce the results of Section 4. The Markov properties (C2*a), equivalent to the condition (C2a), have the property that children with complete neighbourhood do not belong to any conditioning set. In the Markov properties (C3*a), equivalent to the (C3a), the conditional independences are between sets of a partition of $\mathrm{pa}_D(T_h)$ and sub-sets of T_h that contain all the non-children. These properties will play a key role in the main theorems of Section 4.

An alternative condition for (C2a)

Let us consider the family $\mathcal{C}l_h$ of the r_h cliques of the hth component and split the elements $C_{i,h}$ of $\mathcal{C}l_h$ in two sets $C_{i,h} = B_{1i,h} \cup B_{2i,h}$, for $i = 1, \ldots, r_h$, in such way that: if $B_{1i,h}$ is not empty,

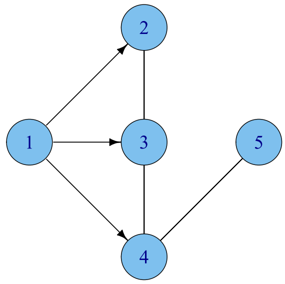

Figure 2. Chain graphs with set of vertices $V = (1, 2, 3, 4, 5)$ and set of edges $E = \{(1, 2); (1, 3); (1, 4); (2, 3); (3, 2); (3, 4); (4, 3); (4, 5); (5, 4)\}$.

$\mathrm{CH}_h \cap B_{1i,h} \neq \varnothing$ and $V_j \in B_{1i,h}$ if and only if $\mathrm{Nb}(V_j) = C_{i,h}$, while $B_{2i,h} = C_{i,h} \setminus B_{1i,h}$. Below we report an example which shows how to decompose the cliques $C_{i,h}$.

Example 6. Let us consider the graph in Figure 2. In the component $T_2 = (2, 3, 4, 5)$, we have the family of the cliques $\mathcal{C}l_2 = \{(2, 3); (3, 4); (4, 5)\}$. In the Table 1, we list elements of Cl_2 split according to the previous rules.

Definition 3. *The condition* (C2*a) *is described by the following list of independences*:

$$B_{1i,h} \perp T_h \setminus C_{i,h} | \operatorname{pa}_D(T_h) \cup B_{2i,h} \tag{3.2}$$

$\forall i = 1, \dots, r_h$ *and* $h = 1, \dots, s$.

$$(V_j \cup \mathbf{B}_{1,V_j,h}) \perp T_h \setminus \mathrm{Nb}(V_j \cup \mathbf{B}_{1,V_j,h}) | \operatorname{pa}_D(T_h) \cup \operatorname{nb}(V_j \cup \mathbf{B}_{1,V_j,h}), \tag{3.3}$$

$\forall V_j \in \bigcup_{i=1}^{r_h} B_{2i,h}$ *and* $h = 1, \dots, s$ *where* $\mathbf{B}_{1,V_j,h} = \operatorname{nb}(V_j) \cap (\bigcup_{i=1}^{r_h} B_{1i,h})$.

By definition the conditioning set $\operatorname{pa}_D(T_h) \cup B_{2i,h}$ of (3.2) can not contain children of T_h with complete neighbourhood. The following lemma shows that the same holds for the conditioning set of (3.3).

Table 1. List of the cliques $C_{i,2}$ with the respective partition (given by $B_{1i,2}$ and $B_{2i,2}$ sets) concerning the component 2 for the graph in Figure 2

$C_{i,2}$	$B_{1i,2}$	$B_{2i,2}$
2, 3	2	3
3, 4	$\varnothing$	3, 4
4, 5	$\varnothing$	4, 5

Lemma 1. *For all vertices* $V_j \in \bigcup_{i=1}^{r_h} B_{2i,h}$, *the set of vertices* $\mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h})$ *is a subset of* $\bigcup_{i=1}^{r_h} B_{2i,h}$.

The proof is reported in Appendix A.

While the conditions (3.2) and (3.3) are a clear consequence of the condition (C2a) in the formula (3.1), the following theorem shows that these conditions are equivalent to the condition (C2a).

Theorem 3.1. *The* (C2*a) *is a list of independences that is equivalent to the list of independences given by* (C2a).

The proof of this theorem is reported in Appendix A.

Example 7. For the graph in Figure 1, the list of conditional independences (3.2) is the single statement $4 \perp 6|2, 3, 5$, while the (3.3) gives no independence statement.

Example 8 (Continuation of Example 6). Applying (3.2), to the graph in Figure 2, we get $2 \perp 4, 5|1, 3$. Applying (3.3) we get $2, 3 \perp 5|1, 4$; $4 \perp 2|1, 3, 5$ and $5 \perp 2, 3|1, 4$.

An alternative condition for (C3b)

The condition (C3b) expresses the relationship between a set of vertices and its parents. On the other hand, the alternative condition, (C3*b), focuses on the relationships between a vertex and its children. Therefore, we define the class PA_h of sets composed by elements having the same children in T_h. Note that, the elements of PA_h are a partition of $\mathrm{pa}_D(T_h)$.

Definition 4. *The class* PA_h *of elements with same children in* T_h, *is*

$$\mathrm{PA}_h = \{\mathcal{A} : \mathrm{ch}(V_i) \cap T_h = \mathrm{ch}(V_j) \cap T_h, \forall V_i, V_j \in \mathcal{A}\}. \tag{3.4}$$

Let us consider the elements of this class partially ordered according to the following rule: $\forall \mathcal{A}, \mathcal{B} \in \mathrm{PA}_h$ *if* $|\mathrm{ch}(\mathcal{B})| < |\mathrm{ch}(\mathcal{A})|$ *then* $\mathcal{A} \prec \mathcal{B}$.

Definition 5. *The new condition* (C3*b) *is defined by the following list of conditional independences.*

$$\mathcal{A} \perp [T_h \setminus \mathrm{ch}(\mathcal{A})] | (\mathrm{pa}_D(T_h) \setminus \mathcal{A}), \qquad \forall \mathcal{A} \in \mathrm{PA}_h, h = 1, \ldots, s. \tag{3.5}$$

Theorem 3.2. *The list of independences* (C3*b) *is equivalent to the list* (C3b).

The proof of this theorem is given in the Appendix A.

Example 9. Referring to the graph in Figure 1, the PA_1 class is empty, since T_1 has no parents. The family PA_2, referring to the component T_2, is composed by the only set (1). Finally, the third component has parent set $\mathrm{pa}_D(T_3) = (2, 3)$. Since $\mathrm{ch}(2) = (4) \neq \mathrm{ch}(3) = (5)$, the class of parents with common children is $\mathrm{PA}_3 = \{(2); (3)\}$. So from the (C3*b), we have $2 \perp 5, 6|3$ and $3 \perp 4, 5|2$.

4. A smooth subclass of GMs II

This section introduces a smooth subclass of GMs II. In order to define this new subclass, we benefit from the known smoothness property of marginal models. In fact, with the help of Theorem 2.1, we study which graphs yield lists of independences that can be represented with marginal models. For this purpose, we follow the next 4 steps:

(1) a class of hierarchical marginal sets $\mathcal{H}$ is introduced;

(2) the list of hierarchical and complete marginal parameters associated to the previous marginal sets is defined;

(3) the list of parameters to set equal to zero according to the formula (2.3) of Theorem 2.1 is given;

(4) the condition (2.4) of Theorem 2.1 is checked.

In the step one, we define a hierarchical class of marginal sets. For every three sets A_i, B_i and C_i associated to an independence $A_i \perp B_i | C_i$ derived from (C1), (C2*a) and (C3*b), the hierarchical class of marginal sets must contain at least the elements $(A_i \cup B_i \cup C_i)$. Thus, according to (C1) the marginal sets:

$$\mathcal{M}_h^1 = \bigcup_{j \le h} T_j, \qquad h = 1, 2, \ldots, s, \tag{4.1}$$

are introduced. According to (C2*a) the marginal sets:

$$\mathcal{M}_h^{2^*a} = T_h \cup \mathrm{pa}_D(T_h), \qquad h = 1, 2, \ldots, s, \tag{4.2}$$

are defined. Finally, according to the third condition (C3*b) the following marginal sets are used:

$$\mathcal{M}_{h,A}^{3^*b} = \mathrm{pa}_D(T_h) \cup \mathrm{NC}_h \cup A \qquad \forall A \in \mathcal{G}_h, h = 1, 2, \ldots, s, \tag{4.3}$$

where $\mathcal{G}_h$ includes the sets $\mathrm{CH}_h \setminus \mathrm{ch}(\mathcal{A})$, for every $\mathcal{A} \in \mathrm{PA}_h$, and all the intersections of these sets.

Note that, by definition, the following relationships always hold:

$$\mathcal{M}_{h,A}^{3^*b} \subseteq \mathcal{M}_h^{2^*a} \subseteq \mathcal{M}_h^1 \qquad \forall A \in \mathcal{G}_h, h = 1, \ldots, s. \tag{4.4}$$

The hierarchical class of marginal sets, for each h, first contains all sets $\mathcal{M}_{h,\mathcal{A}}^{3^*b}$, sorted according to the hierarchical principle (see Definition 1), then it contains the set $\mathcal{M}_h^{2^*a}$ and finally the set $\mathcal{M}_h^1$. Thus, for each component h, we have the following class:

$$\mathcal{H}_{\mathrm{II}}^h = \{\{\mathcal{M}_{h,A}^{3^*b}, \forall A \in \mathcal{G}_h\}, \mathcal{M}_h^{2^*a}, \mathcal{M}_h^1\}. \tag{4.5}$$

Lastly, the family of all marginal sets of the chain graph is given by the following ordered collection:

$$\mathcal{H}_{\mathrm{II}} = \{\mathcal{H}_{\mathrm{II}}^h, h = 1, \ldots, s\}. \tag{4.6}$$

Note that, it may occur that some of the previous sets match. For example, for a given h and A, it could happen that $\mathcal{M}_h^{2^*a}$ is equal to $\mathcal{M}_{h,A}^{3^*b}$, or that $\mathcal{M}_h^{2^*a}$ is equal to $\mathcal{M}_h^1$.

Example 10. Let us consider the graph in Figure 1. From the (C1), we have the marginal sets $\mathcal{M}_1^1 = (1)$, $\mathcal{M}_2^1 = (1, 2, 3)$ and $\mathcal{M}_3^1 = (1, 2, 3, 4, 5, 6)$. From the (C2*a), we get the marginal sets $\mathcal{M}_1^{2^*a} = (1)$, $\mathcal{M}_2^{2^*a} = (1, 2, 3)$ and $\mathcal{M}_3^{2^*a} = (2, 3, 4, 5, 6)$. Finally, let us define the classes PA_3 as $\{(2); (3)\}$ and $\mathcal{G}_3$ as $\{(6); (4); \varnothing\}$. Thus, from the (C3*b) we get $\mathcal{M}_{1,1}^{3^*b} = (1)$, $\mathcal{M}_{3,6}^{3^*b} = (2, 3, 5, 6)$, $\mathcal{M}_{3,4}^{3^*b} = (2, 3, 4, 5)$ and their intersection $\mathcal{M}_{3,\varnothing}^{3^*b} = (2, 3, 5)$. The hierarchical class is $\mathcal{H}_{\mathrm{II}} = \{(1); (1, 2, 3); (2, 3, 5); (2, 3, 4, 5); (2, 3, 5, 6); (2, 3, 4, 5, 6); (1, 2, 3, 4, 5, 6)\}$.

Once the class of marginal sets is defined, we can determine the hierarchical and complete parameters, $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$, $\forall \mathcal{L} \in \mathcal{P}(V)$, where $\mathcal{M}(\mathcal{L})$ is the first marginal set in $\mathcal{H}_{\mathrm{II}}$ containing the set $\mathcal{L}$. At this point, we must select the parameters to constrain to zero. On the basis of (2.3) of Theorem 2.1, these parameters are $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$ where $\mathcal{L} \in D_h^1$ according to the (C1), $\mathcal{L} \in D_h^{2^*a}$, according to the (C2*a) and $\mathcal{L} \in D_h^{3^*b}$ according to the (C3*b). The classes D_h^1, $D_h^{2^*a}$ and $D_h^{3^*b}$ depend on the alternative Markov properties defined in the Section 3.3. The definition of these classes are in Appendix B and here we report only an example.

Example 11 (Continuation Examples 5, 7, 9 and 10). Applying the formula (2.2) to the (C1) independence (in the Example 5), the (C2*a) independence, listed in the Example 7 and the (C3*b) independences (in Example 9), the following classes of interactions, concerning null parameters. are obtained: $D_1^1 = \varnothing$, $D_2^1 = \varnothing$ and $D_3^1 = \{(1, 4); (1, 5); (1, 6); (1, 4, 5); (1, 4, 6); (1, 5, 6); (1, 4, 5, 6); (1, 2, 4); (1, 2, 5); (1, 2, 6); (1, 2, 4, 5); (1, 2, 4, 6); (1, 2, 5, 6); (1, 2, 4, 5, 6); (1, 3, 4); (1, 3, 5); (1, 3, 6); (1, 3, 4, 5); (1, 3, 4, 6); (1, 3, 5, 6); (1, 3, 4, 5, 6); (1, 2, 3, 4); (1, 2, 3, 5); (1, 2, 3, 6); (1, 2, 3, 4, 5); (1, 2, 3, 4, 6); (1, 2, 3, 5, 6); (1, 2, 3, 4, 5, 6)\}$. $D_1^{2^*a} = \{\varnothing\}$, $D_2^{2^*a} = \{\varnothing\}$, $D_3^{2^*a} = \{(4, 6); (2, 4, 6); (3, 4, 6); (4, 5, 6); (2, 3, 4, 6); (2, 4, 5, 6); (3, 4, 5, 6); (2, 3, 4, 5, 6)\}$. Finally, $D_1^{3^*b} = \{\varnothing\}$, $D_2^{3^*b} = \varnothing$ and $D_3^{3^*b} = \{(2, 5); (2, 6); (2, 5, 6); (2, 3, 5); (2, 3, 6); (2, 3, 5, 6); (3, 4); (3, 5); (3, 4, 5); (2, 3, 4); (2,3,5); (2, 3, 4, 5)\}$.

By using Theorem 2.1, we obtain the class of GMs II which is parametrizable with marginal models. In particular, Theorem 4.1 shows when condition (2.4) of Theorem 2.1 is satisfied by the family $\mathcal{H}_{\mathrm{II}}$ of marginal sets and the sets D_h^1, $D_h^{2^*a}$ and $D_h^{3^*b}$.

Theorem 4.1. *A graphical model of type II is a marginal model with parameters*

$$\left\{ \boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \Big\backslash \bigcup_{h=1}^{s} (D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}), \mathcal{M} \in \mathcal{H}_{\mathrm{II}} \right\}$$

if:

(i) $\mathrm{nb}(V_i) \in \mathcal{C}_h$ *for all* $V_i \in \mathrm{CH}_h$, *or if*

(ii) *for all* $V_i \in \mathrm{CH}_h$ *such that* $\mathrm{nb}(V_i) \notin \mathcal{C}_h$, *all sets* K, *such that* $K \notin \mathcal{C}_h$ *and* $K \cap \mathrm{nb}(V_i) \neq \varnothing$, *satisfy* $\mathrm{pa}(K) \supseteq \mathrm{pa}(V_i)$, $h = 1, 2, \ldots, s$.

This theorem shows that the smoothness problem can only arise when there are vertices in the children set CH_h with non-complete neighbourhood and that the smoothness property is assured

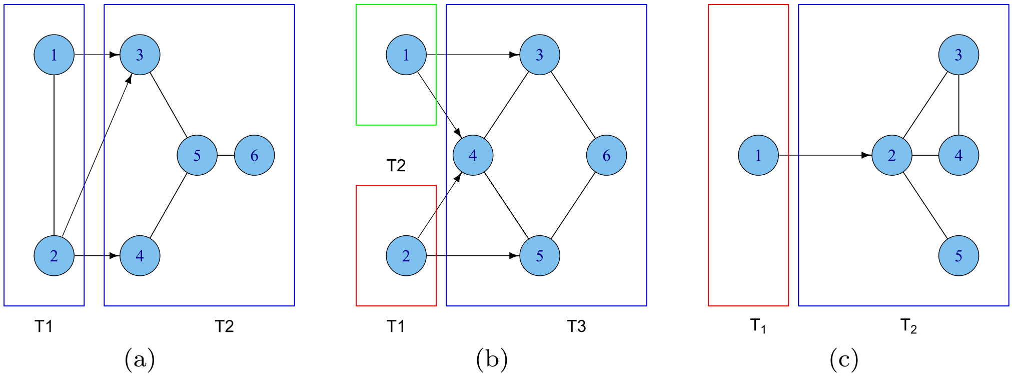

Figure 3. Chain graphs.

if all non-complete sets containing at least one neighbour of a child V_i, such that $\mathrm{nb}(V_i) \notin \mathcal{C}_h$, have parent set containing $\mathrm{pa}(V_i)$. The proof of this theorem is in Appendix C.

Example 12. The graph in Figure 3(a) represents the following list of independences (C2*a) and (C3*b): $3 \perp 4, 6|1, 2, 5$; $4 \perp 3, 6|1, 2, 5$; $6 \perp 3, 4|1, 2, 5$; $1 \perp 4, 5, 6|2$; $2 \perp 5, 6|1$. The marginal class referring to this graph is $\mathcal{H}_{\mathrm{II}} = \{(1, 2); (1, 2, 5, 6); (1, 2, 4, 5, 6); (1, 2, 3, 4, 5, 6)\}$. Since any vertex in $\mathrm{CH}_2 = (3; 4)$ has a complete set of neighbours the Theorem 4.1 holds.

Example 13. The list of conditional independences (C1), (C2*a) and (C3*b) represented in the graph in Figure 3(b) is $1 \perp 2$; $3 \perp 5|1, 2, 4, 6$; $4 \perp 6|1, 2, 3, 5$; $1 \perp 5, 6|2$ and $2 \perp 3, 6|1$. The class of marginal sets is $\mathcal{H}_{\mathrm{II}} = \{(1); (2); (1, 2); (1, 2, 6); (1, 2, 3, 6); (1, 2, 5, 6); (1, 2, 3, 4, 5, 6)\}$. Note that each vertex in $\mathrm{CH}_2 = \{3, 4, 5\}$ has a non-complete set of neighbours thus it is necessary to take into account the class of non-complete sets K, $K \subset T_2$, with at least one neighbour of a child. When $V_i = 3$, for every $K \in \{(4, 6), (4, 5, 6), (3, 4, 6), (3, 4, 5), (3, 5, 6), (3, 4, 5, 6)\}$, it is $\mathrm{pa}(K) = (1, 2)$ and $\mathrm{pa}(3) = (1) \subseteq \mathrm{pa}(K)$. The same holds when $V_i = 5$. Finally, when $V_i = 4$, for every $K \in \{(3, 5), (4, 5, 6), (3, 4, 6), (3, 4, 5), (3, 5, 6), (3, 4, 5, 6)\}$ it is $\mathrm{pa}(K) = \mathrm{pa}(4) = (1, 2)$. Thus the condition (ii) of Theorem 4.1 holds.

Example 14. The graph in Figure 3(c) represents the following list of independences (C2*a) and (C3*b): $3 \perp 5|1, 2, 4$, $4 \perp 5|1, 2, 3$, $5 \perp 3, 4|1, 2$ and $1 \perp 3, 4, 5$. The class of marginal sets referring to this graph is $\mathcal{H}_{\mathrm{II}} = \{(1); (1, 3, 4, 5); (1, 2, 3, 4, 5)\}$. Since $\mathrm{CH}_2 = (2)$ has the non-complete set of neighbours $\mathrm{nb}(2) = (3, 4, 5)$, and since the non-complete set $(3, 4, 5)$, have parent set $\mathrm{pa}(K) = \varnothing$ the conditions of the Theorem 4.1 are not satisfied.

The following example shows that a GM II that satisfies the conditions of Theorem 4.1 is not necessarily equivalent to a GM I.

Example 15. The GM II associated with the graphs in Figure 4 represents the following system of independences (C2*a) and (C3*b): $3 \perp 5|1, 2, 4$; $1 \perp 5|2$; $2 \perp 3|1$. The class of marginal

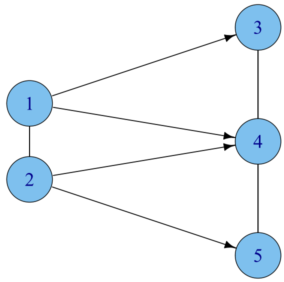

Figure 4. Chain graphs with set of vertices $V = \{1, 2, 3, 4, 5\}$ and set of edges $E = \{(1, 2); (2, 1); (1, 3); (1, 4); (2, 4); (2, 5); (3, 4); (4, 3); (4, 5); (5, 4)\}$.

sets is $\mathcal{H}_{\mathrm{II}} = \{(1, 2); (1, 2, 5); (1, 2, 3); (1, 2, 3, 4, 5)\}$. The vertex 4 is in CH_2 and the set if its neighbours is not complete. As the non-complete sets K in $\{(3, 5), (3, 4, 5)\}$, have parent set $\mathrm{pa}(K) = (1, 2) = \mathrm{pa}(4)$, Theorem 4.1 assures that the GM II is a marginal model. The graph in Figure 4 is an example of a biflag, thus, according to Theorem 6 of [1] there is no GM I which represents the same structure of relationships.

In Theorem 4.1, we introduced a smooth subclass of GMs II which is parametrizable with marginal models. It is legitimate to ask if a chain graph model of type II, that does not satisfy the condition of the Theorem 4.1, can be a smooth model not parametrizable by a marginal model. The following theorem is an answer to the previous question.

Theorem 4.2. *A graphical model of type II, that does not satisfy the condition of the Theorem 4.1, is not smooth.*

The proof of this theorem is in Appendix C. From this theorem, it trivially follows that the only GM II models, that admit a marginal parametrization, are the ones that satisfy the condition of Theorem 4.1.

5. Smooth GMs II applied to a real dataset

We used a data-set from the European Values Study-EVS-(2008) [7] in order to show the ability of the GMs II, parametrizable according to the marginal models, to represent a system of conditional independences of categorical variables. The EVS is a research project on human values in Europe. In particular, the research involves what Europeans think about *family*, *work*, *religion*, *politics* and *society*. We used the GMs II to highlight the dependence of some variables classified as *personal* variables and *opinion* variables on *gender*. To this aim, we used the following variables:

G: *Gender* (“Female”, “Male”);
E: *Employed* (“Yes”, “No”);
C: *Children* (“Yes”, “No”);
T: *Trust in people* (“Yes”, “No”);

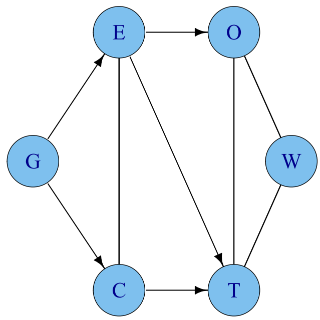

Figure 5. Chain Graph representing the variables in the Northern Islands Case.

O: *Opinion on society* (“High”, “Mean”, “Low”);
W: *Personal perceived well-being level* (“High”, “Low”).

We divided the variables in three groups, each one corresponding to a component in the chain graph. In the first group, we placed only *Gender*. In the second group, there are variables concerning the status of the respondents (*Employed*, *Children*). Finally, the last group regards the variables that consider the opinion of the respondents concerning specific topics (*Personal perceived well-being level*, *Opinion about the society*, *Trust in people*). We represented each group of variables with a component in the chain graph. We fitted GMs II for different European Countries. The two most interesting cases, concerning the Northern Islands (i.e., Ireland, United Kingdom and Island) and Italy are reported below. In both cases, we fitted the saturated marginal model (unconstrained model) corresponding to the complete chain graphical model. We proceeded testing the smooth GMs II obtained by removing arcs from the complete chain graphical model. All models were tested using the Likelihood Ratio test which compares the saturated model with the chosen model. In both cases, we chose the simplest model, with fewer number of arcs, still able to represent the data. For the Northern Islands dataset, we chose the graph in Figure 5. The marginal model corresponding to this graph has a Likelihood ratio test statistic $Gsq = 53.19302$ with 51 degree of freedom and the model displayed in Figure 5 can be retained with a p-value of 0.38976. The Italian case is well described by the graph in Figure 6. Here, the statics test Gsq is 68.84138, with 55 degree of freedom and a p-value equal to 0.0935.

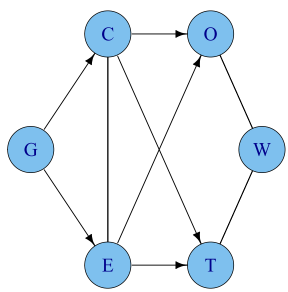

Figure 6. Chain Graph representing the variables in the Italian Case.

Table 2. List of nonredundant conditional independences obtained from the condition (C1), (C2*a) and (C3*b) for the two cases

Northern Islands	Italy
G ⊥ T, W, O|C, E	G ⊥ T, W, O|C, E
W ⊥ C, E	W ⊥ C, E
O ⊥ C|E	O ⊥ T|C, E, W

Note that the graphs in both Figures 5 and 6 satisfy the conditions of Theorem 4.1.

Let us analyse the chosen models. Table 2 on page 876 reports the non-redundant lists of conditional independences obtained from the condition (C1), (C2*a) and (C3*b) for the two models, while Table 3 on page 877 reports the non null parameters describing the relationships among the variables.

In both models, we observe the independence of the opinion variables T, O, W from *gender* given personal variables C and E. Furthermore, we see that variables referring to personal perceived well-being level (W) is jointly independent of both *Employment* (E) and *Children* (C).

In the Northern Islands Case (Figure 5), the last independence concerns *Opinion on the society* (O) that is independent of *Children* (C) given *Employment* (E). Instead, in the Italian case the *Opinion on society* (O) is independent of *Trust in people* (T) given the *Personal perceived well-being level* (W), *Employment* (E) and *Children* variable (C).

Furthermore, looking at the non-null parameters in Table 3 on page 877, it is possible to say that both the relationships between *Trust in people* T and *Children* C and the relationship between *Trust in people* T and *Employment* E in the Northern Islands case are stronger than in the Italian case.

The study of these two datasets clarifies the advantages of CG models of type II. In particular, with reference to the graph in Figure 5, according to the (C3b), we model the independence of W from C and E, without conditioning on the response variables O and T, analogously the independence of O from C is conditional only on the explicative variable E. The same figure, according to the chain graph models of type I (GMs I) proposed by [11], provides independence statements involving all the other variables in the same component such as $W \perp E, C|O, T$ and $O \perp C|W, T, E$. Thus, GMs II allow to study the effect of explicative variables on some response variables independently from the other response variables.

The graph in Figure 6 shows the differences between the GMs II and the chain graph models of type IV (GMs IV) proposed by [5]. In this case, the GMs II, according to the (C2a), model the independences between response variables by conditioning on the explicative variables and the remaining response variables, thus we have $O \perp T|C, E, W$, while according to the GMs IV we will have $O \perp T|C, E$. In this case, the GMs II use log-linear models for the joint distribution of the variables within the same component. In our opinion, the GMs II combine the advantages of both GMs I and GMs IV. First of all, they describe the associations in a chain component in a sensible way as the GMs I do and the GMs IV do not. Moreover, the effects of independent variables are modeled according to the GMs IV way, which is a more natural approach than the one followed by the GMs I.

All the analysis were carried out with the software R with the help of the *hmmm* package, [4].

Table 3. Nonnull parameters concerning the two marginal models

Northern Islands			Italy		
$\mathcal{M}$	$\mathcal{L}$	$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$	$\mathcal{M}$	$\mathcal{L}$	$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$
G	*G*	−0.0449	*G*	*G*	0.0152
CEG	*C*	0.6361	*CEG*	*C*	0.6434
CEG	*E*	0.6491	*CEG*	*E*	0.7051
CEG	*CE*	0.0207	*CEG*	*CE*	−0.0410
CEG	*CG*	0.9245	*CEG*	*CG*	0.4120
CEG	*EG*	−0.1460	*CEG*	*EG*	−0.6676
CEG	*CEG*	−0.6237	*CEG*	*CEG*	−0.3566
WCE	*W*	1.5997	*WCE*	*W*	0.6695
WOCE	*O*	[1.6709; −1.5237]	*WOTCE*	*O*	[1.4606; −3.1637]
WOCE	*WO*	[−0.0045; −0.1318]	*WOTCE*	*T*	−0.6910
WOCE	*OE*	[−0.7183; −0.2680]	*WOTCE*	*WO*	[0.6394; −0.4189]
WOCE	*WOC*	[0.3076; −0.0108]	*WOTCE*	*WT*	−0.0988
WOCE	*WOE*	[0.9925; −0.4202]	*WOTCE*	*OC*	[1.1911; 1.3128]
WOCE	*WOCE*	[−0.3958; 0.1132]	*WOTCE*	*TC*	−0.6064
WOTCE	*T*	−1.5630	*WOTCE*	*OE*	[−0.6684; 1.6665]
WOTCE	*WT*	0.5808	*WOTCE*	*TE*	−1.1285
WOTCE	*OT*	[0.3687; 0.5243]	*WOTCE*	*WOC*	[−0.3753; −0.1285]
WOTCE	*TC*	−21.9010	*WOTCE*	*WTC*	0.0500
WOTCE	*TE*	−19.3440	*WOTCE*	*WOE*	[0.4847; −0.6825]
WOTCE	*WOT*	0.2704	*WOTCE*	*WTE*	1.3874
WOTCE	*WTC*	21.8659	*WOTCE*	*OTC*	[0.7660; −0.9532]
WOTCE	*WTC*	[21.8395; −0.3217]	*WOTCE*	*OCE*	[−0.1567; −2.4855]
WOTCE	*WTE*	20.0330	*WOTCE*	*TCE*	0.8888
WOTCE	*OTE*	[1.5645; −1.4979]	*WOTCE*	*WOTC*	[0.9136; −0.4878]
WOTCE	*TCE*	42.4563	*WOTCE*	*WOTE*	[−0.0037; 1.6923]
WOTCE	*WOTC*	[−21.9093; 1.1838]	*WOTCE*	*WOTCE*	−0.5678
WOTCE	*WOTC*	[−20.3008; 20.3660]			
WOTCE	*WTCE*	−42.6677			
WOTCE	*OTCE*	[−42.5531; 21.0933]			
WOTCE	*WOTCE*	[43.0787; −21.6751]			

Appendix A

Proof of Markov equivalences

Proof of Lemma 1

By definition, $\mathrm{nb}(B_{1i,h}) = B_{2ih}$, $\forall i = 1, \ldots, r_h$, so $\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \subseteq \bigcup_{i=1}^{r_h} B_{2i,h}$ and consequently even $\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \setminus V_j \subseteq \bigcup_{i=1}^{r_h} B_{2i,h}$.

Since the set $\mathbf{B}_{1,V_j,h}$, by definition, is equal to $\mathrm{nb}(V_j) \cap \bigcup_{i=1}^{r_h} B_{1i,h}$, then in the set $\mathrm{nb}(V_j) \setminus \mathbf{B}_{1,V_j,h}$ there is no vertex belonging to $\bigcup_{i=1}^{r_h} B_{1,i}$, thus $\mathrm{nb}(V_j) \setminus \mathbf{B}_{1,V_j,h} \subseteq \bigcup_{i=1}^{r_h} B_{2,i}$.

The lemma follows by noting that $\mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h}) = (\mathrm{nb}(V_j) \setminus \mathbf{B}_{1,V_j,h}) \cup (\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \setminus V_j)$.

Proof of Theorem 3.1

We prove that the independences (C2*a) follow from the list of independences (C2a). When $A = B_{1i,h}$, $B_{1i,h} \subset C_{i,h} \in \mathcal{C}l_h$, the set of neighbours of A is $\mathrm{nb}(A) = \mathrm{nb}(B_{1i,h}) = C_{i,h} \setminus B_{1i,h} = B_{2i,h}$. The neighbourhood of A is $\mathrm{Nb}(A) = C_{i,h}$ and from (C2a) we get $B_{1i,h} \perp T_h \setminus C_{i,h} | \mathrm{pa}_D(T_h) \cup B_{2i,h}$, which is a (3.2) independence. On the other hand, when A is equal to $V_j \cup (\mathbf{B}_{1,V_j,h})$, where $\mathbf{B}_{1,V_j,h} = \mathrm{nb}(V_j) \cap (\bigcup_{i=1}^{r_h} B_{1i,h})$, the set of neighbours of A is $\mathrm{nb}(A) = \mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h})$ and the neighbourhood is $\mathrm{Nb}(A) = \mathrm{Nb}(V_j \cup (\mathrm{nb}(V_j) \cap \mathbf{B}_{1,V_j,h}))$, so the (C2a) implies (3.3).

Now we prove that from the (C2*a) we get the (C2a). For this purpose, we use the equivalence between the (C2a) and the following statement [10]:

$$V_j \perp T_h \setminus \mathrm{Nb}(V_j) | \mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j), \qquad V_j \in T_h. \tag{A.1}$$

Thus, it is sufficient to prove that the (C2*a) implies (A.1). From (3.2), applying the properties of the conditional independences [10], we have:

$$V_j \perp T_h \setminus C_{i,h} | \mathrm{pa}_D(T_h) \cup B_{2i,h} \cup B_{1i,h} \setminus (V_j), \qquad V_j \in B_{1i,h}.$$

If $V_j \in B_{1i,h}$, it holds that $\mathrm{Nb}(V_j) = C_{i,h}$, $B_{2i,h} \cup B_{1i,h} \setminus (V_j) = C_{i,h} \setminus (V_j) = \mathrm{nb}(V_j)$. Thus the previous formula becomes:

$$V_j \perp T_h \setminus \mathrm{Nb}(V_j) | \mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j), \qquad V_j \in B_{1i,h},$$

that is the formula in (A.1). This relationship holds for all $V_j \in \bigcup_{i=1}^{r_h} B_{1i,h}$. The remaining vertices of T_h are considered in the statement (3.3) from which we get:

$$V_j \perp T_h \setminus \mathrm{Nb}(V_j \cup \mathbf{B}_{1,V_j,h}) | \big(\mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathbf{B}_{1,V_j,h}\big), \qquad V_j \in B_{2i,h}. \tag{A.2}$$

Note that, by definition, $\mathbf{B}_{1,V_j,h} \subset \mathrm{nb}(V_j)$ and, for every $B_{1i,h} \subseteq \mathbf{B}_{1,V_j,h}$, the vertex V_j, $V_j \in B_{2i,h}$, is adjacent to each vertex in $\mathrm{nb}(B_{1i,h}) = B_{2i,h}$, since these are complete sets. Thus it is $\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \setminus (V_j) \subset \mathrm{nb}(V_j)$. From the previous two inclusion relations it follows that

$$\begin{aligned}
\mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathbf{B}_{1,V_j,h} &= \big[\mathrm{nb}(V_j) \cup \mathrm{nb}(\mathbf{B}_{1,V_j,h})\big] \setminus (V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathbf{B}_{1,V_j,h} \\
&= \big[\mathrm{nb}(V_j) \setminus \mathbf{B}_{1,V_j,h} \cup \mathbf{B}_{1,V_j,h}\big] \cup \big[\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \setminus (V_j)\big] \\
&= \mathrm{nb}(V_j) \cup \big[\mathrm{nb}(\mathbf{B}_{1,V_j,h}) \setminus (V_j)\big] = \mathrm{nb}(V_j),
\end{aligned} \tag{A.3}$$

and that $\mathrm{Nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathbf{B}_{1,V_j,h} = \mathrm{Nb}(V_j)$. The two equalities above and (A.2) imply that

$$V_j \perp T_h \setminus \mathrm{Nb}(V_j) | \big(\mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j)\big)$$

that is the (A.1), for all $V_j \in \bigcup_{i=1}^{r} B_{2i,h}$.

Proof of Theorem 3.2

The list of independences from the (C3b) implies the list from the (C3*b). Applying the statement (C3b) to $T_h \setminus \mathrm{ch}(\mathcal{A})$, we get

$$T_h \setminus \mathrm{ch}(\mathcal{A}) \perp \mathrm{pa}_D(T_h) \setminus \mathrm{pa}\big(T_h \setminus \mathrm{ch}(\mathcal{A})\big) | \, \mathrm{pa}\big(T_h \setminus \mathrm{ch}(\mathcal{A})\big).$$

Since $\mathcal{A} \subseteq \mathrm{pa}_D(T_h) \setminus \mathrm{pa}(T_h \setminus \mathrm{ch}(\mathcal{A}))$, from the properties of conditional independence it follows $T_h \setminus \mathrm{ch}(\mathcal{A}) \perp \mathcal{A} | \, \mathrm{pa}_D(T_h) \setminus \mathcal{A}$, that is the conditional independence (C3*b) for the set $\mathcal{A}$.

The conditional independences (C3*b) imply the (C3b) ones. Given a set $A \subseteq T_h$, let $\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_w\}$ be the collection of sets of PA_h having $A \subseteq T_h \setminus \mathrm{ch}(\mathcal{A}_i)$. Now we consider independences related to these sets:

$$\begin{cases} \mathcal{A}_1 \perp T_h \setminus \mathrm{ch}(\mathcal{A}_1) | \big(\mathrm{pa}_D(T_h) \setminus \mathcal{A}_1\big), \\ \ldots \\ \mathcal{A}_w \perp T_h \setminus \mathrm{ch}(\mathcal{A}_w) | \big(\mathrm{pa}_D(T_h) \setminus \mathcal{A}_w\big). \end{cases}$$

Since $A \subseteq T_h \setminus \mathrm{ch}(\mathcal{A}_i)$, $i = 1, \ldots, w$, we get:

$$\begin{cases} \mathcal{A}_1 \perp A \Bigg| \left(\mathrm{pa}_D(T_h) \Big\backslash \bigcup_{i=1}^{w} \mathcal{A}_i\right) \cup \left(\bigcup_{i=1}^{w} \mathcal{A}_i \Big\backslash \mathcal{A}_1\right), \\ \ldots \\ \mathcal{A}_w \perp A \Bigg| \left(\mathrm{pa}_D(T_h) \Big\backslash \bigcup_{i=1}^{w} \mathcal{A}_i\right) \cup \left(\bigcup_{i=1}^{w} \mathcal{A}_i \Big\backslash \mathcal{A}_w\right). \end{cases}$$

Using the intersection property of conditional independence, property (P5) of Lauritzen [10], we obtain $\bigcup_{i=1}^{w} \mathcal{A}_i \perp A | \, \mathrm{pa}_D(T_h) \setminus \bigcup_{i=1}^{w} \mathcal{A}_i$ which is the conditional independence (C3b): $A \perp \mathrm{pa}_D(T_h) \setminus \mathrm{pa}(A) | \, \mathrm{pa}(A)$.

Appendix B

Family of interaction sets concerning null parameters

In this Appendix, we define for all the Markov properties (C1), (C2*a) and (C3*b), respectively, the classes of interaction D_h^1, $D_h^{2^*a}$ and $D_h^{3^*b}$, $h = 1, \ldots, s$. Applying formula (2.2) to the previous Markov properties we get the following classes:

$$D_h^1 = \mathcal{P}\left(\bigcup_{j=1}^{h} T_j\right) \Bigg\backslash \left(\mathcal{P}\big(T_h \cup \mathrm{pa}_D(T_h)\big) \cup \mathcal{P}\left(\bigcup_{j=1}^{h-1} T_j\right)\right); \tag{B.1}$$

$$D_h^{2^*a} = \left(\bigcup_{i=1}^{r_h} D_{h,B_{1i,h}}^{2^*a}\right) \cup \left(\bigcup_{V_j \in \bigcup_{i=1}^{r_h} B_{2i,h}} D_{h,V_j}^{2^*a}\right), \qquad \forall h = 1, \ldots, s, \tag{B.2}$$

where $D^{2^*a}_{h,B_{1i,h}}$ is:

$$\mathcal{P}\big(T_h \cup \mathrm{pa}_D(T_h)\big) \setminus \big(\mathcal{P}\big(C_{i,h} \cup \mathrm{pa}_D(T_h)\big) \cup \mathcal{P}\big(T_h \setminus \big(B_{1i,h} \cup \mathrm{pa}_D(T_h)\big)\big)\big) \tag{B.3}$$

and $D^{2^*a}_{h,V_j}$ is:

$$\begin{aligned}&\mathcal{P}\big(T_h \cup \mathrm{pa}_D(T_h)\big)\\ &\quad \setminus \big(\mathcal{P}\big(\mathrm{Nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathrm{pa}_D(T_h)\big) \cup \mathcal{P}\big(T_h \setminus (V_j \cup \mathbf{B}_{1,V_j,h}) \cup \mathrm{pa}_D(T_h)\big)\big)\end{aligned} \tag{B.4}$$

and finally

$$D^{3^*b}_h = \bigcup_{\mathcal{A}\in \mathrm{PA}_h} D^{3^*b}_{h,\mathcal{A}}, \qquad \forall h = 1, \ldots, s, \tag{B.5}$$

where $D^{3^*b}_{h,\mathcal{A}}$ is:

$$\mathcal{P}\big(T_h \setminus \mathrm{ch}(\mathcal{A}) \cup \mathrm{pa}_D(T_h)\big) \setminus \big(\mathcal{P}\big(\mathrm{pa}_D(T_h)\big) \cup \mathcal{P}\big(T_h \setminus \mathrm{ch}(\mathcal{A}) \cup \mathrm{pa}_D(T_h) \setminus \mathcal{A}\big)\big). \tag{B.6}$$

Appendix C

Proof of the main results

The proof of Theorem 4.1 is greatly simplified by the following lemmas.

Lemma 2. *Given the vertex $V_i \in \mathrm{CH}_h$ and the set $K \subseteq T_h$, such that* $\mathrm{pa}(V_i) \subseteq \mathrm{pa}(K)$, *if $K \subseteq A$, with $A \in \mathcal{G}_h$, then $V_i \in A$.*

Proof. The class $\mathcal{G}_h$ is composed by the sets $\mathrm{CH}_h \setminus \mathrm{ch}(\mathcal{A})$ and by their intersections. Thus, for every $A \in \mathcal{G}_h$ there is a family of sets $P_A \subset \mathrm{PA}_h$ such that $A = \mathrm{CH}_h \setminus \bigcup_{\mathcal{A}\in P_A} \mathrm{ch}(\mathcal{A})$ and $\mathrm{pa}(A) \cap (\bigcup_{\mathcal{A}\in P_A} \mathcal{A}) = \varnothing$.

By the assumption of Lemma 2, it follows that $\mathrm{pa}(V_i) \subseteq \mathrm{pa}(K) \subseteq \mathrm{pa}(A) \subseteq \bigcup_{\mathcal{A}\notin P_A} \mathcal{A}$. If $V_i \notin A$, it would be $V_i \in \mathrm{ch}(\mathcal{A})$ for a set $\mathcal{A} \in P_A$ and $\mathrm{pa}(V_i) \nsubseteq \bigcup_{\mathcal{A}\notin P_A} \mathcal{A}$ which is a contradiction. □

Lemma 3. *A graphical model of type II is a marginal model with parameters $\{\boldsymbol{\eta}^{\mathcal{M}}_{\mathcal{L}} : \mathcal{L} \in \mathcal{P}(V) \setminus \bigcup_{h=1}^{s}(D^1_h \cup D^{2^*a}_h \cup D^{3^*b}_h), \mathcal{M} \in \mathcal{H}_{\mathrm{II}}\}$, if* $\mathrm{nb}(V_i) \in \mathcal{C}_h$ *for all $V_i \in \mathrm{CH}_h$.*

Proof. In order to prove this lemma, we apply Theorem 2.1 to the parametrization $\{\boldsymbol{\eta}^{\mathcal{M}}_{\mathcal{L}} : \mathcal{L} \in \mathcal{P}(V), \mathcal{M} \in \mathcal{H}_{\mathrm{II}}\}$.

According to the condition (C1), the parameters to constrain to zero are $\boldsymbol{\eta}^{\mathcal{M}(\mathcal{L})}_{\mathcal{L}}$, $\mathcal{L} \in \bigcup_{h=1}^{s} D^1_h$. We must check if the marginals $\mathcal{M}(\mathcal{L})$ satisfy the condition (2.4) of the Theorem 2.1, that is to say if $\mathrm{pa}_D(T_h) \subseteq \mathcal{M}(\mathcal{L}) \subseteq \bigcup_{j\le h} T_h$, for $\mathcal{L} \in D^1_h$, $h = 1, 2, \ldots, s$. It is easy to see that each element $\mathcal{L} \in D^1_h$ has at least one vertex in T_h and one vertex in $\bigcup_{j=1}^{h-1} T_j \setminus \mathrm{pa}_D(T_h)$. Since for a given h

the only marginal set containing these subsets of vertices is $\mathcal{M}_h^1 = \bigcup_{j \le h} T_j$, the condition (2.4) of the Theorem 2.1 always holds for the independences following from (C1) property.

Regarding the (C2*a), each vector of parameters $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$, with $\mathcal{L} \in \bigcup_{h=1}^{s} D_h^{2^*a}$, must be constrained to zero. In this case, the condition (2.4) of Theorem 2.1 holds if for every $\mathcal{L} \in D_{h,B_{1i,h}}^{2^*a}$, $h = 1, 2, \ldots, s$, $i = 1, 2, \ldots, r_h$, $\mathcal{M}(\mathcal{L})$ satisfies the condition:

$$\mathrm{pa}_D(T_h) \cup B_{2i,h} \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup \mathrm{pa}_D(T_h), \tag{C.1}$$

and, for every $\mathcal{L} \in D_{h,V_j}^{2^*a}$, $V_j \in \bigcup_{i=1}^{r_h} B_{2i,h}$, $h = 1, 2, \ldots, s$, the condition:

$$\mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup \mathrm{pa}_D(T_h). \tag{C.2}$$

When $\mathcal{L} \supseteq \mathrm{CH}_h$ or the set $T_h \cap \mathcal{L}$ has parent set equal to $\mathrm{pa}_D(T_h)$, $\mathcal{M}(\mathcal{L})$ is equal to $\mathcal{M}_h^{2^*a}$ and the conditions (C.1) and (C.2) are both trivially satisfied.

In general, there will be a set $A \in \mathcal{G}_h$, such that $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^*b}$, where $A \supseteq \mathcal{L} \cap \mathrm{CH}_h$ and the right inclusions in (C.1) and (C.2) still hold.

If $\bigcup_{i=1}^{r_h} B_{2i,h}$ is contained in NC_h, also the left inclusions in (C.1) and (C.2) are satisfied because, according Lemma 1, the conditional set $\mathrm{pa}_D(T_h) \cup \mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h})$ in (3.3) is a subset of $\mathrm{pa}_D(T_h) \cup (\bigcup_{i=1}^{r_h} B_{2i,h})$. But this is the case, because by the hypothesis it is $\mathrm{nb}(V_i) \in \mathcal{C}_h$, for all $V_i \in \mathrm{CH}_h$ or equivalently CH_h belongs to $\bigcup_{i=1}^{r_h} B_{1i,h}$, $h = 1, \ldots, s$.

Finally, regarding (C3*b), the relationship

$$\mathrm{pa}_D(T_h) \setminus \mathcal{A} \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \setminus \mathrm{ch}(\mathcal{A}) \cup \mathrm{pa}_D(T_h) \tag{C.3}$$

must hold for every $\mathcal{L}$ belonging to $\bigcup_{\mathcal{A} \in \mathrm{PA}_h} D_{h,\mathcal{A}}^{3^*b}$. The left inclusion in (C.3) holds since $\mathrm{pa}_D(T_h)$ is a subset of each marginal set.

To check the right inclusion in (C.3) note that every $\mathcal{L} \in D_{h,\mathcal{A}}^{3^*b}$ is the union $\mathcal{L} = P \cup B \cup R$ of three sets such that: $P \subseteq \mathrm{pa}_D(T_h)$, $P \cap \mathcal{A} \neq \varnothing$, $B \subseteq \mathrm{CH}_h \setminus \mathrm{ch}(\mathcal{A})$ and $R \subseteq \mathrm{NC}_h$. It follows that $\mathcal{M}(\mathcal{L}) = \mathrm{pa}_D(T_h) \cup A \cup \mathrm{NC}_h$, where A is the smallest set of $\mathcal{G}_h$ containing B. The right inclusion in (C.3) is satisfied because it is $A \subseteq \mathrm{CH}_h \setminus \mathrm{ch}(\mathcal{A})$. □

Remark 1. Note that the assumption, $\mathrm{nb}(V_j) \in \mathcal{C}_h$ if $V_j \in \mathrm{CH}_h$, is used to assure that the list of conditional independences (C2*a) satisfies the condition (2.4) of Theorem 2.1, but is not needed for the independences (C1) and (C3*b).

Proof of Theorem 4.1

According Lemma 3 and Remark 1, we need only to prove that the left inclusions in (C.1) and (C.2) are satisfied if there are children V_i with non-complete set of neighbours. More precisely, the marginal set $\mathcal{M}(\mathcal{L})$ in (C.1) and (C.2) must contain all children with non-complete neighbours set that belongs to the conditioning set.

With respect to a (3.2) independence, the sets $K = \mathcal{L} \cap T_h$, $\mathcal{L} \in D_{h,B_{1x,h}}^{2^*a}$ are non-complete sets such that $K \cap B_{1x,h} \neq \varnothing$. If $\mathrm{pa}(K) = \mathrm{pa}_D(T_h)$ then $\mathcal{M}(\mathcal{L}) = \mathcal{M}_h^{2^*a}$ and the left inclusion of (C.1) holds. Otherwise, let A, $K \subseteq A$, the set of $\mathcal{G}_h$ such that $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^*b}$. We must show that A contains all the children $V_i \in B_{2x,h}$. For every child $V_i \in B_{2x,h}$, its neighbours set $\mathrm{nb}(V_i)$ is not complete and $K \cap \mathrm{nb}(V_i) \neq \varnothing$ because $\mathrm{nb}(V_i) \supset B_{1x,h}$. By assumption, if $\mathrm{nb}(V_i) \notin \mathcal{C}_h$, $\mathrm{pa}(V_i) \subseteq \mathrm{pa}(K)$ and by Lemma 2 it follows that $V_i \in A$.

With respect to a (3.3) type independence, the sets $K = \mathcal{L} \cap T_h$, $\mathcal{L} \in D^{2^*a}_{h,V_j}$ are non-complete sets such that $K \cap (V_j \cup \mathbf{B}_{1V_j,h}) \neq \varnothing$. If $\mathrm{pa}(K) = \mathrm{pa}_D(T_h)$ then $\mathcal{M}(\mathcal{L}) = \mathcal{M}^{2^*a}_h$ and the left inclusion of (C.2) holds. Otherwise, let A, $K \subseteq A$, be the set of $\mathcal{G}_h$ such that $\mathcal{M}(\mathcal{L}) = \mathcal{M}^{3^*b}_{h,A}$. We must show that A contains all the children $V_i \in \mathrm{nb}(V_j \cup \mathbf{B}_{1V_j,h})$.

According Lemma 1, every child $V_i \in \mathrm{nb}(V_j \cup \mathbf{B}_{1V_j,h})$ has a non-complete set of neighbors ($\mathrm{nb}(V_i) \notin C_h$), and if $V_j \in K$ then $K \cap \mathrm{nb}(V_i) \neq \varnothing$ because from $\mathrm{nb}(V_j \cup \mathbf{B}_{1,V_j,h}) \subset \mathrm{nb}(V_j)$ (see formula (A.3)) we have $V_j \in \mathrm{nb}(V_i)$. By assumption $\mathrm{pa}(V_i) \subseteq \mathrm{pa}(K)$, if $\mathrm{nb}(V_i) \notin \mathcal{C}_h$, and by Lemma 2 it follows that $V_i \in A$.

If $V_j \notin K$ and $V_j \in \mathrm{CH}_h$, we use the property $K \cap \mathrm{nb}(V_j) \neq \varnothing$ which follows from $\mathbf{B}_{1V_j,h} \subset \mathrm{nb}(V_j)$. In this case, the assumption $\mathrm{pa}(V_j) \subseteq \mathrm{pa}(K)$ and the Lemma 2 imply that $V_j \in A$. As $K \cup \{V_j\}$ is a non-complete subset of A, containing V_j, the previous argument, in the case $V_j \in K$, proves that $V_i \in A$.

Finally, when $V_j \notin \mathrm{CH}_h$ and $V_j \notin K$, we consider a vertex $V_l \in K$ such that $V_l \in T_h \setminus \mathrm{Nb}(V_j \cup \mathbf{B}_{1V_j,h})$. If $V_l \in \mathrm{nb}(V_i)$, then $K \cap \mathrm{nb}(V_i) \neq \varnothing$ and $\mathrm{pa}(V_i) \subseteq \mathrm{pa}(K)$ imply $V_i \in A$ according to Lemma 2 (remember that $V_i \in \mathrm{nb}(V_j \cup \mathbf{B}_{1V_j,h})$ and V_i is a child such that $\mathrm{nb}(V_i) \notin \mathcal{C}_h$). When $V_l \notin \mathrm{nb}(V_i)$ and it is adjacent to a vertex V_m in $\mathrm{nb}(V_i)$, we use the property $K \cap \mathrm{nb}(V_m) \neq \varnothing$. Note that, as $V_j \notin \mathrm{CH}_h$, it must be $V_m \in \mathrm{CH}_h$ because (V_j, V_m) is a non-complete set with one element of $\mathrm{nb}(V_i)$ and by hypothesis it must be $\mathrm{pa}(V_j, V_m) \supseteq \mathrm{pa}(V_i)$. Since the non-complete set $(V_l, V_i) \subseteq \mathrm{nb}(V_m)$, $\mathrm{nb}(V_m)$ is not complete. Thus, the assumption $\mathrm{pa}(V_m) \subseteq \mathrm{pa}(K)$ and the Lemma 2 imply that $V_m \in A$. Because $K \cup \{V_m\}$ is a non-complete set containing V_m and $V_m \in \mathrm{nb}(V_i)$, according Lemma 2 it must be $V_i \in A$.

This argument can be extended to any vertex $V_l \in (T_h \setminus \mathrm{Nb}(V_j \cup \mathbf{B}_{1,V_j,h})) \cap K$ linked by a path to V_i.

Proof of Theorem 4.2

By hypothesis, there exist a vertex V_j, $\mathrm{nb}(V_j) \notin \mathcal{C}_h$, and an incomplete set N such that $N \cap \mathrm{nb}(V_j) \neq \varnothing$, $\mathrm{pa}(N) \cap \mathrm{pa}(V_j) \neq \mathrm{pa}(V_j)$. Thus if $\mathcal{L} = (N \cup \mathrm{NC}_h) \setminus (V_j)$, $\mathrm{pa}_D(T_h) \setminus \mathrm{pa}(\mathcal{L}) \neq \varnothing$. To satisfy the (C3b) condition $\mathcal{L} \perp \mathrm{pa}_D(T_h) \setminus \mathrm{pa}(\mathcal{L}) | \mathrm{pa}(\mathcal{L})$, the interaction $\boldsymbol{\eta}^{\mathcal{M}}_{\mathcal{L} \cup \mathrm{pa}_D(T_h)}$ must be null in a marginal $\mathrm{pa}(\mathcal{L}) \subseteq \mathcal{M}(\mathcal{L} \cup \mathrm{pa}_D(T_h)) \subseteq \mathcal{L} \cup \mathrm{pa}_D(T_h)$. As N is not complete, there is a set A, $A \subset \mathcal{L}$, such that $V_j \in \mathrm{nb}(A)$, $\mathrm{Nb}(A) \subset T_h$. To satisfy the (C2a) condition $A \perp T_h \setminus \mathrm{Nb}(A) | \mathrm{pa}_D(T_h) \cup \mathrm{nb}(A)$, the interaction $\boldsymbol{\eta}^{\mathcal{M}}_{\mathcal{L} \cup \mathrm{pa}_D(T_h)}$ must be null in a marginal $\mathcal{M} \supset \mathrm{pa}_D(T_h) \cup \mathrm{nb}(\mathcal{A})$. As $V_j \in \mathrm{nb}(A)$ it must be $V_j \in \mathcal{M}$. From $V_j \notin \mathcal{M}$, it follows that the marginal parameters pertaining to the interaction set $\mathcal{L}$ must be null in two different marginal distributions and the non-smoothness property follows from Theorem 3 of [3].

References

[1] Andersson, S.A., Madigan, D. and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Stat.* **28** 33–85. MR1844349

[2] Bartolucci, F., Colombi, R. and Forcina, A. (2007). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. *Statist. Sinica* **17** 691–711. MR2398430

[3] Bergsma, W.P. and Rudas, T. (2002). Marginal models for categorical data. *Ann. Statist.* **30** 140–159. MR1892659

[4] Colombi, R., Giordano, S., Cazzaro, M. and the R Development Core Team (2013). Hierarchical multinomial marginal models. R package version 1.0-1.

[5] Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation. Monographs on Statistics and Applied Probability* **67**. London: Chapman & Hall. MR1456990

[6] Drton, M. (2009). Discrete chain graph models. *Bernoulli* **15** 736–753. MR2555197

[7] European Values Study (2008). In European Values Study, 4th wave, Italy. GESIS Data Archive, Cologne, Germany, ZA4755 Data File Version 1.0.0 (2010-11-30), DOI:10.4232/1.10031.

[8] Evans, R.J. and Richardson, T.S. (2013). Marginal log-linear parameters for graphical Markov models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 743–768. MR3091657

[9] Forcina, A., Lupparelli, M. and Marchetti, G.M. (2010). Marginal parameterizations of discrete models defined by a set of conditional independencies. *J. Multivariate Anal.* **101** 2519–2527. MR2719878

[10] Lauritzen, S.L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. New York: Oxford Univ. Press. MR1419991

[11] Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. MR0981437

[12] Marchetti, G.M. and Lupparelli, M. (2011). Chain graph models of multivariate regression type for categorical data. *Bernoulli* **17** 827–844. MR2817607

[13] Nicolussi, F. (2013). Marginal parameterizations for conditional independence models and graphical models for categorical data. Ph.D. thesis, Univ. of Milano Bicocca. Available at http://hdl.handle.net/10281/43679.

[14] Rudas, T., Bergsma, W.P. and Németh, R. (2010). Marginal log-linear parameterization of conditional independence models. *Biometrika* **97** 1006–1012. MR2746171

[15] Wermuth, N. and Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 687–717. MR2088296

Received October 2014 and revised July 2015