

Renáta Németh, Tamás Rudas

## On the application of discrete marginal graphical models\*

The final, definitive version of this paper has been published in *Sociological Methodology*  
August 2013 vol. 43 no. 1 70-100  
by SAGE Publications Ltd, All rights reserved. © Renáta Németh, Tamás Rudas  
doi:10.1177/0081175013481960  
<http://smx.sagepub.com/content/43/1/70.full.pdf+html>

### Renáta Németh

Department of Statistics, Faculty of Social Sciences  
Eötvös Loránd University, Budapest  
nemethr@caesar.elte.hu

### Tamás Rudas

Department of Statistics, Faculty of Social Sciences  
Eötvös Loránd University, Budapest  
rudas@tarki.hu

**Renáta Németh** is Assistant Professor in the Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University. Her main research areas are categorical data analysis and survey methodology with applications in the social sciences including social mobility and social inequalities in health. She has published in methodological and applied journals, including *Biometrika*, *Bulletin of Sociological Methodology*, *Journal of Epidemiology and Community Health* and *American Journal of Public Health*.

**Tamás Rudas** is Professor and Head in the Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University. He is also an Affiliate Professor in the Department of Statistics, University of Washington. Professor Rudas is an elected Fellow of the European Academy of Sociology and President of the European Association of Methodology. His main interests include categorical data analysis and survey methodology.

### Keyword

graphical models, marginal models, marginal log-linear parameters, intergenerational mobility, welfare system

\*

### Corresponding author

Renáta Németh  
nemethr@caesar.elte.hu  
Department of Statistics, Faculty of Social Sciences  
Eötvös Loránd University, Budapest  
Pázmány Péter sétány 1/A, H-1117 Budapest, Hungary

## **Acknowledgements**

The European Union and the European Social Fund have provided financial support to this project under the grant agreement TÁMOP-4.2.1/B-09/1/KMR. The authors thank Wicher Bergsma, Antonio Forcina, Thomas Richardson, Péter Róbert, and L. Andries van der Ark for their helpful comments. Some of the results reported in this paper appeared in the doctoral dissertation of the first author, to which the second author served as advisor. Direct correspondence to Renáta Németh at [nemethr@caesar.elte.hu](mailto:nemethr@caesar.elte.hu).

## **Abstract**

Graphical models are defined by general and possibly complex conditional independence assumptions and are well suited to model direct and indirect associations and effects that are of central importance in many problems of sociology. Such relevance is apparent in research on social mobility. Graphical models are discussed in the marginal modeling framework, and a general version of path models for categorical data is also introduced. These models are applied to the social status attainment process and, as substantive results, new insights into the difference between liberal and conservative welfare systems are gained. To help the use of these models, all details of the analyses are posted on a website so that readers can easily modify the analyses to their own data and models.

# 1 Introduction

A graphical model (see, e.g. Lauritzen and Wermuth 1989, Frydenberg 1990, Whittaker 1990, Cox and Wermuth 1996, Lauritzen 1996, Andersson et al. 2001, Richardson 2003, Wermuth and Cox 2004, and Drton 2009) is defined over a graph whose vertices correspond to the variables of interest and the missing edges of the graph encode conditional independence restrictions. Since these conditional independence restrictions for certain kinds of graphs do not involve all variables, rather certain marginal distributions, graphical models based on graphs are, in general, marginal models.

The present paper helps social scientists to apply graphical models by (a) giving an intuitive overview of the most important properties of graphical models, (b) providing this overview from the perspective of marginal modeling (Bergsma and Rudas 2002), which leads to a deeper insight than other approaches, including easily interpretable parameters, (c) concentrating on discrete variables, which have many interesting characteristics not shared by continuous data and (d) providing all details of the analyses presented on an accompanying website (anonymous version for the review of the manuscript: [discrete-graphical-models.site40.net](http://discrete-graphical-models.site40.net)), so that readers may modify the files to perform similar analyses on their own data and using their own models. Many of the analyses presented concern comparisons of various welfare regimes. The analyses use the `cmm` function of R (Bergsma and van der Ark, 2009). The main methodological contribution of the paper is presenting graphical models as marginal models and providing tools for their application within this framework, and the introduction of general path models for categorical data.

Although defining a model by drawing a graph with undirected lines to represent associations or with arrows to represent effects, or perhaps with both lines and arrows, may seem a very intuitive way of thinking about relationships among the variables, to associate a coherent meaning with the graphs, one has to apply certain rules, called Markov properties. As a simple example, a status attainment model may assume that the father's education does not have a direct effect on the son's social position, only an indirect effect through the son's educational level. This is a conditional independence assumption, specified in the graph model by a Markov property, between the first two variables, given the third one. Several more relevant examples of the application of graphical models to problems in social mobility research will be given throughout the paper.

It will be illustrated (see, e.g., Example 5) that it is the Markov properties that define the model and it may happen that different graphs imply the same conditional independences (Markov equivalence). These Markov properties are specific to categorical data. The choice between continuous or categorical data is sometimes conceptual but sometimes is only implied by choices made during data collection. A celebrated example of a conceptual choice is social position. In American sociological thinking, social position is traditionally conceptualized as social prestige measured by continuous variables (e.g. Blau and Duncan 1967, Treiman 1970). In Europe, to the contrary, social position is defined as social class, not necessarily of a fully ordered kind, and is being treated as a nominal variable (see, e.g., the EGP class scheme, Erikson and Goldthorpe 1992). These choices have far reaching consequences on the kinds of analyses that are usually performed with such data. While the continuous approaches usually seek simplicity and structure in some kind of a linear relationship, the categorical approach allows for much finer modeling. Indeed, as far as the conditional independence structure, or more generally, the structure of conditional associations among variables is concerned, categorical data allow for richer structure than continuous data do under the almost automatic assumption of multivariate normality. This is illustrated in the next example.

*Example 1*

Spouses tend to be similar with respect to education, socio-economic status, personality etc. Spouse similarity is a widely investigated topic in sociology, psychology and epidemiology (see, e.g., Watkins and Meredith 1981). The question of interest here is whether similarity in cognitive abilities changes over time during marriage. Let the hypothesis be that the cognitive abilities of the male ( $M$ ) and female partner ( $F$ ) are independent of the length of marriage ( $L$ ):  $M \perp L, F \perp L$ . However, similarity between spouses may be stronger (positive association) if marriage is longer, that is  $MF \not\perp L$ . The following hypothetical data illustrate this possibility ( $F$  and  $M$ : 1 - low, 2 - high,  $L$ : 1 – short, 2 – long):

	$L=1$	$L=2$		$L=1$	$L=2$		$L=1$	$F=1$	$F=2$		$L=2$	$F=1$	$F=2$
$M=1$	60	30	$F=1$	80	40	$M=1$	50	10	$M=1$	27	3		
$M=2$	40	20	$F=2$	20	10	$M=2$	30	10	$M=2$	13	7		

Note that the first two tables are marginal, the last two are conditional. The corresponding odds ratios (see Rudas 1998) are

$$OR(M, L) = 1, \quad OR(F, L) = 1,$$

$$OR(M, F | L = 1) = 1.67 < OR(M, F | L = 2) = 4.85,$$

that is, the independence assumptions and the assumption of increasing similarity hold. If the three variables had a joint normal distribution, the independences  $M \perp L, F \perp L$ . would also imply that the association between  $M$  and  $F$  is independent of  $L$ , that is, similarity would not increase with time. ■

The exposition in this paper assumes that the reader is familiar with the fundamentals of log-linear modeling (Rudas 1998, Agresti 2002). Sections 2, 3 and 0 present graphical models of increasing complexity and illustrate their applications for associations, effects, and both associations and effects, respectively. Section 5 defines path models derived from discrete graphical models that impose a very simple structure that is well represented by graphs and has readily interpretable parameters. The theory is illustrated by several worked out examples, some of which give new insight into certain aspects of social mobility (see, e.g., Example 8).

## 2 Associations only – undirected graphs

Undirected graphical models (abbreviated as UG models, also called Markov random fields) are defined over a graph whose vertices are connected with lines (undirected edges). The underlying assumption is that all variables are considered “on an equal footing” (a term used by Cox and Wermuth 1996). UGs can be used to study symmetric associations conditioned on all remaining variables.

The graph encodes a set of conditional independence relations, known as the Markov property (MP) associated with the graph. An MP translates missing edges between vertices into conditional independence statements between variables. The conditional independence properties embodied in an undirected graph may be formulated on three different levels (Lauritzen 1996); they differ in the size of variable sets claimed to be conditionally independent. The first property is about two variables, the second is about one variable and a set of variables, and the third is about two sets of variables.

*Pairwise Markov property (P):* Any two non-adjacent variables are conditionally independent given all other variables. Two variables are non-adjacent if there is no line between them.

*Local Markov property (L):* A variable is conditionally independent of all variables non-adjacent to it, given all variables adjacent to it.

*Global Markov property (G):* Let  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  disjoint subsets of variables.  $\mathcal{A}$  and  $\mathcal{B}$  are conditionally independent given  $\mathcal{C}$ , if  $\mathcal{C}$  separates them.  $\mathcal{C}$  separates  $\mathcal{A}$  and  $\mathcal{B}$  if every path from a variable in  $\mathcal{A}$  to a variable in  $\mathcal{B}$  goes through a variable in  $\mathcal{C}$ .

Though the three properties above seem to give different interpretation to graphs, assuming a positive joint distribution, they are equivalent (Lauritzen 1996). Throughout the paper, positivity is assumed, hence  $P$ ,  $L$  and  $G$  are equivalent. However, it may be beneficial to keep in mind all of them. For example, it may be useful to know that if two variables are not connected, then they are conditionally independent given all others ( $P$ ), but also given the adjacent vertices of one of them ( $L$ ) or given any of their separating sets ( $G$ ).

The Markov properties show that the presence of a particular edge has no intrinsic meaning, because its interpretation depends on the presence or absence of other edges. Consider, for example, variables  $A$ ,  $B$ , and  $C$ , with edges  $AB$  and  $BC$ . The meaning of the  $AB$  edge depends on whether  $A$  and  $C$  are also connected. If they are, the  $AB$  edge means that  $A$  and  $B$  may have an association conditioned on  $C$ , and the conditional association may be different at different levels of  $C$ . If they are not, this conditional association may be present but cannot be affected by  $C$ .

The next example illustrates the equivalence of the MPs.

#### *Example 2*

Consider, again, the three variables  $M$ ,  $F$ ,  $L$  (cognitive abilities of the male or female partner, length of marriage) as in Example 1. To represent the models in which  $M \perp L | F$  or  $F \perp L | M$  holds, the  $ML$  or  $FL$  lines should be deleted, respectively, from the complete graph. The graph from which both  $ML$  and  $FL$  edges are deleted represents the model

- $(M \perp L | F) \cap (F \perp L | M)$ , according to  $P$ ,
- $(L \perp MF) \cap (F \perp L | M) \cap (M \perp L | F)$ , according to  $L$ ,
- $(L \perp MF) \cap (L \perp F) \cap (L \perp M)$ , according to  $G$ .

With the help of the general properties of conditional independence (see Lauritzen 1996:29) these can be proved to be also equivalent to the simple multiple independence  $L \perp MF$ . ■

A real-valued function of the probabilities of a distribution is called a parameter, while a set of parameters is a parameterization if it is one-to-one (and, thus, invertible) mapping from the distributions onto the parameter values. Statistical models may be defined by restrictions on the parameters. The non-restricted, free parameters measure the strength of effects allowed in the model and parameterize (characterize) all distributions in the model. Conditional independence (CI) models, like the one considered above, can be defined by setting the values of appropriate log-linear parameters (see e.g., Agresti 2002) to zero.

For example, the model  $A \perp B | C$  may be parameterized by setting the  $AB$  and  $ABC$  effects to zero:

$$\lambda_{ij*}^{ABC} = 0, \quad \lambda_{ijk}^{ABC} = 0, \quad (1)$$

for all  $i, j$  and  $k$ , where the  $ABC$  superscript denotes that the parameters are calculated within the  $ABC$  three-dimensional table. The subscript denotes the indices of the variables to which the parameter pertains. The omitted variables are denoted by an asterisk. We shall refer to the variables in the superscript as the marginal and to the variables in the subscript as the effect. In this paper, the  $\lambda$  parameters are calculated using the usual effect coding (Agresti 2002). Some of the restrictions in (1) are redundant, since the sum of the parameters corresponding to the same effect is equal to zero. In the following, this notation refers only to non-redundant restrictions by omitting the last level of the variables, e.g.  $\lambda_{ij*}^{ABC} = 0$  abbreviates  $\lambda_{ij*}^{ABC} = 0, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1$ .

The free parameters not restricted by the model are

$$\lambda_{***}^{ABC}, \quad \lambda_{i**}^{ABC}, \quad \lambda_{*j*}^{ABC}, \quad \lambda_{**k}^{ABC}, \quad \lambda_{*jk}^{ABC}, \quad \lambda_{i*k}^{ABC}.$$

In general, if  $\mathcal{A}, \mathcal{B}$  and  $\mathcal{C}$  are sets of variables, define, with  $\mathbb{P}(\cdot)$  denoting the set of all subsets,

$$\mathbb{D}(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \mathbb{P}(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}) \setminus (\mathbb{P}(\mathcal{A} \cup \mathcal{C}) \cup \mathbb{P}(\mathcal{B} \cup \mathcal{C})). \quad (2)$$

Then  $\mathcal{A} \perp \mathcal{B} \mid \mathcal{C}$  can be defined by setting those parameters to zero which are associated with any of the effects that are elements of  $\mathbb{D}(\mathcal{A}, \mathcal{B}, \mathcal{C})$ . For example, with  $\mathcal{A} = A, \mathcal{B} = DE$  and  $\mathcal{C} = C$ ,  $A \perp DE \mid C$  can be parameterized as

$$\lambda_{i**m}^{ACDE} = \lambda_{i*l*}^{ACDE} = \lambda_{i*lm}^{ACDE} = \lambda_{ik*m}^{ACDE} = \lambda_{ikl*}^{ACDE} = \lambda_{iklm}^{ACDE} = 0, \quad (3)$$

where those effects are restricted to zero which contain  $A$  and at least one of  $D$  and  $E$ .

*Example 2* (continued)

The log-linear parameterization of the model in Example 2 can be based, for example, on the conditional independences  $M \perp L \mid F$  and  $F \perp L \mid M$ :

$$\lambda_{*jk}^{FML} = \lambda_{i*k}^{FML} = \lambda_{ijk}^{FML} = 0,$$

while the free parameters, parameterizing the distributions in the model are:

$$\lambda_{i**}^{FML}, \lambda_{*j*}^{FML}, \lambda_{**k}^{FML}, \lambda_{ij*}^{FML}. \quad \blacksquare$$

The next example shows a more complex UG and also illustrates that although a log-linear parameterization is possible, substantive interest may require different parameters to be used.

*Example 3*

Spouses tend to be similar with respect to not only cognitive abilities (C) but education (E) and socio-economic status (S) as well (e.g. Watkins and Meredith 1981). Consider the UG model in *Figure 1* for these characteristics.

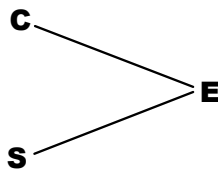


Figure 1 An undirected graph model for the association of cognitive abilities (C), education (E) and socio-economic status (S).

The UG can be translated as  $S \perp C | E$ , which is a plausible assumption. Note that this model does not imply that  $C \not\perp E | S$  or  $C \not\perp E$ . If one is interested in whether the edge  $CE$  may be deleted, the models with and without the edge  $CE$  need to be compared on the basis of the observed data.

Spouse similarity may be due to partner selection factors and/or the effects of marriage and common environment. In order to exclude the effect of the latter factors, only newly married couples are considered in the partner selection model of Figure 2. Letters with/without prime denote the male partner's/female partner's characteristics, respectively.

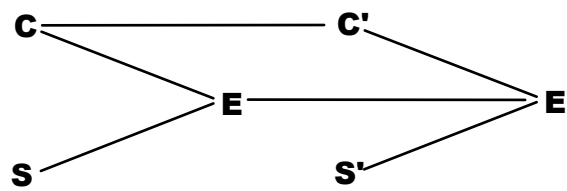


Figure 2 Expanding the model in Figure 1 with partner's characteristics

It is a natural expectation that  $S \perp C | E$  and  $S' \perp C' | E'$ , encoded by the initial graph, still hold. In the new graph  $E$  separates  $S$  and  $C$ , while  $E'$  separates  $S'$  and  $C'$ , hence these statements still really hold in the new model.

The missing edge between  $S$  and  $S'$  can be translated by the global MP as  $S \perp S' | EE'$ . That is, a woman with, for example, university degree may tend to select a partner taking into account his education (the  $EE'$  edge is present), but the potential spouses' socio-economic status does not affect her decision additionally. In other words, socio-economic status affects partner selection only *indirectly*, through education. Regarding  $S$  and  $S'$ , the global MP implies other conditional independences as well. For example  $S \perp S' | E$  also holds.

If the edge between  $C$  and  $C'$  was also deleted, then  $EE'$  would separate  $CS$  and  $C'S'$ , hence  $CS \perp C'S' | EE'$  would hold in the corresponding model. That is, both cognitive abilities and socio-economic status would affect partner selection only indirectly, through education.

The model of Figure 2 can be easily parameterized within the  $CSEC'S'E'$  table with log-linear parameters. However, if one is interested in estimating the strength of association between educational levels of spouses given their socio-economic statuses, but not given their cognitive abilities, the analysis of  $EE'SS'$  marginal table is relevant. The  $\lambda_{ij**}^{EE'SS'}$  parameter in this table is a log-linear parameter which is not computed from the entire contingency table rather from a marginal of it. The next section shows parameterizations using marginal log-linear parameters.

### 3 Effects only - directed acyclic graphs

UGs can be used to study symmetric associations. However, in some cases associations are not symmetric, but represent relations of response and dependence. Such situations motivate introducing directed acyclic graph models (DAG models, also called Bayesian networks).

DAG models may be illustrated with a graph whose vertices are connected with arrows (directed edges) under the constraint that no directed cycles are present.

In the case of DAGs we give only the pairwise and local MPs. For the global MP see, e.g., Lauritzen (1996). In a directed graph, parents are those vertices from whom arrows point to the given vertex, while descendants are the vertices that are reachable from the given vertex through a directed path.

*Directed Pairwise Markov property (DP):* If variables  $A$  and  $B$  are such that  $B$  is neither a parent nor a descendant of  $A$ , then  $A$  is independent of  $B$  given the other non-descendants of  $A$ .

*Directed Local Markov property (DL):* A variable is conditionally independent of its non-descendants (excluding its parents), given its parents.

Arrows of a DAG have no individual meaning, only the graph structure as a whole has an interpretation. However, directed MPs are sensitive to the direction of the arrows, in an intuitively plausible way: parents have direct effect on their children, grandparents affect their grandchildren only indirectly etc. The next example illustrates this with a DAG model of a classical problem in sociology.

#### Example 4

Figure 3 presents a status attainment model. The variables are  $E$  - Education,  $O$  - Occupation,  $I$  - Income,  $F$  - Father's education,  $G$  - Father's occupation. There is a time order among the variables here, so the directions of the arrows are unambiguous.

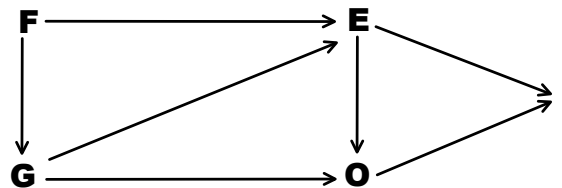


Figure 3 Directed acyclic graph model for status attainment

The Local Markov property gives that

$$O \perp F | EG, \quad I \perp FG | EO \quad (4)$$

that is, the father's education does not affect the offspring's occupation directly, however, taking into account edges  $FE$  and  $EO$ , it may have an indirect effect, through education. Similarly, social origin does not affect income directly, possibly only through education and occupation. ■

The original source of the graph in Figure 3 is Duncan et al. (1968), but it was also used by Treiman (1970), whose industrialization hypothesis was formulated within this model. They obtained the graph through path analysis and used continuous variables to operationalize social status. Path modeling and graphical modeling interpret missing arrows in different ways. Discussion of the relationships between graphical modeling and structural equation modeling (SEM), which is an extension of classical path analysis, has been given by Wermuth (1988), Whittaker (1990), Cox and Wermuth (2001), and Wermuth (2003) for the Gaussian case. The SEM approach has been extended to handle categorical data in many ways. For a summary of the relationship between categorical SEM and DAGs, see Bergsma et al.(2009), Chapter 5.1.



DAG models restrict marginals, hence the usual log-linear parameterization cannot be used to parameterize them. A general method of obtaining a proper parameterization is based on a well-numbering of the variables. A well-numbering is an order of the vertices, in which if  $A$  precedes  $B$ , then  $A$  is a nondescendant of  $B$ . Such an ordering is always possible. If the variables of the graph are well-numbered, then a variable  $A$  is conditionally independent of those nondescendants (excluding the parents of  $A$ ) that precede  $A$ , given the parents of  $A$  (Lauritzen et al. 1990). The set of these new CIs is equivalent to the set of the original CIs read-off by the MPs.

Such a parameterization was proposed by Rudas et al. (2006) and is based on the  $n$  (=number of variables) marginals containing the first  $k$  ( $k=1, \dots, n$ ) elements of the well-numbering. All effects appear within the marginal that comes first from among those that contain it. This is illustrated now on the model in Example 4.

*Example 4 (continued)*

A well-numbering of the variables is  $F, G, E, O, I$ , so the marginals are  $F, FG, FGE, FGEO, FGEOI$ . The CIs implied by the well-numbering are identical with the original CIs of (4) The log-linear parameterization of the  $FGEOI$  table is obtained as:

$$\begin{aligned} & \lambda_i^F, \lambda_{*j}^{FG}, \lambda_{ij}^{FG}, \lambda_{**k}^{FGE}, \lambda_{i*k}^{FGE}, \lambda_{*jk}^{FGE}, \lambda_{ijk}^{FGE}, \lambda_{***l}^{FGEO}, \lambda_{i**l}^{FGEO}, \lambda_{*j*l}^{FGEO}, \lambda_{**kl}^{FGEO}, \lambda_{ij*l}^{FGEO}, \lambda_{i*kl}^{FGEO}, \\ & \lambda_{*jkl}^{FGEO}, \lambda_{ijkl}^{FGEO}, \lambda_{****m}^{FGEOI}, \lambda_{i***m}^{FGEOI}, \lambda_{*j**m}^{FGEOI}, \lambda_{**k*m}^{FGEOI}, \lambda_{***lm}^{FGEOI}, \lambda_{ij**m}^{FGEOI}, \lambda_{i*k*m}^{FGEOI}, \lambda_{i**lm}^{FGEOI}, \lambda_{ijk*m}^{FGEOI}, \quad (5) \\ & \lambda_{ij*lm}^{FGEOI}, \lambda_{i*klm}^{FGEOI}, \lambda_{*jk*m}^{FGEOI}, \lambda_{*j*lm}^{FGEOI}, \lambda_{*jklm}^{FGEOI}, \lambda_{**klm}^{FGEOI}, \lambda_{ijklm}^{FGEOI}. \end{aligned}$$

To define  $O \perp F \mid EG$ , all the parameters are set to zero which are associated with any of the effects that are elements of  $\mathbb{D}(O, F, EG)$  within the  $FGEO$  marginal:

$$\lambda_{i**l}^{FGEO}, \lambda_{ij*l}^{FGEO}, \lambda_{i*kl}^{FGEO}, \lambda_{ijkl}^{FGEO}, \quad (6)$$

while  $I \perp FG \mid EO$  can be parameterized with restrictions on any of the effects being elements of  $\mathbb{D}(I, FG, EO)$  within the  $FGEOI$  table:

$$\begin{aligned} & \lambda_{i***m}^{FGEOI}, \lambda_{*j**m}^{FGEOI}, \lambda_{ij**m}^{FGEOI}, \lambda_{i*k*m}^{FGEOI}, \lambda_{i**lm}^{FGEOI}, \lambda_{ijk*m}^{FGEOI}, \lambda_{ij*lm}^{FGEOI}, \lambda_{i*klm}^{FGEOI}, \lambda_{*jk*m}^{FGEOI}, \lambda_{*j*lm}^{FGEOI}, \lambda_{*jklm}^{FGEOI}, \quad (7) \\ & \lambda_{ijklm}^{FGEOI}. \end{aligned}$$

From the perspective of the  $FGEOI$  table, the parameters in (6) are marginal log-linear parameters. Marginal log-linear parameters can be interpreted as measuring average conditional association among the variables involved in the effect, conditioned on all other variables in the marginal and then averaged over all possible categories of the conditioning variables. For an in-depth discussion of how the marginal log-linear parameters can be interpreted, see Bergsma and Rudas (2003) and Rudas and Bergsma (2004).

An important characteristic of models and parameterizations is smoothness. This aids the interpretation of models, and guarantees the applicability of standard asymptotic theory. It is needed to compute maximum likelihood estimates and to test model fit, as will be done in this paper. Rudas et al. (2010) identified hierarchical conditional independence models with parameters having a simple combinatorial property that implies that the model and its parameterization are smooth. The hierarchy property can be checked easily: it holds if the marginals in which the model is defined can be arranged in a non-decreasing order (i.e., no marginal contains one which comes later in the sequence) so that every effect is parameterized in the earliest marginal containing it. For to the status attainment model of Figure 3, the parameterization in (5) is hierarchical, hence smooth.

Another characteristic of parameterizations important for interpretation is variation independence. Variation independence means that the joint range of the parameters is the combination of the separate ranges of the parameters involved. If the parameters are not variation independent, they cannot be interpreted without paying attention to the other parameters, moreover, improperly combined parameters may lead to non-existing models. Bergsma and Rudas (2002) gave a simple combinatorial condition called ordered decomposability for hierarchical marginal log-linear parameterizations to be variation independent. The parameter set of (5) is ordered decomposable, so the parameters are variation independent. Parameterizations produced by the well-numbering algorithm described above are always hierarchical and ordered decomposable.

Bergsma and Van der Ark (2009) developed the *R* package *cmm* to fit categorical marginal models. A model specification with no redundant restriction is necessary to apply the fitting algorithm and to calculate the degrees of the freedom of chi-squared statistics to test goodness-of-fit. If the parameterization is hierarchical, then the specification does not have redundancies, and the likelihood-ratio statistic has an asymptotic chi-squared distribution with the number of degrees of freedom being equal to the number of restricted parameters (Bergsma and Rudas 2002).

*Example 4* (continued)

The model is now applied to the International Social Survey Programme data (ISSP Research Group 1999). For the analysis, all respondents who actually worked and for whom all variables were recorded are included. Occupation variables *G* and *O* have three levels according to the country-specific terciles of the Standard International Occupational Prestige Scale (1: top third, 3: bottom third). Income (*I*) is binary (above/below country-specific median), father's education (*F*) and respondent's education (*E*) are also binary (respondent: with/without college degree; father: secondary school completed/not completed). It can be easily calculated that altogether 42 non-redundant parameters are restricted in (7). The number of degrees of freedom of the likelihood-ratio test is therefore equal to 42.

The model is fitted separately to data from different welfare systems, because welfare states can be thought to increase or decrease the impact of social origins on destinations (DiPrete and Grusky 1990). Nations are classified into five welfare-state types – conservative, liberal, social-democratic, Mediterranean and post-socialist – adding Mediterranean and post-socialist to the welfare capitalism model of Esping-Andersen (1990). Mediterranean was argued to be included as a distinct type by e.g., Leibfried (1992), Ferrera (1996) and Bonoli (1997). ISSP countries with data on each of the five variables that are representatives of the five types of welfare regimes were included: conservative (France, West-Germany), social-democratic (Sweden, Norway), liberal (United States, Australia, New Zealand), Mediterranean (Portugal, Spain) and post-socialist (Hungary, Czech Republic, Slovakia, Slovenia, Latvia, East-Germany, Poland). There are some empty cells in the five-dimensional *FGEIOI* tables. Since the fitting algorithm does not converge when empty cells are present (odds ratios involving zero cells are not defined), the observed zeroes were replaced by a small constant (1.0e-10).

The results are reported in Table 1. Details of the fitting procedure can be found on the article webpage.

\*\*\*Insert Table 1 near here \*\*\*

According to the *p* values, the model fits reasonably well for all regimes except the post-socialist one. However, the BIC shows only a slightly worse fit for the latter regime type. A detailed analysis of the post-socialist regimes (see the article webpage) shows that, according to the likelihood-ratio statistics, only the arrow *GI* could be deleted from the complete graph,

but not *FO* and *FI*. That is, father’s education has a direct effect on the offspring’s occupation and income. This is possibly a consequence of the negative educational discrimination during state socialism: children with highly educated fathers who suffered negative discrimination nevertheless inherited the cultural capital, and they found other ways to make achievements.

Making sense of the above results involves interpretation of the parameter estimates. Estimates of the free parameters are given on the article webpage for each regime separately. In case of post-socialist regimes, the adjusted standardized residuals are of particular interest. These measure deviations between observed and fitted cell frequencies, and deeper insight can be gained into the reason for lack of fit by analyzing them.

As was mentioned earlier, good fit of a model containing an arrow does not imply the presence of that arrow, therefore the size and the standard error of the effects are also of interest. As Table 2 shows, in the case of liberal regimes, the *GO* effect, which is calculated conditioned and averaged over *FE*, is not significant in any of its categories. Within the *FGEO* marginal, effects *FGO* and *FGEO* have zero parameters, while *GEO* is also non-significant (see the article webpage), that is, the *GO* effect is zero not only on an average but also conditioned on either *F* or *E* or both. In other words, occupational prestige ascription does not operate directly in liberal regimes. Indeed, if the *GO* arrow is deleted from the graph for liberal regimes (for the details see the article webpage), the increase in likelihood-ratio statistic is not significant (10.1 on 8 degrees of freedom, BIC=-299).

A look at the last two columns of Table 2 reveals a different picture for conservative regimes. There are significant, strong effects here, with signs denoting that entry into high-prestige jobs is easier for those with high-prestige father even when controlled for education. The estimate of the effect *GEO* (see the article webpage) refines these findings.  $\lambda_{*313}^{FGEO}$  and  $\lambda_{*321}^{FGEO}$  are near significant negative and positive effects, respectively, showing that disadvantage of persons with low-prestige father is eased if the father completed secondary school, and vice versa, their disadvantage is enhanced if the father did not even finish his secondary education.

\*\*\*Insert Table 2 near here \*\*\*

■

The examples below illustrate the scope and the limitation of directional interpretation of DAGs.

*Example 5*

In some cases, a comparison of model fit permits conclusions on the direction of a particular effect. Consider three variables *F* – having Friends, *H* - Happiness and *S* - Subjective general health, and let the arrow  $F \rightarrow S$  be present, along with  $S \leftarrow H$  or  $S \rightarrow H$ , as in Figure 4.



Figure 4 Directed acyclic graph models for the association of happiness (H), subjective health (S) and having friends (F).

The first graph implies marginal independence between *F* and *H*, while the second graph implies conditional independence between them. That is, in this case it can be decided based on the data, which directions to choose.

This is illustrated with data from the European Social Survey (Norwegian Social Science Data Services 2008) for France. The variables are operationalized as  $F$  - “Anyone to discuss intimate and personal matters with”: yes/no,  $H$  - “How happy are you”: rather happy/rather unhappy) and  $S$  - “How is your health in general?": fair or good / bad). The model tests give a likelihood-ratio statistics of 4.80 and 1.68, on 1 and 2 degrees of freedom, respectively, which means the first model fits poorly but the second model fits reasonably well. Note that this result, in itself, does not imply a claim that someone’s subjective health would determine the individual’s happiness in France and even less that there would be a causal effect between these variables. General rules of the parameterization of DAG models are described later in this section. Details of this example are on the article webpage, where other DAG variants can be also found, fitted to data from several countries. For example, none of the two models fit the data from the Netherlands and Great Britain, but a third model,  $F \rightarrow H \rightarrow S$ , fits well the Dutch data. ■

In general, data cannot determine the direction of an effect. As an extreme example, consider the graphical representation of a spurious relationship ( $A \leftarrow B \rightarrow C$ ) and an intervening relationship ( $A \rightarrow B \rightarrow C$ ). They suggest logically entirely different association structures between  $A$  and  $C$ , however, the two graphs are Markov-equivalent implying the same model:  $A \perp C \mid B$ . Additionally, reversing both arrows of the intervening relationship graph the same conditional independence is obtained, again. The example illustrates the fact that arrows of a DAG have no actual meaning: by reversing one or both arrows of the intervening relationship graph, the model remains the same, showing that the intuitive notion of causality conveyed by these arrows is unfounded. If  $A \perp C \mid B$  holds, each of the three graphs is appropriate. However, as always, substantive knowledge may make some of them less relevant than others. A DAG may be Markov-equivalent not only with another DAG but also with an UG or with a chain graph, see Section 0.

These examples warn against naïve interpretations of graphs and are also relevant for data-driven approaches for identifying a best-fitting graph. The resulting graph is to be interpreted solely on the basis of conditional independences, other characteristics may not be justified by the data. If more than one best-fitting Markov-equivalent graphs turn up, the equivalences can be exploited in the interpretation (some examples are in Wermuth and Lauritzen 1990), or, if background knowledge is available, the graphs can be considered in the order of their plausibility.

## 4 Both associations and effects – chain graphs

UG and DAG models allow modeling exclusively either associations (lines) or effects (arrows). Chain graphs (CGs) with lines and arrows allow a distinction between response or intermediate response and explanatory variables (Wermuth and Lauritzen 1990) and generalize UGs and DAGs. The statistical properties of CG models were investigated by Lauritzen (1996), Cox and Wermuth, (1996), Frydenberg (1990), Andersson et al. (2001), Drton (2009). There are different competing approaches in defining the Markov properties associated with them, classified by Drton (2009) as Type I - IV properties. In the sequel only block-recursive Type-IV chain graph models will be discussed, also investigated by Marchetti and Luparelli (2010) and Rudas et al. (2010).

A chain graph may have both arrows and lines under the constraint that semi-directed circles do not exist. This implies that the vertices can be partitioned into components such that vertices within a component are reachable from each other through an undirected path. Markov properties of Type-IV (Wermuth and Cox 1992) consider each variable conditionally

on other variables in the preceding components (explanatory variables), but ignore the variables in the same component (intermediate variables).

The block-recursive Type-IV (in the following: block-recursive) Markov property is defined by Properties 1, 2 and 3 below. Property 1 is defined over the DAG formed by the chain components analogously to the directed local MP. Properties 2 and 3 interpret lack of lines and of arrows, respectively.

*Property 1.* Any component is conditionally independent of its non-descendant components (excluding its parent components), given its parent components.

*Property 2.* A subset  $\mathcal{B}$  of a component  $\mathcal{C}$  is conditionally independent of all non-adjacent variables within  $\mathcal{C}$ , given the parent components of  $\mathcal{C}$ .

*Property 3.* A subset  $\mathcal{B}$  of a component  $\mathcal{C}$  is conditionally independent of those variables in the parent components of  $\mathcal{C}$  which are themselves not parents of  $\mathcal{B}$ , given the parents of  $\mathcal{B}$ .

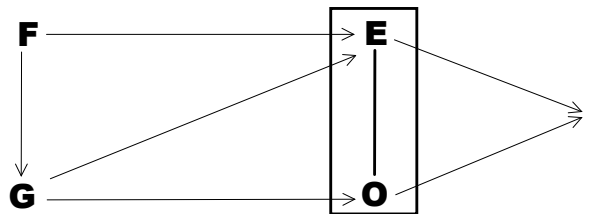
The next example shows a CG model that is obtained as a modification of a DAG model.

*Example 6*

The intergenerational mobility process of Figure 5 with Education and Occupation considered on an equal footing would be reasonable if lifelong learning was a widespread phenomenon. A plausible formulation of this is to omit  $E$  from the conditioning set of the independence of  $O$  and  $F$ , since  $E$  does not precede  $O$  anymore. The chain graph in Figure 5, under the block-recursive interpretation, yields that

$$O \perp F \mid G, \quad I \perp FG \mid EO \tag{8}$$

(Boxes are drawn around non-singleton components.) ■



*Figure 5 Chain graph model for status attainment*

Just like in the case of DAGs, one should be warned against seemingly plausible, but in fact fallacious interpretations which also affect model selection and model interpretation. Some main points (the first two hold for UGS and DAGs as well):

- a) Edges in the graph do not correspond to, but only allow direct associations or effects.
- b) The interpretation should be based solely on the conditional independences, see the case of Markov-equivalent graphs.
- c) Type-IV CG models are not generalizations of UG models in the sense that a graph with only lines would give necessarily the same set of independences interpreted as a UG model or as a CG model. The reason is that Type-IV Markov properties define conditioning sets by ignoring variables in the same component. However, CG models of any type are real generalizations of DAG models.

d) A line in a CG model does not correspond to an arrow with unspecified direction, or to the existence of a latent common parent or child. Lauritzen and Richardson (2002) show this for Type-I CG models but their findings can be extended to Type-IV models, see the article webpage.

A method for smooth hierarchical parameterizations of block-recursive Type-IV models was given by Rudas et al. (2010). Let  $C_1 \dots C_T$ , be a well-numbered order of the  $T$  components. Then, the marginals of the parameterization are the sets

- $PA(C_k) \cup \mathcal{B} : \mathcal{B} \subseteq C_k, k = 1 \dots T$ , in a hierarchical order, where  $PA(C_k)$  are the parent-components of  $C_k$ , and, as the last marginal
- $C_1 \cup \dots \cup C_k, k = 1 \dots T$ .

Each effect has its parameter in the first marginal containing it. The parameters set to zero to define the model are those associated with the effects in

$$\left\{ \mathbb{D}(\mathcal{B}, C_k \setminus \mathcal{B} \setminus nb(\mathcal{B}), PA(C_k)) : \mathcal{B} \subseteq C_k \right\} \cup \left\{ \mathbb{D}(\mathcal{B}, PA(C_k) \setminus pa(\mathcal{B}), pa(\mathcal{B})) : \mathcal{B} \subseteq C_k \right\} \cup \cup \mathbb{D}(C_k, PRE(C_k) \setminus PA(C_k), PA(C_k)),$$

for all  $C_k$ , where  $\mathbb{D}(\cdot)$  was defined in (2),  $nb(\mathcal{B})$  and  $pa(\mathcal{B})$  are the vertices from which lines go or arrows point to  $\mathcal{B}$ , respectively, and  $PRE(C_k)$  is the set of components that precede  $C_k$  in the well-numbering.

Ordered decomposability implying variation independence (Bergsma and Rudas 2002) is not generally ensured by the parameterization procedure but can be easily checked for a particular parameterization.

The above method produces smooth hierarchical parameterizations for block-recursive CG models. There is no method developed yet to obtain such parameterization for other types of CG models. See Rudas et al. (2010) and Forcina et al. (2010) for related results.

The next example illustrates the procedure of parameterizing block-recursive CG models.

*Example 6 (continued)*

A hierarchical parameterization of the mobility model can be given using the marginals  $F, FG, FGO, FGE, FGEO, FGEOI$ . Here

$$\mathbb{D}(O, PA(O) \setminus pa(O), pa(O)) = \mathbb{D}(O, F, G)$$

and

$$\mathbb{D}(I, PRE(I) \setminus PA(I), PA(I)) = \mathbb{D}(I, FG, EO)$$

contain the effects with which the following zero valued parameters are associated to define the model:

$$\lambda_{i^*k}^{FGO}, \lambda_{ijk}^{FGO}, \lambda_{i^*m}^{FGEOI}, \lambda_{j^*m}^{FGEOI}, \lambda_{ij^*m}^{FGEOI}, \lambda_{i^*k^*m}^{FGEOI}, \lambda_{i^*lm}^{FGEOI}, \lambda_{ijk^*m}^{FGEOI}, \lambda_{ij^*lm}^{FGEOI}, \lambda_{i^*klm}^{FGEOI}, \lambda_{j^*k^*m}^{FGEOI}, \lambda_{j^*lm}^{FGEOI}, \lambda_{i^*jklm}^{FGEOI}, \lambda_{ij^*klm}^{FGEOI}.$$

The remaining non-zero effects parameterize the distributions within the model. Ordered decomposability holds, as well, thus the parameters are variation independent.

This model was fitted to the ISSP data, and the results are reported in Table 3, details can be found on the article webpage. Both the LR-test and the BIC suggest that the model does not fit to the post-socialist data, while the conservative regimes show poor fit according to the LR-test. The present CG model and the DAG model discussed in Example 4 can be compared by BIC, and the DAG model turns out to fit better for each regime. ■

\*\*\*Insert Table 3 near here \*\*\*

The next example illustrates that whether two variables are to be connected by a line or an arrow, may or may not be derived from the analysis of data.

*Example 7*

Consider the three CG models in Figure 6 describing development of depression from working conditions through work stress. The variables are  $C$  = control over work pace, amount of work and others' work,  $T$  = able to complete work in working time,  $S$  = work stress,  $D$  = depression and  $I$  = inadequate social support. The three models differ only in the type of association between  $D$  and  $I$ .

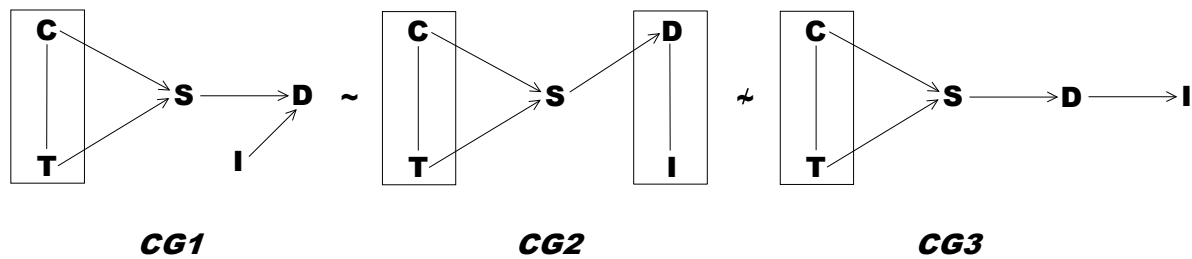


Figure 6 Chain graph models for the association of control over work ( $C$ ), being able to complete work in working time ( $T$ ), work stress ( $S$ ), depression ( $D$ ) and inadequate social support ( $I$ ).

The three graphs imply the following set of CIs, respectively:

$$CG1: D \perp CT \mid SI, I \perp CTS \tag{9}$$

$$CG2: DI \perp CT \mid S, I \perp S \tag{10}$$

$$CG3: D \perp CT \mid S, I \perp CTS \mid D \tag{11}$$

Properties (9) and (10) can be seen to be equivalent using the properties of conditional independence (see Lauritzen 1996: 29) however, their parameterizations implied by the above procedure are different and the better interpretable one should be chosen. The CIs in (11) define a model not equivalent to the other two, that is, the direction of  $DI$  association may be partially judged on the basis of data.

The models can be easily parameterized by the procedure described above. The parameterizations are given on the article webpage; they are hierarchical by definition and ordered decomposability holds as well so the parameters are variation independent. The webpage contains details of fitting these models to real data. ■

## 5 Path models

Further meaningful restrictions on graphical models may lead to models that are not graphical. E.g., if mother's and father's social status are both considered in an intergenerational mobility model, a plausible question is whether they have a joint effect on

the offspring's status over and above their separate effects, more formally, whether the corresponding second-order effect exists. Rudas et al. (2006), inspired by the modified path models of Goodman (1973), defined discrete path models by restricting discrete DAG models, setting higher than first-order effects in the hierarchical marginal log-linear parameterization to zero. Path models may be similarly derived from UG and CG models.

These models have only effects related to edges, and are analogous to those used in Gaussian path analysis since a multivariate normal distribution allows only two-way interactions. Path models are marginal log-linear models and, in addition to having a straightforward interpretation, inherit all the desirable statistical properties of the initial graphical model through an implied parameterization. Consider for example CG1 in Example 7. One may be interested in whether the predictors of depression act independently. The path model restricts also  $\lambda_{CTS}^{CTS}$ ,  $\lambda_{SDI}^{SDI}$  to zero, implying that  $C$  and  $T$  have only separate effects on  $D$  and also  $S$  and  $I$  have separate effects. The parameterization remains smooth and variation independent.

The semantics of path models depends heavily on the parameterization of the initial graphical model. E.g., in the case of CGs, if there is no complete ordering among the components, then different well-numberings yielding different parameterizations exist. The usual log-linear parameterization of UG models is unique, therefore only one path model can be derived from them. However, as we have shown in Example 3, a UG model may need to be reparameterized by marginal parameters more suitable to the research question. In this case the corresponding path model may change as well.

The final example applies a path model derived from a DAG to the status attainment problem.



*Example 8*

The model is derived from the DAG in Example 4. In addition to the zero parameters listed in (6) and (7), the path model restricts to zero  $\lambda_{ijk}^{FGE}$ ,  $\lambda_{*jkl}^{FGEO}$ ,  $\lambda_{**klm}^{FGEIO}$  from among the free parameters. The model assumes that the factors influencing social status have only separable effects. The parameterization inherited from that of the DAG remains smooth and variation independent.

The path model is fitted to the conservative, social-democratic, liberal and Mediterranean regimes to which the DAG model fitted well (see the article webpage for details). The path model is nested within the DAG model, hence the likelihood-ratio statistic can be used for comparison of model fit. The results are given in Table 4.

\*\*\*Insert Table 4 near here \*\*\*

The path model fits quite well to the liberal and conservative regimes, and in their cases, according to both  $G^2$  and BIC, the path model performs better than the DAG model. It's fit is also acceptable in the case of the other regimes.

Parameter estimates are given in Figure 7. For binary variables, only the non-redundant parameters pertaining to the smaller values are presented, while for variables with three categories each parameter is given. The parameters on the arrows are arranged in two matrices, where C denotes conservative, L denotes liberal regimes. A parameter estimate in the  $i$ th row and  $j$ th column of a matrix of the arrow  $AB$  pertain to the  $i$ th level of  $A$  and the  $j$ th level of  $B$ . The asterisks \*, \*\*, and \*\*\* indicate significance at the  $p < .05$ ,  $p < .01$ , and  $p < .001$  levels, respectively.

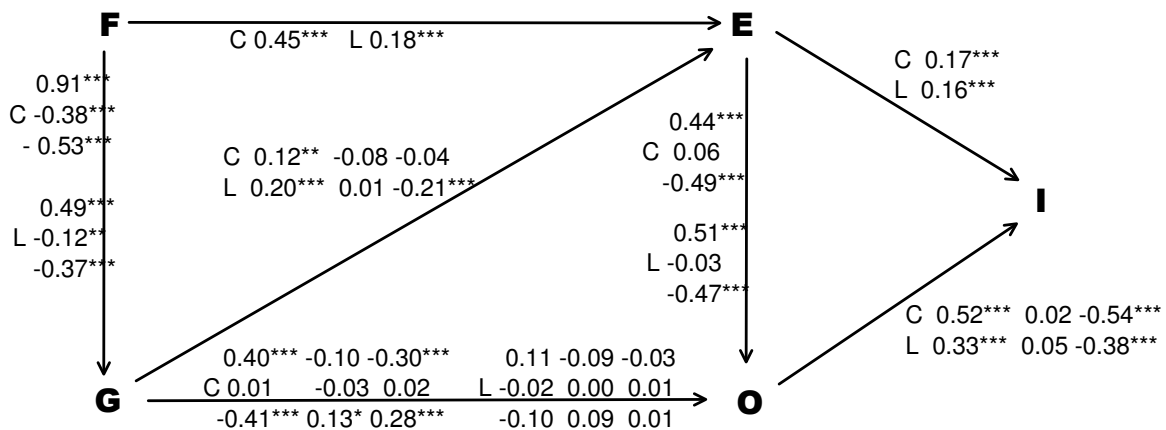


Figure 7 Parameter estimates for the path model

The results show, among others, that mobility operates through different channels in different welfare regimes. In liberal regimes, one sees significant GE, EO and non-significant GO association. In conservative regimes G has a strong direct effect on O, while its effect on E is weak. That is, in liberal regimes education mediates the association between origin and occupation, while in conservative regimes, the origin-occupation association is rather direct.

Social mobility in contemporary mobility research is often measured by the unidiff parameter of two-dimensional tables of respondent's by father's occupation (Erikson and Goldthorpe 1992, Breen 2004). When explaining cross-country differences in the value of the unidiff parameter, country-level (contextual) variables like type of the welfare regime or educational

policy are used, see e.g., Beller and Hout (2006). As our analysis shows, graphical modeling may add a more micro-level (individual) perspective to these results. ■

## 6 References

- Agresti, Alan. 2002. *Categorical data analysis*. 2nd ed., Hoboken, New Jersey: Wiley.
- Andersson, Steen A., David Madigan and Michael D. Perlman. 2001. "Alternative Markov properties for chain graphs." *Scandinavian Journal of Statistics* 28(1):33–85.
- Beller, Emily and Michael Hout. 2006. "Welfare states and social mobility: How educational and social policy may affect cross-national differences in the association between occupational origins and destinations." *Research in Social Stratification and Mobility*, 24(4):353–365.
- Bergsma, Wicher P., Marcel Croon and Jacques A. Hagenaars. 2009. "*Marginal models for dependent, clustered and longitudinal categorical data*." New York: Springer.
- Bergsma, Wicher P. and Tamás Rudas. 2002. "Marginal models for categorical data." *Annals of Statistics*, 30(1):140-159
- Bergsma, Wicher P. and Tamás Rudas. 2002. "Modeling conditional and marginal association in contingency tables." *Annales de la Faculte des Sciences de Toulouse*, 11(4):455-468.
- Bergsma, Wicher P. and L. Andries van der Ark. 2009. *cmm: Categorical marginal models*. R package version 0.3. [www.cmm.st](http://www.cmm.st)
- Blau, Peter M and Otis D. Duncan. 1967. *The American Occupational Structure*. New York: Wiley.
- Bonoli, Giuliano. 1997. „Classifying Welfare States: a Two-dimension Approach.” *Journal of Social Policy*, 26(3):351–372.
- Breen, Richard (editor). 2004. *Social Mobility in Europe*. Oxford: Oxford University Press.
- Cox, David R. and Nanny Wermuth. 1996. *Multivariate Dependencies – Models, Analysis and Interpretation*. London: Chapman & Hall.
- Cox, David R. and Nanny Wermuth. 2001. "Some statistical aspects of causality." *European Sociological Review*, 17(1):65–74.
- DiPrete, Thomas A. and David B. Grusky. 1990. "Structure and trend in the process of stratification for American men and women." *American Journal of Sociology*, 96(1):107–143.
- Mathias Drton. 2009. "Discrete chain graph models." *Bernoulli*, 15(3):736-753.
- Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux*. Oxford: Calderon.
- Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. New Jersey: Princeton University Press.
- European Social Survey (ESS), Round 4 Data*. 2008. Data file edition 3.0. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.
- Ferrera, Maurizio. 1996. „The »Southern« Model of Welfare in Social Europe.” *Journal of European Social Policy*, 6(1):17–37.
- Forcina, Antonio, Monia Lupparelli and Giovanni M. Marchetti. 2010. "Marginal parameterizations of discrete models defined by a set of conditional independencies." *Journal of Multivariate Analysis*, 101(10):2519-2527.

- Frydenberg, Morten. 1990. "The chain graph Markov property." *Scandinavian Journal of Statistics*. 17(4):333-353.
- Goodman, Leo A. 1973. "The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach." *Biometrika*, 60(1):179-192.
- International Social Survey Programme (ISSP), Social inequality III*, Study No. ZA3430. 1999. ISSP Research Group, distributor: GESIS Cologne Germany.
- Lauritzen, Steffen L. 1996. *Graphical Models*. Oxford: Clarendon Press
- Lauritzen, Steffen L., A. Philip Dawid, Birgitte N. Larsen and Hanns-Georg Leimer. 1990. "Independence properties of directed Markov fields." *Networks*, 20(5):491-505.
- Lauritzen, Steffen L. and Thomas S. Richardson. 2002. „Chain graph models and their causal interpretation (with discussion).” *Journal of the Royal Statistical Society*, Series B, 64(3):321-361.
- Lauritzen, Steffen L. and Nanny Wermuth. 1989. "Graphical models for association between variables, some of which are qualitative and some quantitative." *Annals of Statistics*, 17(1):31-57.
- Leibfried, Stephan. 1992. „Towards a European welfare state? On Integrating Poverty Regimes into the European Community.” Pp. 245-279 in *Social Policy in a Changing Europe* editd by Zs. Ferge and J. E. Kolberg. Frankfurt am Main: Campus Verlag.
- Marchetti, Giovanni M. and Monia Lupparelli. 2010. "Chain graph models of multivariate regression type for categorical data." *Bernoulli*, Forthcoming.
- Richardson, Thomas S. 2003. "Markov Properties for Acyclic Directed Mixed Graphs." *Scandinavian Journal of Statistics* 30(1):145-157.
- Rudas, Tamás. 1998. *Odds ratios in the Analysis of Contingency Tables*. Thousand Oaks: Sage.
- Rudas, Tamás and Wicher P. Bergsma. 2004. "On Applications of Marginal Models to Categorical Data." *Metron* 62(1):15-37.
- Rudas, Tamás, Wicher P. Bergsma and Renáta Németh. 2006. "Parameterization and estimation of path models for categorical data." Pp. 383-394 in *COMPSTAT 2006 Proceedings in Computational Statistics* edited by A. Rizzi and M. Vich. Heidelberg: Physica.
- Rudas, Tamás, Wicher P. Bergsma and Renáta Németh. 2010. "Marginal log-linear parameterization of conditional independence models." *Biometrika*, 97(4): 1006-1012
- Treiman, Donald J. 1970. "Industrialization and Social Stratification." Pp. 207–234 in *Social Stratification, Research and Theory for the 1970s* edited by E. O. Laumann. Indianapolis: Bobbs-Merill,
- Watkins, Maria P. and William Meredith. 1981. "Spouse similarity in newlyweds with respect to specific cognitive abilities, socioeconomic status, and education." *Behavior Genetics* 11(1):1-21.
- Wermuth, Nanny and Steffen L. Lauritzen. 1990. "On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Chain Models." *Journal of the Royal Statistical Society. Series B* 52(1):21-50.

Wermuth, Nanny and David R. Cox. 1992. "Graphical models for dependencies and associations." Pp. 235-247 in *Proceedings of the 10th Symposium on Computational Statistics, vol. 1*, edited by Y. Dodge and J. Whittaker. Heidelberg: Physica.

Wermuth, Nanny and David R. Cox. 2004. "Joint response graphs and separation induced by triangular systems." *Journal of the Royal Statistical Society B*, 66(3):687-717.

Whittaker, Joe. 1990. *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

## 7 Tables

<b>Regime type</b>	<b>G<sup>2</sup></b>	<b>p (df=42)</b>	<b>BIC</b>	<b>n</b>
Conservative	48.0	0.244	-247	1124
Social-democratic	49.9	0.189	-241	1007
Liberal	46.6	0.290	-252	1236
Mediterranean	45.2	0.338	-243	954
Post-socialist	94.6	0.000	-239	2831

*Table 1.* Fit statistics for the model in *Figure 1*

Parameter	Liberal regimes		Conservative regimes	
	Estimate	Standard error	Estimate	Standard error
$\lambda_{*1*1}^{FGEO}$	0.098	0.067	0.420	0.075
$\lambda_{*1*2}^{FGEO}$	-0.017	0.066	-0.010	0.077
$\lambda_{*1*3}^{FGEO}$	-0.081	0.066	-0.410	0.068
$\lambda_{*2*1}^{FGEO}$	-0.063	0.062	-0.103	0.081
$\lambda_{*2*2}^{FGEO}$	-0.021	0.061	-0.043	0.079
$\lambda_{*2*3}^{FGEO}$	0.084	0.061	0.147	0.068
$\lambda_{*3*1}^{FGEO}$	-0.035	0.064	-0.317	0.082
$\lambda_{*3*2}^{FGEO}$	0.038	0.063	0.053	0.075
$\lambda_{*3*3}^{FGEO}$	-0.003	0.065	0.264	0.065

Table 2. Estimated values of the GO effect

<b>Regime type</b>	<b>G<sup>2</sup></b>	<b>p (df=36)</b>	<b>BIC</b>	<b>n</b>
Conservative	60.1	0.007	-193	1124
Social-democratic	40.8	0.268	-208	1007
Liberal	50.8	0.05	-206	1236
Mediterranean	47.2	0.101	-200	954
<i>Post-socialist</i>	138.6	0.000	-148	2831

*Table 3* Fit statistics for the model in *Figure 5*

Regime type	Path model		BIC	Path model   DAG model	
	G <sup>2</sup>	p (df=50)		Difference in G <sup>2</sup>	p (df=8)
Conservative	57.0	0.230	-294	9.08116	0.335496
Social-democratic	66.6	0.058	-279	16.7937	0.03233
Liberal	58.5	0.192	-297	11.91966	0.154825
<i>Mediterranean</i>	67.4	0.051	-276	22.19558	0.004566

*Table 4* Fit of the path model and its relative fit given the DAG model