

In: (Wright, J., ed.) International Encyclopedia of the Social and Behavioral Sciences, 2nd ed., **10**, Elsevier, Oxford, 341–350; also available on arXiv1407.7783

Graphical Markov Models: Overview

Nanny Wermuth* and D.R. Cox**

**Mathematical Statistics, Chalmers University of Technology, Gothenburg, Sweden and Medical Psychology & Medical Sociology, Gutenberg-University, Mainz, Germany e-mail: wermuth@chalmers.se, and **Nuffield College, Oxford University, Oxford, UK e-mail: david.cox@nuffield.ox.ac.uk*

Abstract We describe how graphical Markov models emerged in the last 40 years, based on three essential concepts that had been developed independently more than a century ago. Sequences of joint or single regressions and their regression graphs are singled out as being the subclass that is best suited for analyzing longitudinal data and for tracing developmental pathways, both in observational and in intervention studies. Interpretations are illustrated using two sets of data. Furthermore, some of the more recent, important results for sequences of regressions are summarized.

1 Some general and historical remarks on the types of model

Graphical models aim to describe in concise form the possibly complex interrelations between a set of variables so that key properties can be read directly off a graph. The central idea is that each variable is represented by a node in a graph. Any pair of nodes may become coupled, that is joined by an edge. Coupled nodes are also said to be adjacent. For many types of graph, a missing edge represents some form of conditional independence between the pair of variables and an edge present can be interpreted as a corresponding conditional dependence. Because the conditioning set may be empty, or may contain some or all of the other variables, a variety of types of graph have been developed and are used to represent different types of structure.

A particularly important distinction is between directed and undirected edges. In the former an arrow indicates the direction of dependence of a response on an explanatory variable, the latter is also called a regressor. If, on the other hand, two variables are to be interpreted on an equal standing then the edge between them is typically undirected.

For instance, systolic and diastolic blood pressure are treated as being on equal standing because they are two aspects of a single phenomenon, namely a blood pressure wave.

Graphical Markov models started to be developed after 1970 as special subclasses of log-linear models for contingency tables and of joint Gaussian distributions, where conditional independence constraints are imposed such that conditioning is on all the other variables; see Darroch et al. (1980), Wermuth (1976, 1980). These models are typically represented by undirected graphs with edges that are full lines, called nowadays concentration graphs. The same types of graph were used by the physicist Gibbs (1902) to describe for two systems of particles having the same node sets, one as more complex whenever its nodes have more edges, that is larger numbers of ‘nearest neighbors’.

The first extension was to situations in which the variables can be arranged recursively, that is in an ordered sequence, so that each variable turns into a single response to variables in its past and may be explanatory only to other variables in its future; see Wermuth (1980), Wermuth & Lauritzen (1983). This led for single responses to what are now called directed acyclic graphs, for sequences of joint or single responses to the so-called regression graphs, and to distributions said to be generated over graphs.

For such generated Gaussian distributions, these models include as a subclass the path analyses of the geneticist Wright (1923, 1934). Wright had studied, for his data, sequences of exclusively linear regressions as possible stepwise generating processes. In his graphs, each missing arrow corresponds to the vanishing of the partial correlation coefficient given all remaining variables in the past of a given response, which in Gaussian distributions is a conditional independence constraint. For other types of generated distributions, an important issue is to identify parametric consequences of conditional independences; as for instance for the so-called CG-regressions, see Lauritzen & Wermuth (1989), Edwards & Lauritzen (2001).

In general, the vanishing of a correlation coefficient may coexist with a nonlinear dependence. Conversely, a substantial partial correlation can occur in spite of conditional independence; for an example, see Wermuth & Cox (1998). Currently, one knows also for jointly symmetric binary variables, generated over a directed acyclic graph, that an arrow vanishes in the graph if and only if there is a corresponding zero partial correlation; see Wermuth, Cox & Marchetti (2009). Zero partial correlations given all other variables may reflect conditional independences more generally, provided the distribution is generated over even more specialized types of graph; see Loh & Wainwright

(2013), Wermuth, Marchetti & Zwiernik (2014), Wermuth & Marchetti (2014).

The next extensions were to variables of any type so that a missing edge corresponds to a conditional independence. This exploits a proposal by the probabilist Markov (1912): many types of seemingly complex joint distribution may be much simplified by conditional independences. Important issues are defining sets of independences for a graph and criteria to derive all implied independences so that an independence structure captured by the graph becomes well understood; see here Section 4.

Further developments include different types of chain graph models; see Cox & Wermuth (1993), Drton (2009). The name ‘chain graph’ reflects a full ordering of the variables into a sequence of joint or single responses and possibly a set of remaining variables that is to capture the context of a study or properties of individuals given at the baseline and hence regarded as given. For parameterizations of Gaussian distributions with different types of chain graph, see Wermuth, Wiedenbeck & Cox (2006).

Statistical monographs documenting the early development of what are now called graphical Markov models are by Whittaker (1990), Edwards (1995), Lauritzen (1996), Cox & Wermuth (1996). For surveys of probabilistic aspects, see Pearl (1988) and Studený (2005). For graphical models used in expert systems, see Cowell et al. (1999) and for machine learning, see Wainwright & Jordan (2008).

Special theoretical results, not touched upon here, have been derived under the assumption that a given probability distribution satisfies all independence constraints represented by a graph and no more. Such distributions have been said to be ‘faithful to the graph’; see Spirtes, Glymour & Scheines (1993).

The existence of a family of faithful distributions means that one can choose a member at random, constrained just by the independences specified by a graph, and this distribution captures precisely the independence structure of the graph. But such an existence result is of no relevance when other constraints are known for subsets of parameters. For instance, the whole family of Gaussian distributions is faithful to concentration graphs, but there is a simple subfamily with additional parameter constraints that is always ‘unfaithful’ to concentration graphs, see Wermuth (2012), Section 2.5.

More importantly, for data analyses, in which it is crucial to respect known important dependencies, such an existence is also of no relevance. In most research settings, much is known about subsets of parameters because the direction and strength of some dependences derive from substantive theory or from results in previous empirical studies.

2 Sequences of regressions and data generating processes

Directed acyclic graphs are a common subclass of all currently known types of chain graph. Distributions generated over a directed acyclic graph result in a stepwise fashion in terms of univariate conditional distributions, also called recursive single-response regressions. These may be compatible with causal interpretations and therefore appear attractive, at first sight, for substantive research that is driven by causal hypotheses.

It is however well understood, from randomized clinical trials in particular, that real interventions often result in modifying more than a single response and possibly other features. For instance, a medication that is to reduce blood pressure, will typically affect both systolic and diastolic pressure. Hence models that are to be compatible with causal interpretations should contain response variables that may respond at the same time to a change in relevant explanatory variables. This is possible with joint response regressions.

Nevertheless, much of the current literature on causal modeling is based on directed acyclic graphs and on only virtual interventions; see Pearl (2009, 2014 in press). With such a virtual intervention, those changes in single response are recorded that result by fixing the values of some variables. This may for instance be treatments. One removes arrows pointing to these variables in a given graph and estimates treatment effects given these additional independence constraints. Implicit is the assumption that a given joint family of distribution remains unchanged, otherwise.

These recorded changes in virtual interventions, even though they are often called ‘causal effects’, may tell next to nothing about actual effects in real interventions with, for instance, completely randomized allocation of patients to treatments. In such studies, independences result by design and they lead to missing arrows in well-fitting graphs; see for example Figure 9 below, in the last subsection.

Sequences of regressions in joint and single responses are the most attractive chain graph models for observational as well as intervention studies. For these models, residual associations among the responses may often be regarded as secondary features. But, standard univariate estimation methods for regression coefficients may lead to distorted estimates whenever two associated joint responses on equal standing have disjoint subsets of regressors, a situation that has become known as seemingly unrelated regressions. When strong residual dependences remain after one has been regressing each response component separately on two different sets of regressors, then distorted parameter esti-

mates may result, see Haavelmo (1943), Zellner (1963), Drton & Richardson (2004).

For known types of chain graph other than regression graphs, recursive generating processes may not exist since response components are conditioned on other joint responses on equal standing. While substantive theory may occasionally support such models, this type of conditioning is typically counterintuitive whenever one main aim of an analysis is to identify consequences of interventions or early predictors for joint responses. But now, graphical criteria are available to decide whether the regression graph of a given sequence of joint or single regressions, defines the same independence structure as another chain graph, that is whether they are Markov equivalent; see Wermuth & Sadeghi (2012), Section 7, and here the subsection: Markov Equivalence of Regression Graphs.

For applications, arguably the most important developments concern conditions under which the regression graph can be used to trace development, to predict implied dependences in addition to implied independences, as well as graphical criteria for possible confounding effects when some variables are ignored or subpopulations are studied.

3 Regression graphs as partial summaries of data analyses

We give here two examples of data for which corresponding detailed statistical analyses have been published; see the Appendix in Wermuth & Sadeghi (2012) for the first and the Appendix in Wermuth & Cox (2013) for the second set of data.

In the regression graphs derived for both examples, the dashed lines capture remaining associations among two joint responses and full lines, for the context variables, reflect that conditioning is on all the context variables simultaneously. Estimation is for both data sets by local modeling of each response alone on its past and of response pairs on the union of their directly explanatory variables. When there are only independence constraints concerning responses and no seemingly unrelated regressions, local maximum-likelihood (m.l) estimates lead directly to proper m.l. estimates for the joint distribution; see Cox (2006) for such types of general principle.

In both following examples, data analyses lead to traceable regressions; see Wermuth (2012). This means that for a given ordered sequence of single and joint responses, the resulting regression graph permits one not only to read off independences, but more importantly for most applications, sequences of directed edges show pathways of development.

Tracing of paths is always possible for joint Gaussian distributions whenever each edge in its regression graph represents a substantial conditional dependence; see here Section 4. This may not be possible with graphs of some structural equation models; see Wermuth (1992). More importantly, sufficient conditions for traceability using just the graph, are now known for other types of distributions generated over regression graphs.

Data on chronic pain treatment

For some medical data, Figure 1 shows a first ordering of the variables. This ordering was derived in discussions by physicians and statisticians, prior to the analyses.

We are grateful to Judith Kappesser, now Department of Psychology, Liebig-University of Giessen, for letting us use her data. They are for 201 chronic pain patients who have been given a three-week stationary treatment at a chronic pain clinic. Two main research question are: which development is most influential for success or failure of treatment and is it necessary to include information on the patient’s site of main pain?

The response of primary interest is self-reported success of treatment, measured three months after discharge by a score summarizing several aspects of the illness. Among the context variables, capturing features that cannot be modified, were age, gender, lower or higher level of formal schooling and the number of previous other illnesses. Only those are shown in Figure 1 that had sizeable dependences to the responses of main interest.

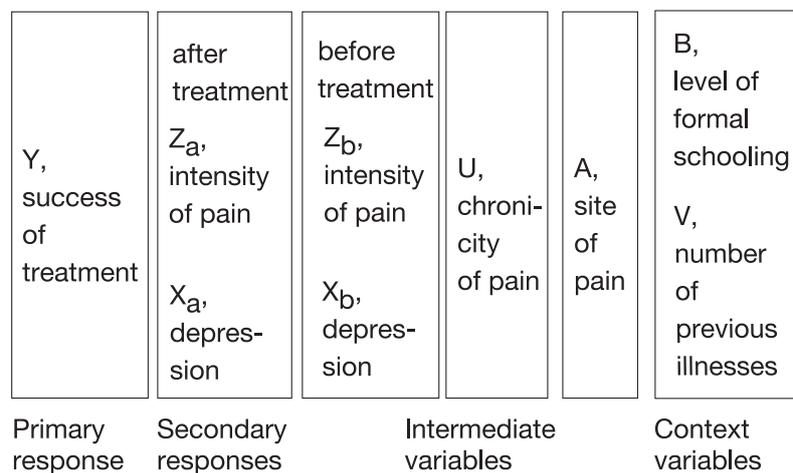


Figure 1: Initial ordering of variables in the chronic pain study for a sequence of regressions. Variables within a same box are treated on an equal standing. Variables in any one box are considered conditionally given all variables their past, listed within boxes to their right.

There are a number of intermediate variables. Before and after a three-week stationary treatment, questionnaire scores are available of depression and of intensity of

pain ranging from ‘no pain’ to ‘pain as strong as imaginable’. The chronification score incorporates different aspects, such as time since onset of pain, spreading of pain, use of pain relievers, the patient’s pain treatment history. Main site of pain has here two categories: ‘back pain’ and pain on the ‘head, face, or neck’.

The regression graph of Figure 2 summarizes some aspects of the statistical analyses. Throughout, symbol $\perp\!\!\!\perp$ means independence, symbol \pitchfork means dependence and symbol $|$ is to be read as ‘given’. The graph shows, in particular, which of the variables are needed so that for any given response, adding one more of the potentially explanatory variables does not improve prediction. Site of pain is an important intermediate variable since it is a node along a direction-preserving path of arrows pointing from level of schooling to treatment success, hence should be part of any future study of chronic pain.

Some of the directions and type of dependences, which cannot be read off the graph, are as follows. Patients with many years of formal schooling (13 years or more) are more likely to be head-pain patients, the others are more likely to be back-pain patients, possibly because more of them have jobs involving heavy physical work. Back-pain patients have higher scores of pain chronicity, reach higher stages of intensity of pain before treatment and report higher intensity of pain after treatment.

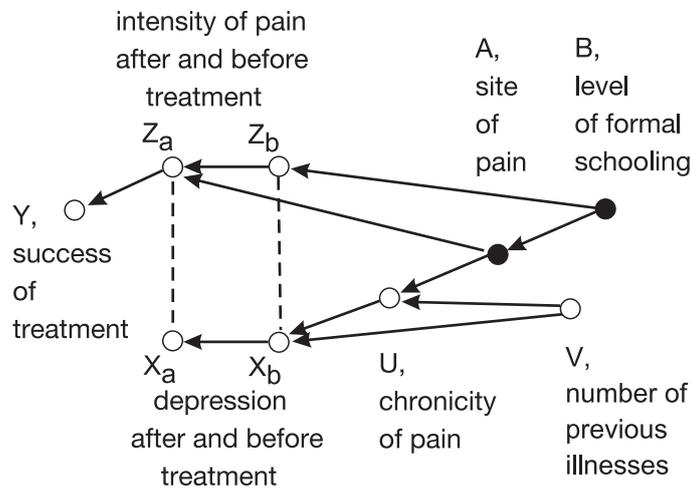


Figure 2: Well-fitting sequences of regressions which have a statistically significant relation for each edge present in the graph. Discrete variables are drawn as dots, continuous ones as circles. For instance, the two context variables are marginally independent, written as $B \perp\!\!\!\perp V$ and response Z_b is conditionally independent of V given A, B, U ; $Z_b \perp\!\!\!\perp V | A, B, U$, also A is marginally dependent on B ; written as $A \pitchfork B$, while U depends on V given A ; $U \pitchfork V | A$. **Note:** Full arrows in regression graphs are sometimes also drawn as dashed-line arrows and dashed lines as so-called arcs, which are full lines with arrow-heads at both ends

Never captured by the graph alone are nonlinear dependences, here of Y on Z_a .

Treatment success, Y , is low whenever higher levels of intensity of pain remain after treatment, Z_a . But at relatively lower levels of Z_a , treatment is clearly the more successful the lower the intensity of pain at discharge, Z_a . The model fits the data well since for each response taken separately, no indication was found that adding variables, further nonlinear or interactive effects would improve prediction and no strong dependences remained among the residuals of the joint responses on equal standing.

One important path of development is that patients with shorter formal schooling are more likely to get chronic back-pain and patients with chronic back-pain get help too late and respond less well to the type of treatment offered in the chronic pain clinic. This suggests as possible interventions to modify the type of treatment for the back-pain patients or, much more ambitiously, to raise the general level of formal schooling.

Data on child development

Here we use for data of 347 families participating in the ‘Mannheim study of children at risk’. We are grateful to Manfred Laucht, Central Institute of Mental Health, Mannheim, for permitting a reanalysis of the data.

The study started with a random sample of more than 100 newborns from the general population of children born near Mannheim, Germany. This sample was completed to give roughly equal subsamples, in each of nine level combinations of the two types of adversity at birth, categorized to be at levels ‘no, moderate or high’; see Laucht, Esser & Schmidt (1997). In other words, there was heavy oversampling of children at risk for motoric or cognitive deficits in later years and a random sample served as a control group, which provides, in particular, comparable norm values.

The recruitment of families stopped with 362 children. All measurements were reported in standardized form using the mean and standard deviation of the starting random sample. Of the 362 German-speaking families who entered the study when their first, single child was born without malformations or any other severe handicap, 347 families still participated when their child reached the age of 8 years.

There are joint responses of main and of secondary interest at age 8 and at age 4 and a half years. Each of these contains two components of possible deficits: cognitive or motoric; see the derived graph in Figure 3.

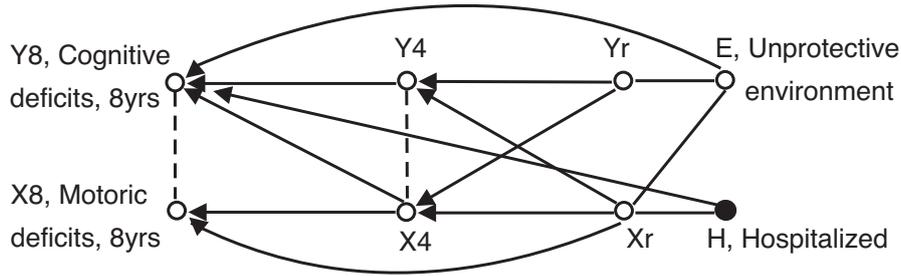


Figure 3: A well-fitting regression graph in the child development study; arrows point from regressors in the past to responses in the future; dashed lines are for dependent responses given their past; full lines are for dependent early risk factors given the remaining context variables.

$$\begin{aligned}
 Y_8 &: Y_4 + X_4^2 + E + H & X_8 &: X_4^2 + X_r \\
 Y_4 &: Y_r + X_r^2 & X_4 &: Y_r + X_r^2 \\
 Y_r &: E^2 & X_r &: E + H
 \end{aligned}$$

Note: In this notation supplementing the graph, every square term implies that also a main effect is included in a regression; see discussions of Wilkinson's notation in McCullagh and Nelder (1983).

One main difference to previous analyses is that we averaged three different assessments of each of two types of risk: recorded at birth, at 3 months and at two years. In both cases, this can be justified by the six observed pairwise correlations being all nearly equal; see e.g. Wermuth (2013). These averaged scores are risks at two years for motoric deficits, labelled X_r , and for cognitive deficits, labelled Y_r . Two possible additional risks were identified at three months after birth, a score called unprotective environment, E , and a binary variable H , which records whether the child had to be hospitalized during the first three months after birth.

The graph shows in particular for motoric deficits at age eight years of the child, X_8 , that no arrows are pointing to it from risks at 3 months, E, H , or from any variable for the cognitive side, Y_8, Y_4, Y_r , given information on the more recent motoric deficit, X_4 and the risk for motoric deficits, X_r . By contrast, cognitive deficits at age 8 years, Y_8 depend directly on unprotective environment at 3 months, on being hospitalized up to 3 months, as well as on motoric deficits at 2 years, given both more recent deficits, Y_4, X_4 . A tentative interpretation is given using Figure 7 in the next to last subsection.

The coding of variables implies that all dependences are positive so that deficits accumulate with each additional regressor. Some of the effects are accelerated compared to linear dependences. This is not reflected in the graph alone, but in Wilkinson's notation which adds to the graph sums of highest-order effects; see legend of Figure 3.

4 Questions regarding applications and statistical research

Here, we first list questions that arise in general when new statistical models are applied. We then give partial answers and references for sequences of regressions.

- (i) Are studies available in which the new models have been used fruitfully?
- (ii) Can well-fitting models be derived from data and be tested as hypotheses formulated for future data?
- (iii) What additional possibilities do the models offer to gain better insight into given research questions, especially to an understanding of development over time?

As to (i), case studies with sequences of regressions, which include joint responses, continue to accumulate. In addition to the two examples summarized here in the previous section; see for instance Cox & Wermuth (1996), Cheung & Andersen (2003), Hardt et al. (2004), Smith (2009), Hardt, Herke & Schier (2011), Marchetti et al. (2011), Schier et al. (2014 in press), Solis-Trapala et al. (2014 submitted).

As to (ii), one most attractive feature of sequences of regressions is that their fitting requires typically no new estimation techniques. Standard methods are often available; for linear regressions, see Weisberg (2014), for categorical responses, see Cox (1972), McCullagh & Nelder (1989), Andersen & Skovgaard (2010). To screen for nonlinear or interactive effects, see Cox & Wermuth (1994). Regression graphs represent hypotheses about generating processes which can be tested in future studies.

When seemingly unrelated regressions are hypothesized, direct estimation methods are based for discrete variables on generalized linear models; see Marchetti and Lupparelli (2011), for joint Gaussian distributions, estimation is possible with structural equation models; see e.g. Bollen (1989). But as mentioned before, when strong residual dependences remain in Gaussian seemingly unrelated regressions, especially for small sample sizes, estimates may be far from the population values.

One may also regard seemingly unrelated regressions as so-called reduced models and embed them in larger so-called covering models for which estimation is again standard; see Cox & Wermuth (1990). Often m.l. estimates in the covering model are not far from those obtained by separate regressions and are then good approximations to the m.l. estimates in the reduced model.

As to (iii), a number of new results have been obtained in the last years. We concentrate on the following few, discussed in the next three sections:

- a) graphical conditions for the Markov equivalence of two regression graphs, to be used for possible alternative interpretations or for different types of fitting algorithm,
- b) conditions and criteria for deriving implied independences and implied dependences from a given regression graph,
- c) ways of distinguishing different types of confounding that are implied by a generating process when some variables are unobserved, that is marginalized over, or a population subgroup is studied, that is when there is conditioning on the levels of some variables present in the generating process.

Markov equivalence of regression graphs

Two different graphs are Markov equivalent if they define the same independence structure. The independence structure of a regression graph, with a given set of nodes and an ordering for the responses, is determined by its list of missing edges for node pairs i, k . Such a graph has possibly two types of undirected edge, full lines between context nodes and dashed lines between responses on equal standing, as well as arrows starting at a regressor and pointing to a response; see Figure 1 to recall the meaning of these terms.

A missing full i, k -line between two context nodes means conditional independence given all remaining context nodes and a missing dashed i, k -line between two response nodes on equal standing means conditional independence given other nodes in the past of i and k . If a possible i, k -arrow pointing from k to i is missing, this means conditional independence given all other nodes in the past of i except for k .

A V-configuration in such a graph, often called just a V, consists of three nodes with two uncoupled nodes, called the outer nodes, i, k . Both outer nodes are coupled to an inner node, \circ . There are two types of Vs in regression graphs, named collision and transmitting Vs. In each of the three possible types of collision Vs:

$$i \text{---} \circ \longleftarrow k, \quad i \longrightarrow \circ \longleftarrow k, \quad i \text{---} \circ \text{---} k,$$

the two edges are removed when the inner node is marginalized over. In each of the remaining transmitting Vs, both edges are removed by conditioning on the inner node.

Directed acyclic graphs consist exclusively of arrows so that there can be only the middle V above, where two uncoupled regressors point to a common response. Graphs

of only full lines, often called concentration graphs, have only transmitting Vs, while graphs of only dashed lines, often called covariance graphs, have only collision Vs.

These three types of graph are subclasses of regression graphs. The names of the last two derive from parameters in joint Gaussian distributions. For such a distribution, zero off-diagonal elements, σ_{ik} , in the covariance matrix capture a marginal independence, and zero off-diagonal elements, σ^{ik} , in the concentration matrix, the inverse of the covariance matrix, capture a conditional independence given all the remaining nodes.

Two different types of regression graph, which have the same node set and the same set of missing edges but different types of edge for some node pairs, are Markov equivalent if and only if their sets of collision Vs coincide, even though for any given V, there may be any of the three types of collision Vs above; for a proof, see Wermuth and Sadeghi (2012). One example is shown in Figure 4.



Figure 4: A regression graph that is Markov equivalent to a concentration graph.

The graph on the left of Figure 4, resulted from a retrospective study, with more variables than reported here. Questions about their childhood were answered by 283 adult females when visiting a general practitioner for some minor health problems; see Hardt (2008). No nonlinear or interactive effects were detected.

The well-fitting graph contains two binary variables, level of formal schooling and severe sexual abuse during childhood, as well as three quantitative measurements. Age in years is recorded directly, the other variables being derived from questionnaires. Family status indicates the recalled social standing of the family during early childhood. Family distress includes psychological disturbances and alcohol or drug problems of the parents.

From the Markov equivalence to the concentration graph on the right of Figure 4, one knows, for instance, directly that sexual abuse is independent of the level of formal schooling given knowledge about the family status. This implication may be derived from the defining independences of the graph on the left or be based on the following separation result in graph theory. If in an undirected graph with three disjoint subsets α, β, c of the given set of nodes, every path between α and β has a node in c , then α

is separated from β by c , that is removal of set c leaves α and β disconnected. For concentration graphs, this separation implies $\alpha \perp\!\!\!\perp \beta | c$; see e.g. Lauritzen et al. (1990).

Figure 5 shows instead a regression graph on the left that is Markov equivalent to a covariance graph on the right. It is the simplest case of the seemingly unrelated regressions with regressors labelled 3, 4 and responses labelled 1, 2. For instance, it follows from the graph definitions that edge $1 \text{---} 2$ means $1 \pitchfork 2 | 3, 4$ on the left and $1 \pitchfork 2$ on the right.



Figure 5: A regression graph that is Markov equivalent to a covariance graph.

As discussed in the next section, to transfer such equivalence results for graphs to a distribution generated over a regression graph, this distribution has to mimic some properties of joint Gaussian distributions; see also Wermuth and Sadeghi (2012).

Implied independences in directed acyclic graphs

Work on reading all implied independences off a directed acyclic graph started with a path-braking, complex theorem, called d-separation; see Pearl (1988), Geiger, Verma & Pearl (1990). For the reduction of d-separation to the above described separation criterion for undirected graphs; see Lauritzen et al. (1990).

A third criterion is based on matrix representations of graphs, named their edge matrices. Square edge matrices contain zeros for missing edges, ones for edges present and ones along the diagonal to extend the graph theoretic notion of adjacency matrices in such a way that some sums or products of edge matrices describe transformations of graphs; see Wermuth, Wiedenbeck & Cox (2006). It gives the most direct criterion in that it leads to a matrix with the dimensions of two disjoint node subsets α and β : if this is a matrix of zeros only, then it is established that the graph implies a desired conditional independence.

In particular, let four disjoint subsets α, β, c, m partition the node set of a directed acyclic graph, let further $a = \alpha \cup m$, and $b = \beta \cup c$, then the transformation to the edge matrix $\mathcal{P}_{a|b}$ parallels the transformation of parameters in linear sequences of single-response regressions to the matrix $\Pi_{a|b}$ of least-squares regression coefficients with Y_a as response and Y_b as regressor. For joint Gaussian distributions, the submatrix of $\Pi_{a|b}$ for

rows α and columns β is zero, $\Pi_{\alpha|\beta.c} = 0$, if and only if $\alpha \perp\!\!\!\perp \beta|c$. In general, $\mathcal{P}_{\alpha|\beta.c}$, the corresponding submatrix of \mathcal{P}_{ab} is zero if the given graph implies $\alpha \perp\!\!\!\perp \beta|c$.

For directed acyclic graphs, the three types of criteria have been proven to be equivalent; see Marchetti & Wermuth (2009). For combining independences in sequences of regressions, the generated distribution has to satisfy additional conditions that mimic some properties of Gaussian distributions.

Independence properties of joint Gaussian distributions have been nicely summarized in the information theory literature, together with other properties that hold for all probability distributions; see Lněnička and Matúš (2007), Definition 1. To combine pairwise conditional independences in general sequences regression just as in the Gaussian case, two properties are needed; see Sadeghi and Lauritzen (2014). We call the presence of the intersection property the ‘downward combination’ and the presence of the composition property the ‘upward combination’ of pairwise independences.

A graphical illustration of these two properties is given for just three nodes i, h, k in Figure 6. In this simplest case, upward and downward combination mean, respectively,

$$(i \perp\!\!\!\perp h \text{ and } i \perp\!\!\!\perp k) \implies (i \perp\!\!\!\perp h, k) \implies (i \perp\!\!\!\perp h|k \text{ and } i \perp\!\!\!\perp k|h),$$

$$(i \perp\!\!\!\perp h|k \text{ and } i \perp\!\!\!\perp k|h) \implies (i \perp\!\!\!\perp h, k) \implies (i \perp\!\!\!\perp h \text{ and } i \perp\!\!\!\perp k).$$

Essential for both properties in the above two equations is the first implication. The second implication holds in all probability distributions and just serves to motivate the chosen names. For instance, upward combination means above to move from two simple to the two more complex conditional independence statements; expressed more generally: it means moving to increased conditioning sets.

On the left of Figure 6 is a complete regression graph with joint response $\{h, k\}$, on the right a complete directed acyclic graph with the single responses ordered as (i, h, k)

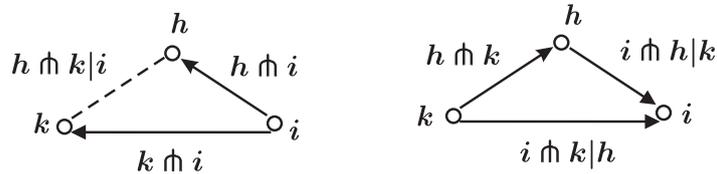


Figure 6: Two regression graphs where removal of the hi -edge and the ki -edge leads to $i \perp\!\!\!\perp h, k$, provided independences combine upward (left) and downward (right).

A sufficient condition for the intersection property is to have a strictly positive distribution; see San Martin, Mouchard & Rolin (2005) for extensions. Sufficient for the

composition property is that to every nonlinear effect there is also a linear one and that to every interactive effect there are also main effects. Such linear or main effects have to be substantial to be relevant for a model fit; see Németh and Rudas (2013) for an example when this property does not hold in a sociological context. In such instances, models more complex than graphical models are needed, such as those included for discrete variables in the class of marginal models; see Bergsma and Rudas (2002).

Deciding on implied dependences

Edge matrix results for directed acyclic graphs have been extended to the more general types of regression graph and, more importantly, to inferring that a generating regression graph implies the conditional dependence $\alpha \pitchfork \beta|c$ if $\mathcal{P}_{\alpha|\beta.c} \neq 0$; see Wermuth (2012).

To permit conclusions about dependences implied by sequences of regressions, the generated distribution has to have, in addition to upward and downward combination of independences, non-vanishing parameters to each edge present in the graph and it has to be dependence-inducing, a property also called singleton transitivity. Distributions with all above mentioned properties have been called traceable, since their regression graph can be used to study pathways of development.

In any specific application, there may be special parametric constellations that lead to path cancellations and hence to preventing us from recognizing dependences that are implied by the given generating process for other parametric constellations. This cannot happen, when all variables are coded, as in the Mannheim study, to have higher values for stronger adversities, for higher risks and for increased deficits and only linear dependences are observed or dependences that are accelerated compared to the linear effects. In addition, prediction of the strength or direction of induced associations may vary with the measures of dependence used; see Jiang, Ding & Geng (2014 in press).

For linear dependences among three variables, there are already three types of possible parameters: covariances, concentrations and regression coefficients. A detailed notation is needed to understand their relations:

$$\sigma_{12|3} = \sigma_{12} - \sigma_{13}\sigma_{23}/\sigma_{33}, \quad (1)$$

$$\sigma^{23.1} = \sigma^{23} - \sigma^{12}\sigma^{13}, \quad (2)$$

$$\beta_{1|3} = \beta_{1|3.2} + \beta_{1|2.3}\beta_{2|3}, \quad (3)$$

These three nice recursive properties were derived by different authors. Here, $\sigma_{12|3}$ denotes the conditional covariance of 1, 2 given 3; see Anderson (1958), $\sigma^{23.1}$ the con-

centration of 2, 3 marginalized over 1; see Dempster (1969), and $\beta_{1|3.2}$ the population coefficient of 2 when regressing 1 on both 2 and 3; see Cochran (1938) for the property.

The right-hand sides of the above equations contain, for trivariate Gaussian distributions, the parameters representing dependence in different graphs for three nodes. These parameters are in equation (1) for a covariance graph, in (2) for a concentration graph and in (3) for a directed acyclic graph: with 1 as response to 2, 3 and 2 as a response to 3; see also Figure 8, last section.

The coefficient of 3 when regressing 1 only on 3, denoted by $\beta_{1|3}$ can be regarded as the result of two paths: of $\beta_{1|3.2}$, an arrow pointing from 3 to 1, and of a sequence of two arrows starting at 3 and pointing to 1 via 2 with effect $\beta_{1|2.3}\beta_{2|3}$. Sometimes, the first is called the direct effect of 3 on 1, the second the indirect effect and $\beta_{1|3}$ the overall effect.

If for instance $0 = \sigma_{12} = \sigma_{13}$ in equation (1), then $\sigma_{12|3} = 0$ and consequently also $\beta_{1|3.2} = 0$ in equation (3) so that 1 has no dependence on 2 and 3 jointly. This is the Gaussian case of the more general notion of combining independences upwards since here $(1 \perp\!\!\!\perp 2 \text{ and } 1 \perp\!\!\!\perp 3) \implies 1 \perp\!\!\!\perp 2, 3$.

With more variables, the equations (1) to (3) generalize, by adding everywhere a larger conditioning set in equations (1) and (3), to get, for instance, $\sigma_{12|c}, \sigma_{12|3c}$ and $\beta_{1|3.c}, \beta_{1|3.2c}$, and by adding a larger marginalizing set in equation 2 to get $\sigma^{23.m}, \sigma^{23.1m}$. The three measures of dependence relate for a node set consisting of 1, 2, 3, c, m as

$$\beta_{1|3.c} = \sigma_{13|c} / \sigma_{33|c} = -\sigma^{13.m} / \sigma^{11.m},$$

derived via the sweep operator by Dempster (1969), and explaining above why $0 = \sigma_{12|3} = \beta_{1|2.3}$ and $0 = \sigma^{23.1} = \beta_{2|3}$ when all variances such as $\sigma_{33|c}$ and all precisions, such as $\sigma^{11.m}$, are nonzero in the two single-response regressions.

For joint Gaussian distributions, the dependence-inducing property can be proven in its simplest form using equation (1): for $1 \pitchfork 3$ and $2 \pitchfork 3$, one may have at most $\sigma_{12|3} = 0$ or $\sigma_{1|2} = 0$ but never both. In the case of $\sigma_{12|3} = 0$, the induced correlation, $\rho_{12} = \sigma_{12} / \sqrt{\sigma_{11}\sigma_{22}}$, has also been named a ‘spurious correlation’, since the dependence between the two variables 1 and 2 can be ‘explained away’ by conditioning on their common neighbor 3. For three variables, the dependence-inducing property can be written as:

$$(1 \pitchfork 3 \text{ and } 2 \pitchfork 3) \implies \text{at most } 1 \perp\!\!\!\perp 2|3 \text{ or } 1 \perp\!\!\!\perp 2 \text{ but never both.}$$

It has been proven that binary variables are dependence inducing; see Simpson

(1951). But there are constellations of counts for which both independences $1 \perp\!\!\!\perp 2|3$ and $1 \perp\!\!\!\perp 2$ seem to hold and a decision would have to be based on outside information.

Independence-predicting and independence-preserving graphs

The results for deciding on implied independences and dependences, based on an induced edge matrix, have been extended to derive regression graphs implied by a given generating process when, for instance, the order of the responses or the conditioning sets or the marginalizing sets are changed; see Wermuth & Cox (2004).

Then the so-called independence-predicting graphs, may be derived by using the partial closure operator (Wermuth, Wiedenbeck & Cox, 2006) as well as sums and products of edge matrices, but each of these operations has also an intuitive translation into closing special types of paths in graphs; see Wermuth (2012).

For instance, when in a study different from the Mannheim one, the same variables were available except for cognitive deficits after 4 years, Y_8, Y_4 , one predicts, from the starting graph in Figure 3 and an unchanged order of the remaining variables, the regression graph in Figure 7. In this case, it is just a subgraph of Figure 3 obtained by removing nodes and edges of Y_8 and Y_4 , no additional dependences are induced.

By contrast, when X_8 and X_4 are not available, all remaining variables turn out to be directly explanatory for Y_8 . One consequence of this finding is that it is more important for psychologists and psychiatrists to evaluate also motor development than it is for physicians to take also cognitive development into account.

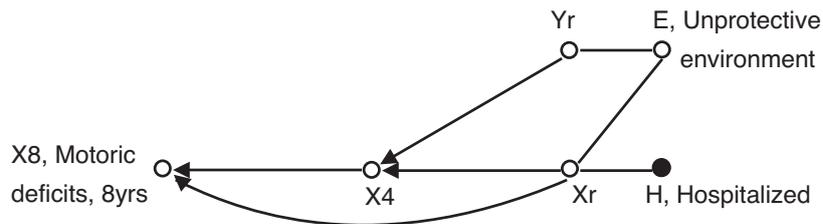


Figure 7: The regression graph induced by Figure 3 after marginalizing over Y_8 and Y_4 .

Typically, one cannot use such an induced regression graph to condition on or marginalize over more nodes and still see what the starting graph would have implied for this case, while this is possible with the so-called independence-preserving graphs. Starting from directed acyclic graphs, three types of such graphs have been proposed, the MC-graphs by Koster (2002), the maximal ancestral graphs by Richardson and Spirtes (2002) and the summary graphs by Wermuth (2011). For their relations and proofs of

Markov equivalence, see Sadeghi (2013). To construct these types of graph within the program environment R, see Sadeghi and Marchetti (2012).

Summary graphs may be derived for the more general regression graphs and, if obtained after marginalizing only, they provide graphical criteria to detect confounding.

Graphical representations of distortions

For traceable regressions, the regression graph represents an independence structure and a partly specified dependence structure. The graph can be viewed as a hypothesis of how sequences of single or joint responses generate a joint distribution when, for instance, corresponding point estimates of parameters are taken as the population values. For reliable interpretations of estimated effects, it is important to prevent, if at all possible, any important, sizeable distortion of the actual population parameters.

Different types of distortions of a treatment effect may be avoided by using regression graphs and appropriate study designs. To see this, we show in Figure 8 simple cases of under-conditioning, of over-conditioning and of direct confounding by using a node crossed by two lines, $\cancel{\circ}$, to indicate that the variable is to be marginalized over and a node surrounded by a square, \square , to indicate that the variable is to be conditioned on.

For instance, by using label 3 for treatment and label 1 for the outcome or response to 2 and 3, the graph on the left of Figure 8 shows one important intermediate variable, labelled 2, which could be, say, compliance of the patients to treatment.

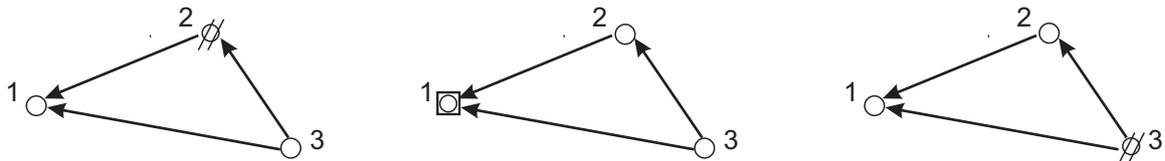


Figure 8: Left: under-conditioning for $1 \perp 3 | 2$ by ignoring an intermediate variable, middle: over-conditioning for $2 \perp 3$ by selecting levels of the common response 1 and right: direct confounding for $1 \perp 2 | 3$ by ignoring the common explanatory variable 3 for 1 and 2; generating a double edge $1 \rightleftarrows 2$ in the summary graph obtained by marginalizing over node 3.

In particular, when differential compliance is ignored, a distorted overall effect results which then coincides with the result of a so-called intention-to-treat analysis. Conditioning on the common response in the middle graph distorts the simple dependence of 2 on 3 and is also named a selection bias. More complex cases of such over-conditioning may occur, for instance, when there is an effect for $2 \leftarrow 3$ and, in addition, a path like $2 \rightarrow \square \leftarrow \cancel{\circ} \rightarrow \square \leftarrow 3$.

The simplest case of direct confounding, shown on the right of Figure 8, is also called the presence of an unmeasured confounder. It has a longer history than graphical models; see Vandembroucke (2002). It is avoided when there is a successful, fully randomized allocation of individuals to treatment levels, since in that case, all effects on treatment, observed or unobserved, are removed. This design leads also to a removal of all incoming arrows to the treatment variable in a regression graph and to the absence of any double edge in corresponding summary graphs, where the only possible double edge is $\circ \leftarrow \leftarrow \circ$.

Another source of distortion, named indirect confounding, may lead to strong dependence reversal compared to the population dependence; see Wermuth and Cox (2008). A first example is due to Robins and Wasserman (1997), shown here in Figure 9, left and the summary graph after marginalizing over U in Figure 9, right

Notice that in this regression graph there are no unmeasured confounders of $Y \leftarrow T_p$ and there is no over-conditioning and no under-conditioning for this dependence of Y on T_p . But A is intermediate between Y and T_p . By conditioning on just the regressors T_r and T_p of Y that remain when U is unobserved, one conditions implicitly also on their past, hence on A and thereby activates the path $Y \text{---} A \leftarrow T_p$ to distort the conditional dependence $Y \pitchfork T_p | T_r, U$ as it is present in the generating graph.

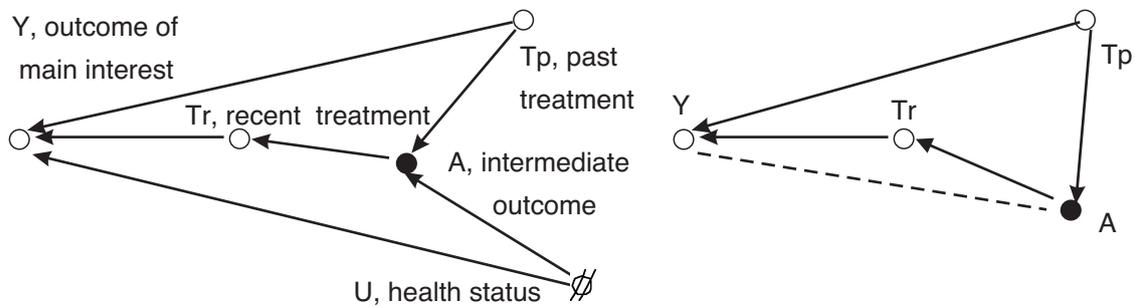


Figure 9: Left: full randomization for T_p leads to removal of $T_p \leftarrow U$, randomized allocation of individuals to selected levels of a more recent treatment, T_r , given results of the intermediate outcome A , leads to removal of $T_r \leftarrow U$ and of $T_r \leftarrow T_p$ and to the presence of $T_r \leftarrow A$; right: the summary graph obtained by marginalizing over U ; path $Y \text{---} A \leftarrow T_p$ explaining the indirect confounding of $Y \leftarrow T_p$ when response Y is regressed on only T_p and T_r .

More generally, indirect confounding will result for $Y \leftarrow T$ in a regression graph, when in the summary graph obtained after marginalizing over all unobserved variables, one or more of the following two types of paths are present; see Wermuth (2011):

$$Y \text{---} \circ \text{---} \circ \dots \circ \text{---} \circ \text{---} T \quad \text{or} \quad Y \text{---} \circ \text{---} \circ \dots \circ \text{---} \circ \leftarrow T,$$

where every inner node along the path is an intermediate variable between Y and T and ... denotes a possible continuation of the same type of neighboring edges.

Indirect confounding, which can be much stronger than direct confounding, appears to have been largely ignored so far in the literature, not only in the statistical one, but also in the current literature on causal modeling, based on virtual interventions using directed acyclic graphs or structural equations; see Pearl (2014 in press), Richardson and Robins (2013). Only direct confounding and selection bias are frequently considered.

The specification of generating processes via sequences of single or joint response regressions lead to regression graphs that may be needed for useful causal interpretations and to corresponding summary graphs that help to avoid possible mistaken interpretations when some variables are hidden, that is latent or unobserved, and some responses are conditioned on.

Bibliography

- Andersen, P.K. & Skovgaard, L.T. (2010). *Regression with Linear Predictors* Springer, New York.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. (3rd ed., 2003), Wiley, New York.
- Bergsma, W. & Rudas, T. (2002). Marginal models for categorical data. *Annals of Statistics* **30**, 140–159.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.
- Cheung, S.Y. & Andersen, R. (2003). Time to read: family resources and educational outcomes in Britain. (2003). *Journal of Comparative Family Studies* **34**, 413–433.
- Cochran, W. G. (1938). The omission or addition of an independent variate in multiple linear regression. *Supplement of the Journal of the Royal Statistical Society* **5**, 171–176.
- Cox, D.R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society, Series C* **21**, 113–120.
- Cox, D.R. (2006). *Principles of Statistical Inference*. Cambridge University Press, Cambridge.
- Cox, D.R. & Wermuth, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika* **77**, 747–761.
- Cox, D.R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science* **8**, 204–218; 247–277.
- Cox, D.R. & Wermuth, N. (1994). Tests of linearity, multivariate normality and adequacy of linear scores. *Journal of the Royal Statistical Society, Series C* **43**, 347–355.
- Cox, D.R. & Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation*. Chapman and Hall (now CRC), London.
- Cowell R.G., Dawid A.P., Lauritzen S.L. & Spiegelhalter D.J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York.
- Darroch, J.N., Lauritzen, S.L. & Speed, T.P. (1980). Markov fields and log-linear models for contingency tables. *Annals of Statistics* **8**, 522–539.
- Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading Mass.

- Drton, M. (2009). Discrete chain graph models. *Bernoulli* **15**, 736–753.
- Drton, M. & Richardson, T.S. (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model *Biometrika* **91**, 383–392.
- Edwards, D. (1995). *Introduction to Graphical Modelling*. (2nd ed. 2000) Springer, New York.
- Edwards, D. & Lauritzen, S.L. (2001). The TM algorithm for maximising a conditional likelihood function. *Biometrika* **88**, 961–972.
- Geiger, D., Verma, T.S. & Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20**, 507–534.
- Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics*. (reprinted by Dover Publications, New York, 1960).
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11**, 1–12.
- Hardt, J., Petrak, F., Filipas, D. & Egle, U.T. (2004). Adaption to life after surgical removal of the bladder – an application of graphical Markov models for analysing longitudinal data. *Statistics in Medicine* **23**, 649–666.
- Hardt, J., Sidor, A., Nickel, R., Kappis, B., Petrak, F., & Egle, U. T. (2008). Childhood adversities and suicide attempts: a retrospective study. *Journal of Family Violence* **23**, 713–718.
- Hardt, J., Herke, M., & Schier, K. (2011). Suicidal ideation, parent-child relationships, and adverse childhood experiences: A cross-validation study using a Graphical Markov Model. *Child Psychiatry & Human Development* **42**, 119–133.
- Jiang, Z., Ding, P. & Geng, Z. (2014). Qualitative evaluation of associations by the transitivity of the association signs. *Statistica Sinica*, in press.
- Koster, J. (2002). Marginalising and conditioning in graphical models. *Bernoulli* **8**, 817–840.
- Laucht M., Esser G., & Schmidt M.H. (1997). Developmental outcome of infants born with biological and psychosocial risks. *Journal of Child Psychology and Psychiatry*. **38**, 843–853.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Lauritzen, S.L., Dawid, A.P., Larsen, B. & Leimer, H.G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491–505.
- Lauritzen, S.L. & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics* **17**, 31–57.
- Loh, P.L. & Wainwright, M.J. (2013). Structure estimation for discrete graphical models: generalized covariance matrices and their inverses *Annals of Statistics* **41**, 3022–3049.
- Lněnička, R. & Matúš, F. (2007). On Gaussian conditional independence structures. *Kybernetika* **43**, 323–342.
- Marchetti, G.M. & Lupporelli, M. (2011). Chain graph models of multivariate regression type for categorical data. *Bernoulli*, **17**, 827–844.
- Marchetti, G.M., Vannini, I., Gottard, A. & Vignoli, D. (2011). Regression graph models: an application to joint modelling of fertility intentions among childless couples. In: *Proceedings 26th International Workshop on Statistical Modelling*. Conesa D., Forte A., Lopez-Quilez A. & Munoz F. (eds). 358–363. .
- Marchetti, G.M. & Wermuth, N. (2009). Matrix representations and independencies in directed acyclic graphs. *Annals of Statistics* **47**, 961–978.
- Markov, A.A. (1912). *Wahrscheinlichkeitsrechnung*. (German translation of 2nd Russian edition: Markoff, A.A., 1908). Teubner, Leipzig.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall (now CRC), London.

- Németh, R. & Rudas, T. (2013). On the application of discrete marginal graphical models. *Sociological Methodology* **43**, 70–100.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* Morgan Kaufmann, San Mateo.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge Univ. Press, Cambridge.
- Pearl, J. (2014) Trygve Haavelmo and the emergence of causal calculus. *Econometric Theory*, in press.
- Richardson, T.S. & Robins, J.M. (2013). Single world intervention graphs: a primer. *Second UAI Workshop on Causal Structural Learning*, Bellevue, Washington.
- Richardson, T.S. & Spirtes, P. (2002). Ancestral Markov graphical models. *Annals of Statistics* **30**, 962–1030.
- Robins, J. & Wasserman, L. (1997). Estimation of effects of sequential treatments by reparametrizing directed acyclic graphs. In: *Proceedings of the 13th Annual Conference on UAI*. Geiger, D. & Shenoy, O. (eds.) Morgan and Kaufmann, San Francisco, 409–420.
- Sadeghi, K. (2013). Stable mixed graphs. *Bernoulli*. **19**, 2330–2358.
- Sadeghi, K. & Lauritzen, S.L. (2014). Markov properties for mixed graphs. *Bernoulli*. **20**, 676–696.
- Sadeghi K. & Marchetti, G.M. (2012). Graphical Markov models with mixed graphs in R. *The R Journal*, **4**, 65–73.
- San Martin E., Mochart M. & Rolin, J.M. (2005). Ignorable common information, null sets and Basu’s first theorem. *Sankhya* **67**, 674–698.
- Schier, K., Herke, M., Nickel, R., Egle, U.T., & Hardt, J. (2014). Long-term sequelae of emotional parentification: A cross-validation study using sequences of regressions. *Journal of Child and Family Studies*, in press.
- Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*. **13**, 238–241.
- Smith, R.B. (2009) Issues matter: a case study of factors influencing voting choices. *Case Studies in Business, Industry and Government Statistics* **2**, 127–146.
- Solis-Trapala, I., Ward K. , Webb, R.E.B, Shoenmakers, I., Prentice, A., Goldberg G.R. (2014). The role of non-mechanical physiological factors in skeletal phenotype using sequences of regressions. Submitted
- Spirtes, P., Glymour C. & Scheines R. (1993). *Causation, Prediction and Search*. Springer, New York.
- Studený, M. (2005). *Probabilistic Conditional Independence Structures*. Springer, London.
- Vandenbroucke J,P. (2002) The history of confounding. *Social and Preventive Medicine* **47**, 216–224.
- Wainwright, M.J. & Jordan, M.I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* **1**, 1–305.
- Weisberg, S. (2014). *Applied Linear Regression*. (4th ed.) Wiley, New York.
- Wermuth, N. (1976). Analogies between multiplicative models for contingency tables and covariance selection. *Biometrics* **32**, 95–108.
- Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *Journal of the American Statistical Association* **75**, 963–97.
- Wermuth, N. (1992). On block-recursive regression equations (with discussion). *Brazilian Journal of Probability and Statistics* **6**, 1–56.
- Wermuth, N. (2011). Probability models with summary graph structure. *Bernoulli*, **17**, 845–879.

- Wermuth, N. (2012). Traceable regressions. *International Statistical Review*. **80**, 415–438.
- Wermuth, N. (2013). Comment: are marginal models needed? *Sociological Methodology*. **43**, 114–119.
- Wermuth, N. & Cox, D.R. (1998). Statistical dependence and independence. *Encyclopedia of Biostatistics*. P. Armitage & T. Colton (eds), Wiley, New York, 4260–4264.
- Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *Journal of the Royal Statistical Society, Series B* **66**, 687–717.
- Wermuth, N. & Cox, D.R. (2008). Distortions of effects caused by indirect confounding. *Biometrika* **95**, 17–33.
- Wermuth, N. & Cox, D.R. (2013). Concepts and a case study for a flexible class of graphical Markov models. In: *Robustness and Complex Data Structures. Festschrift in honour of Ursula Gather*. Becker, C., Fried, R. & Kuhnt, S. (eds.), Springer, Heidelberg, 331–350; also on ArXiv: 1303.1436.
- Wermuth, N., Cox, D.R. & Marchetti, G.M. (2009). Triangular systems for symmetric binary variables. *Electronic J. of Statist.* **3**, 932–955.
- Wermuth, N. & Lauritzen, S.L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70**, 537–552.
- Wermuth, N. & Marchetti, G.M (2014). Star graphs induce tetrad correlations: for Gaussian as well as for binary variables. *Electronic Journal of Statistics* **8**, 253–273.
- Wermuth, N., Marchetti, G.M & Zwiernik, P. (2014). Binary distributions of concentric rings. *J. Multiv. Analysis* **130**, 252–260; also on ArXiv: 1311.5655
- Wermuth N. & Sadeghi, K. (2012). Sequences of regressions and their independences (invited paper with discussion). *TEST* **21**, 215–279.
- Wermuth, N., Wiedenbeck, M. & Cox, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Mathematics* **46**, 883–901.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.
- Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics* **8**, 239–255.
- Wright, S. (1934). The method of path coefficients. *Annals of Statistics* **5**, 161–215.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* **57**, 348–368.