Statistical Science
1993, Vol. 8, No. 3, 204-283

Linear Dependencies Represented by Chain Graphs

D. R. Cox and Nanny Wermuth

Abstract. Various special linear structures connected with covariance matrices are reviewed and graphical methods for their representation introduced, involving in particular two different kinds of edge between the nodes representing the component variables. The distinction between decomposable and nondecomposable structures is emphasized. Empirical examples are described for the main possibilities with four component variables.

Key words and phrases: Chain model, conditional independence, covariance selection, decomposable model, linear structural equation, multivariate analysis, path analysis.

1. INTRODUCTION

This paper has three broad objectives. The first is to illustrate the rich variety of special forms of association and dependence that can arise even with as few as three or four variables. The second is to show the value of graphical representation in clarifying these dependencies; for this we introduce graphs with two different kinds of edge and some further features which are also new. The third objective is to show the importance in interpretation of the distinction between decomposable and nondecomposable models.

A series of examples will be used in illustration, partly to show that many of the special structures do indeed arise in applications and partly to show in outline the implications for interpretation, although reference to the subject matter literature is necessary for a full account. Most of the examples arise from recent investigations at University of Mainz. For purposes of exposition we have chosen examples with at most four variables; that is, we have simplified by omitting mention of variables which analysis had shown to have no bearing on the points at issue.

We confine the discussion to those problems with essentially linear structure in which the interrelationships are adequately captured by the covariance matrix of the variables. Of course in applications, checks for nonlinearities and outliers are required, and these have been done for all examples whenever we had access to the raw data.

D. R. Cox is Warden, Nuffield College, Oxford OX1 1NF, United Kingdom. Nanny Wermuth is Professor, Psychologisches Institut, Johannes Gutenberg-Universität Mainz, Postfach 3980, D-55099 Mainz, Germany.

The need to discuss special structures arises partly because the relations of marginal independence and conditional independence expressed thereby are often of substantive interest and partly because in a *saturated model* with p component variables, that is, one in which the covariance matrix is unrestricted other than to being positive definite, there are $1/2p(p-1)$ correlations, and reduction of dimensionality may be desirable to avoid a superabundance of parameters.

There are strong connections with, in particular, the long history of work in path analysis in genetics, in simultaneous equations in econometrics and linear structural models in psychometrics and with the body of recent work applying graph-theoretic ideas to the study of systems of conditional independencies arising especially in the study of expert systems.

In Section 2 we review some general properties of linear regression systems as related to the covariance matrix of the variables and stress the distinction between multivariate regression and block regression and between decomposable and nondecomposable structures. In Section 3 we introduce the main conventions useful in a graph-theoretic representation of the independency relations that may hold; in Section 4 we discuss relations with previous work, and in Section 5 we give a series of empirical examples for four variables. The paper concludes in Section 6 with some general discussion. The emphasis throughout is on the structure and interpretation of the various models rather than on the procedures for fitting.

2. SOME PROPERTIES OF COVARIANCE MATRICES

It is convenient to set out some properties of systems of linear least squares regressions derivable from a covariance matrix. These are full regression equations in a multivariate normal distribution. There is throughout the usual interplay between relatively weak second-order properties of least squares regression and the strong properties derivable from an assumption of multivariate normality, such as that zero correlation or zero partial correlation implies independence or conditional independence.

We consider first the $p \times 1$ vector $Y = (Y_1, \ldots, Y_p)^T$ with mean $E(Y) = \mu$. We denote the positive definite covariance matrix by $\mathrm{cov}(Y) = \Sigma$, and its inverse, the concentration matrix, therefore by Σ^{-1}; the diagonal elements of Σ are the variances (σ_{ii}), those of Σ^{-1} are the precisions (σ^{ii}). The off-diagonal elements of Σ are the covariances (σ_{ij}), those of Σ^{-1} are the concentrations (σ^{ij}). A marginal correlation ρ_{ij} is expressible via elements of the covariance matrix, in a way similar to that in which a partial correlation, $\rho_{ij.k}$, given all of the remaining variables $k = \{1, \ldots, p\}\backslash\{i, j\}$, is expressible via elements of the concentration matrix:

$$\rho_{ij} = \sigma_{ij}\left(\sigma_{ii}\sigma_{jj}\right)^{-1/2}, \rho_{ij.k} = -\sigma^{ij}\left(\sigma^{ii}\sigma^{jj}\right)^{-1/2}.$$

This implies in particular that in the usual notation (Dawid, 1979a) for independence,

$$Y_i \perp\!\!\!\perp Y_j, \quad \text{if and only if } \sigma_{ij} = 0,$$
$$Y_i \perp\!\!\!\perp Y_j \mid Y_k, \quad \text{if and only if } \sigma^{ij} = 0,$$

where as above $k = \{1, \ldots, p\}\backslash\{i, j\}$.

To study regression models, we partition Y into Y_a and Y_b, $p_a \times 1$ and $p_b \times 1$, respectively, $p_a + p_b = p$, and call the two parts response and explanatory variables. Let the covariance matrix and the concentration matrix be conformally partitioned:

$$(1) \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \cdot & \Sigma_{bb} \end{pmatrix}, \Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} \\ \cdot & \Sigma^{bb} \end{pmatrix},$$

then the covariance matrix Σ_{bb} of the explanatory variables and correspondingly their concentration matrix $\Sigma_{bb}^{-1} = \Sigma^{bb.a} = \Sigma^{bb} - (\Sigma^{ab})^T(\Sigma^{aa})^{-1}\Sigma^{ab}$ do not contain parameters needed to specify a standard regression model of Y_a on Y_b. Instead, their observed counterparts are taken as fixed or indeed sometimes are fixed by sampling design.

We now distinguish between a multivariate regression and a block regression. To simplify the notation we shall without essential loss of generality take often $E(Y) = 0$. We describe the distinct parameters in the two types of regression models, that is, the two ways of parametrizing the conditional distribution of Y_a given Y_b. For a *multivariate regression* of Y_a on Y_b, that is, for $Y_a = \Pi_{a|b}Y_b + \varepsilon_a$ with $E(\varepsilon_a) = 0, E(\varepsilon_a Y_b^T) = 0$, the regression equation parameters $\Pi_{a|b}$ and the residual variance $\mathrm{var}(\varepsilon_a)$ can be written in a matrix as $(\Sigma_{aa.b}, \Pi_{a|b})$, where

$$(2) \quad \begin{aligned} \Pi_{a|b} &= \Sigma_{ab}\Sigma_{bb}^{-1}, \\ \mathrm{var}(\varepsilon_a) &= \Sigma_{aa.b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^T. \end{aligned}$$

In a saturated multivariate regression (2) each component of Y_a is regressed separately on the full set of components Y_b.

On the other hand in a saturated *block regression* each component of Y_a is regressed not only on Y_b but also on all remaining components of Y_a. Then the regression equation parameters are instead proportional to the elements of the matrix $(\Sigma^{aa}, \Sigma^{ab})$ (Wermuth, 1992). The reason is that the expected value of a component Y_i of Y_a given all remaining variables of Y can be obtained by taking expectations in

$$(3) \quad \Sigma^{aa}Y_a + \Sigma^{ab}Y_b = \omega_a$$

where $E(\omega_a) = 0$, $\mathrm{var}(\omega_a) = \Sigma^{aa}$ and dividing the ith equation by the concentration σ^{ii}. Equation (3) is derived from a block triangular decomposition of the concentration matrix, $\Sigma^{-1} = A^T T^{-1} A$, where

$$(4) \quad \begin{aligned} A &= \begin{pmatrix} I_{aa} & (\Sigma^{aa})^{-1}\Sigma^{ab} \\ 0 & I_{bb} \end{pmatrix}, \\ T^{-1} &= \begin{pmatrix} \Sigma^{aa} & 0 \\ 0 & \Sigma^{bb.a} \end{pmatrix}, \end{aligned}$$

as the first p_a equations of $(T^{-1}A)(Y - E(Y)) = \omega$. The residuals ω have zero mean and covariance matrix T^{-1}.

For a block regression, the resulting coefficient of variable Y_j in the ith equation is minus a partial regression coefficient given all remaining variables of Y, that is, given all remaining response and explanatory variables. On the other hand, in a multivariate regression the coefficient of Y_j in the ith equation is a partial regression coefficient given all remaining variables of Y_b, that is, given all remaining explanatory variables. To express this distinction more formally, we write a partial regression coefficient $\beta_{ij.d}$ for $\{1, \ldots, p\} = a \cup b = \{\{i, j\}, d, g\}$ in terms of elements of the conditional covariance matrix of (Y_i, Y_j) given Y_d and of elements of the concentration matrix of (Y_i, Y_j), having marginalized over Y_g, as

$$\beta_{ij.d} = \frac{\sigma_{ij.d}}{\sigma_{jj.d}} = -\frac{\sigma^{ij.g}}{\sigma^{ii.g}}.$$

Note that in the case of a block regression g is empty and d is the set of all remaining variables of Y, that is, $d = (a \cup b)\backslash\{i, j\}$, while in the case of a multivariate regression $d = b\backslash\{j\}$, and $g = a\backslash\{i\}$. Note further that

$$(5) \quad Y_i \perp\!\!\!\perp Y_j \mid Y_d, \quad \text{if and only if } \beta_{ij.d} = 0.$$

To judge the relative strength of the dependence of a response on several explanatory variables, it is sometimes useful to compare the standardized regression coefficients, that is, $\beta_{ij.d}^* = \beta_{ij.d}\sigma_{jj}^{\frac{1}{2}}\sigma_{ii}^{-\frac{1}{2}}$.

One of the major distinctions between multivariate regression and block regression lies in the meaning of the relation between two components Y_i and Y_j, both within Y_a, and in the meaning of the relation of a component Y_i from Y_a to a component Y_j from Y_b. To describe this in detail it is useful to recall how a partial regression coefficient relates to a partial correlation coefficient

$$\beta_{ij.d} = \rho_{ij.d}\sqrt{\frac{\sigma_{ii.d}}{\sigma_{jj.d}}} = \rho_{ij.d}\sqrt{\frac{\sigma^{jj.g}}{\sigma^{ii.g}}}.$$

Thus, in a block regression, that is, where $d = (a \cup b)\backslash\{i, j\}$, the relation between Y_i from Y_a and Y_j is measured essentially by the partial correlation given all remaining variables of Y, no matter whether Y_j is from Y_a or it is from Y_b. By contrast in a multivariate regression, that is, where $d = b\backslash\{j\}$, the measure of the relation of Y_i from Y_a to Y_j from Y_b is proportional to the partial correlation given the variables in Y_b other than Y_j; the correlation between Y_i and Y_j both within Y_a is given all variables in Y_b. Thus, a larger set of variables is considered simultaneously in block regression if compared with the corresponding multivariate regression. Written in matrix notation their parameters are related by

$$\text{(6)} \qquad \Pi_{a|b} = -(\Sigma^{aa})^{-1}\Sigma^{ab}, \Sigma_{aa.b} = (\Sigma^{aa})^{-1}$$

$$\text{(7)} \qquad \Sigma^{ab} = -(\Sigma_{aa.b})^{-1}\Pi_{a|b}, \Sigma^{aa} = (\Sigma_{aa.b})^{-1}.$$

Some of the special models we shall consider correspond to specifying some elements of regression equations to be zero, that is, to structures that appear simplified if compared with the saturated model. The choice between block regression and multivariate regression is then largely determined by the research questions and by a decision as to which of the two parametrizations permits a simpler description of the relations. For instance, in each of Examples 1, 2 and 7 of the empirical examples of Section 5 we can think of two variables as joint responses, $Y_a = (Y, X)^T$, and of two variables as explanatory, $Y_b = (V, W)^T$. A simplifying description is possible with block regression but not with multivariate regression in Example 1, while a simpler structure results with multivariate regression than with block regression in Examples 2 and 7.

If not only the conditional distribution of Y_a given Y_b is of interest, but the marginal relations among component variables within Y_b as well, we are led to a simple type of regression chain model: we specify the joint density via

$$f_{ab} = f_{a|b}f_b,$$

and make a choice for $f_{a|b}$ among a multivariate and a block regression.

A specification of the joint distribution of Y_a, Y_b by a saturated *multivariate regression chain model* has $(\Sigma_{aa.b}, \Pi_{a|b})$ as parameters for the conditional distribution of Y_a given Y_b and Σ_{bb} for the marginal distribution of Y_b. With a saturated *block regression chain model* the parameters are the regression coefficients obtained as described above from $(\Sigma^{aa}, \Sigma^{ab})$ and the concentration matrix $\Sigma^{bb.a} = \Sigma_{bb}^{-1}$.

Considering, for instance, a multivariate regression chain model instead of a multivariate regression model can lead to a simpler structure. This is the case in Example 7 but not in Example 2 of Section 5 since the explanatory variables can be taken to be marginally uncorrelated in the former but not in the latter.

In the next more complex regression chain model the joint density of three (vector) variables Y_a, Y_b and Y_c is specified via

$$f_{abc} = f_{a|bc}f_{b|c}f_c,$$

that is, via a regression of Y_a on Y_b and Y_c, a regression of Y_b on Y_c and the marginal distribution of Y_c. This would be an adequate approach if the components of Y_a are the response variables of primary interest having Y_b and Y_c as potential explanatory variables, if Y_b plays the role of an intermediate variable containing potentially explanatory components for Y_a and possible responses to Y_c and, finally, if Y_c consists of explanatory variables whose joint distribution is to be analyzed.

A particularly important family of regression chains are the *univariate recursive regressions* in which, for a given ordering of the components of $Y = (Y_1, \ldots, Y_q)^T$, we define the model via the regression of Y_r on $Y_{r+1}, \ldots, Y_p$ for $r = 1, \ldots, q$; $q \le p-1$. An independence hypothesis is said to be *decomposable* if it specifies one or more of the regression coefficients in such a system to be zero. Early descriptions of univariate recursive regressions have been given by Wright (1921, 1923) with an emphasis on applications in genetics and by Tinbergen (1937) for the study of business cycles.

By contrast a *nondecomposable independence hypothesis* consists of a set of k independence relations for k distinct variable pairs that cannot, in its entirety, be reexpressed in terms of vanishing coefficients in the above form: that is, no ordering of the variables would produce a decomposable independence hypothesis with the same implications from the same distributional assumption. The following arguments apply provided that there are no so-called forbidden states, that is, states of zero probability (Dawid, 1979a).

For instance, for a trivariate normal distribution of Y, Z, X the hypothesis $Y \perp\!\!\!\perp X \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ corresponds to zero concentrations for pairs (Y, X) and (X, Z) and it implies $X \perp\!\!\!\perp (Y, Z)$. This hypothesis can be reexpressed by $Y \perp\!\!\!\perp X \mid Z$ and $X \perp\!\!\!\perp Z$ corresponding to $\beta_{yx.z} = \beta_{xz} = 0$ in a univariate recursive system for $(Y, X, Z)^T$. Thus the hypothesis is decomposable even though initially not expressed in that form. On the other hand, no ordering of the variables would permit us to specify the hypothesis $Y \perp\!\!\!\perp X$ and $Z \perp\!\!\!\perp U$ as zero restrictions in a univariate recursive regression system. Thus the hypothesis is nondecomposable. Further examples for nondecomposable hypotheses are discussed in Section 5.

They arise in applications with four or more variables, as we shall see below, but suffer from a number of disadvantages both in terms of the difficulty of fitting, but more importantly, in terms of indirectness of interpretation. The need for such models was noted by Haavelmo (1943) who pointed out substantive research questions about relations which form a system of equations to be fulfilled simultaneously, but which are not a system of univariate recursive regressions. His subject matter example is as follows: consumption in an economy per year depends on total income, investment per year depends on consumption and total income is the sum of consumption and investment. A slightly simplified version of Haavelmo's argument for the simultaneous treatment of equations is given in Section 4. As a consequence of his results, the class of linear structural equations was developed to study simultaneous relations. It is mainly discussed in econometrics (Goldberger, 1964), in psychometrics (Jöreskog, 1973) and in sociology (Duncan, 1969); it includes univariate recursive regression systems and multivariate regressions as a subclass but, in general, a zero coefficient in a structural equation does not correspond to an independence relation. More generally the graphical representations to be introduced in Section 3 are equivalent to those used in path analysis and in discussions of structural equations only in rather special cases. We deal with this important point further in Section 4.

A representation in terms of univariate recursive regressions combines several advantages. First, and most importantly, it describes a stepwise process by which the observations could have been generated and in this sense may prove the basis for developing potential causal explanations. Second, each parameter in the system has a well-understood meaning since it is a regression coefficient: that is, it gives for unstandardized variables the amount by which the response is expected to change if the explanatory variable is increased by one unit and all other variables in the equation are kept constant. As a consequence, it is also known how to interpret each additional zero restriction: in the case of jointly normal variables, each added restriction introduces a further conditional independence, and it is known how parameters are modified if variables are left out of a system (Wermuth, 1989). Third, general results are available for interpreting structures, that is, for reading all implied independencies directly off a corresponding graph (Pearl, 1988; Lauritzen et al., 1990) and for deciding from the graphs of two distinct models whether they are equivalent (Frydenberg, 1990a). Fourth, an algorithm exists (Pearl and Verma, 1991; Verma and Pearl, 1992) which decides for arbitrary probability distributions and an almost arbitrary list of conditional independence statements whether the list defines a univariate recursive system; if it does, a corresponding directed acyclic graph is drawn. Fifth, the analysis of the whole association structure can be achieved with the help of a sequence of separate univariate linear regression analyses (Wold, 1954).

The word *causal* is used in a number of different senses in the literature; for a review see Cox (1992). Glymour et al. (1987) and Pearl (1988) have developed valuable procedures for finding relatively simple structures of conditional independencies which they define to be causal. We prefer to restrict the word to situations where there is some understanding of an underlying process. From this perspective it is unrealistic to think that causality could be established from a single empirical study or even from a number of studies of similar form. We aim, however, by introducing appropriate subject matter considerations into the empirical analysis, to produce descriptions and summaries of the data which point toward possible explanations and which in some cases of univariate recursive systems could be consistent with a causal explanation.

3. SOME GRAPHICAL REPRESENTATIONS

With only three component variables, the number of possible special independency models is fairly small but with four and more components there is a quite rich and potentially confusing variety of special cases to be considered. Graphical representation helps clarify the various possibilities, and it is convenient to introduce the key ideas and conventions in terms of three variables.

A systematic account of graphical methods by Whittaker (1990) emphasizes undirected graphs, that is, systems in which all variables are treated on an equal footing. Here we use largely directed graphs to emphasize relations of response and dependence; it is fruitful also to allow two different kinds of edge between the nodes of a graph and to introduce some additional special features.

First we introduce, where appropriate, a distinction between the response variables of primary interest, one or more levels of intermediate response variables, and explanatory variables, all in general with several component variables. The distinction between variable types is usually introduced on a priori subject matter considerations, for example via the temporal ordering of the variables. Sometimes, however, there are several such provisional interpretations and some may be suggested by the data under analysis. The distinction between variable types is expressed in the graphs via (c) below.

The following conventions have been used in constructing the graphs in this paper and are illustrated in their simplest form in Figures 1–3 for three variables:

(a) each continuous variable is denoted by a node, a circle;

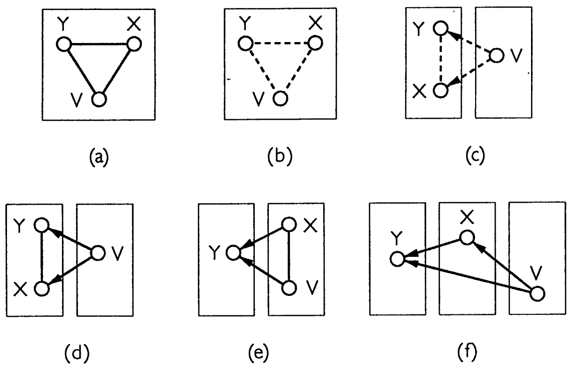

FIG 1. *Six distributionally equivalent ways of specifying a saturated model for three variables. (a) Joint distribution of Y, X, V with three substantial concentrations: (b) joint distribution of Y, X, V with three substantial covariances; (c) multivariate regression chain model with regressions of Y on V and of X on V and with correlated errors; (d) block regression chain model with regressions of Y on X, V and of X on Y, V; (e) univariate regression of Y on X, V and joint distribution of X, V; (f) univariate recursive regression system with Y as response to X, V; X as intermediate response to V. For instance, graph (e) with double lines round the right-hand box would represent the standard linear model for regression of Y on fixed explanatory variables X, V.*

(b) there is at most one connecting line between each pair of nodes, an edge;

(c) variables are graphed in boxes so that variables in one box are considered conditionally on all boxes to the right (in line with the notation $P(A \mid B)$ for the probability of A given B) so that the response variables of primary interest are in the left-hand box and its explanatory variables are in boxes to the right;

(d) if full lines are used as edges, each variable is considered conditionally on other variables in the same box (as well as those to the right), whereas if dashed lines are used variables are considered ignoring other response variables in the same box, that is, marginally with respect to response variables in the same box;

(e) the absence of an edge means that the corresponding variable pair is conditionally independent, the conditioning set being as specified in (d);

(f) variables in the same box are to be regarded

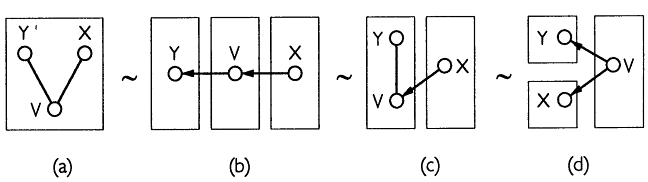

FIG. 2. *Four distributionally equivalent ways of specifying $Y \perp\!\!\!\perp X \mid V$; (a) covariance selection model for Y, X, V having parameters $\rho_{yv.x} \neq 0$, $\rho_{zv.y} \neq 0$, and $\rho_{yx.v} = 0$; (b) univariate recursive regression model with $\beta_{yv.x} \neq 0$, $\beta_{yx.v} = 0$, $\beta_{vx} \neq 0$; (c) block regression chain model with Y, V as joint responses to X and with independent parameters $\rho_{yv.x} \neq 0$, $\beta_{yx.v} = 0$, $\beta_{vx.y} \neq 0$; (d) two independent regressions of Y on V and of X on V with $\beta_{yv} \neq 0$, $\beta_{xv} \neq 0$, $\rho_{yx.v} = 0$.*

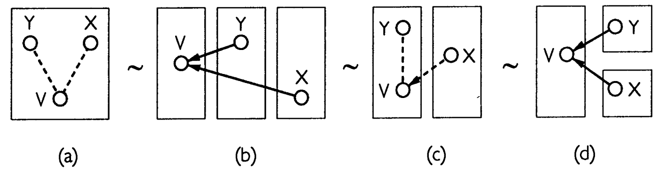

FIG. 3. *Four distributionally equivalent ways of specifying $Y \perp\!\!\!\perp X$; (a) linear structure in covariances with $\rho_{yv} \neq 0$, $\rho_{xv} \neq 0$, $\rho_{yx} = 0$; (b) univariate recursive regression model with $\beta_{vx.y} \neq 0$, $\beta_{vy.x} \neq 0$, $\beta_{yx} = 0$; (c) multivariate regression chain model with $\rho_{yv.x} \neq 0$, $\beta_{vx} \neq 0$, $\beta_{yx} = 0$; (d) multiple regression of V on two independent regressors Y, X, with $\beta_{vy.x} \neq 0$, $\beta_{vx.y} \neq 0$, $\rho_{yx} = 0$.*

in a symmetrical way, for instance as both response variables, and connected by undirected edges (lines without arrowheads, for correlations), whereas relations between variables in different boxes are shown by directed edges (arrows, for regression coefficients) such that an arrow points from the explanatory variable to the response;

(g) graphs drawn with boxes represent substantive research hypotheses (Wermuth and Lauritzen, 1990) in which the presence of an edge means that the corresponding partial correlation is large enough to be of substantive importance. This corresponds to the notion that the model being represented is the simplest appropriate one in the sense that relations considered to be unimportant are not part of the model; graphs obtained by removing the boxes represent statistical models in which a connecting edge places no such constraint on the correlation, that is, it could also be a zero correlation;

(h) a row of unstacked boxes implies an ordered sequence of (joint) responses and (joint) intermediate responses, each together with their explanatory variables. Boxes are stacked if no order is to be implied, in order to indicate independence of several (joint) variables conditionally on all boxes to the right;

(i) if the right-hand box has two lines around it, then the relations among variables in this box are regarded as fixed at their observed levels; this is to indicate a regression model instead of a regression chain model, the latter containing parameters also for those components which are exclusively explanatory.

In the present paper we use only graphs with edges of one type, that is, either all full lines or all dashed lines. It would be possible to have mixture of the two types of edge in the same graph, for example provided that all the edges within one block are of the same type and all the edges directed at a particular block are of the same type.

In a sense the distinction between full and dashed edges serves a double purpose. The distinction between full and dashed arrows from one box to another determines the different conditioning sets used in the various regression equations under consideration. The distinction between full and dashed lines within a box specifies whether it is the concentration or the covariance matrix of the residuals that is the focus of interest. In this sense the nature of the edges corresponds to the parameters of interest.

The joint distribution of all variables is in the present context specified by the vector of means and the covariance or the concentration matrix. However any such given matrix may correspond to a number of models with quite different interpretations in the light of the distinction between types of variable as response, intermediate response or explanatory variable. A complete graph, that is, one in which all edges are present, represents a saturated model, that is, in the present context a model without any specified independence relations.

To stress the distinction between the multivariate regression and block regression contained in Figure 1, we write the corresponding equations explicitly. The multivariate regression equations implied by Figure 1c are

$$E(Y \mid V = v) - \mu_y = \beta_{yv}(v - \mu_v),$$
$$E(X \mid V = v) - \mu_x = \beta_{xv}(v - \mu_v),$$

with

$$\operatorname{cov}(\varepsilon_{y.v}, \varepsilon_{x.v}) = \rho_{yx.v}(\sigma_{yy.v}\sigma_{xx.v})^{1/2}.$$

By contrast the block regression equations implied by Figure 1d are

$$E(Y \mid X = x, V = v) - \mu_y$$
$$= \beta_{yx.v}(x - \mu_x) + \beta_{yv.x}(v - \mu_v),$$
$$E(X \mid Y = y, V = v) - \mu_x$$
$$= \beta_{xy.v}(y - \mu_y) + \beta_{xv.y}(v - \mu_v),$$

with

$$\beta_{yx.v} = \rho_{yx.v}(\sigma_{yy.v} / \sigma_{xx.v})^{1/2}, \beta_{xy.v} = \rho_{yx.v}(\sigma_{xx.v} / \sigma_{yy.v})^{1/2},$$
$$\operatorname{cov}(\varepsilon_{y.xv}, \varepsilon_{x.yv}) = -\rho_{yx.v}(\sigma_{yy.xv}\sigma_{xx.yv})^{1/2},$$

where the conditional variance of the variable given all remaining variables is the reciprocal value of a precision, for example, $\sigma_{yy.xv} = 1 / \sigma^{yy}$. Relations between the sets of parameters in the two types of regressions are given by Equations (6) and (7).

4. RELATIONS WITH PREVIOUS WORK

We illustrate the distinction between the graphical chain models of the present paper and structural equation models via two examples. Suppose first that X and Y are standardized to mean zero and variance one and denote their correlation coefficient by ρ. Then

$$Y = \rho X + \varepsilon_y, \ X = \rho Y + \varepsilon_x,$$

where $(\varepsilon_y, \varepsilon_x)$ are residuals from linear regression equations. That is, the coefficients ρ in these equations have an interpretation as regression coefficients. Direct calculation shows that

$$\operatorname{cov}(\varepsilon_y, \varepsilon_x) = \operatorname{cov}(Y - \rho X, X - \rho Y) = -\rho(1 - \rho^2),$$

which is nonzero unless $\rho = 0$. That is, the two regression equations imply correlated residuals except for degenerate cases.

On the other hand, if we were to adopt

$$Y - \rho X = \varepsilon_y, \ X - \rho Y = \varepsilon_x$$

as structural equations with uncorrelated residuals, then another direct calculation shows that the regression of Y on X is

$$E(Y \mid X = x) = \frac{E(YX)}{E(X^2)}x = \frac{\operatorname{var}(\varepsilon_y) + \operatorname{var}(\varepsilon_x)}{\operatorname{var}(\varepsilon_y) + \rho^2 \operatorname{var}(\varepsilon_x)}\rho x$$

which is not ρx, again unless $\rho = 0$. That is, the coefficients in these structural equations do not have an interpretation as regression coefficients, as was noted by Haavelmo (1943).

To make the related point that missing edges in the graphical representation of linear structural equations (Van de Geer, 1971) do not in general have the independency interpretation of chain graphs, consider the following two structural equations

$$Y + \gamma_{yx}X + \gamma_{yv}V = \varepsilon_y,$$
$$\gamma_{xy}Y + X + \gamma_{xw}W = \varepsilon_x,$$

illustrated in Figure 4. For correlated errors $(\varepsilon_y, \varepsilon_x)$, a count of parameters shows that this represents a saturated model; that is, it allows an arbitrary covariance matrix for $(Y, X, V, W)^T$. That is, in particular, the missing edges between V and X, and between W and Y do not imply independencies, conditional or unconditional. For some further discussion of possibilities for interpreting the parameters in this model see Wermuth (1992) and Goldberger (1992). For linear structural equations in general, the interpretation of equation pa-

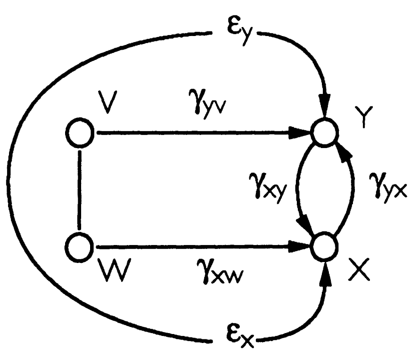

FIG. 4. *Graphical representation of two structural equations in which the missing edges for (V, X) and (W, Y) do not correspond to independencies and do not restrict the covariance matrix for (Y, X, W, V)^T.*

rameters, be they present or missing, has to be derived from scratch for each model considered.

However, an interpretation in terms of independencies is available also for structural equations, whenever such a model is *distributionally equivalent* to one of the chain graph models, that is, if the same joint distribution holds for the two types of models, possibly specified in two distinct ways, *and* the parameter vectors of the two models are in one-to-one correspondence.

Three classes or families of models can be identified to have this property. These are models that have a representation by a chain graph which is:

[1] a *covariance graph*, that is, a single box graph in which all present edges are undirected dashed lines, as in Figures 1b and 3a;

[2] a *multivariate regression graph*, that is, a two-box graph in which all present edges are dashed, being lines within and arrows between boxes, as in Figures 1c and 3c and in which the right-hand box has two lines around it, the distribution of its components being fixed.

[3] a *univariate recursive regression graph*, that is, a graph of $q + 1$ boxes, q of them with a single response variable and the right-hand box with $p - q$ additional explanatory variables, as in Figures 1f, 2b and 3b. In addition the right-hand box has two lines around it to indicate that only the conditional distribution of $Y_1, \ldots, Y_q$ given the remaining variables is the model of interest.

The conventions (a) to (i) for constructing chain graphs imply for univariate recursive regression graphs that arrows have the same interpretation no matter whether they are all dashed or whether they are all full arrows. That is whenever there are no proper joint responses in a model then dashed and full edge arrows are interpreted in the same way.

To distinguish better between dashed and full-edge graphs when their interpretation differs we suggest speaking further of:

[4] a *concentration graph*, that is, a single box graph in which all edges are undirected full lines, as in Figures 1a and 2a;

[5] a *block regression graph*, that is, a two-box graph in which all present edges are full, being lines within and arrows between boxes, as in Figures 1d and 2c and in which the right-hand box has two lines around it.

Then, a *multivariate regression chain graph* can be viewed as a combination of a (sequence of) graph(s) [2] with [1] and a *block regression chain graph* as a combination of a (sequence of) graph(s) [5] with [4]. More general chain graphs with both types of edges result as further combinations of these four building blocks.

Univariate recursive regression graphs are essentially identical to the *directed acyclic graphs* used in work on expert systems (Pearl, 1988). One of the latter results from one of the former by replacing the complete undirected graph of the explanatory variables by an acyclic orientation, that is, by a univariate recursive regression graph in arbitrary order of the nodes and by discarding all boxes.

To investigate distributional equivalence it is helpful to use the notion of a skeleton graph introduced by Verma and Pearl (1992). A *skeleton graph* is obtained from our Figures by removing boxes and arrows and ignoring the type of edge. For instance, the skeleton graphs in Figures 2a to 2d are all the same. If the skeletons differ then the corresponding models cannot be equivalent. But if the skeletons are the same, then the graphs may still imply different independencies, as in Figures 2 and 3.

Distributional equivalence to a model of univariate recursive regressions is closely tied to our notion of a nondecomposable independence hypothesis. We speak of a *decomposable model* if it is distributionally equivalent to a model of univariate recursive regressions and of a nondecomposable model otherwise. Thus, all saturated chain models for linear relations considered in this paper are decomposable, since they all specify the same joint distribution (Figure 1). A nonsaturated model is decomposable if and only if it contains *not* even one nondecomposable independence hypothesis. In complex cases, such a model may contain large sections that are decomposable and in analysis and interpretation account can be taken of that.

This notion of a decomposable model coincides with the notion of a decomposable graph when this graph has undirected full edges, that is, when it is a concentration graph. For variables with a joint normal distribution a concentration graph specifies a covariance selection model (Dempster, 1972). Such a model is decomposable if and only if the concentration graph is triangulated, that is, if it does not contain a chordless n-cycle for $n \geq 4$ (Wermuth, 1980; Speed and Kiiveri, 1986). A sequence of nodes $(a_1, \ldots, a_n)$ is said to form a *chordless n-cycle* in a chain graph if only consecutive nodes and the endpoints of the sequence are connected by edges and a chordless cycle in a sequence of four or more variables characterizes a nondecomposable independence hypothesis in concentrations. An example is Form (i) for (Y, X, V, W) discussed in Section 5. A special well-studied example of a decomposable covariance selection model is represented by a *chordless n-chain* in concentrations, that is, sequence of nodes $(a_1, \ldots, a_n)$ for which only consecutive nodes of the sequence are connected by edges. This is a Markov

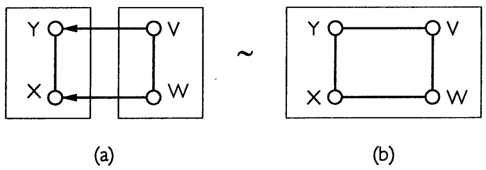

FIG. 5. *Block regression chain model (a) and covariance selection model (b) both specifying the nondecomposable hypothesis* (i): $Y \perp\!\!\!\perp W \mid (X, V)$ *and* $X \perp\!\!\!\perp V \mid (Y, W)$.

chain model. An example is Form (vi) for (Y, X, V, W) discussed in Section 5.

Figures 1–3 show that not only full-edge but also dashed-edge chain graph models can be decomposable, that is, distributionally equivalent to a model of univariate recursive regressions. We characterize situations in which this is not possible for four variables in the next section.

5. SOME EMPIRICAL EXAMPLES

We now introduce eight special kinds of independence hypothesis for four variables, together with their associated graphs, and illustrate most of them via empirical examples. All involve two or more independency conditions. The special structures we shall consider are as follows, the first three and the last two being nondecomposable:

(i) $Y \perp\!\!\!\perp W \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (Y, W)$,

(see Figures 5a and 5b) called the chordless four-cycle in concentrations and which correspond to the vanishing of two elements in the concentration matrix, and hence to a special case of the covariance selection models (Dempster, 1972). It can also be viewed as a chordless four-cycle in a block regression chain model with joint responses Y, X and joint explanatory variables V, W. Next we consider

(ii) $Y \perp\!\!\!\perp W \mid V$ and $X \perp\!\!\!\perp V \mid W$,

called a chordless four-cycle in a multivariate regression chain model (see Figure 6a) and which contains regressions of Y and X on V and W, being a special case of the seemingly unrelated regressions of Zellner (1962);

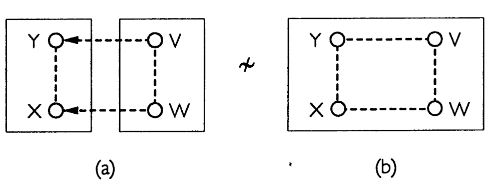

FIG. 6. *Multivariate regression chain model (a) specifying the nondecomposable hypothesis* (ii): $Y \perp\!\!\!\perp W \mid V$ *and* $X \perp\!\!\!\perp V \mid W$ *and a linear in covariances structure (b) specifying the nondecomposable hypothesis* (iii): $Y \perp\!\!\!\perp W$ *and* $X \perp\!\!\!\perp V$.

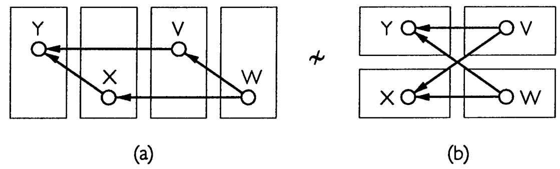

FIG. 7. *Univariate recursive regressions (a) specifying* (iv): $Y \perp\!\!\!\perp W \mid (X, V)$ *and* $X \perp\!\!\!\perp V \mid W$ *and independent multiple regressions with independent explanatory variables (b) specifying* (v): $Y \perp\!\!\!\perp X \mid (V, W)$ *and* $V \perp\!\!\!\perp W$.

(iii) $Y \perp\!\!\!\perp W$ and $X \perp\!\!\!\perp V$,

called the chordless four-cycle in correlations (see Figure 6b), a special case of covariance matrices with linear structure (Anderson, 1973).

These may be contrasted with a decomposable model based on a recursive sequence of univariate regressions with Y as response to X, V, W, with X as response to V, W and with V as response to W and having restrictions on the same two variable pairs (see Figure 7a)

(iv) $Y \perp\!\!\!\perp W \mid (X, V)$ and $X \perp\!\!\!\perp V \mid W$.

Four further cases, the first two decomposable, the last two not, are

(v) $Y \perp\!\!\!\perp X \mid (V, W)$ and $V \perp\!\!\!\perp W$,

two independent regressions of Y and X on two independent regressors V and W (see Figure 7b);

(vi) $Y \perp\!\!\!\perp (V, W) \mid X$ and $X \perp\!\!\!\perp W \mid V$,

called a chordless four-chain in concentrations or a Markov chain (see Figures 8a and 8b), that is, a chordless four-chain in a system of univariate recursive regressions again with Y as response to X, V, W, with X as response to V, W and with V as response to W and having response Y and explanatory variable W as chain endpoints;

(vii) $Y \perp\!\!\!\perp W$ and $X \perp\!\!\!\perp V$ and $V \perp\!\!\!\perp W$,

called a chordless four-chain in covariances (see Figures 9a and 9b) or a chordless four-chain in a multivariate regression chain model with Y, X as joint responses and having explanatory variables V, W as chain endpoints;

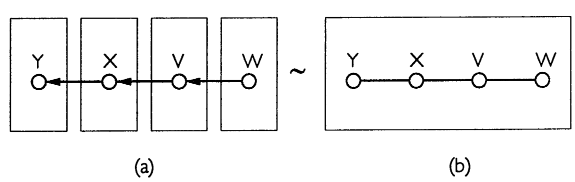

FIG. 8. *Univariate recursive regressions (a) and covariance selection model both specifying the decomposable hypothesis* (vi): $Y \perp\!\!\!\perp (V, W) \mid X$ *and* $X \perp\!\!\!\perp W \mid V$.

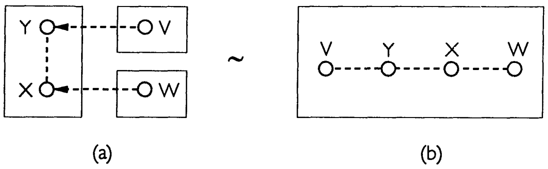

FIG. 9. *Multivariate regression chain model (a) and a linear in covariances structure (b) both specifying the nondecomposable hypothesis* (vii): $Y \perp\!\!\!\perp W$ *and* $X \perp\!\!\!\perp V$ *and* $V \perp\!\!\!\perp W$.

(viii) $$Y \perp\!\!\!\perp W \mid (X, V) \text{ and } X \perp\!\!\!\perp V \mid (Y, W) \text{ and } V \perp\!\!\!\perp W;$$

called a chordless four-chain in a block regression chain model with Y, X as joint responses and having explanatory variables V, W as chain endpoints. The corresponding chain graph has the same shape as the graph in Figure 9a, but dashed lines and arrows are replaced by full lines and arrows.

For our present purpose we give for each empirical example correlations and standardized concentrations showing these as the lower and upper triangle, respectively, such as in Table 1. This allows direct detection of linear marginal independencies between pairs of variables, as shown by very small marginal correlations, that is, standardized covariances, and linear conditional independencies between pairs of variables given all remaining variables, as shown by very small partial correlations, that is, standardized concentrations.

For a formal analysis, consistency of data with a particular structure would be examined via a likelihood ratio test or its equivalent, typically comparing a maximum likelihood fit of the constrained model with that of a saturated model. For the present purposes, however, it is enough to rely on informal comparisons of marginal correlations, partial correlations or standardized regression coefficients, although such dimensionless measures are not in general appropriate for comparing different studies.

Example 1 [*Table* 1, *Figure* 5, *Form* (i)]. Emotions as dispositions or traits of a person and emotions as states, that is, as evoked by particular situations, are notions central to research on stress and on strategies to cope with stressful events. Questionnaires with which the state-trait versions of the emotions anxiety and anger are measured have been developed by Spielberger et al. (1970, 1983). We obtained data for 684 female college students from C. Spielberger on the variables Y, state anxiety; X, state anger; V, trait anxiety and W, trait anger; summaries are displayed in Table 1.

The upper corner of Table 1 shows close agreement with the Form (i): $Y \perp\!\!\!\perp W \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (Y, W)$, see also Figures 5a and 5b. This nondecomposable model has the simple interpretation that prediction of either state variable is not further improved by adding the other trait variable to the remaining two explanatory variables but it does not directly suggest a stepwise process by which the data might have been generated.

Example 2 [*Table* 2, *Figure* 6a, *Form* (ii)]. From a study of the status and reactions of patients awaiting a particular kind of operation (Slangen, Kleeman and Krohne, 1992) we obtained as basic information for 44 female patients: Y, the ratio of systolic to diastolic blood pressure; X, the diastolic blood pressure; both measured in logarithmic scale; V, body mass, that is, weight relative to height, and W, age. Table 2 shows substantial correlations except for a small marginal correlation of pair (Y, W) and a small partial correlation of pair (X, V). These are not to be directly interpreted if—as appears reasonable—each of the blood pressure variables is regarded as a potential response to body mass and age. Instead, the standardized regression coefficients in a saturated multivariate regression of Y, X on V, W display possible independencies of interest. They show close agreement with Form (ii): $Y \perp\!\!\!\perp W \mid V$ and $X \perp\!\!\!\perp V \mid W$, see also Figure 6a, with standardized regression coefficients

$$\begin{pmatrix} \hat{\beta}^*_{yv.w} & \hat{\beta}^*_{yw.v} \\ \hat{\beta}^*_{xv.w} & \hat{\beta}^*_{xw.v} \end{pmatrix} = \begin{pmatrix} 0.486 & 0.040 \\ 0.037 & -0.275 \end{pmatrix}$$

and from Table 2 correlated errors since $\hat{\rho}_{yx.vw} = -0.566$. This nondecomposable model gives as interpretation that diastolic blood pressure increases just with age

TABLE 1
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half) means and standard deviations for n = 684 students

Variable	Y State anx	X State ang	V Trait anx	W Trait ang
Y: = State anxiety	1	0.45	0.47	**−0.04**
X: = State anger	0.61	1	**0.03**	0.32
V: = Trait anxiety	0.62	0.47	1	0.32
W: = Trait anger	0.39	0.50	0.49	1
Mean	18.87	15.23	21.20	23.42
Standard deviation	6.10	6.70	5.68	6.57

Data for Example 1 to Form (i): $Y \perp\!\!\!\perp W \mid (X, V)$ and $X \perp\!\!\!\perp V \mid (Y, W)$ and to Figures 5a and 5b.

TABLE 2
Observed marginal correlations (lower half), observed partial correlations given all remaining variables (upper half), means and standard deviations for n = 44 patients

Variable	Y Lratio bp	X Lsyst. bp	V Body mass	W Age
Y: = Log (syst/diast) bp	1	−0.566	−0.241	0.300
X: = Log diastolic bp	−0.544	1	−0.107	0.491
V: = Body mass	−0.253	0.336	1	0.572
W: = Age	−0.131	0.510	0.608	1
Mean	0.453	4.29	0.379	29.52
Standard deviation	0.091	0.13	0.060	10.59

Data for Example 2 to Form (ii): $Y \perp\!\!\!\perp W \mid V$ and $X \perp\!\!\!\perp V \mid W$ and to Figure 6a.

after controlling for an increase in body mass and that the ratio of systolic to diastolic blood pressure is higher the lower the body mass for persons of the same age. But again, the model does not directly suggest a stepwise process by which the data could have been generated.

Example 3 [*Table* 3, *Figure* 6b, *Form* (iii)]. In a study of strategies to cope with stressful events Kohlmann (1990) collected data for 72 students replying to a German and an American questionnaire. They are both intended to capture two similar strategies: Y, cognitive avoidance and V, blunting are thought of as strategies to reduce emotional arousal and X, vigilance and W, monitoring as strategies to reduce insecurity. The data in Table 3 agree well with Form (iii): $Y \perp\!\!\!\perp W$ and $X \perp\!\!\!\perp V$, see also Figure 6b, but not with (i) because in this case the marginal correlations but not the partial correlations are small.

It is plausible to see strong positive correlations between both pairs of similar strategies, a moderate negative correlation between each set of competing strategies measured one way and no correlation between a strategy measured with one questionnaire and the competing strategy measured with the other questionnaire. However, this structure again cannot be reexpressed with zero regression coefficients in any system of recursive univariate regressions; that is, it does not have a direct explanation as a process by which the data could have been generated.

Pairs of forms from the above special cases (i) to (iv) are mutually exclusive whenever the correlations of all variable pairs other than the two constrained pairs (Y, W) and (X, V) are substantial although with limited data it is of course possible that several different simplified structures are consistent with the data. An exception where two different sets of the above conditions may hold simultaneously is provided by (i) and (iii); that is, a chordless four-cycle in concentrations and in correlations can occur together if a very special structure is present, that is if the marginal correlations in the population satisfy orthogonalities such as

$$(8) \qquad \begin{aligned} \rho_{yw} &= 0, \rho_{xv} = 0, \\ \rho_{yv}\rho_{vw} + \rho_{yx}\rho_{xw} &= 0, \\ \rho_{yv}\rho_{yx} + \rho_{vw}\rho_{xw} &= 0. \end{aligned}$$

The next set of data is an example of this special case.

Example 4 [*Table* 4, *Figures* 5b *and* 6b, *Forms* (i) *and* (iii)]. In a study of effects of working conditions on the manifestation of hypertension, Weyer and Hodapp (1979) report the correlations among the four potential influencing variables displayed in Table 4 for 106 healthy employees. The variables, which are measured with questionnaires, are Y, nervousness; X, stress at work; V, satisfaction with work and W, hierarchical status at work. The observations agree well with both (i): $Y \perp\!\!\!\perp W \mid (X,V)$ and $X \perp\!\!\!\perp V \mid (Y,W)$ (see also Figure 5b) and with (iii): $Y \perp\!\!\!\perp W$ and $X \perp\!\!\!\perp V$ (see also Figure 6b). There is no immediate interpretation; however, one

TABLE 3
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half), means and standard deviations for n = 72 students

Variable	Y Cogn. avoid.	X Vigilance	V Blunting	W Monitoring
Y: = Cognitive Avoidance	1	−0.30	0.49	0.21
X: = Vigilance	−0.20	1	0.21	0.51
V: = Blunting	0.46	**0.00**	1	−0.25
W: = Monitoring	**0.01**	0.47	−0.15	1
Mean	17.49	12.57	3.71	10.40
Standard deviation	6.77	6.39	2.12	3.07

Data for Example 3 to Form (iii): $Y \perp\!\!\!\perp W$ and $X \perp\!\!\!\perp V$ and to Figure 6b.

TABLE 4
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half) for n = 106 healthy employees

Variable	*Y* Nervous	*X* Stress	*V* Satisf.	*W* Hier. Stat.
Y: = Nervousnous	1	0.33	0.26	**0.00**
X: = Stress at work	0.34	1	**0.06**	0.30
V: = Satisfaction with work	0.27	**0.04**	1	−0.35
W: = Hierarchical status	**0.01**	0.29	−0.34	1

Data for Example 4 to Forms (i) and (iii) and to Figures 5b and 6b simultaneously.

explanation for this special structure is that a different combination of the questionnaire items of X, V would lead to variables X^*, V^* such that the much simpler structure $(X^*, Y) \perp\!\!\!\perp (V^*, W)$ holds (Cox and Wermuth, 1992a). For the special structure (8) both the canonical correlations and the transformation matrix to obtain X^*, V^* can be expressed in closed form.

Example 5 [*Table* 5, *Figure* 7b, *Form* (v)]. For an analysis of aggregate economic data von der Lippe (1977) computed growth rates for 24 postwar years in Germany for Y, employment; X, capital gains; V, private consumption and W, exports. The correlation structure suggests that knowing the change in capital gain does not help in predicting the change in employment for given change levels of the demand side, that is, consumption and export (Wermuth, 1979); in addition, changes in consumption were not correlated with changes in exports. This implies two independent responses to two independent explanatory variables or close agreement to Form (v): $Y \perp\!\!\!\perp X \mid (V, W)$ and $V \perp\!\!\!\perp W$; see also Figure 7b.

Example 6 [*Table* 6, *Figure* 8, *Form* (vi)]. In a conditioning experiment with 48 subjects (Zeiner and Schell, 1971), one purpose was to examine discrimination between a noxious and an innocuous stimulus in two periods of a conditioning experiment with Y, a long-interval discriminatory response (6–10 seconds); X, a short-interval discriminatory response (1–5 seconds) in the light of earlier responses: V, the strongest response in the first interval and W, the response to an innocuous stimulus before the experiment itself; all responses are measured as skin resistance. The correlations displayed in Table 6 suggest (Hodapp and Wermuth, 1983, p. 384) a Markov structure (vi) in which $Y \perp\!\!\!\perp (V,W) \mid X$ and $X \perp\!\!\!\perp W \mid V$, see also Figures 8a and 8b, and thus in which the long-interval discriminatory response depends directly only on the short-interval discriminatory response; this short-interval response is directly dependent on the strongest response in the short interval and the latter is well predicted by just the response to an innocuous stimulus before the experiment.

Example 7 [*Table* 7, *Figure* 9, *Form* (vii)]. From an

TABLE 5
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half) of growth rates for n = 24 postwar years in Germany

Variable	*Y* Employment	*X* Capital gain	*V* Consumption	*W* Export
Y: = Employment	1	**−0.11**	0.68	0.55
X: = Capital gain	0.47	1	0.50	0.43
V: = Consumption	0.67	0.55	1	−0.51
W: = Export	0.44	0.39	**0.04**	1

Data for Example 5 to Form (v): $Y \perp\!\!\!\perp X \mid (V, W)$ and $V \perp\!\!\!\perp W$ and to Figure 7b.

TABLE 6
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half) for n = 48 subjects

Variable	*Y* Long	*X* Short	*V* Strong	*W* Innoc
Y: = Long int. discriminatory response	1	0.70	**−0.04**	**−0.12**
X: = Short int. discriminatory response	0.72	1	0.29	**0.14**
V: = Strongest short interval response	0.30	0.54	1	0.62
W: = Response to innocuous stimulus	0.19	0.43	0.71	1

Data for Example 6 to Form (vi): $Y \perp\!\!\!\perp (V,W) \mid X$ and $X \perp\!\!\!\perp V \mid W$ and to Figures 8a and 8b.

TABLE 7
Observed marginal correlations (lower half) and observed partial correlations given two remaining variables (upper half), means and standard deviations for n = 39 diabetic patients

Variable	Y GHb	X Knowledge	V Duration	W Fatalism
Y: = Glucose control, GHb	1	−0.431	−0.407	−0.262
X: = Knowledge, illness	−0.344	1	−0.111	−0.517
V: = Duration, illness	−0.404	**0.042**	1	−0.028
W: = Fatalism, illness	**−0.071**	−0.460	**0.060**	1
Mean	10.02	33.18	147.05	20.13
Standard deviation	2.07	7.86	92.00	5.75

Data for Example 7 to Form (vii): $Y \perp\!\!\!\perp V$, $Y \perp\!\!\!\perp W$, and $X \perp\!\!\!\perp V$ and to Figures 9a and 9b.

investigation of determinants of blood glucose control (Kohlmann et al., 1991), we have data for 39 diabetic patients, who had at most 10 years of formal schooling. The variables considered are Y, a particular metabolic parameter, the glycosylated hemoglobin GHb; X, a score for particular knowledge about the illness, V, the duration of illness in months, and W, a questionnaire score measuring the patients external attribution to "chance" of the occurence of events related to the illness; an attitude called external fatalism. The correlations in Table 7 suggest a structure of the Form (vii), that is, with $Y \perp\!\!\!\perp W, X \perp\!\!\!\perp V$, and $V \perp\!\!\!\perp W$, see also Figures 9a and 9b. One interpretation is that duration of illness and external fatalism are independent explanatory variables in two seemingly independent regressions, where metabolic adjustment is better (low values of GHb) the longer the duration of the illness, knowledge about the illness is lower the higher the external fatalism of a person, and after conditioning on duration and fatalism the metabolic adjustment is still better the higher the knowledge ($\hat{\rho}_{yx.vw} = -0.431$).

6. DISCUSSION

There are a number of general issues arising from the special cases discussed in the previous section, especially the extension to more than four component variables and to models with other than only linear dependencies; for the latter see Cox and Wermuth (1993).

Graphs with, in our notation, full edges have an elegant connection with the theory of Markov random fields which allows general properties to be deduced. See Lauritzen (1989) for a survey of these topics and Isham (1981) for a review of Markov random fields in a broader context. Graphs with dashed edges, or possibly graphs with mixtures of dashed and full edges, do not have the same general features, and it is an open question as to what exactly can be said about them in generality.

There are four types of nondecomposable independence hypotheses illustrated in Section 4 for four variables, namely:

(a) *Nondecomposable hypotheses in block regression chain models [Form* (i), *Example* 1, *Table* 1, *Figure* 5a *and Form* (viii)]. In a block regression chain model the components, even in the simplest case, are divided into responses $Y_a = (Y, X)$ and explanatory variables $Y_b = (V, W)$ with a full directed arrow unless the corresponding regression coefficient in (3) is zero and a full undirected line for the explanatory variables unless they are marginally uncorrelated. For four variables a nondecomposable independence hypothesis in a block regression chain model is characterized by a chordless four-chain in the full edge chain graph, with the two ends of the sequence being explanatory variables, that is, for (V, Y, X, W) in our examples. Figure 5a with Form (i) gives an example of the four-cycle which contains the described four-chain, while Form (viii) leads to an example of the chordless four-chain;

(b) *Nondecomposable hypotheses in concentrations [Form* (i), *Example* 1, *Table* 1, *Figure* 5b]. Models of zero concentrations, that is, the covariance selection models of Dempster (1972), differ from block regression models – from (a) – in treating all variables on an equal footing, that is, having them in the same box where all edges are full undirected lines unless the corresponding variables are partially uncorrelated given the remaining component variables. For four variables a nondecomposable hypotheses in concentrations is characterized by a chordless four-cycle in the associated undirected graph of full edges, that is, in the concentration graph. Figure 5b with Form (i) gives an example of a chordless four-cycle in concentrations for (V, Y, X, W).

(c) *Nondecomposable hypotheses in multivariate regression chain models [Form* (ii), *Example* 2, *Table* 2, *Figure* 6a *and Form* (vii), *Example* 7, *Table* 7, *Figure* 9a]. In multivariate regression chain models the components are – as for (a) – even in the simplest case divided into responses $Y_a = (Y, X)$ and explanatory variables $Y_b = (V, W)$ with a dashed directed arrow unless the corresponding regression coefficient in (2) is zero, a dashed undirected line for the responses unless they are partially uncorrelated given the explanatory variables, and a dashed undirected line for the explanatory variables unless they are marginally uncorrelated. For four variables a nondecomposable independence hypothesis in a multivariate regression chain model is characterized by a chordless four-chain in the dashed edge chain graph with the two ends of the sequence being explanatory variables, that is, for (V, Y, X, W) in our examples. Figure 6a with Form (ii) gives an example of the four-cycle which contains the described four-chain, while Figure 9a with Form (vii) gives an example of the four-chain. Both are seemingly unrelated regressions (Zellner, 1962) together with a specification for the distribution of the explanatory variables.

(d) *Nondecomposable hypotheses in covariances* [*Form* (iii), *Example* 3, *Table* 3, *Figure* 6b *and Form* (vii), *Example* 7, *Table* 7, *Figure* 9b]. Models of zero covariances, that is, models for hypotheses linear in covariances (Anderson, 1973), have – as in (b) – a single block of variables. All edges are dashed undirected lines unless the corresponding variables are marginally uncorrelated. For four variables a nondecomposable independence hypothesis in covariances is characterized by a chordless four-chain in the associated undirected graph of dashed edges, that is, in the covariance graph. Figure 6b with Form (iii) gives an example of the four-cycle which contains a chordless four-chain, while Figure 9b with Form (vii) gives an example of the four-chain.

Models which contain even a single nondecomposable independence hypothesis cannot be distributionally equivalent to a model of univariate recursive regressions. Our examples illustrate that such nondecomposable structures arise in a number of different contexts. There is need to identify them and to find explanations of how they could have been generated. Criteria for establishing nondecomposability for more than four variables are not yet published for general dashed-edge chain graphs, while for full-edge chain graphs such criteria were given by Lauritzen and Wermuth (1989) and for undirected dashed line graphs by Pearl and Wermuth (1993).

We have in this paper concentrated on the kinds of special structure that can arise, especially on their specification and interpretation, rather than on the details of fitting and assessing model adequacy. Under normal-theory assumptions maximum-likelihood fitting and testing for nondecomposable models will call for iterative procedures. A rather general asymptotically efficient noniterative procedure based on embedding the model to be fitted in a saturated model is available (Cox and Wermuth, 1990) either for direct use or as a starting point for iteration (Jensen, Johansen and Lauritzen, 1991). Several issues are important for iterative algorithms. Is there a global maximum or are there several local maxima? Which conditions guarantee the existence of maximum-likelihood estimates? What are the convergence properties of an algorithm? Again, more is known for models represented by full-edged graphs (Speed and Kiiveri, 1986; Frydenberg and Edwards, 1989; Frydenberg and Lauritzen, 1989; Edwards, 1992) than for models with dashed edge graphs. Some of the latter may be fitted with algorithms suitable for linear structural equations; for a discussion of different alternatives see Lee, Poon and Bentler (1992).

For mixtures of discrete and continuous variables, models corresponding to chain graphs with full edges have been intensively studied (Lauritzen and Wermuth, 1989; Lauritzen, 1989; Frydenberg, 1990b; Wermuth and Lauritzen, 1990; Cox and Wermuth, 1992b; Wermuth, 1993), but for models corresponding to chain graphs with dashed edges or possibly mixtures of dashed and full edges the extensions to discrete and mixtures of discrete and continuous variables remain to be developed.

The issue of model choice in the analysis of data has too many ramifications to be discussed satisfactorily in the present paper; some different suitable strategies for analyses with a moderate number of variables are discussed in Wermuth and Cox (1992). In general, if there is sufficient substantive knowledge to give a firm indication both of the nature of the variables and of the independencies expected, then model choice consists largely of testing the adequacy of the proposed model, in particular in examining the supposedly zero correlations, concentrations and regression coefficients. The less the guidance from subject matter considerations, the more tentative will be the conclusions about model structure, but the broad principles of variable selection in empirical regression discussed, for example, by Cox (1968) and Cox and Snell (1974), will apply. In particular, where a number of different models of roughly equal complexity give satisfactory fits to the data, all should be incorporated in the conclusions, unless a choice can be made on subject matter grounds.

There are many aspects of the study of multiple dependencies and associations not addressed in the present paper. In particular the role of latent or hidden variables in clarifying the interpretation of relatively complex structures has not been dealt with, nor has the related matter of the effect of errors of observations in possibly distorting dependencies. Finally, we reemphasize the point made in Section 3 that a key argument for aiming for univariate recursive regressions consistent with subject matter knowledge is that they suggest a stepwise process by which the data might have been generated.

ACKNOWLEDGMENTS

We are grateful to Morten Frydenberg for his helpful comments, to the referees for very constructive suggestions and to the British-German Academic Research Collaboration Programme for supporting our joint work.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135-141.

COX, D. R. (1968). Regression methods; notes on some aspects of regression analysis (with discussion). *J. Roy. Statist. Soc. Ser. A* **131** 265-279.

COX, D. R. (1992). Causality; some statistical aspects. *J. Roy. Statist. Soc. Ser. A* **155** 291-301.

COX, D. R. and SNELL, E. J. (1974). The choice of variables in observational studies. *J. Roy. Statist. Soc. Ser. C* **23** 51-59.

COX, D. R. and WERMUTH, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika* **77** 747-761.

COX, D. R. and WERMUTH, N. (1992a). On the calculation of derived variables in the analyses of multivariate responses. *J. Multivariate Anal.* **42** 167-172.

COX, D. R. and WERMUTH, N. (1992b). Response models for mixed binary and quantitative variables. *Biometrika* **79** 441-461.

COX, D. R. and WERMUTH, N. (1993). Some recent work on methods for the analysis of multivariate observational data in the social sciences. In *Conference Proceedings of the 7th International Conference on Multivariate Analysis, Pennsylvania State Univ., May 1992*. North-Holland, Amsterdam. To appear.

DAWID, A. P. (1979a). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 1-31.

DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157-175.

DUNCAN, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin* **72** 177-182.

EDWARDS, D. (1992). *Graphical Modelling with MIM*. Manual, Univ. Copenhagen.

FRYDENBERG, M. (1990a). The chain graph Markov property. *Scand. J. Statist.* **17** 333-353.

FRYDENBERG, M. (1990b). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* **18** 790-805.

FRYDENBERG, M. and EDWARDS, D. (1989). A modified iterative proportional scaling algorithm for estimation in regular exponential families. *Comput. Statist. Data Anal.* **8** 143-153.

FRYDENBERG, M. and LAURITZEN, S. L. (1989). Decomposition of maximum-likelihood in mixed interaction models. *Biometrika* **76** 539-555.

GLYMOUR, C., SCHEINES, R., SPIRTES, P. and KELLY, K. (1987). *Discovering Causal Structure*. Academic, New York.

GOLDBERGER, A. S. (1964). *Econometric Theory*. Wiley, New York.

GOLDBERGER, A. S. (1992). Models of substance; comment on "On block recursive linear regression equations," by N. Wermuth. *Revista Brasileira de Probabilidade e Estatística* **6** 46-48.

HAAVELMO, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11** 1-12.

HODAPP, V. and WERMUTH, N. (1983). Decomposable models: A new look at interdependence and dependence structures in psychological research. *Multivariate Behavioral Research* **18** 361-390.

ISHAM, V. (1981). An introduction to spatial point processes and Markov random fields. *Internat. Statist. Rev.* **49** 21-43.

JENSEN, S. T., JOHANSEN, S. and LAURITZEN, S. L. (1991). Globally convergent algorithms for maximizing a likelihood function. *Biometrika* **78** 867-878.

JÖRESKOG, K. G. (1973). A general method for estimating a linear structural equation system. In *Structural Equation Models in the Social Sciences* (A. S. Goldberger and O. D. Duncan, eds.) 85-112. Seminar Press, New York.

KOHLMANN, C.-W. (1990). *Streßbewältigung und Persönlichkeit*. Huber, Bern.

KOHLMANN, C. W., KROHNE, H. W., KÜSTNER E., SCHREZENMEIR, J., WALTHER, U. and BEYER, J. (1991). Der IPC-Diabetes-Fragebogen: ein Instrument zur Erfassung krankheitsspezifischer Kontrollüberzeugungen bei Typ-I-Diabetikern. *Diagnostica* **37** 252-270.

LAURITZEN, S. L. (1989). Mixed graphical association models (with discussion). *Scand. J. Statist.* **16** 273-306.

LAURITZEN, S. L., DAWID, A.P., LARSEN, B. and LEIMER, H. G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491-505.

LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31-57.

LEE, S.-Y., POON, W.-Y. and BENTLER, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika* **57** 89-105.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufman, San Mateo, CA.

PEARL, J. and VERMA, T. S. (1991). A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning* (J. A. Allen, R. Fikes and E. Sandewall, eds.). Morgan Kaufman, San Mateo, CA.

PEARL, J. and WERMUTH, N. (1993). When can an association graph admit a causal interpretation? In *Conference Proceedings of the 4th International Workshop on Artificial Intelligence and Statistics, Fort Lauderdale, Florida*. To appear.

SLANGEN, K., KLEEMANN, P. P. and KROHNE, H. W. (1992). Coping with surgical stress. In *Attention and Avoidance; Strategies in Coping with Aversiveness* (H. W. Krohne, ed.) 321-348. Springer, New York.

SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138-150.

SPIELBERGER, C. D., GORSUCH, R. L. and LUSCHENE, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Palo Alto, CA.

SPIELBERGER, C. D., RUSSELL, S. and CRANE, R. (1983). Assessment of anger. In *Advances in Personality Assessment* (J. N. Butcher and C. D. Spielberger, eds.) **2** 159-187. Erlbaum, Hillsdale, NJ.

TINBERGEN, J. (1937). *An Econometric Approach to Business Cycle Problems*. Hermann, Paris.

VAN DE GEER, J. P. (1971). *Introduction to Multivariate Analysis for the Social Sciences*. Freeman, San Francisco.

VERMA, T. S. and PEARL, J. (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In *Uncertainty in Artificial Intelligence* (D. Dubois, M. P. Wellman, B. D'Ambrosio and P. Smets, eds.) **8** 323-330. Morgan Kaufmann, San Mateo, CA.

VON DER LIPPE, P. (1977). Beschäftigungswirkung durch Umverteilung? *WSI-Mitteilungen* **8** 505-512.

WERMUTH, N. (1979). Datenanalyse und multiplikative Modelle. *Allgemeines Statistisches Archiv* **63** 323-339.

WERMUTH, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.* **75** 963-972.

WERMUTH, N. (1989). Moderating effects in multivariate normal distributions. *Methodika* **3** 74-93.

WERMUTH, N. (1992). On block-recursive regression equations (with discussion). *Revista Brasileira de Probabilidade e Estatística* 6 1–56.

WERMUTH, N. (1993). Association structures with few variables: characteristics and examples. In *Theory and Methods for Population Health Research* (K. Dean, ed.) 181–202. Sage, London.

WERMUTH, N. and COX, D. R. (1992). Graphical models for dependencies and associations. In *Computational Statistics, Proceedings of the 10th Symposium on Computational Statistics, Neuchâtel.* (Y. Dodge and J. Whittaker, eds.) 1 235–249. Physica, Heidelberg.

WERMUTH, N. and LAURITZEN, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. Roy. Statist. Soc. Ser. B* 52 21–72.

WEYER, G. and HODAPP, V. (1979). Job-stress and essential hypertension. In *Stress and Anxiety* (I. G. Sarason and C. D. Spielberger, eds.) 6 337–349. Hemisphere, Washington, D.C.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.

WOLD, H. O. (1954). Causality and econometrics. *Econometrica* 22 162–177.

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research* 20 557–585.

WRIGHT, S. (1923). The theory of path coefficients: A reply to Niles' criticism. *Genetics* 8 239–255.

ZEINER, A. R. and SCHELL, A. M. (1971). Individual differences in orienting, conditionality, and skin resistance responsitivity. *Psychophysiology* 8 612–622.

ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* 57 348–368.