Revista Brasileira de Probabilidade e Estatística (1992), **6**, pp. 1–56.

ON BLOCK-RECURSIVE LINEAR REGRESSION EQUATIONS

Nanny Wermuth

Psychological Institute
University of Mainz
Staudinger Weg 9
D-6500 Mainz
Federal Republic of Germany

Summary

The main purpose of this paper is to clarify relations and distinctions between several approaches suggested in the statistical literature for analysing structures in correlation matrices, i.e. of relations among observable, quantitative variables having exclusively linear associations. Block-recursive regression equations are derived as the key to understanding the relation between two main approaches, between graphical chain models for continuous variables on the one hand and linear structural equations discussed in the econometric and in the psychometric literature on the other hand. Their relations to other model classes such as covariance selection, multivariate linear regression, and path analysis are discussed.

Key words: Conditional independence; covariance selection; decomposable model; graphical chain model; linear structural equations; multivariate regression; nondecomposable independence hypothesis; path analysis.

1. Introduction

The first suggestions of how to describe and study structures of correlated variables were given by Wright (1921, 1923, 1934) and called path analysis.

He defined what we would now call reduced models (Cox and Wermuth, 1990), with the help of zero restrictions on parameters of linear equations and proposed methods for estimating path coefficients and implied correlations, i.e. for estimating the equation parameters and the correlation matrix under the assumptions of the reduced model. Furthermore, he suggested to evaluate the goodness of fit of a reduced model by judging how much the correlations implied by the model deviated from those actually observed. However he did not state conditions under which his proposed estimates of equation parameters and of the correlation matrix could be derived from general principles.

Tukey (1954) established that Wright's estimates of equation parameters can be justified by the method of least squares if applied to a single multiple regression equation expressed in standardized variables, i.e. to a linear equation having residuals which are uncorrelated with the explanatory variables and having observed variables which all have mean zero and variance one. The same holds for a recursive sequence of such equations (Wold, 1954). They involve a set of variables $\{Y_1, \ldots, Y_q\}$ in which Y_r is specified by linear regression on $\{Y_{r+1}, \ldots, Y_q\}$ for $r = 1, \ldots, q-1$, error terms in the different equations being uncorrelated. Wermuth (1980) showed that Wright's proposed estimate of the correlation matrix is not, in general, a maximum likelihood estimate for recursive regressions of jointly normal variables, but only for those which define a decomposable correlation matrix. These are arguments for viewing Wright's path analysis as the first description of univariate recursive regressions and of decomposable covariance selection models (Dempster, 1972). In the latter class of models the joint association structure can be decomposed into structures of several proper subsets of variables so that complex tasks – like describing its features – can be made simpler by splitting them into several smaller ones.

Any model defined in terms of univariate recursive regressions combines several advantages. First, it describes a stepwise process by which the data could have been generated and in this sense it may even 'prove the basis

for developing causal explanations' (Cox, 1991). Second, each parameter in the system has a well understood meaning since it is a regression coefficient, i.e. it indicates for unstandardized variables by which amount 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: each added restriction introduces a further conditional linear 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, i.e. 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 the corresponding models are distributionally equivalent (Frydenberg, 1990), i.e. whether they specify the same joint distribution and the same set of restrictions. Fourth, the analysis of the whole structure can be achieved with the help of a sequence of separate univariate linear regression analyses. Fifth, an algorithm exists (Pearl and Verma, 1991, Verma and Pearl, 1992) which decides for an arbitrary list of conditional independence statements whether it defines a univariate recursive system and, if it does, a corresponding directed acyclic graph is drawn.

There are, however, substantive research questions – like the one discussed by Haavelmo (1943) and like the one described below in Section 6 – which require to look at linear equations simultaneously. In that case some of the nice features of univariate recursive regressions are necessarily lost. The common feature of such more complex models is that they do not prescribe a recursive process from which the data could be generated. If independencies hold simultaneously of a form which cannot be conveniently formulated by zero restrictions on individual parameters of any system of univariate recursive regressions we name them *nondecomposable independencies.*

The simplest three distinct types of nondecomposable independence hypotheses occur for four variables. For Y_1, Y_2, Y_3, Y_4 they can be written in the

notation for independencies introduced by Dawid (1979) as:

$$\begin{array}{lll} (i): & Y_1 \perp\!\!\!\perp Y_4 \,|\, (Y_2, Y_3) & \text{and} \quad Y_2 \perp\!\!\!\perp Y_3 \,|\, (Y_1, Y_4), \\ (ii): & Y_1 \perp\!\!\!\perp Y_4 \,|\, Y_3 & \text{and} \quad Y_2 \perp\!\!\!\perp Y_3 \,|\, Y_4, \\ (iii): & Y_1 \perp\!\!\!\perp Y_4 & \text{and} \quad Y_2 \perp\!\!\!\perp Y_3. \end{array}$$

The first hypothesis (i) states conditional independence of pairs (Y_1, Y_4) and (Y_2, Y_3) given all of the remaining variables. The second hypothesis (ii) postulates conditional independence of (Y_1, Y_4) given Y_3 and of (Y_2, Y_3) given Y_4. The last hypothesis (iii) just means marginal independence of both pairs, of (Y_1, Y_4) and of (Y_2, Y_3). None of these hypotheses can be formulated in terms of zero restrictions on a univariate recursive system, but each of them results most conveniently as a zero restriction model in one – but not in the other two – of the following model classes: hypothesis (i) as a graphical chain model (Lauritzen and Wermuth, 1989), hypothesis (ii) as a linear structural equation model (Goldberger, 1964; Jöreskog,1973), and hypothesis (iii) as a covariance matrix with linear structure (Anderson, 1973).

To see this and to clarify more general distinctions and similarities between different approaches to studying such nondecomposable independence hypotheses we summarize a number of known facts about regression and correlation coefficients in Section 2 and introduce systems of block-recursive regressions in Section 3. In Section 4 relations of block-recursive regressions to graphical chain models, to path analyses, and to covariance selection (Dempster, 1972) are discussed, while Section 5 treats relations of block-recursive regressions to linear structural equations. Finally, in Section 6 a research question corresponding to the nondecomposable independence hypothesis of type (i) is described and results for some data are reported.

2. Notation and facts

We assume a multivariate normal distribution for a q-dimensional random vector $X = (Y_1, \ldots, Y_q)^T$ partitioned into three subvectors X_a, X_b, X_c and let

the mean μ, the positive definite covariance matrix Σ and the concentration matrix Σ^{-1} be partitioned accordingly

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{pmatrix}, \quad \Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ \Sigma^{ba} & \Sigma^{bb} & \Sigma^{bc} \\ \Sigma^{ca} & \Sigma^{cb} & \Sigma^{cc} \end{pmatrix}.$$

Diagonal elements of Σ are called variances (σ_{ii}), those of Σ^{-1} precisions (σ^{ii}). Off-diagonal elements of Σ are covariances (σ_{ij}), those of Σ^{-1} are concentrations (σ^{ij}).

From Dempster (1969, Chapter 4) we derive all of the following facts. We know that Σ_{bb} is the covariance matrix of X_b in the marginal distribution of X_b, while Σ^{bb} is the concentration matrix of X_b in the conditional distribution of X_b given $X_a = x_a$, $X_c = x_c$. The partial covariance matrix of both of X_a and X_b given $X_c = x_c$ and the partial concentration matrix of X_b and X_c in their joint marginal distribution are, respectively,

$$\begin{pmatrix} \Sigma_{aa.c} & \Sigma_{ab.c} \\ \Sigma_{ba.c} & \Sigma_{bb.c} \end{pmatrix}, \quad \begin{pmatrix} \Sigma^{bb.a} & \Sigma^{bc.a} \\ \Sigma^{cb.a} & \Sigma^{cc.a} \end{pmatrix}$$

where

$$\Sigma_{aa.c} = \Sigma_{aa} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{ca}, \quad \Sigma^{bb.a} = \Sigma^{bb} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ab},$$

$$\Sigma_{bb.c} = \Sigma_{bb} - \Sigma_{bc}\Sigma_{cc}^{-1}\Sigma_{cb}, \quad \Sigma^{cc.a} = \Sigma^{cc} - \Sigma^{ca}(\Sigma^{aa})^{-1}\Sigma^{ac},$$

$$\Sigma_{ab.c} = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}, \quad \Sigma^{bc.a} = \Sigma^{bc} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ac}.$$

Furthermore, the conditional covariance matrix X_a given both of $X_b = x_b$, $X_c = x_c$ and the concentration matrix in the marginal distribution of X_c are

$$\Sigma_{aa.bc} = \Sigma_{aa.c} - \Sigma_{ab.c}\Sigma_{bb.c}^{-1}\Sigma_{ba.c}, \quad \Sigma^{cc.ab} = \Sigma^{cc.a} - \Sigma^{cb.a}(\Sigma^{bb.a})^{-1}\Sigma^{bc.a},$$

respectively.

Known relations between (partial) covariance and (partial) concentration matrices are

$$\Sigma_{aa.bc} = (\Sigma^{aa})^{-1}, \quad \Sigma_{bb.c} = (\Sigma^{bb.a})^{-1}, \quad \Sigma_{cc} = (\Sigma^{cc.ab})^{-1}. \tag{2.1}$$

and parameters in regressions can be expressed in terms of either covariances or concentrations as follows.

If we take a multivariate linear regression of X_a on both of X_b and X_c i.e.

$$X_a = \Pi_{a|bc} \begin{pmatrix} x_b \\ x_c \end{pmatrix} + V_a$$

with

$$\mathrm{E}(V_a) = 0, \ \ \mathrm{Cov}\{V_a(X_b^T X_c^T)\} = 0, \ \ \mathrm{Var}(V_a) = \Sigma_{aa.bc},$$

where each variable is measured in deviation from its mean, then the equation parameters $\Pi_{a|bc}$ can be written as

$$\begin{aligned} \Pi_{a|bc} &= (\Sigma_{ab.c}\Sigma_{bb.c}^{-1} \quad \Sigma_{ac.b}\Sigma_{cc.b}^{-1}) \\ &= -(\Sigma^{aa})^{-1}(\Sigma^{ab} \ \Sigma^{ac}). \end{aligned}$$

Similarly, the equation parameters $\Pi_{b|c}$ in a multivariate regression of X_b on X_c alone can be written as

$$\Pi_{b|c} = \Sigma_{bc}\Sigma_{cc}^{-1} = -(\Sigma^{bb.a})^{-1}\Sigma^{bc.a}. \tag{2.2}$$

The matrix $\Pi_{b|c}$ has elements $\beta_{ij.d}$ where the index $i \in b$ refers to a regressand, i.e. to a response variable $X_{\{i\}}$, the index $j \in c$ refers to the regressor, i.e. to the explanatory variable $X_{\{j\}}$, and the set $d = c\backslash\{j\}$ to the remaining regressors X_d in a univariate linear regression of $X_{\{i\}}$ on X_c. Thus, each matrix of regression coefficients from a multivariate linear regression contains as elements partial regression coefficients which are identical to equation parameters in separate univariate linear regressions.

Each partial regression coefficient $\beta_{ij.d}$ can be expressed in terms of covariances or concentrations as

$$\beta_{ij.d} = \frac{\sigma_{ij.d}}{\sigma_{jj.d}} = -\frac{\sigma^{ij.g}}{\sigma^{ii.g}}, \tag{2.3}$$

where $d = c\backslash\{j\}$ is again the index set of the remaining regressors and $g = \{1,\ldots,q\}\backslash\{\{i\} \cup c\}$ is the index set of variables being neither regressand nor

regressor in the univariate regression. Note that the apparent similarity in the dot-notation does not imply a similar interpretation: in a partial covariance the variables to the right of the dot are the variables conditioned upon (in (2.3): X_d), while in a partial concentration the variables to the right of the dot are those over which one has marginalized (in (2.3): X_g).

Similarly to (2.3) we have for the partial correlation coefficient

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

For saturated models discussed so far, i.e. for unrestricted regressions, relations completely analogous to (2.1), (2.2), and (2.3) hold for observed quantities. More precisely, let $x^l = (y_1^l, \ldots, y_q^l)^T$ be column l of the q by n matrix of n observations on the random vector X. Then $\hat{\Sigma} = (\sum_{l=1}^n (x^l - \bar{x})(x^l - \bar{x})^T)/n$ with $\bar{x} = (\sum_{l=1}^n x^l)/n$ is the observed unrestricted covariance matrix. If, for instance, $\hat{\Pi}_{b|c}$ is defined in terms of submatrices of $\hat{\Sigma}$ just as $\Pi_{b|c}$ in (2.2) is defined in terms of submatrices of Σ, then its elements coincide with least squares regression coefficients obtained from separate univariate linear regressions of $X_{\{i\}}$ for $i \in b$ on X_c.

A similar result holds for reduced multivariate regression models defined in terms of zero restrictions of the type

$$\Pi_{a|bc} = (0 \quad \Pi_{a|c}) = -(\Sigma^{aa})^{-1}(0 \quad \Sigma^{ac}),$$

i.e. those which have been called 'general linear hypotheses'. To obtain maximum-likelihood estimates under such a hypothesis, again, just separate univariate least squares regression analyses are needed (Anderson, 1958, p. 210), those of $X_{\{i\}}$ for $i \in a$ on X_c.

Some other reduced multivariate regression models require simultaneous analyses. The simplest one is the nondecomposable independence hypothesis (*ii*) of Section 1, i.e. $Y_1 \perp\!\!\!\perp Y_4 \,|\, Y_3$ and $Y_2 \perp\!\!\!\perp Y_3 \,|\, Y_4$. With $X_b = (Y_1 \;\; Y_2)^T$, $X_c = (Y_3 \;\; Y_4)^T$ it is equivalent to restricting the equation parameters:

$$\Pi_{b|c} = \begin{pmatrix} \beta_{13.4} & \beta_{14.3} \\ \beta_{23.4} & \beta_{24.3} \end{pmatrix}$$

to satisfy $\beta_{14.3} = \beta_{23.4} = 0$, since in multivariate normal distributions a variable pair is conditionally independent if and only if its corresponding conditional covariance is zero, i.e.

$$X_{\{i\}} \perp\!\!\!\perp X_{\{j\}} \,|\, X_{\{d\}} \text{ if and only if } \sigma_{ij.d} = 0.$$

It follows from (2.3) that the variable pair $(X_{\{i\}}, X_{\{j\}})$ is conditionally independent given $X_{\{d\}}$ if and only if the corresponding partial regression coefficient $(\beta_{ij.d})$ or the corresponding concentration $(\sigma^{ij.g})$ is zero.

3. Block-recursive linear regression equations

In this section, model classes are derived which consist of systems of linear block-recursive equations. First, saturated or complete equations of block-recursive concentration equations are defined as one-to-one transformations of a concentration matrix, i.e. of the inverse covariance matrix (Section 3.1). Second, saturated or complete block-recursive linear regression equations are obtained by one-to-one transformations from block-recursive concentration equations (Section 3.2). Third, two equivalent classes of reduced models are defined by imposing sets of zero restrictions on the equation parameters of either complete system (Section 3.3). Finally, it is shown (Section 3.4) that any set of zero restrictions – of the type described in (3.4) below – on the parameters of a complete block-recursive system can be partitioned into subsets of zero restrictions affecting separate components of the concentration matrix of all variables. This property assures for each reduced model in these model classes that the number of the degrees of freedom coincides with its number of zero restrictions on equation parameters.

3.1 Complete block-recursive concentration equations

We take again a q-dimensional random vector $X = (Y_1, \ldots, Y_q)^T$ to be partitioned into three subvectors X_a, X_b, X_c. We thereby assume that in an

application a dependence chain $\mathcal{C} = (a, b, c)$ has been prespecified from subject matter considerations. It gives an ordered partitioning of $\{1, \ldots, q\}$. The chain elements a, b, and c define with $a \cup b \cup c$, $b \cup c$, and c three sets of *concurrent variables* which determine, in particular, how an equation parameter is to be interpreted: each is a partial association given all of the remaining concurrent variables. In general there may of course be more than three chain elements.

To each positive definite covariance matrix Σ of X there exists - for a given order on the variables specified with the dependence chain $\mathcal{C} = (a, b, c)$ - a unique upper block-triangular decomposition of the concentration matrix Σ^{-1} as

$$\Sigma^{-1} = A^T T^{-1} A \tag{3.1}$$

with

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

This is a direct extension of results on successive orthogonalisation discussed by Dempster (1972, Chapter 4.2). The diagonal block matrices of T^{-1} relate to (partial) covariance matrices as stated in (2.1). The diagonal block matrices of A are identity matrices, while the upper off-diagonal blocks relate to matrices of regression coefficients as stated in (2.2).

From (3.1) a *complete system of block-recursive concentration equations* is defined by taking

$$B^*(X - \mu) = W^* \tag{3.2}$$

with

$$B^* = T^{-1}A = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ 0 & \Sigma^{bb.a} & \Sigma^{bc.a} \\ 0 & 0 & \Sigma^{cc.ab} \end{pmatrix}, \quad \mu = \mathrm{E}(X).$$

From this definition together with (3.1) the mean and variance of the residuals W^* are obtained as

$$\mathrm{E}(W^*) = B^*(\mathrm{E}(X) - \mu) = 0,$$

$$\mathrm{Var}(W^*) = T^{-1}A\,\mathrm{Var}(X)A^TT^{-1} = T^{-1}.$$

Thus, the q variances and the $q(q-1)/2$ distinct association parameters of the complete system of equations (3.2) are the diagonal and upper off-diagonal elements in B^*.

3.2 Complete block-recursive regression equations

A *complete system of block-recursive regressions* results from (3.2) after dividing each single equation in the system by its corresponding diagonal element of B^*. This gives

$$B(X-\mu) = W \tag{3.3}$$

with

$$\begin{aligned} X_{\{i\}} &= \mathrm{E}(X_{\{i\}} \mid X_{a\cup b\cup c\setminus\{i\}}) + W_{\{i\}}\,, \qquad i \in a, \\ &= \mathrm{E}(X_{\{i\}} \mid X_{b\cup c\setminus\{i\}}) + W_{\{i\}}\,, \qquad i \in b, \\ &= \mathrm{E}(X_{\{i\}} \mid X_{c\setminus\{i\}}) + W_{\{i\}}\,, \qquad i \in c, \end{aligned}$$

where the expectations in each equation are conditional given all of the remaining concurrent variables.

It follows from the definition of a partial regression coefficient as a quotient of a concentration to a precision in equation (2.3) that each equation parameter in (3.3), that is each element of B, is in fact a partial regression coeffficient.

For instance, with $a = \{1,2\}$, $b = \{3,4\}$, $c = \emptyset$ we have

$$B^* = \begin{pmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} & \sigma^{14} \\ \sigma^{21} & \sigma^{22} & \sigma^{23} & \sigma^{24} \\ 0 & 0 & \sigma^{33.12} & \sigma^{34.12} \\ 0 & 0 & \sigma^{43.12} & \sigma^{44.12} \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & -\beta_{12.34} & -\beta_{13.24} & -\beta_{14.23} \\ -\beta_{21.34} & 1 & -\beta_{23.14} & -\beta_{24.13} \\ 0 & 0 & 1 & -\beta_{34} \\ 0 & 0 & -\beta_{43} & 1 \end{pmatrix}, \quad B\mu = \begin{pmatrix} \alpha_{1.234} \\ \alpha_{2.134} \\ \alpha_{3.4} \\ \alpha_{4.3} \end{pmatrix},$$

where the α's are intercepts.

3.3 Incomplete block-recursive regression equations

An *incomplete system of block-recursive regressions* is a complete system (3.3) with some of the equation parameters restricted to be zero. More precisely, zero restrictions on elements b_{rs} of B are defined with any set

$$I \subseteq \tilde{I} = \{(r,s) \mid 1 \leq r \neq s \leq q\} \tag{3.4}$$

such that for $(r,s) \in I$ we have $b_{rs} = b_{sr} = 0$ if (r,s) is a position in one of the block-matrices of B along the diagonal and we have only $b_{rs} = 0$ if (r,s) is a position in one of the off-diagonal block-matrices of B. An example is the hypothesis (i) of Section 1, which translates for jointly normal variables into $\beta_{14.23} = \beta_{23.14} = 0$.

Similarly, an *incomplete system of block-recursive concentration equations* is a complete system (3.2) with some of its equation parameters restricted to be zero. From the definitions of B, B^* together with equation (2.3) and the positive definiteness of Σ it follows that to each incomplete system of block-recursive regression equations there is a corresponding incomplete system of concentration equations such that zero restrictions on B^* coincide with those on B, and conversely.

3.4 Implications of incomplete block-recursive regression equations for the covariance matrix

Each set of zero restrictions (3.4) having k elements introduces exactly k restrictions on Σ. This follows from the recursive parametrisation $\{\varphi_a, \varphi_b, \varphi_c\}$ of Σ^{-1}:

$$\Sigma^{-1} = \varphi_a(\Sigma^{aa})^{-1}\varphi_a^T + \varphi_b(\Sigma^{bb.a})^{-1}\varphi_b^T + \varphi_c(\Sigma^{cc.ab})^{-1}\varphi_c^T \tag{3.5}$$

with

$$\varphi_a = \begin{pmatrix} \Sigma^{aa} \\ \Sigma^{ba} \\ \Sigma^{ca} \end{pmatrix}, \quad \varphi_b = \begin{pmatrix} 0_{ab} \\ \Sigma^{bb.a} \\ \Sigma^{cb.a} \end{pmatrix}, \quad \varphi_c = \begin{pmatrix} 0_{ac} \\ 0_{bc} \\ \Sigma^{cc.ba} \end{pmatrix},$$

which gives Σ^{-1} as the sum of three q by q matrices, of

$$\begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ \Sigma^{ba} & \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ab} & \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ac} \\ \Sigma^{ca} & \Sigma^{ca}(\Sigma^{aa})^{-1}\Sigma^{ab} & \Sigma^{ca}(\Sigma^{aa})^{-1}\Sigma^{ac} \end{pmatrix},$$

$$\begin{pmatrix} 0_{aa} & 0_{ab} & 0_{ac} \\ 0_{ba} & \Sigma^{bb.a} & \Sigma^{bc.a} \\ 0_{ca} & \Sigma^{cb.a} & \Sigma^{cb.a}(\Sigma^{bb.a})^{-1}\Sigma^{bc.a} \end{pmatrix}, \quad \begin{pmatrix} 0_{aa} & 0_{ab} & 0_{ac} \\ 0_{ba} & 0_{bb} & 0_{bc} \\ 0_{ca} & 0_{cb} & \Sigma^{cc.ab} \end{pmatrix}.$$

To each Σ^{-1} there corresponds one set $\{\varphi_a, \varphi_b, \varphi_c\}$ and each set $\{\varphi_a, \varphi_b, \varphi_c\}$ defines one Σ^{-1}. The distinct elements of the φ's coincide with the distinct elements of the matrix B^* in the block-recursive system of concentration equations (3.2) and zero restrictions on concentrations in one of the φ's do not affect the elements in another φ. Therefore, we call the elements of $\{\varphi_a, \varphi_b, \varphi_c\}$ the *separate φ-components* of Σ^{-1}.

The considerations in this section can be extended to dependence chains with more than three elements by introducing an appropriate more complex notation. This leads to the following proposition.

Proposition 3.1. *For a given dependence chain any positive definite covariance matrix Σ can be represented uniquely in terms of the coefficient matrix B of X and the covariance matrix $\mathrm{Cov}(W)$ of residuals in a system of block-recursive regressions and conversely. Each set of zero restrictions (3.4) on the regression coefficients in such a system can be partitioned into sets of zero restrictions affecting the separate φ-components of the concentration matrix Σ^{-1}.*

4. Relations to path analysis, to covariance selection, and to graphical chain models

4.1 Path analysis

If one disregards Wright's suggestions for estimating parameters and for a

causal interpretation of equation parameters then the key ideas of path analysis for observable variables (Wright, 1921, 1923, 1934) are

- to represent a system of linear relations in terms of graphs containing vertices for variables, lines for symmetric associations and arrows for directed associations;

- to specify a hypothesis about the system by assuming certain relations between variable pairs to be only indirect, corresponding to missing direct connections between pairs of vertices in the graph and to zero restrictions on linear equation parameters;

- to assess the goodness-of-fit of a hypothesis by comparing observed correlations with those implied by the hypothesis;

- to evaluate the relative importance of the direct relations in terms of appropriate standardised measures of association.

The derivation of block-recursive equations with the help of a block-triangular decomposition of the concentration matrix is a direct extension of a method used previously (Wermuth, 1980) to justify univariate recursive equations, i.e. as obtainable by a triangular decomposition of the concentration matrix. Univariate recursive systems, i.e. Wright's path analysis can be viewed as a special case of block-recursive systems when each block contains just a single variable.

For these reasons a complete equation system (3.3) can be called a *saturated block-recursive path analysis model.* The term saturated reminds us that just a reparametrisation of but no hypothesis on the covariance matrix is of interest. A dependence chain, prespecified from subject-matter considerations, determines which type of regression coefficients are the parameters in the system. An incomplete equation system (3.3) and (3.4) having $I \neq \emptyset$ can be called a *reduced block-recursive path analysis model.*

Though Wright's rules for computing estimates of equation parameters and of correlations do, in general, not apply to the block-recursive models his important concept of 'indirect relations' has exactly the same meaning as in the recursive univariate case: each indirect relation implies a particular zero partial correlation (compare (3.4), (3.3), and (2.3)). The other key ideas are closely adhered to as well (Wermuth and Lauritzen, 1990).

4.2 Covariance selection

Covariance selection has been proposed by Dempster (1972) as an application of 'the principle of parsimony in parametric model fitting, which suggests that parameters should be introduced only sparingly, and only when the data indicate that they are required'. In block-recursive regressions covariance selection plays two quite different roles: it is used to justify the parameter estimates – originally obtained by maximum-likelihood estimation for jointly normal variables – in a distribution-free context by minimizing the generalized residual variance (Wilks, 1932) and it gives a class of models which is equivalent to degenerate block-recursive regressions, i.e. to those in which the dependence chain has just one element.

Dempster showed how to derive the maximum-likelihood estimate $\tilde{\Sigma}$ for the covariance matrix of a multivariate normal distribution from the observed covariance matrix $\hat{\Sigma}$ subject to restricting a subset of concentrations in Σ^{-1} to be zero. More precisely, a zero pattern in Σ^{-1} is specified with a set

$$J \subseteq \tilde{J} = \{(r,s) \mid 1 \leq r \neq s \leq q\}$$

such that for $(r,s) \in J$ we have $\sigma^{rs} = \sigma^{sr} = 0$. The maximum was shown to be unique if it exists and to be obtained when the determinant of the concentration matrix $\hat{\Sigma}^{-1}$ is minimized *and* when, at the same time, the determinant of the observed inverse correlation matrix $\hat{P}^{-1}$ is minimized. The reason is that

$$\mathrm{Det}(\hat{\Sigma}^{-1}) = (\hat{\sigma}_{11}\hat{\sigma}_{22}\ldots\hat{\sigma}_{qq})\,\mathrm{Det}(\hat{P}^{-1})$$

and that in the estimation procedure of covariance selection the variances are kept fixed at their observed levels. Furthermore, all covariances corresponding to unrestricted concentrations are kept fixed at their observed levels so that that the maximum-likelihood estimate has the property

$$\tilde{\sigma}^{rs} = \tilde{\sigma}^{sr} = 0 \qquad \text{for } (r,s) \in J$$

$$\tilde{\sigma}_{rs} = \tilde{\sigma}_{sr} = \hat{\sigma}_{rs} \qquad \text{for } (r,s) \in \tilde{J} \setminus J.$$

The convergence of several iterative algorithms to solve the likelihood equations has been proven by Speed and Kiiveri (1986). One of these algorithms is available as a Fortran routine (Wermuth and Scheidt, 1977), another is implemented in a program for more general models, introduced as hierarchical mixed interaction models, by Edwards (1990).

With covariance selection the problem of minimizing the generalized residual variance, i.e. $\mathrm{Det}(\mathrm{Var}(W^*))$ and $\mathrm{Det}(\mathrm{Var}(W))$ in incomplete block-recursive equations is solved. To see this consider first systems consisting of a single block. Then (3.1) to (3.3) give directly

$$\mathrm{Det}(\mathrm{Var}(W^*)) = (\rho^{11}\rho^{22}\ldots\rho^{qq})^2\,\mathrm{Det}(\mathrm{Var}(W)) = \mathrm{Det}(\Sigma^{-1}),$$

where ρ^{ii}'s are the diagonal elements of P^{-1} and where we have deleted hats to simplify notation. Next, take systems consisting of three blocks. Then it follows from the special triangular form of A in (3.1) that $\mathrm{Det}(A) = 1$ and that therefore the determinant of the residual covariance matrix $\mathrm{Var}(W^*) = T^{-1}$ equals the determinant of Σ^{-1}. Furthermore, the block-diagonal form of T and (2.1) give

$$\begin{aligned}\mathrm{Det}(\Sigma^{-1}) &= \mathrm{Det}(\Sigma^{aa})\mathrm{Det}(\Sigma^{bb.a})\mathrm{Det}(\Sigma^{cc.ab}) \\ &= \mathrm{Det}(\Sigma^{-1}_{aa.bc})\mathrm{Det}(\Sigma^{-1}_{bb.c})\mathrm{Det}(\Sigma^{-1}_{cc})\end{aligned}$$

so that the minimum of $\mathrm{Det}(\Sigma^{-1})$ subject to restrictions (3.4) is reached when the determinants of the three inverse conditional covariance matrices are minimized separately – provided the restrictions (3.4) can be partitioned to affect exactly one of the conditional matrices.

Two arguments are needed here. First, since the association parameters of the block-recursive concentration equations (3.2) are elements of

$$(\Sigma^{aa}\ \Sigma^{ab}\ \Sigma^{ac}),\quad (\Sigma^{bb.a}\ \Sigma^{bc.a}),\quad (\Sigma^{cc.ab}),$$

the restrictions (3.4) can be partitioned correspondingly into three sets of restrictions. The first set of restrictions affects concentrations involving just variables in a of the concurrent variables given by $a \cup b \cup c$, the second affects concentrations involving just variables in b of the concurrent variables given by $b \cup c$, and the third affects concentrations of variables given by c. Second, to mimimize the determinant of the concentration matrix of a set of concurrent variables subject to having zero concentrations involving only response variables is to mimimize the determinant of the inverse conditional covariance matrix of the responses given the remaining concurrent variables. For instance, for $b \cup c$, the response variables are $X_{\{i\}}$ with $i \in b$ and concentrations involving only response variables are the concentrations of $X_{\{i\}}$ and $X_{\{j\}}$ with $i \in b$ and $j \in b \cup c$, then the last claim follows since

$$\mathrm{Det}\begin{pmatrix} \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{cb} & \Sigma_{cc} \end{pmatrix}^{-1} = \mathrm{Det}\begin{pmatrix} \Sigma^{bb.a} & \Sigma^{bc.a} \\ \Sigma^{cb.a} & \Sigma^{cc.a} \end{pmatrix} = \mathrm{Det}(\Sigma^{-1}_{bb.c})\,\mathrm{Det}(\Sigma^{-1}_{cc})$$

and the elements of Σ_{cc} remain fixed with no restrictions on the corresponding concentrations, on those in $\Sigma^{cc.a}$.

The arguments in this section can be directly extended to dependence chains with J elements, which specify corresponding J sets of concurrent variables and J sets of concurrent influences: for the chain $\mathcal{C} = (\mathcal{C}_1, \ldots, \mathcal{C}_J)$ the concurrent variables are $\mathcal{C}^{(j)} = \mathcal{C}_j \cup \ldots \cup \mathcal{C}_J$ for $j = 1, \ldots, J$ while the concurrent influences for $\mathcal{C}^{(j)}$ are $\mathcal{C}^{(j+1)}$ except for $\mathcal{C}^{(J)}$, where it is the empty set. This leads to the following proposition.

Proposition 4.1. *The generalized residual variance in incomplete block-recursive regressions equations (3.3) and (3.4) with J blocks is minimized with J separate covariance selections on covariance matrices of the corresponding*

J sets of concurrent variables: for each given set of concurrent variables no zero concentrations occur for pairs in which both variables are concurrent influences.

In the case where the variables have a joint normal distribution, the estimates of equation parameters obtained by minimizing the generalized residual variance coincide with maximum-likelihood estimates.

4.3 Graphical chain models for continuous variables

Graphical chain models were introduced by Lauritzen and Wermuth (1984, 1989) as a class of models for the analysis of relationships among variables, some of which are qualitative and some quantitative. They are sometimes called CG-*chain models* to remind us that conditional Gaussian regressions are assumed for the conditional distributions of the response variables in the chain.

The investigated relations are dependencies, i.e. directed associations, or symmetric associations taking into account that for many substantive research questions some variables are regarded as responses, some as potential explanatory variables, i.e. influences, some as being intermediate, i.e. playing the role of both influences and responses and some as being on equal footing so that symmetric associations are of interest. In a model with a dependence chain $\mathcal{C} = (a, b, c)$, X_a contains responses, X_b intermediate variables and X_c influences; variables within each of the subsets a, b, c are viewed as being on equal footing.

For CG-chain models with only continuous variables the joint distribution is multivariate normal. Its density f can be written in terms of so-called canonical characteristics (g, h, K) as

$$\log f(x) = g + h^T x - \tfrac{1}{2} x^T K x, \tag{4.1}$$

where

$$h = \Sigma^{-1}\mu, \quad K = \Sigma^{-1}, \quad g = \text{a normalising constant.}$$

Corresponding to the chain $\mathcal{C} = (a, b, c)$, and to the three sets of concurrent variables, those with indices in $a \cup b \cup c$, $b \cup c$, and c, the joint density is factorized into (joint) conditional and (joint) marginal normal densities as

$$f = f_{a|bc}\, f_{b|c}\, f_c.$$

Each reduced model results from restrictions specified with a set

$$\breve{I} \subseteq \tilde{I} = \{(r,s)) \mid 1 \leq r \neq s \leq q\} \tag{4.2}$$

such that for $(r,s) \in \breve{I}$ we have $X_{\{r\}}$ conditionally independent of $X_{\{s\}}$ given all of their remaining concurrent variables. Such a model has been called a graphical chain model since the set $\breve{I}$ can be interpreted as a set of missing edges in an associated chain graph.

The relevant canonical parameters in a graphical chain model for only continuous variables having dependence chain $\mathcal{C} = (a, b, c)$ are identical to the parameters of a block-recursive system of concentration equations (5). This can be seen from the canonical parameters of the former, from

$$\Sigma^{aa}, \Sigma^{ab}, \Sigma^{ac}, h^a; \quad \Sigma^{bb.a}, \Sigma^{bc.a}, h^{b.a}; \quad \Sigma^{cc.ab}, h^{c.ab}$$

obtained from the canonical parameters in joint normal distributions of the three sets of concurrent variables where

$$\begin{pmatrix} h^a \\ h^b \\ h^c \end{pmatrix} = \Sigma^{-1} \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix}, \quad \begin{pmatrix} h^{b.a} \\ h^{c.a} \end{pmatrix} = \begin{pmatrix} \Sigma^{bb.a} & \Sigma^{bc.a} \\ \Sigma^{cb.a} & \Sigma^{cc.a} \end{pmatrix} \begin{pmatrix} \mu_b \\ \mu_c \end{pmatrix}, \quad h^{c.ba} = \Sigma^{cc.ab} \mu_a \,,$$

as well as from the parameters of the latter being

$$B^* = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} & \Sigma^{ac} \\ 0 & \Sigma^{bb.a} & \Sigma^{bc.a} \\ 0 & 0 & \Sigma^{cc.ab} \end{pmatrix}, \quad B^* \mu = \begin{pmatrix} h^a \\ h^{b.a} \\ h^{c.ab} \end{pmatrix}.$$

Thus, a graphical chain model with $\breve{I} \neq \emptyset$ just gives an incomplete system of block-recursive concentration equations with $I = \breve{I}$, and conversely. From

such considerations the following result can be derived for dependence chains with any number of elements.

Proposition 4.2. *The class of block-recursive regressions for variables having a joint normal distribution is equivalent to the class of graphical chain models for continuous variables.*

Important general results are available for this class: as for univariate recursive systems it is known how to read off a graph all independencies implied by a reduced model and how to judge from the graphs whether two distinct models are equivalent (Frydenberg, 1990). This permits, in particular, to decide from the graph of each reduced model whether it is equivalent to a univariate recursive system or not.

5. Linear structural equations

5.1 Derivation from multivariate regression equations

After the need for simultaneous analyses in some economic applications had been demonstrated (Havelmoo, 1943) linear structural equations were developed to deal with such situations and discussed mainly within the econometric literature (see, e.g., Goldberger, 1964).

In order to relate linear structural equations to block-recursive regression equations we describe first their formal relation to multivariate regression equations for X_d given $X_c = x_c$, where X_c may be fixed or random

$$X_d = \Pi_{d|c} x_c + V_d. \tag{5.1}$$

Linear structural equations can be viewed as resulting from (5.1) by premultiplying the residual matrix V_d with an *arbitrary nonsingular matrix* Γ^*_{dd} (compare with Goldberger, 1964; p. 297) giving

$$\Gamma^*_{dd}(X_d - \Pi_{d|c} x_c) = U^*_d.$$

In this context the multivariate regression equations (5.1) are called the *reduced form* of the linear structural equations – not to be confused with the notion of a reduced model.

It is common to use a normalized version obtained by dividing each of the equations by the corresponding diagonal element of Γ^*_{dd}. This defines a matrix Γ_{dd} having ones along the diagonal and gives the *linear structural equations*

$$\Gamma_{dd}X_d + \Gamma_{dc}x_c = U_d \tag{5.2}$$

with equation parameters

$$\Gamma_{dd}, \quad \Gamma_{dc} = -\Gamma_{dd}\Pi_{d|c}$$

and covariance matrix

$$\mathrm{Var}(U_d) = \Omega_{dd} = \Gamma_{dd}\Sigma_{dd.c}\Gamma^T_{dd}.$$

5.2 Properties of some linear structural equations

It is well known (Goldberger, 1964, p.316) that without further restrictions the equations (5.2) are over-parametrized so that not all of its parameters can be estimated from the observed covariance matrix. In the econometric literature such systems are called 'not identifiable'. We prefer to say that they do not define a statistical model.

Concern had been voiced early (Liu, 1960) that the necessity to introduce restrictions on the structural equations (5.2) to remove overparametrization could lead to misspecifications with all the undesirable consequences. In particular, the interpretation of equation parameters and of association structures defined with linear structural equations depends crucially on the chosen set of restrictions.

To elaborate on this in relatively simple situations we consider just two equations in four variables, i.e. the case with $X_d = (Y_1\ Y_2)^T$ and $X_c =$

$(Y_3\ Y_4)^T$. The equations (5.2) can then be written as

$$y_1 + \gamma_{12}y_2 + \gamma_{13}y_3 + \gamma_{14}y_4 = u_1, \qquad (5.3)$$

$$y_2 + \gamma_{21}y_1 + \gamma_{23}y_3 + \gamma_{24}y_4 = u_2,$$

with equation parameters

$$\Gamma_{dd} = \begin{pmatrix} 1 & \gamma_{12} \\ \gamma_{21} & 1 \end{pmatrix}, \quad \Gamma_{dc} = \begin{pmatrix} \gamma_{13} & \gamma_{14} \\ \gamma_{23} & \gamma_{24} \end{pmatrix}$$

and with covariance matrix

$$\mathrm{Var}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ . & \omega_{22} \end{pmatrix}.$$

Two appropriately chosen restrictions are necessary to express the nine distinct parameters of (5.3) uniquely in terms of the seven free parameters of the unrestricted covariance matrix Σ. Thereby the three distinct elements of the covariance matrix of X_c are taken as given in this conditional model.

A saturated statistical model, i.e. a *complete system of linear structural equations* results from (5.3) if restrictions have been specified which remove over-parametrization but do not introduce restrictions on the covariance matrix. Such systems are referred to as being 'both identifiable and simple' in the econometric literature (Malinvaud, 1966; p. 561).

As examples for possible different saturated models and associated different meanings of equation parameters in linear structural equations three different cases are discussed.

Case A: Saturated multivariate regression equations result from (5.3) with the restrictions $\gamma_{12} = \gamma_{21} = 0$. The equation parameters are then

$$\Gamma_{dd} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Gamma_{dc} = \begin{pmatrix} -\beta_{13.4} & -\beta_{14.3} \\ -\beta_{23.4} & -\beta_{24.3} \end{pmatrix}.$$

Case B: Saturated univariate recursive regression equations result from (5.3) with the restrictions $\gamma_{21} = \omega_{12} = 0$. The equation parameters are then

$$\Gamma_{dd} = \begin{pmatrix} 1 & -\beta_{12.34} \\ 0 & 1 \end{pmatrix}, \quad \Gamma_{dc} = \begin{pmatrix} -\beta_{13.24} & -\beta_{14.23} \\ -\beta_{23.4} & -\beta_{24.3} \end{pmatrix}.$$

Case C: A further set of saturated equations results from (5.3) with the restrictions $\gamma_{14} = \gamma_{23} = 0$. The equation parameters are then

$$\Gamma_{dd} = \begin{pmatrix} 1 & -\frac{\beta_{14.3}}{\beta_{24.3}} \\ -\frac{\beta_{23.4}}{\beta_{13.4}} & 1 \end{pmatrix}, \Gamma_{dc} = \begin{pmatrix} -(\beta_{13.4} - \gamma_{12}\beta_{23.4}) & 0 \\ 0 & -(\beta_{24.3} - \gamma_{21}\beta_{14.3} \end{pmatrix}. \tag{5.4}$$

This is a saturated model with parameters which are difficult to interpret and which are defined only for covariance matrices having $\sigma_{24.3} \neq 0$ and $\sigma_{13.4} \neq 0$. But, it has further undesirable features.

One is that a single zero restriction on an equation parameter can introduce more than one restriction on the reduced form and hence on the covariance matrix. For instance, setting $\gamma_{24} = 0$ in (5.4) implies $\beta_{14.3} = \beta_{24.3} = 0$. Another is that the interpretation of a reduced model can remain ambiguous until more is specified about the system. An example is the reduced model obtained from setting $\gamma_{21} = 0$ in (5.4). In the econometric literature such a model is referred to as being 'identifiable but not simple', since its parameters can all be estimated but it implies restrictions on Σ. For this particular reduced model the equation parameters expressed in terms of regression coefficients are

$$\Gamma_{dd} = \begin{pmatrix} 1 & -\frac{\beta_{14.3}}{\beta_{24.3}} \\ 0 & 1 \end{pmatrix}, \quad \Gamma_{dc} = \begin{pmatrix} -\beta_{13.4} & 0 \\ 0 & -\beta_{24.3} \end{pmatrix},$$

while the coefficients in a regression of X_d on X_c expressed in terms of the structural equation parameters are

$$\Pi_{d|c} = \begin{pmatrix} -\gamma_{13} & \gamma_{12}\gamma_{24} \\ 0 & -\gamma_{24} \end{pmatrix}.$$

The latter shows that $\beta_{23.4} = 0$, i.e. that one restriction implied by this reduced model for the covariance matrix is $\sigma_{23.4} = 0$.

An interpretation of this structure depends on the value of one of the

unrestricted parameters in the system, on γ_{12}. More precisely, we get

$$\text{(I)} \quad \sigma_{14.3} = \sigma_{23.4} = 0 \quad \text{for } \gamma_{12} = 0,$$

$$\text{(II)} \quad \sigma_{14.23} = \sigma_{23.4} = 0 \quad \text{for } \gamma_{12} = -\sigma_{12.34}/\sigma_{22.34},$$

$$\text{(III)} \quad \sigma_{23.4} = 0 \quad \text{otherwise.}$$

The first (I) follows directly from the reduced form parameters expressed in terms of the structural parameters, i.e. from $\beta_{14.3} = \gamma_{12}\gamma_{24}$, while the second (II) follows from the recursion formula for regression coefficients, from $\beta_{14.23} = \beta_{14.3} - \beta_{12.34}\beta_{24.3}$, from $\gamma_{12} = -(\beta_{14.3}/\beta_{24.3})$ and from $\beta_{12.34} = \sigma_{12.34}/\sigma_{22.34}$.

Now (I) and (II) specify structures which are disjoint for all covariance matrices having $\sigma_{12.34} \neq 0$ and $\sigma_{24.3} \neq 0$. Consequently, for such cases the change in the interpretation from (I) to (II) means a change between two mutually exclusive hypotheses.

No such problem occurs if the value of γ_{12} is known, but then we are back in one of the subclasses (Cases A or B) of linear structural equations, having well-defined parameters since (I) can be seen as a reduced multivariate regression model, (II) as a reduced univariate recursive regression model, (III) can be specified as a zero restriction model within either one of these two subclasses.

The two examples to Case C suggest that it may be difficult to judge for any given substantive research question, whether a linear structural equation model provides an appropriate framework: unless one has extremely strong prior knowledge or the model belongs to a well-studied subclass like univariate recursive systems or multivariate regression the interpretation of its parameters and of the whole structure has to be derived from scratch in each application.

Interesting theoretical questions are in general unanswered for linear structural equations as well, such as:

- Are there other subclasses of linear structural equation models than univariate recursive regressions and multivariate regressions in which the parameters relate in a simple way to regression coefficients?

- When are two linear structural equation models equivalent?
- Which moderation effects are to be expected for single parameters and for the interpretation of a structure if one marginalises over a variable in the structural equations or if a variable is not observed?

5.3 Distinctions and relations to block-recursive equations

Block-recursive linear regression equations can be derived from the multivariate regression model (5.1) in a similar way as the linear structural equations (5.2). Block-recursive concentration equations result from (5.1) by premultiplying the residual matrix V_d with a *positive definite matrix* B^*_{dd} being determined from a block-triangular decomposition of the concentration matrix of the residuals V_d, i.e. from a block-triangular decomposition of $\Sigma^{-1}_{dd.c}$.

More precisely, let $d = a \cup b$, then the unique upper block-triangular decomposition of $\Sigma^{-1}_{dd.c}$ corresponding to the partitioning (a, b) of d is

$$\Sigma^{-1}_{dd.c} = A^T_{dd} T^{-1}_{dd} A_{dd} \tag{5.5}$$

with

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

Then the equations obtained from $B^*_{dd} V_d$ with $B^*_{dd} = T^{-1}_{dd} A_{dd}$ are identical to the equations for X_i with $i \in d$ in (3.2). This follows with

$$B^*_{dd} V_d = B^*_{dd}(X_d - \Pi_{d|c} x_c) = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} \\ 0 & \Sigma^{bb.a} \end{pmatrix} \begin{pmatrix} X_a \\ X_b \end{pmatrix} + \begin{pmatrix} \Sigma^{ac} \\ \Sigma^{bc.a} \end{pmatrix} X_c$$

since

$$-B^*_{dd}\Pi_{d|c} = (T^{-1}_{dd} A_{dd})(\Sigma^{dd})^{-1}\Sigma^{dc} = (A^{-T}_{dd}\Sigma^{-1}_{dd.c})(\Sigma_{dd.c}\Sigma^{dc}) = A^{-T}_{dd}\Sigma^{dc}$$

and

$$A^{-T}_{dd}\Sigma^{dc} = \begin{pmatrix} I_{aa} & 0 \\ -\Sigma^{ba}(\Sigma^{aa})^{-1} & I_{bb} \end{pmatrix} \begin{pmatrix} \Sigma^{ac} \\ \Sigma^{bc} \end{pmatrix} = \begin{pmatrix} \Sigma^{ac} \\ \Sigma^{bc.a} \end{pmatrix}.$$

The total set of concentration equations in (3.2) is then obtained by simply adding a complete system of one block for Σ_{cc} if X_c is random and regression equations result from (3.2) as described in Section 3.2. Thus, block-recursive regressions can be viewed as linear structural equations in which a particular meaning is attached to the equation parameters.

Considerations of this type lead to the following proposition.

Proposition 5.1. *Linear structural equations are equivalent to block-recursive regression equations if and only if both of the following conditions hold:*
(1) the structural equations can be reordered to get Γ_{dd} to be of upper block-diagonal form so that the blocks define sets of concurrent variables and
(2) each nonzero off-diagonal element in row i and column j of Γ_{dd} is the partial regression coefficient of X_j in a linear regression of X_i on X_j and on all of their remaining concurrent variables.

To summarize some of the distinctions and similarities between structural equations and block-recursive equations we note first that both

- can be viewed as extensions of Wright's path analysis to study structures more complex than univariate recursive equations;
- contain univariate recursive systems as a subclass;
- permit to formulate different types of nondecomposable independence hypotheses;
- can be extended to include latent variables.

We note further that for block-recursive regression equations

- there is no overparametrization and, consequently, there are no problems of identification;
- each equation parameter is a regression coefficient and hence interpretations of a structure in terms of linear independencies can be directly

derived and general results are available to read directly off the graph of a model which independencies are implied;

- covariance selection models form a subclass, but multivariate regressions with nondecomposable independence hypotheses have no simple equivalent formulation as block-recursive equations, i.e. they cannot be conveniently expressed in terms of zero restrictions on parameters of any block-recursive equations.

Similarly, we note that for linear structural equations

- multivariate regressions form a subclass but covariance selection models with nondecomposable independence hypotheses have no simple equivalent formulation as linear structural equations;

- no general results are available which help in the interpretation of structures defined with linear structural equations unless the equations are at the same time either univariate recursive regressions or multivariate regressions.

For characterizations of nondecomposable independence hypotheses within the classes of covariance selection, block-recursive regressions and multivariate regressions and for discussions of how they could have been generated by univariate recursive equations in the observed variables *and* some appropriately chosen latent variables see Wermuth and Cox (1991).

6. A substantive research question

A research question from psychological research requiring a simultaneous analysis of linear equations concerns emotions as a disposition or trait of a person and emotions as a state evoked by particular situations. These notions are central to research on stress and on strategies to cope with stressful events.

Questionnaires with which state and trait of the emotions anxiety and anger are to be measured have been developed by Spielberger et al. (1970, 1983).

Associations of a linear type are considered to be appropriate descriptions of pairwise relations between the variables Y_1 := State anxiety, Y_2 := State anger, Y_3 := Trait anxiety, and Y_4 := Trait anger. Expectations regarding the correlation structure of the four variables are as follows:

- all marginal correlations are positive and of moderate size, while no partial correlation is negative;
- emotions in particular situations (states) are influenced by the dispositions (traits) of a person and not conversely;
- if either state variable is predicted in terms of the other three variables then there is no direct dependence on the other trait variable.

The hypotheses concerning only indirect dependencies can be stated in terms of conditional expectations as

$$\mathrm{E}(Y_1 \mid Y_2 = y_2, Y_3 = y_3, Y_4 = y_4) = \alpha_{1.234} + \beta_{12.34}y_2 + \beta_{13.24}y_3 + \beta_{14.23}y_4,$$

$$\mathrm{E}(Y_2 \mid Y_1 = y_1, Y_3 = y_3, Y_4 = y_4) = \alpha_{2.134} + \beta_{21.34}y_1 + \beta_{23.14}y_3 + \beta_{24.13}y_4$$

with

$$\beta_{14.23} = \beta_{23.14} = 0.$$

From C. Spielberger we obtained data displayed in Table 6.1 on 684 female college students for the state-trait versions of the variables anxiety and anger. They show an excellent fit of the observations to the hypothesized structure: the partial correlations of (Y_1, Y_4) and (Y_2, Y_3) have observed values close to zero, while they are fixed to be zero under the reduced model. The marginal correlations except for the two pairs (Y_1, Y_4) and (Y_2, Y_3) are – under this reduced model – forced to agree with the observed ones. For the two pairs where deviations between the observed and estimated correlations can occur, these deviations are small, i.e. we have observed $\hat{\rho}_{14} = .39$, $\hat{\rho}_{23} = .47$ and we have estimated under the model $\tilde{\rho}_{14} = .42$, $\tilde{\rho}_{23} = .45$.

Table 6.1

Observed marginal correlations (lower half) and observed partial correlations given all remaining variables (upper half), and further data summaries, $n = 684$

Variable	Y_1 State anx	Y_2 State ang	Y_3 Trait anx	Y_4 Trait ang
Y_1:= State anxiety	1	.45	.47	**−.04**
Y_2:= State anger	.61	1	**.03**	.32
Y_3:= Trait anxiety	.62	.47	1	.32
Y_4:= Trait anger	.39	.50	.49	1
Mean	18.87	15.23	21.20	23.42
Standard Deviation	6.10	6.70	5.68	6.57

Table 6.2

Test results for zero partial correlation of each variable pair given all of its remaining concurrent variables

Indices of pairs	Indices of concurrent variables	Value of chi-square statistic	Degrees of freedom	Corresponding fractile or p-value
(1,2)	1234	153.90	1	<. 001
(1,3)	1234	171.51	1	<. 001
(1,4) :=(S-anx, T-ang)	1234	1.22	1	**.268**
(2,3) :=(S-ang, T-anx)	1234	0.33	1	**.572**
(2,4)	1234	78.04	1	< .001
(3,4)	34	189.73	1	< .001

These impressions are supported on the one hand by the overall likelihood-ratio test for the goodness-of-fit since the value of likelihood ratio chi-square statistic on 2 degrees of freedom is 2.1 and on the other hand by the tests statistics reported in Table 2 for tests of zero partial correlation of each pair taken separately. The conclusions about the association structure of these four variables were confirmed by very similar unpublished results for a sample of 588 male college students and they could not have been formulated conveniently by zero restrictions on individual parameters of traditional linear structural equations.

Acknowledgements

I want to thank Sir David Cox, Steffen Lauritzen, Christel Jennen-Steinmetz, and referees for their most helpful comments on earlier versions of this work.

(Received May 1991. Revised October 1991.)

References

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

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

Cox, D.R. (1991) Causality; some statistical aspects. *J. Roy. Statist. Soc. Ser. A*, **155**, 291-301.

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

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

Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Reading: Addison-Wesley.

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

Edwards, D. (1990). Hierarchical mixed interaction models (with discusssion). *J. Roy. Statist. Soc. Ser. B* **52**, 3-20, 51-72.

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

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

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

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

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. (1984). Mixed interaction models. Res. Rep. R-84-8. Institute of Electronic Systems, Aalborg University.

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

Liu, T.-C. (1960). Underidentification, structural estimation, and forecasting. *Econometrica* **28**, 855-865.

Malinvaud, E. (1966). *Statistical Methods of Econometrics.* Amsterdam: Rand McNally.

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

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

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 Lushene, R.E. (1970). Manual for the State-Trait Anxiety Inventory. Palo Alto: Consulting Psychologists Press.

Spielberger, C.D. Russell, S. and Crane, R. (1983). Assessment of anger. In J.N. Butcher and C.D. Spielberger (eds.) *Advances in Personality Assessment.* (Vol. 2), 159-187. Hillsdale: Erlbaum.

Tukey, J.W. (1954). Causation, regression, and path analysis. In: O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush (eds.). *Statistics and Mathematics in Biology*, 35-66. Ames: The Iowa State College Press.

Verma, T.S. and Pearl, J. (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In: Proceedings, 8th. Conference on Uncertainty in Artificial Intelligence, Stanford, 323-330. San Mateo: Morgan Kaufman.

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* **III**, 74-93.

Wermuth, N. and Cox, D.R. (1991). Explanations for nondecomposable hypotheses in terms of univariate recursive regressions including latent variables. Berichte zur Stochastik und verwandte Gebiete 91-2, University of Mainz

Wermuth, N. and Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discusssion). *J. Roy. Statist. Soc. Ser. B* **52**, 21-72.

Wermuth, N. and Scheidt, E. (1977). Fitting a covariance selection model to a matrix, Algorithm 105. *J. Roy. Statist. Soc. Ser. C* **26**, 88-92.

Wilks, S.S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24**, 471-494.

Wold, H.O. (1954). Causality and econometrics, *Econometrica* **22**, 162-177.

Wright, S. (1921). Correlation and causation. *J. Agricult. Res.* **20**, 557-585.

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. *Ann. Math. Statist.* **5**, 161-215.

DISCUSSION

Gerhard Arminger

Bergische Universität Gesanthochschule Wuppertal
Wuppertal - Germany

Ms. Wermuth is to be congratulated on clarifying the similarities and distinctions of two important models used in statistical analysis, that is between graphical chain models for quantitative variables on one hand and the linear structural models on the other hand. The central contribution is the introduction of block recursive regression equations and showing that the block recursive structure can be obtained from a linear structural equation system by restricting the matrix Γ^*_{dd} of the structural equation system to be positive definite and to be a function of the covariance matrix $\sum_{dd.c}$ of the residuals V_d. Consequently, she can derive the conditions under which one or the other model may be more meaningful in applications. Altogether I have found the paper quite stimulating and it has given me a new outlook on the many possibilities of modeling a covariance structure.

I will first consider some of the differences between the approaches which are more of a philosophical nature but may also help in understanding when one should use one or the other approach. Second, some of the statistical assumptions that seem to me overly restrictive in the graphical chain model for estimation purposes are discussed.

The basic approach of linear structural models written as

$$\tilde{y} = \tilde{\alpha} + \tilde{B}\tilde{y} + \tilde{\Gamma}\tilde{x} + \tilde{\epsilon}\,,$$

is to concentrate on modelling the effects of the endogenous variables $\tilde{y}$ on each other, captured in the matrix $\tilde{B}$ and the effects of $\tilde{x}$ on $\tilde{y}$, captured in the matrix $\tilde{\Gamma}$. The disturbances are supposed to have $E\tilde{\epsilon} = 0$ and $V(\tilde{\epsilon}) = \tilde{\Omega}$ and to be uncorrelated with $\tilde{x}$. (The notation "$\sim$" is used to avoid confusion with Wermuth's notation). Of course, if $\tilde{B} = 0$, the model corresponds to multivariate regression and if $\tilde{B}$ is lower triangular, it corresponds to a recursive system, that is, path analysis. However, certain hypotheses in economics and other fields, often derived from a dynamic equation system equilibrium, make the formulation of $\tilde{B}$ as a general matrix where $(I - \tilde{B})$ is invertible and meaningful. The estimation of $\tilde{\alpha}, \tilde{B}, \tilde{\Gamma}$ and $\tilde{\Omega}$ can only be based on the reduced form:

$$\tilde{y} = (I - \tilde{B})^{-1}\tilde{\alpha} + (I - \tilde{B})^{-1}\tilde{\Gamma} + (I - \tilde{B})\tilde{\epsilon} = \alpha^* + \Pi^* x + \epsilon^*$$

with

$$V(\epsilon^*) = (I - \tilde{B})^{-1}\tilde{\Omega}(I - \tilde{B})^{-1^T} = \tilde{\Sigma} .$$

Given the information by the mean structure $E(\tilde{y}|x) = \alpha^* + \Pi^*\alpha^* + \Pi^*\tilde{x}$ and the conditional covariance structure $V(\tilde{y}|x) = V(\epsilon^*) = \tilde{\Sigma}$, the parameters $\tilde{\alpha}, \tilde{B}, \tilde{\Gamma}$ and $\tilde{\Omega}$ have to be restricted in one way or the other; these restrictions are called identification restrictions. The restrictions actually used will of course depend on substantive issues, but one should note that restrictions on any of the parameters $\tilde{\alpha}, \tilde{B}, \tilde{\Gamma}$ and $\tilde{\Omega}$ can be used. In applications, one is usually interested in what happens in $\tilde{y}_i$, if $\tilde{y}_j$ or $\tilde{x}_l$ changes by one unit. Hypotheses about $\tilde{\Omega}$ or about the joint unconditional covariance matrix of $\tilde{y}$ and $\tilde{x}$ are usually not of interest. Therefore restrictions are typically placed on $\tilde{\alpha}, \tilde{B}$ and $\tilde{\Gamma}$ rather than on $\tilde{\Omega}$ or on $V(\tilde{y})$.

Also, if one uses structural modelling, one is usually not interested in independence or conditional independence of variables and one is not willing to assume that the variables in $\tilde{x}$ or in $\tilde{\epsilon}$ are (multivariate) normal. The focus in estimating simultaneous equation systems is therefore to estimate $\tilde{\alpha}, \tilde{B}, \tilde{\Gamma}$ under the weakest possible assumption about $\tilde{\epsilon}$ which turns out to be that

$E(\tilde{\epsilon}|x) = 0$ given first order identification, that is $E(\tilde{y}|\tilde{x},\theta_1) = E(\tilde{y}|\tilde{x},\theta_2)$ implies $\theta_1 = \theta_2$ for all $\tilde{x}$, where θ_1 and θ_2 are two parameters vectors. This can be achieved for instance by using the pseudo ML method of Gourieroux, Monfort and Trognon (1984).

The goal of graphical modelling is quite different. The central idea is not the estimation of a multivariate mean structure under certain hypotheses as in econometrics but the idea of (conditional) independence which in the multivariate normal distribution is the same as zero (partial) correlation. Hence, the modeler is concerned with structures defined with respecto to the inverse of the covariance matrix of all variables that are considered. The interesting fact is, however, that path analysis can be derived from the linear structural model as well as the graphical chain model, although the distributional assumptions are weaker for the estimation from the structural model yet. The identification issue that arises in the linear structural model also arises in the graphical chain model but it is solved in a prescribed way. This is immediately seen from the derivation of the matrix B^*_{dd} from the unique upper block-triangular decomposition of $\sum^{-1}_{dd.c}$ which is equal to $(I - \tilde{B})^T\tilde{\Omega}^{-1}(I - \tilde{B})$ in the linear structural model. While in the linear structural model $\tilde{B}, \Gamma$ and $\tilde{\Omega}$ are first modelled independently of each other and then restrictions are used to achieve identification, the matrix B^*_{dd} in the graphical model is a bijective function of the unique decompositon of $\sum^{-1}_{dd.c}$ and therefore identified.

In Wermuth's paper the estimation issue is not discussed. Judging from the introduction, section 4.2. and section 4.3 and the example, the estimation is based on the assumption that the joint distribution of all quantitative variables in the model is normal. Such an assumption, however, would be unacceptable for econometrics and many other fields. However, if one is willing to give up the assumption of normality and instead of looking at (conditional) independence considers zero (partial) correlation one can estimate graphical chain models using either pseudo ML methods for mean and covariance structures (Arminger and Sobel, 1990) or a minimum distance estimator (MDE)

for covariance matrices discussed by Shapiro (1986).

Here, I will consider only briefly the minimum distance estimation. Let S the empirical covariance matrix of a vector y and let s be the vectorized lower triangular of S. For samples of size n, it is assumed that s is asymptotically multivariate normal, that is $\sqrt{n}(s - \sigma(\theta)) \overset{A}{\sim} \mathcal{N}(0, W)$ where $\sigma(\theta)$ is the vectorized form of the expected covariance matrix $\sum(\theta)$ and θ is the vector of model parameters of $\sum$. The matrix W is the variance-covariance matrix of $\sqrt{n}s$. The parameter vector θ is estimated in such a way that the Mahalanobis distance

$$Q(\theta) = n(s - \sigma(\theta))^T \hat{W}^{-1}(s - \sigma(\theta))$$

is minimized. $\hat{W}$ is a consistent estimate of W which can be obtained from the first four empirical moments of y. As shown by Shapiro (1986), $Q(\theta)$ is asymptotically χ^2 distributed with $p - q$ degrees of freedom if $\sum(\theta)$ is correctly specified; p is the number of elements of s and q is the number of free parameters in θ. The asymptotic distribution of the MDE θ is then:

$$\sqrt{n}(\hat{\theta} - \theta) \overset{A}{\sim} \mathcal{N}\left(0, \left(\frac{\partial\sigma(\theta)}{\partial\theta} W^{-1} \frac{\partial\sigma(\theta)}{\partial\theta^T}\right)^{-1}\right) .$$

To embed the block recursive models discussed by Wermuth we note that hypotheses formulated in B of equation (3.3) are equivalent to hypotheses formulated in $B^\star$ in equation (3.2) and that hypotheses about $B^\star$ can be formulated as hypotheses about $\sum^{-1}$ as shown in equation (3.1) or (3.5). Hence, one can define a vector of parameters θ as the vectorized lower triangular of $\sum^{-1}$ and write $\sigma(\theta)$ as a possibly fairly complicated function of the elements of $\sum^{-1}$ which may be restricted.

Programming such a function is much simplified of one uses a powerful matrix language such as GAUSS. In fact, a general program for mean and covariance structure analysis like MECOSA (Schepers and Arminger 1992) which is written in GAUSS may be used directly to estimate and test block recursive equation models such as the one used in the example.

References

Arminger, G. and Sobel, M.E. (1990). Pseudo Maximum Likelihood Estimation of Mean and Covariance Structures With Missing Data. *Journal the American Statistical Association.* **85**, 195-302.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica* **52**, 681-700.

Schepers, A. and Arminger, G. (1992). MECOSA - *User Guide: A Program for the Analysis of Mean and Covariance Structures with Non-Metric Dependent Variables*, Bergische Universität Wuppertal, Federal Republic of Germany.

Shapiro, A. (1986). Asymptotic Theory of Overparameterized Structural Models. *Journal of the American Statistical Association* **81**, 142-149.

D.R. Cox

Nuffield College
Oxford - U.K.

Traditional multivariate analysis of continuous variables is largely based on means and covariance matrices, i.e. to some extent implicitly or explicitly on the multivariate normal distribution. There are broadly at least three limitations to such methods, two of them somewhat opposing one another. First the methods are frequently invariant under rather large families of linear transformations of the component variable, and this may be in conflict with specific information about the nature of the individual variables preservation of whose identity is important. The second limitation is that, especially when the number p of component variables is large, the specification in terms of $p + \frac{1}{2}p(p+1)$ parameters may be too highly parameterized, i.e. any structure in the covariance matrix may be hard to discern from among so many entries.

Finally, somewhat, by contrast with the previous point, the methods do not allow for interdependencies that are nonlinear or "interactive", for example

for the possibility that the slope of the relation between Y_1 and Y_2 depends on the levels of other variables. (Of course we can study such effects by regression analysis but that requires a separation into response and explanatory variables).

Professor Wermuth's paper addresses the first and second of these points and I know that in other work she is concerned about the final one, the possibility of nonlinerarity and other such effects.

One value of the paper lies in establishing the relation between important families of models that are at least partly substantively based and further in the results towards the end of the paper which show clearly the difficulties that can arise in interpreting the parameters in linear structural systems.

The use of the more substantively based models described in the paper is likely in general to be preferable to other more empirical ways of dealing with an over-abundance of parameters. A more empirical approach involves a search for such regularities as (a) sets of essentially zero correlations, (b) blocks of variables with essentially equal correlation within each block and essentially equal cross-correlation between all pairs of variables in two given blocks, (c) simplifications based on the inverse covariance matrix, corresponding particularly to (a), the detection of zeros. This isolates zero partial correlations of pairs of variables given all other variables, the covariance-selection method of Dempster. While at first sight the use of the inverse matrix is less direct than the use of marginal correlations, i.e. the correlation matrix itself, the inverse has the advantage of being linked with notions of conditional independence and therefore is often more likely to point towards specific models with a substantive interpretation, in particular via the class of graphical chain models, and hence to the whole range of ideas discussed in this paper.

Jan de Leeuw

Departments of Mathematics and Psychology
University of California
Los Angeles - U.S.A.

Graphical Models, etc.

The theory of hierarchical models, decomposable models, graphical models, chain models, or path models, has changed, and will continue to change, the face of multivariate statistical analysis. The accomplishments so far have been impressive. There are important and convincing systematizations of early work by Wright, Goodman, and Dempster in the writings of Wermuth, Speed, Lauritzen, and Cox. The contributions of Pearl, Shafer, Glymour and others provide rich connections with the field of artificial intelligence, and with computational philosophy of science.

Of course such a glowing introduction must necessarily be followed by words of criticism and/or caution. We concentrate on the practical usefulness of graphical model techniques in actual data analysis.

Weak Spots

The theory developed so far is not complete. First, the connections with econometric simultaneous equation theory, and with the related LISREL/EQS systems in psychometrics and sociometrics, have not been worked out in detail. Graphical chain models correspond with (block-)recursive path models with uncorrelated errors. The errors are the only latent (unobserved) variables. In the simultaneous equations models in the Haavelmo tradition there is no longer simple recursiveness. Wermuth's paper, the one we are discussing here, makes an attempt to close this first gap. Her point of view is that block recursive

regression equations are the key to understanding the relationship between the approaches. She only discusses "errors-in-equations" models, and we shall argue that these models are not the most interesting generalization of simple path models. It also seems to me that quite a few of the attractive properties of block models get lost in the generalization. This means that the theory is not very robust.

Second, one of the satisfactory aspects of these models is that categorical and numerical variables can be dealt with in one and the same formulation. But this can only be done by assuming multivariate normality for the numerical variables, and then translating conditional independence into vanishing partial correlations. Although parts of the theory can be based on partial correlations only, we lose the connection with maximum likelihood, which requires a full specification of the model. Unfortunately, real variables are not either multinomial or multinormal. They are most often, in my experience, somewhere in between these extremes. Multinomial allows for too many parameters, multinormal for not enough parameters. We shall come back to this below.

Thirdly, there is a gap between theory and practice. In quantitative genetics, in econometrics, in educational science, and in sociology, path models and simultaneous equation models have been, let's say, disappointing. The collective works of David Freedman illustrate this more forcefully than I ever could. Having graphical models is technologically a big step ahead, but we should be careful not to fall into the LISREL trap. A generation of social scientists has been misguided by an appealing metaphor, wrapped in an authoritarian black box, decorated with forbidding equations. As far as I can see, the outcome has been, let's say, disappointing.

Linear Structural Equations

The material in the first four parts of Wermuth's paper is an excellent and useful review of the various facts and results that are known about block recursive path models. The fifth part is an original comparison of block recursive models with general simultaneous or structural equation models. This part we shall discuss somewhat more in detail.

In the full model, not all parameters can be identified from the first and second order moments. Wermuth (p.18) seems to equate "not identifiable" with "not defining a statistical model". But this is misleading, because the model defines a unique manifold in the space of covariance matrices, even if this manifold is not described with a minimal number of parameters. Here, as in the linear case, it helps to think of the parametrization as just incidental, and of the subset of the covariance cone as the model we are trying to estimate. This also takes away the fear that identification could lead to misspecification. If it does, it is by definition not just identification. Identifying a model means describing it in terms of the smallest number of parameters possible. Also, giving the impression that identification problems come from overparametrization (p.22) is somewhat misleading. They tend to come from underparametrization, of course.

I am not an econometrician, but were I one, I would be somewhat unhappy about Wermuth's account of simultaneous equation modeling. The modern econometric literature on identification is not even mentioned. Hsiao (1983) gives a fairly recent overview, while even more recent material is reviewed in Bekker and Pollock (1986) or Bekker and Dijkstra (1990). There are now symbolic computation methods to investigate identification and equivalence of simultaneous equation models.

Block recursive models were already used as basis for discussing indentification by Fisher in his classical book (1966). There have also been interesting discussions in the econometric literature (between Wold and Bassman, for ins-

tance) about causal interpretation of coefficients in non-recursive models, as compared to block-recursive models. This anticipates some of the distinctions mentioned on page 22-23. While it is true the LISREL framework has been accepted quite uncritically by many social scientists, it is also true there has been a great deal of critical discussion of the algebraic and methodological properties of simultaneous equation models in the econometric research literature. This is not obvious from Wermuth's paper, because she mainly refers to some excellent, but quite ancient, textbooks.

Latent Variables

Wermuth does not mention latent variables, but it is difficult to see how any discussion of linear structural models can be complete without them. In the situations in which such models are typically applied (with some exceptions in theoretical population genetics) there are errors in variables that cannot be ignored. This is even true in engineering and systems theory, where there currently is a lot of interest for latent variable models. I happen to think that factor analysis, with the closely related models of true score theory, errors in variable theory, and latent trait theory, are the most interesting contributions of the social sciences to data analysis. The fact that these techniques have been misused almost to extinction does not take away this basic fact.

Related to the idea of latent variables is the idea of optimal scaling. A latent variable only exists because of its position in the path diagram. In factor analysis, for intance, we say that there exists a variable z such that the observed y_j are independent given z. We do not observe z, it is missing, but we can still test the consequences of our assumptions. If we want to, we can also "estimate" z.

One step further along the road is the basic indicator idea, which is that we never observe the variables we are interested in. The variables in

the path model are theoretical constructs, we only observe indicators for these constructs. The indicators are related to the constructs by small factor models. This is the seductive metaphor I referred to earlier. It seems the only sensible way to model errors-in-variables, and potentially it is a great way to link theory and observation. Unfortunately, in many applications the theory component is missing, and theories are "constructed" by heuristic search over path models. This is the nightmarish part of latent variable modeling. It is a reasonable class of techniques, much more reasonable than block recursive modeling of the observed variables, but I have not seen many convincing applications, and I have seen quite a few hair-raising ones.

Nevertheless I agree with Goldberger (1972), Griliches (1974), Wold (1982), and Aigner et al. (1984) that there is no way around errors-in-variables, and that the errors-in-equations models in the Haavelmo tradition are simply too narrow to be of interest in most social science situations. It is perhaps true, that block-recursive models are more basic and "a key to understanting". They are elegant, they are simple, but they have little practical relevance because of the omnipresence of measurement error. In a sense this means that the vanishing tetrad is of more importance than the vanishing partial correlation coefficient.

Between Multinomial and Multinormal

Relationships between variables can be pictured as arrow diagrams. These qualitative diagrams can be translated into quantitative statements about the joint distribution of the variables in various ways. One translation uses conditional independence, another uses vanishing partial correlations.

In the path diagrams some variables are quantitative, some are qualitative, some are latent and some are manifest. In a number of special cases we know how to integrate all four types of variables into a single model, and large

steps have been made towards a general approach. I think, however, that in order to build realistic models we need intermediate types of variables. Intermediate both between continuous normal and discrete multinomial, but also intermediate between latent and manifest.

Some of these intermediate types have been studied in probit and logit models, in which binary variables are regressed on continuous latent variables for which they are indicators. If we generalize these ideas, which is done in LISCOMP (1982), then we have a kernel with a path model for the theoretical constructs, and we have nonliner regressions relating the constructs to the indicators. Indicators now can be nominal, ordinal, numerical, binary, truncated, censored, and so on. The basic idea of having conditional independence of the indicators given the construct still applies, and the basic inheritance from factor analysis is that constructs are continuous variables.

Building the likelihood function for such models is not very difficult. A fairly complete review is in the book by Bartholomew (1988). But maximizing the likelihood can be very costly and pratically impossible. New developments (such as Gibbs sampling) continue to make the boundaries of computability wider and wider, but in the end the empty cell problem and the complicated integrals that are bound to appear in the likelihood function often defeat us.

If we want to avoid going to a fully specified model, using likelihood methods, then we can use the alternatives proposed by Gifi (1991) or De Leeuw (1988), (1986). These optimal scaling techniques can perhaps be best understood by using the concept of *linearizable regressions.* Variables can be of mixed type (nominal, ordinal, numerical). We do not make specific distributional assumptions, only that the variables can be transformed in such a way that all bivariate regressions become linear. We then apply an optimal scaling technique (any reasonable one will do), which will recover the linearizing transformations. And we fit the errors-in-variables model to the optimally scaled variables. Statistical theory for this two-step approach is developed in De Leeuw (1988).

Summary

Wermuth's paper limits itself, quite appropriately, to a class of statistical models about which we can say a great deal mathematically. It is unfortunate, but significant, that these models may have considerable normative value but little descriptive value. The example analyzed in Section 6 of the paper is not typical for the types of problems analysed with path analysis. But even for this tiny example the blunt statement that "associations of a linear type are considered to be appropriate descriptions of pairwise relations between the variables" gives rise to many doubts. "By whom ?", for instance. Or, "In what sense ?" In social science applications of this sort there is no royal road around measurement error, around ordinal variables, and around nonlinear regression. If one does not know how to model these (and generally we don't), then descriptive techniques should be used.

Block recursive models have great advantages over LISREL type models. Both are elegant in their formulation, but LISREL rapidly takes you into a swamp of identification, interpretation, and computation problems. Block recursive models are not only elegant in their formulation, they are elegant in the analysis. My thesis is that here, as everywhere, elegance comes at a terrible price. They will only be applicable to small, selected data sets, in which we refuse to look at some of the basic assumptions. As far as I can see, the robustnes properties which have saved the t-test from oblivion, do not apply here. The most surprising result of the paper, for me, is the brevity of Section 5, indicating how few results carry over from the block recursive case.

References

Albert Gifi (1991). *Nonlinear multivariate Analysis.* Wiley, Chichester, Great Britain.

Arthur S. Goldberger (1972). Structural equation models in the social sciences. *Econometrika*, **40**, 979-1001.

Bengt O. Muthém (1982). Some categorical variable models with latent variable. In Karl G. Joreskog and Herman O. Wold, editors, *Systems under indirect observation: causality, structure, prediction*, 65-79. North Holland Publishing Co, Amsterdam, The Netherlands.

Cheng Hsiao (1983). Identification. In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics*, volume I, chapter 4, pages 224-283. North Holland Publishing Co, Amsterdam, The Netherlands.

David J. Bartholomew (1987). *Latent variable models and factor analysis.* Griffin, London, Great Britain.

Dennis J. Aigner, Cheng Hsiao, Arie Kapteyn, and Tom Wansbeek (1984). Latent variable models in econometrics. In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics*, volume II, chapter 23, pages 1323-1393. North Holland Publishing Co, Amsterdam, The Netherlands.

Franklin M. Fisher (1966). *The identification problem in econometrics.* McGraw-Hill, New York, New York.

Herman O. Wold (1982). Soft modeling: the basic desing and some extensions. In Karl G. Joreskog and Herman O. Wold, editors, *Systems under indirect observation: causality, structure, prediction*, 65-79. North Holland Publishing Co, Amsterdam, The Netherlands.

Jan de Leeuw (1986). Multivariate analysis with optimal scaling. Calcutta, India. Indian Statistical Institute.

Jan de Leeuw (1988). Multivariate analysis with linearizable regressions. *Psychometrika.*

Paul A. Bekker and Theo K. Dijkstra (1990). On the nature and number of contraints on the reduced form as implied by the structural form. *Econometrika*, **58**, 507-514.

Paul A. Bekker and Stephen Pollock (1986). Identification of linear stochastic models wich convariance restrictions. *Journal of Econometrics*, **31**, 179-208.

Zvi Griliches (1974). Errors in variables and other unobservables. *Econometrika*, **42**, 971-988.

Arthur S. Goldberger

Department of Economics
University of Wisconsin
Madison - Wisconsin, USA

By focusing on algebraic manipulation rather than scientific substance, Wermuth presents a misleading picture of the linear structural equation models used in economics.

Readers who have had a course in economics will recall the textbook description of the market's determination of the quantity and price of a good. On one side of the market, the quantity of the good that consumers will buy is a decreasing function of its price. On the other side of the market, the price that producers will charge is an increasing function of the quantity. The intersection of the demand curve and supply curve determines an equilibrium quantity and price.

Elaborating this story to allow for other determinants of demand and supply, an economist might arrive at a model like this:

$$\begin{aligned} y_1 &= \alpha_1\, y_2 + \alpha_2\, x_1 + u_1 \\ y_2 &= \alpha_3\, y_1 + \alpha_4\, x_2 + u_2. \end{aligned} \tag{1}$$

Here y_1 = quantity, y_2 = price, x_1 = income, x_2 = wage rate, u_1 = demand shock, u_2 = supply shock. The first equation says that household demand depends on price and household income (which are observable) and an unobserved factor. The systematic part of this equation, $\alpha_1 y_2 + \alpha_2 x_1$, may be interpreted as the expected quantity demanded if y_2 and x_1 were fixed. So the α_1 and α_2 parameters have natural meaning for the economist. Similarly for the second, supply, equation. The exclusions of x_2 from the demand equation and of x_1 from the supply equation reflect the economist's understanding of household and producer behavior.

The point of this model is the same as the point of the classroom discussion: the endogenous variables quantity and price are not fixed, but rather are jointly determined by the two sides of the market, a determination that becomes explicit in the solved, or reduced, form of the model:

$$y_1 = \pi_1 x_1 + \pi_2 x_2 + v_1$$
$$y_2 = \pi_3 x_1 + \pi_4 x_2 + v_2, \tag{2}$$

with $\pi_1 = \theta\alpha_2$, $\pi_2 = \theta\alpha_1\alpha_4$, $\pi_3 = \theta\alpha_2\alpha_3$, $\pi_4 = \theta\alpha_4$, $\theta = 1/(1 - \alpha_1\alpha_3)$, $v_1 = \theta(u_1 + \alpha_1 u_2)$, $v_2 = \theta(\alpha_3 u_1 + u_2)$. The shocks u_1 and u_2 are assumed independent of the exogenous variables x_1 and x_2.

Now the structural model (1) above is, apart from notational changes, Wermuth's system (5.3) with $\gamma_{14} = \gamma_{23} = 0$, that is her Case C. (The normality assumption she adopts is not relevant here, nor is it needed in many economic settings). And yet Wermuth announces that its parameters are difficult to interpret, and that the model has further undesirable features related to restrictions on observable covariances. The difficulties, as far as I can see, arise from her failure to pay attention to the substance of the phenomenon being modelled by the economist.

It is fair to ask why economists like to formulate structural models in this manner rather than starting with a multivariate regression. The answer is developed in a book that she cites, Goldberger (1964, pp. 380-388), and at a more elementary level in Goldberger (1991, Chapter 31). For the perspective of a statistician, see Anderson (1991).

One will not get far by imagining that (1) is obtained by an arbitrary algebraic rearrangement of (2), as Wermuth suggests in Section 5.1. I should add that she misrepresents my position, by citing Goldberger (1964, p. 297) as the source for the idea that linear structural equations (like (1)) can be viewed as resulting from an arbitrary nonsingular linear transfomation of the reduced form (multivariate regression like (2)). Goldberger (1964, p. 297) says nothing of the kind, indeed makes no mention of arbitrary linear transforma-

tions. On page 383, the notion of linear transformations is introduced, and is soon disposed of:

"In formulating a structural model, the econometrician, after all, does not take arbitrary linear combinations of endogenous variables. Rather he attempts to construct operational counterparts of the behavioral and technological equations of economic theory." (p. 386)

It is possible that I have misunderstood Wermuth's article. Perhaps she had no real objections to the use of linear structural models where appropriate, but merely wished to persuade readers that there are research topics and research hypotheses (such as that in her Section 6) that do not fall in that category. If so, I would agree.

References

Anderson, T. W. (1991). Trygve Haavelmo and simultaneous equation models. *Scandinavian Journal of Statistics*, **18**, 1-29.

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

Goldberger, A. S. (1991). *A Course in Econometrics.* Cambridge, Mass.: Harvard University Press.

Steffen L. Lauritzen

Institute for Electronic Systems
The University of Aalborg
Denmark

First I wish to congratulate the author for having made this detailed analysis of relations between different parametrizations of a multivariate normal distribution and, in particular, the various interpretations associated with

these.

My favourite aspect of the paper is concerned with the connection between graphical chain models, defined by conditional independence restrictions, and models based on demanding zero restrictions in structural equation models. Usally, however, I prefer to argue about this in a manner which is slightly different from the arguments given in the paper. Below this argument is given in the hope that some readers could find it illuminating. For simplicity I assume that the means are equal to zero and that we are considering equations for a system without pure exogenous variables such that the entire system only involves two groups of variables a and b, using the notation from the paper.

As described in Section 5 of the paper, a system of linear structural equations for the variables is given by

$$\Lambda Y = U$$

where Y is the vector of variables, partitioned into groups $Y = (Y_a, Y_b)$ (corresponding to the dependence chain associated with the system). In general the system of equations is heavily overparametrized since the distribution of Y is determined by its covariance matrix alone, whereas the system of equations above, as stated so far, both have the covariance matrix of the 'residuals' U and the elements of the matrix Λ as varying parameters.

Uniqueness can be obtained by requiring Λ to be upper diagonal with all diagonal elements equal to one, and simultaneously insisting on a diagonal covariance matrix of U. As also pointed out in the paper, the parametrization becomes unique with this convention, the off-diagonal elements of Λ turn out to be partial regression coefficients in a univariate recursive system of regressions, and U is the vector of residuals from this regression.

It is in no way obvious whether a similarly elegant parametrization exists that reflects a block-structure (a, b) - rather than a complete ordering - of the variables Y. The key to this parametrization is given in Section 3. Demand Λ

to be upper block-triangular, such that $\Lambda_{ba} = 0$ in the partioning of Λ as

$$\Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}.$$

Demand further that the covariance matrix Φ of the 'residuals' U be block-diagonal, in the special case studied with two blocks. Finally - and this is the surprising bit - assume the blocks in Φ to be equal to the corresponding blocks in Λ - in particular implying that the diagonal blocks of Λ are symmetric. In this way, Λ and the covariance matrix $\sum$ of Y are in one-to-one correspondence and, by choosing this special parametrization, the author obtains that *any zero-restriction in the parameters of the equation system can be interpreted by conditional independence*, essentially leading to Proposition 4.2 of the paper.

What has here been called Λ is B^* of the paper. To identify the equations as regression equations, it is convenient to divide each single equation by the corresponding diagonal element of Λ, thereby making the remaining elements equal to partial regression coefficients and obtaining B. But note that one could not just multiply each block with the inverse of the corresponding block in Λ. This would mess up any pattern of zeros in the off-diagonal elements in an uncontrollable way.

I find the identification of this relation between block-recursive structural equation systems and conditional independence models interesting and illuminating and congratulate the author with a fine piece of work.

REPLY TO DISCUSSION

More than twenty years ago Arthur Dempster suggested as a possible topic for a Ph.D. thesis that I look into the relations of his covariance selection to path analysis and to structural equations. Because I didn't think I could say anything sensible within reasonable time I didn't choose this topic. And,

I am glad that I didn't, otherwise I would still be waiting for my degree now; some answers are included in the present paper.

The hardest and the most fascinating part of this paper was for me to see why one has to worry about identification with linear structural equations but not with graphical chain models, i.e. Sections 3 and 4. It is nice that David Cox, Steffen Lauritzen and Gerhard Arminger appear to agree that the results concerning the relations of these two model classes are the most important aspect of the paper. Though block-recursive equations had been singled out before by Fisher (1966) to point at simplifications of the identification problem and for instance by Anderson (1991) to point at simplifications in estimation it appears to not have been observed previously that a particular choice of block parameters leads to a unique parametrization and to systematic parts of the equations which are special conditional expectations (3.3) or simple multiples thereof (3.2).

In my view this interpretation of each equation parameter as a regression coefficient is the main common feature of block-regressions, of multivariate regressions, of univariate recursive regressions and of chain models defined from these elements (Cox and Wermuth, 1992a). It is a feature not shared by general structural equations, for instance by the model described as (1) below. This distinction explains, why a statement like 'discarding of structural modeling as a research method can hardly be taken seriously, for this would require eliminating all of its special cases as well' (Bentler, 1987; p.155) is invalid: multivariate regression and univariate recursive regressions are special cases of general structural equations but they have important properties which distinguish them favourably from the general structural equations. A different evaluation of this fact is given by Jöreskog and Sörbom (1978, p.1), who say that 'because each equation in the model represents a causal link rather than a mere empirical association, the structural parameters do not, in general, coincide with coefficients of regressions among observed variables. The structural parameters represent relatively unmixed, invariant and autonomous features

of the mechanism that generate the observable variables.

I am primarily concerned about interpretations and about possibilities for integration of substantive knowledge into model formulations. This leads to a major concern that parameters have a clear interpretation and that restrictions placed on parameters also have a clear meaning. If no general results are available we may need detailed algebraic manipulations to arrive at the interpretation and to see, for instance, that the equation parameters $(\alpha_i, i = 1, \ldots, 4)$ in Arthur Goldberger's saturated structural equations (1):

$$\begin{pmatrix} y_1 - \alpha_1 y_2 - \alpha_2 x_1 \\ y_2 - \alpha_3 y_1 - \alpha_4 x_2 \end{pmatrix} = \begin{pmatrix} 1 & -\alpha_1 \\ -\alpha_3 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} -\alpha_2 & 0 \\ 0 & -\alpha_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \tag{1}$$

i.e. in $\Gamma_{dd} y + \Gamma_{dc} x = u$ do not have the meaning he attributes to them unless there are further constraints.

More precisely, his claim that the systematic part '$\alpha_1 y_2 + \alpha_2 x_1$, may be interpreted as the expected quantity demanded if y_2 and x_1 were fixed' is inconsistent with his specification that (1) are saturated equations (Case C in the paper), i.e. that they have the same number of independent parameters (four equation parameters and three parameters in the covariance matrix of the u's) as there are free parameters in the covariance matrix of the variables (four variances and six covariances minus the three parameters in the covariance matrix of the variables conditioned upon, the x's) and that therefore (1) also does not place restrictions on the reduced form equations (2).

The latter are obtained from (1) after premultiplication by Γ_{dd}^{-1} and they are multivariate regression equations:

$$\begin{aligned} y_1 &= \pi_1 x_1 + \pi_2 x_2 + v_1 = \mathrm{E}(Y_1 \mid X_1 = x_1, X_2 = x_2) + v_1 \\ y_2 &= \pi_3 x_1 + \pi_4 x_2 + v_2 = \mathrm{E}(Y_2 \mid X_1 = x_1, X_2 = x_2) + v_2. \end{aligned} \tag{2}$$

That is, if the specification of (1) is completed by taking (v_1, v_2) in (2) as random variables of zero mean, covariance matrix Σ, and each of the v_i as uncorrelated with (x_1, x_2), then we may calculate the expected quantity de-

manded if y_2 and x_1 are given and it is not in general $\alpha_1 y_2 + \alpha_2 x_1$. In fact, if we start with the reduced form equations (2) and write

$$\delta_1 = \frac{\text{cov}(v_1, v_2)}{\text{var}(v_2)}, \quad \delta_2 = \pi_1 - \delta_1 \pi_3, \quad \delta_3 = \pi_2 - \delta_1 \pi_4, \tag{3}$$

then δ_1 is known to define the partial regression coefficient of y_2 in the regression of Y_1 on y_2, x_1, x_2 and

$$\mathrm{E}(Y_1 \mid Y_2 = y_2, X_1 = x_1, X_2 = x_2) = \delta_1 y_2 + \delta_2 x_1 + \delta_3 x_2, \tag{4}$$

since

$$\begin{aligned} \pi_1 x_1 + \pi_2 x_2 &= \mathrm{E}(Y_1 \mid X_1 = x_1, X_2 = x_2) \\ &= \mathrm{E}_{Y_2 \mid X_1 = x_1, X_2 = x_2} \mathrm{E}(Y_1 \mid Y_2 = y_2, X_1 = x_1, X_2 = x_2) \\ &= (\delta_2 + \delta_1 \pi_3) x_1 + (\delta_3 + \delta_1 \pi_4) x_2. \end{aligned}$$

After writing further

$$\mathrm{E}(X_2 \mid Y_2 = y_2, X_1 = x_1) = c y_2 + d x_1$$

we obtain by taking expectations in (4) over X_2 given y_2, x_1 that

$$\begin{aligned} \mathrm{E}(Y_1 \mid Y_2 = y_2, X_1 = x_1) &= (\delta_1 + \delta_3 c) y_2 + (\delta_2 + \delta_3 d) x_1 \\ &\neq \alpha_1 y_2 + \alpha_2 x_1. \end{aligned} \tag{5}$$

For instance, the coefficient of x_1 in (5), i.e. in the expected quantity demanded given y_2 and x_1, is

$$\begin{aligned} \delta_2 + \delta_3 d &= (\pi_1 - \delta_1 \pi_3) + (\pi_2 - \delta_1 \pi_4) d \\ &= \{\alpha_2 (1 - \delta_1 \alpha_3) + \alpha_4 d (\alpha_1 - \delta_1)\} / (1 - \alpha_1 \alpha_3). \end{aligned}$$

This will equal α_2 only if there are very special restrictions on the parameters, so we have a contradiction. We have shown that if (2) are unrestricted reduced form equations then the parameters in (1) cannot have the meaning Arthur Goldberger claims they have.

One possibility to obtain equations with his interpretation is to demand that

$$\begin{aligned} \mathrm{E}(Y_1 \mid Y_2 = y_2, X_1 = x_1, X_2 = x_2) &= \mathrm{E}(Y_1 \mid Y_2 = y_2, X_1 = x_1) \\ \mathrm{E}(Y_2 \mid Y_1 = y_1, X_1 = x_1, X_2 = x_2) &= \mathrm{E}(Y_2 \mid Y_1 = y_1, X_2 = x_2), \end{aligned} \quad (6)$$

but this leads to a nondecomposable block-regression with blocks (y_1, y_2) and (x_1, x_2) and places two restrictions on the corresponding reduced form equations, i.e. if we have equations (1) with the desired interpretation then the corresponding reduced form equations are restricted. For instance, the systematic part in one of these block-regression equations is (4) with $0 = \delta_3$ or $\pi_2 = \delta_1 \pi_4$ and δ_1, the coefficient of x_2, is not an independent parameter; compare (3). There is an analogous argument for the second equation in the block-regression corresponding to (6). Thus a model with Arthur Goldberger's interpretation of the parameters is a special graphical chain model, a nondecomposable block-regression with five independent parameters and hence two restrictions on the reduced form equations. This has consequences for the estimation.

Efficient estimates of the equation parameters in such a block-regression are not directly obtainable with software for linear structural equations, but they can be obtained for instance with David Edward's program MIM or by adapting any program for covariance selection. They are maximum-likelihood estimates if the four variables have a joint normal distribution and they minimize Wilks' generalized residual variance (compare Section 4), otherwise. It will be interesting to see how - if at all - they differ from the pseudo maximum-likelihood and minimum distance estimates discussed by Gerhard Arminger.

Jan de Leeuw mentions desirable future results for graphical chain models. Some of these have been obtained but have not yet or only recently appeared in print. For instance, models with both qualitative and quantitative responses have been compared regarding different distributional assumptions (Cox and Wermuth, 1992b) including, in particular, various probit

and logit models. Methods for detecting and modeling nonlinear relations in chain models for quantitative variables have been provided (Cox and Wermuth, 1992c,d). Results in several applications with six and more variables have been described (Wermuth and Cox, 1992). In each of these graphical chain models helped to integrate substantive knowledge into the analysis and permitted to split the analysis of all component variables into analyses of sequences of small sets of variables.

Chain models with latent variables are important though not yet far developed. This is one of our next goals. The starting point will not be general linear structural equations, but models with parameters that are regression coefficients, i.e. in particular univariate recursive regressions, multivariate regression chain models, and block-regression chain models.

I thank the discussants warmly for their comments, and Arthur Goldberger for providing an example from economics in which structural equations give a false, but graphical chain models a correct specification of the desired interpretation.

Additional references

Anderson, T.W. (1991). Trygve Haavelmo and simultaneous equation models. *Scand. J. Statist.*, **18**, 1-19.

Bentler, P.M. (1987) Structural modeling and the scientific method: comment on Freedman's critique. *J. Educ. Statist.*, **12**, 151-157.

Cox, D.R. and Wermuth, N. (1992a). Linear dependencies represented by chain graphs. *Berichte zur Stochastik und verw.* Gebiete, 92-93. Universität Mainz.

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. (1992c). Some recent work on methods for the analysis of multivariate observational data in the social sciences. Invited lecture, 7th International Conference on Multivariate Analysis, Penn State, 1992. *Conference Proceedings.*

Cox, D.R. and Wermuth, N. (1992d). Tests of linearity, multivariate normality, and the adequacy of linear scores. *Berichte zur Stochastik und verw.* Gebiete, 92-6. Universität Mainz.

Fisher, F.M. (1966). *The identification problem in econometrics.* New York: Mc Graw-Hill.

Jöreskog, K.G. and Sörbom, D. (1978). LISREL IV. Analysis of Linear Structural Relationships by Maximum-Likelihood (2nd ed). Chicago: International Educational Services.

Wermuth, N. and Cox, D.R. (1992). Graphical models for dependencies and associations. In: Y. Dodge and J. Whittaker (eds.). *Computational Symposium Statistics*, 235-250. Heidelberg: Physica-Verlag.