Linear Recursive Equations, Covariance Selection, and Path Analysis

NANNY WERMUTH*

By defining a reducible zero pattern and by using the concept of multiplicative models, we relate linear recursive equations that have been introduced by econometrician Herman Wold (1954) and path analysis as it was proposed by geneticist Sewall Wright (1923) to the statistical theory of covariance selection formulated by Arthur Dempster (1972). We show that a reducible zero pattern is the condition under which parameters as well as least squares estimates in recursive equations are one-to-one transformations of parameters and of maximum likelihood estimates, respectively, in a decomposable covariance selection model. As a consequence, (a) we can give a closed-form expression for the maximum likelihood estimate of a decomposable covariance matrix, (b) we can derive Wright's rule for computing implied correlations in path analysis, and (c) we can describe a search procedure for fitting recursive equations.

KEY WORDS: Reducible zero patterns; Multiplicative or decomposable models; One-to-one transformations of covariance matrices; Linear recursive equations; Covariance selection; Path analysis.

1. INTRODUCTION

A renewed interest in studying systems of linear equations has been documented in many subject matter areas, in genetics (Li 1975), in sociology and in economics (Blalock 1971; Goldberger and Duncan 1973; Heise 1975, Duncan 1975), and in psychology (Hodapp 1978). The purpose of this article is to present a covariance selection approach to linear recursive equations and to path analysis. Since multivariate normality is an assumption for covariance selection models, we restrict our discussion of linear recursive equations and of path analysis to that situation.

The inferential aspects of path analysis can be described as relating to two distinct purposes: One is to estimate parameters in certain systems of linear equations (see (4.1)), the other is to estimate the parameters of the corresponding correlation matrix (see (4.2)). Wold's (1954) result for linear recursive equations is known to lead to Wright's (1934) rule (4.1) for estimating the equation parameters, but it does not imply the rule (4.2) for estimating the correlation matrix. We use a simple example (2.6) in which the second path analysis rule (4.2) does not give the maximum likelihood estimate (MLE) of the correlation matrix, even though the first path analysis rule (4.1) leads to the MLE's of the equation parameters. We then find a sufficient condition under which both path analysis rules define MLE's. This condition is a reducible zero pattern in the equation parameters or equivalently in the parameters of a covariance selection model.

To this end, in Section 2 we derive linear recursive equations as a triangular reduction of a covariance matrix. We show that a reducible zero pattern for the regression coefficients implies a corresponding covariance selection model in which the inverse covariance matrix has the same reducible zero pattern, and conversely. In Section 3 we show that for all multiplicative models the variables can be reordered so that the inverse covariance matrix has a reducible zero pattern. We state closed-form expressions for the MLE of such decomposable covariance matrices. From these we obtain in Section 4 Wright's rule for computing implied correlations in path analysis. Finally, we suggest in Section 5 how the model search among multiplicative models (Wermuth 1976b) can be modified to become a search for fitting recursive equations. We illustrate this with a set of sociological data.

2. LINEAR RECURSIVE EQUATIONS

By a system of linear recursive equations with independent errors is understood a set of equations (Wold 1954, 1960; Goldberger 1964) of the following form:

$$\begin{aligned} Z_1+a_{12}Z_2+a_{13}Z_3+\ldots+a_{1k}Z_k+a_{1,k+1}Z_{k+1}+\ldots+a_{1p}Z_p&=U_1\\ Z_2+a_{23}Z_3+\ldots+a_{2k}Z_k+a_{2,k+1}Z_{k+1}+\ldots+a_{2p}Z_p&=U_2\\ \vdots\qquad\qquad\vdots\qquad\qquad&\vdots\\ Z_k+a_{k,k+1}Z_{k+1}+\ldots+a_{kp}Z_p&=U_k\,, \end{aligned} \tag{2.1}$$

where Z_i is a dependent or endogeneous variable for $i = 1, \ldots, k$ and is fixed or exogeneous for $i = k + 1, \ldots, p$. The system is called *recursive*, since each of the endogeneous variables Z_i may depend on $Z_{i+1}, \ldots, Z_p$ but not on $Z_1, \ldots, Z_{i-1}$. The errors U_i are assumed to be jointly normally distributed with $U_i \sim N(0, t_i)$ and $E(U_iU_j) = 0$, for $i \neq j$. The system is called *incomplete* if some of the regression coefficients a_{ij} are restricted to be zero (and there are no other types of restrictions), and it is called *complete* if there are no restrictions. It follows

* Nanny Wermuth is Professor of Medical Statistics and of Psychological Methods, Psychologisches Institut, Johannes Gutenberg-Universität, Postfach 3980, 6500 Mainz, Federal Republic of Germany. Volker Hodapp's insistence on the importance of the topic has led to this article. I wish to thank him, Terry Speed, and Steffen Lauritzen for stimulating discussions. Further, I thank the referees and Klaus Miescke for their suggestions on how to improve the article.

© Journal of the American Statistical Association
December 1980, Volume 75, Number 372
Theory and Methods Section

from a result by Wold (1960; see also Malinvaud 1966) that the MLE's of the parameters in complete as well as in all incomplete systems can be obtained by the method of least squares estimation applied to each equation separately, regressing variable Z_i on all Z_j having nonzero regression coefficients in the ith equation.

In order to relate the estimation of equation parameters to the estimation of a covariance matrix, we do not regard $Z_{k+1}, \ldots, Z_p$ as fixed but as random variables. We assume that $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)^T$ follows an origin-centered, nondegenerate normal distribution with covariance matrix $\mathbf{\Sigma}$ such that the elements in positions (i, j) of $\mathbf{\Sigma}$ and of its inverse $\mathbf{\Sigma}^{-1}$ are the covariance σ_{ij} and the concentration (see Dempster 1969) σ^{ij}, respectively, of the variables Z_i and Z_j. Further, we let $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^T$ denote a sample of size n on variable Z_i so that $\mathbf{z}_i^T\mathbf{z}_j$, the sample inner product of variables Z_i and Z_j, is the element in position (i, j) of the matrix $n\mathbf{S}$, where $\mathbf{S}$ is a sample covariance matrix with $s_{ij} = \mathbf{z}_i^T\mathbf{z}_j/n$.

If system (2.1) is appended by $(p - k)$ complete recursive equations in arbitrarily ordered variables $Z_{k+1}, \ldots, Z_p$, we can write a new recursive system in matrix notation as

$$\mathbf{AZ} = \mathbf{U} \quad \text{with} \quad \mathbf{U} = (U_1, \ldots, U_p)^T \quad \text{and} \quad \mathbf{U} \sim N(\mathbf{0}, \mathbf{T}) \ , \tag{2.2}$$

where $\mathbf{A}$ is the triangular matrix of regression coefficients with elements zero below the diagonal and elements unity along the diagonal, $\mathbf{T}$ is the diagonal variance matrix of the errors $\mathbf{U}(E(\mathbf{UU}^T) = \mathbf{T})$, and $E(\mathbf{ZZ}^T) = \mathbf{\Sigma}$. System (2.1) then corresponds to the first k equations in system (2.2); Proposition 1 shows that in this larger system, the relationships among variables $Z_{k+1}, \ldots, Z_p$ can still be regarded as left unanalyzed, since a complete system of recursive equations is just a triangular reduction of an unrestricted covariance matrix. To state Proposition 1, we first need a definition and results (see Dempster 1969) for the triangular reduction of a positive definite covariance (PDC) matrix. For each PDC matrix $\mathbf{\Sigma}$, there exists a triangular decomposition $\mathbf{B}$ such that $\mathbf{B\Sigma B}^T = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{B}$ is a triangular matrix with elements zero below the diagonal and elements unity along the diagonal. The triangular reduction $(\mathbf{B}, \mathbf{D})$ can be constructed by a process of successive orthogonalization from $\mathbf{\Sigma}$, and it is a unique, one-to-one transformation of $\mathbf{\Sigma}$, if the order for orthogonalizing $\mathbf{\Sigma}$ is fixed. Using the Beaton (1964) sweep operator (also discussed in Dempster 1969) for this process, we can express the elements of $\mathbf{D}$ and $\mathbf{B}$ in terms of concentrations and covariances. We assume that $\mathbf{\Sigma}$ is swept in the order $p, p - 1, \ldots, 1$ or equivalently that $\mathbf{\Sigma}^{-1}$ is swept in the order $1, 2, \ldots, p$.

$$\begin{aligned} d_{11} &= \sigma_{11.2,\ldots,p} = 1/\sigma^{11} \\ d_{ii} &= \sigma_{ii.i+1,\ldots,p} = 1/\sigma^{ii.1,\ldots,i-1} \\ d_{pp} &= \sigma_{pp} \qquad = 1/\sigma^{pp.1,\ldots,p-1} \ ; \quad \text{for} \quad 2 \le i \le p-1 \end{aligned} \tag{2.3}$$

$$\begin{aligned} b_{ij} &= -\sigma_{ij.i+1,\ldots,j-1,j+1,\ldots,p}/\sigma_{jj.i+1,\ldots,j-1,j+1,\ldots,p} \\ b_{p-1,p} &= -\sigma_{p-1,p}/\sigma_{pp}, \quad \text{for} \quad 1 \le i < j \le p-1 \end{aligned} \tag{2.4}$$

$$\begin{aligned} b_{1j} &= \sigma^{1j}/\sigma^{11}, \quad \text{for} \quad 2 \le j \le p \\ b_{ij} &= \sigma^{ij.1,\ldots,i-1}/\sigma^{ii.1,\ldots,i-1}, \quad \text{for} \quad 2 \le i < j \le p \ ; \end{aligned} \tag{2.5}$$

where $\sigma_{ij.i+1,\ldots,j-1,j+1,\ldots,p}$ is the element in position (i, j) obtained from sweeping $\mathbf{\Sigma}$ on rows and columns $i + 1, \ldots, j - 1, j + 1, \ldots, p$, that is, the residual covariance of Z_i and Z_j after the components due to variables $Z_{i+1}, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_p$ have been removed. Similarly, $\sigma^{ij.1,\ldots,i-1}$ is the element in position (i, j) after sweeping $\mathbf{\Sigma}^{-1}$ on rows and columns $1, \ldots, i - 1$; it is the concentration of Z_i and Z_j in the marginal joint distribution of the variables $Z_i, Z_{i+1}, \ldots, Z_p$.

Proposition 1: The following two statements are equivalent:

1. A system of linear recursive equations $(\mathbf{A}, \mathbf{T})$ is complete, and
2. The covariance matrix $\mathbf{\Sigma}$ of $\mathbf{Z}$ is unrestricted.

If a system $(\mathbf{A}, \mathbf{T})$ is complete, then the regression coefficients and their least-squares estimates in the ith equation are identical to the elements in the ith row of the triangular decomposition of $\mathbf{\Sigma}$ and $\mathbf{S}$, respectively, obtained by successively orthogonalizing $\mathbf{\Sigma}$ and $\mathbf{S}$ in the order $p, p - 1, \ldots, 1$.

The proof in the direction (1) to (2) follows by noting that (2.2) implies the covariance matrix $\mathbf{\Sigma} = E(\mathbf{ZZ}^T) = \mathbf{A}^{-1}\mathbf{T}(\mathbf{A}^T)^{-1}$. Conversely, set $\mathbf{U} = \mathbf{BZ}$; then $\mathbf{U} \sim N(\mathbf{0}, \mathbf{B\Sigma B}^T)$ with $\mathbf{B\Sigma B}^T = \mathbf{D}$ (see Anderson 1958) so that $\mathbf{U}$ is distributed as required in the specification of the recursive system (2.2). Further, $U_i = Z_i + \sum_{j=i+1}^{p} b_{ij}Z_j$, and its variance is d_{ii}. Thus we can identify a_{ij} with b_{ij} and t_i with d_{ii}, and no restrictions on $\mathbf{\Sigma}$ implies no restrictions on $\mathbf{A}$ and $\mathbf{T}$. To complete the proof, replace $\mathbf{\Sigma}$ by $\mathbf{S}$ and the covariances in (2.4) by their sample equivalents; we obtain therein the usual expression for the least squares coefficients regressing Z_i on $Z_{i+1}, \ldots, Z_p$ except for the necessary change in the sign.

In incomplete recursive systems the matrix with equation parameters $\mathbf{A}$ (and the MLE $\hat{\mathbf{A}}$) is (up to signs) still identical to the triangular decomposition of some restricted covariance matrix $\mathbf{\Sigma}^*$ (and its MLE $\hat{\mathbf{\Sigma}}^*$). However, $\hat{\mathbf{\Sigma}}^*$ does not, in general, coincide with the observed covariance matrix $\mathbf{S}$ in all positions (i, j) having nonzero equation parameters a_{ij}. In the following two examples that are used to point at two distinct situations, we partition $\hat{\mathbf{\Sigma}}^*$ (and similarly $\mathbf{S}$) as

$$\hat{\mathbf{\Sigma}}^* = \begin{bmatrix} \hat{\sigma}_{11}^* & \hat{\sigma}_{12}^* & \hat{\sigma}_{13}^* & \hat{\sigma}_{14}^* \\ & & \hat{\mathbf{\Sigma}}_{(1)}^* & \end{bmatrix}.$$

Example 1: For an incomplete system in two equations, four variables, and only one zero restriction on a_{23}, we

get the matrix of equation parameters $\mathbf{A}$, its MLE $\hat{\mathbf{A}}$, and the MLE $\hat{\mathbf{\Sigma}}^*$ (see Appendix) as

$$\mathbf{A}=\begin{pmatrix}1 & a_{12} & a_{13} & a_{14}\\ 0 & 1 & 0 & a_{24}\\ 0 & 0 & 1 & a_{34}\\ 0 & 0 & 0 & 1\end{pmatrix};$$

$$\hat{\mathbf{A}}=\begin{pmatrix}1 & -s_{12.34}/s_{22.34} & -s_{13.24}/s_{33.24} & -s_{14.23}/s_{44.23}\\ 0 & 1 & 0 & -s_{24}/s_{44}\\ 0 & 0 & 1 & -s_{34}/s_{44}\\ 0 & 0 & 0 & 1\end{pmatrix};$$

$$\hat{\mathbf{\Sigma}}_{(1)}{}^*=\begin{pmatrix}s_{22} & s_{24}s_{34}/s_{44} & s_{24}\\ & s_{33} & s_{34}\\ & & s_{44}\end{pmatrix} \tag{2.6}$$

$$(\hat{\sigma}_{12}{}^*\hat{\sigma}_{13}{}^*\hat{\sigma}_{14}{}^*) = (s_{12}\ \ s_{13}\ \ s_{14})(\mathbf{S}_{(1)})^{-1}\hat{\mathbf{\Sigma}}_{(1)}{}^*\ ;$$

$$\hat{\sigma}_{11}{}^* = s_{11} - (s_{12}\ \ s_{13}\ \ s_{14})(\mathbf{S}_{(1)})^{-1}\cdot[\mathbf{I} - \hat{\mathbf{\Sigma}}_{(1)}{}^*(\mathbf{S}_{(1)})^{-1}](s_{12}\ \ s_{13}\ \ s_{14})^T\ .$$

The fact to be noted here is that with one zero restriction on the equation parameters, five elements in the MLE $\hat{\mathbf{\Sigma}}^*$ deviate from elements in the observed covariance matrix (unless by mere accident s_{23} is identical to $s_{24}s_{34}/s_{44}$).

Example 2: For an incomplete system of two equations, four variables, and two zero restrictions on a_{12} and on a_{23}, we get $\mathbf{A}$, its MLE $\hat{\mathbf{A}}$, and the MLE $\hat{\mathbf{\Sigma}}^*$ as

$$\mathbf{A}=\begin{pmatrix}1 & 0 & a_{13} & a_{14}\\ 0 & 1 & 0 & a_{24}\\ 0 & 0 & 1 & a_{34}\\ 0 & 0 & 0 & 1\end{pmatrix};$$

$$\hat{\mathbf{A}}=\begin{pmatrix}1 & 0 & -s_{13.4}/s_{33.4} & -s_{14.3}/s_{44.3}\\ 0 & 1 & 0 & -s_{24}/s_{44}\\ 0 & 0 & 1 & -s_{34}/s_{44}\\ 0 & 0 & 0 & 1\end{pmatrix};$$

$$\hat{\mathbf{\Sigma}}^*=\begin{pmatrix}s_{11} & -(\hat{a}_{13}\hat{\sigma}_{23}{}^*+\hat{a}_{14}s_{24}) & s_{13} & s_{14}\\ & s_{22} & s_{24}s_{34}/s_{44} & s_{24}\\ & & s_{33} & s_{34}\\ & & & s_{44}\end{pmatrix}. \tag{2.7}$$

In this example the two zero restrictions on the equation parameters imply that the MLE $\hat{\mathbf{\Sigma}}^*$ deviates from the observed covariance matrix in exactly the two positions (i, j) that correspond to the two zero equation parameters a_{ij}.

In the remaining part of this section, we show (a) that a reducible zero pattern in the equation parameters is a necessary and sufficient condition under which the zero pattern in equation parameters is identical to the zero pattern in the inverse covariance matrix, so that (b) the MLE $\hat{\mathbf{\Sigma}}^*$ coincides with $\mathbf{S}$ in all positions (i, j) that correspond to nonzero equation parameters.

Let $\tilde{I} = \{(i, j) \mid 1 \leq i < j \leq p\}$, and let $\mathbf{\Sigma}_{(1,\ldots,k)}$ and $\mathbf{B}_{(1,\ldots,k)}$ be the submatrices of a pdc matrix $\mathbf{\Sigma}$ and its triangular decomposition $\mathbf{B}$, which remain after deleting rows and columns $1, 2, \ldots, k$.

Definition 1: We call $I \subseteq \tilde{I}$ *reducible* if for each $(i, j) \in I$ and $h = 1, \ldots, i-1$, we have $(h, i) \in I$ or $(h, j) \in I$ or both.

Definition 2: For $I \subseteq \tilde{I}$ we define a matrix $\mathbf{M}$ to have zero structure with respect to I if all elements of $\mathbf{M}$ are zero exactly in positions $(i, j) \in I$. For convenience, we term a *reducible zero pattern* any case in which a matrix $\mathbf{M}$ has zero structure with respect to a reducible I.

Now, we can state

Proposition 2: For every reducible $I \subseteq \tilde{I}$ and every $k \in \{0, 1, \ldots, p-2\}$, the following statements are equivalent:

1. $\mathbf{B}_{(1,\ldots,k)}$ has zero structure with respect to $I_{(1,\ldots,k)}$, and
2. $(\mathbf{\Sigma}_{(1,\ldots,k)})^{-1}$ has zero structure with respect to $I_{(1,\ldots,k)}$, where $I_{(0)} = I$ and $I_{(1,\ldots,k)}$ is obtained from I by deleting all pairs (i, j) with $i \in \{1, \ldots, k\}$.

We prove first that (1) implies (2). Since $\mathbf{\Sigma}^{-1} = \mathbf{B}^T(\mathbf{B}\mathbf{\Sigma}\mathbf{B}^T)^{-1}\mathbf{B} = \mathbf{B}^T\mathbf{D}^{-1}\mathbf{B}$, we can write the elements of $\mathbf{\Sigma}^{-1}$ explicitly in terms of the elements of $\mathbf{B}$ and $\mathbf{D}$ as

$$\sigma^{11}=1/d_{11}\ ;\quad \sigma^{1j}=b_{1j}/d_{11}\ ,$$

$$\sigma^{ij}=1/d_{jj}+\sum_{h=1}^{j-1} b_{hj}{}^2/d_{hh}\ , \qquad \text{for}\quad 2\leq j\leq p\ ; \tag{2.8}$$

$$\sigma^{ij}=b_{ij}/d_{ii}+\sum_{h=1}^{i-1} b_{hi}b_{hj}/d_{hh}\ , \quad \text{for}\quad 2\leq i<j\leq p\ .$$

For each PDC $\mathbf{\Sigma}$ it is known that all variances σ_{ii} and all residual variances d_{ii} are nonzero. Thus if $\mathbf{B}$ has a reducible zero pattern, then for each (i, j) with $b_{ij} = 0$, either $b_{hi} = 0$ or $b_{hj} = 0$ (or both) for all $h < i$, so that each term in the sum of products defining σ^{ij} in (2.8) is zero, which completes the proof in the one direction for $k = 0$. For $k > 0$ note that the element in position $(i-k, j-k)$ of $(\mathbf{\Sigma}_{(1,\ldots,k)})^{-1}$ is $\sigma^{ij.1,\ldots,k}$, the concentration of Z_i and Z_j in the joint marginal distribution of $Z_{k+1}, \ldots, Z_p$, which may from (2.8) be written as

$$\sigma^{ij.1,\ldots,k} = b_{ij}/d_{ii} + \sum_{h=k+1}^{i-1} b_{hi}b_{hj}/d_{hh}\ ,$$

$$\text{for}\quad 2 \leq i < j \leq p\ . \tag{2.9}$$

Since $I_{(1,\ldots,k)}$ obtained from a reducible I is reducible, the same arguments as those presented for $\mathbf{B}$ and I apply to $\mathbf{B}_{(1,\ldots,k)}$ and $I_{(1,\ldots,k)}$, and the proof in the direction (1) to (2) is complete.

Conversely, let $\mathbf{\Sigma}^{-1}$ have a reducible zero pattern and

note from (2.3) and (2.5) that $b_{ij} = d_{ii}\sigma^{ij.1,\ldots,i-1}$ so that we may for $2 \leq i < j \leq p$ rewrite (2.8):

$$\sigma^{ij.1,\ldots,i-1} = \sigma^{ij} - \sum_{h=1}^{i-1} b_{hi}b_{hj}/d_{hh} , \qquad (2.10)$$

where $b_{hi}b_{hj}/d_{hh} = \sigma^{hi.1,\ldots,h-1}\sigma^{hj.1,\ldots,h-1}d_{hh}$ is the amount subtracted from element in position (i, j) at step h of sweeping $\boldsymbol{\Sigma}^{-1}$ in the order $1, \ldots, i-1$. From (2.8) we know for each $(1, j) \in I$ that $\sigma^{1j} = 0$ implies $b_{1j} = 0$. Further, a reducible zero pattern in $\boldsymbol{\Sigma}^{-1}$ implies for $(i, j) \in I_{(1)}$ that $\sigma^{1i}\sigma^{1j}d_{11} = b_{1i}b_{1j}/d_{11} = 0$, since $(1, i)$ and/or $(1, j)$ is contained in I so that also $\sigma^{ij.1} = 0$; it implies for $(i, j) \in I_{(1,2)}$ that $\sigma^{2i.1}\sigma^{2j.1}d_{22} = b_{2i}b_{2j}/d_{22} = 0$, since $(2, i)$ and/or $(2, j)$ is contained in $I \cap I_{(1)}$ so that $\sigma^{ij.12} = 0$. Continuing in this way, we know for $(i, j) \in I_{(1,\ldots,i-2)}$ that $b_{i-1,i}b_{i-1,j}/d_{i-1,i-1} = 0$, since $(i-1, i)$ and/or $(i-1, j)$ is contained in $I \cap I_{(1)} \cap \ldots \cap I_{(1,\ldots,i-2)}$ so that $\sigma^{ij.1,\ldots,i-1} = 0$. Thus if $\boldsymbol{\Sigma}^{-1}$ has a reducible zero pattern, then for each $(i, j) \in I$, we get $\sigma^{ij} = \sigma^{ij.1} = \sigma^{ij.12} = \ldots = \sigma^{ij.1,\ldots,i-1} = b_{ij} = 0$. This shows first that (2) implies (1) for $k = 0$ and second that $(\boldsymbol{\Sigma}_{(1,\ldots,k)})^{-1}$ has zero structure with respect to $I_{(1,\ldots,k)}$. The proof for $k > 0$ then follows from rewriting (2.9) in the same manner as (2.10) from (2.8) and by noting that all terms in the obtained sum have just proved to be zero for $(i, j) \in I_{(1,\ldots,k)}$.

In proposition 3 we relate incomplete recursive equations $(\mathbf{A}, \mathbf{T})$, where $\mathbf{A}$ has a reducible zero pattern, to covariance selection. Note again that the result applies to all systems (2.1) with k' or more equations, where k' is such that for all $i > k'$ there is no zero pattern in the marginal distribution of $Z_{i+1}, \ldots, Z_p$, that is, $I_{(1,\ldots,i)} = \emptyset$. In that case we can supplement (2.1) by a triangular decomposition of the unrestricted covariance matrix of the variables $Z_{k'+1}, \ldots, Z_p$ to obtain the last rows of $\mathbf{A}$ (see Proposition 1). A reducible zero pattern in the regression coefficients (in (2.1) and in (2.2)) means that for each $a_{ij} = 0$ it follows that $a_{hi}a_{hj} = 0$ for $h = 1, \ldots, i-1$.

Covariance selection (Dempster 1972) provides the theory for getting the MLE $\hat{\boldsymbol{\Sigma}}^*$ for the covariance structure $\boldsymbol{\Sigma}^*$ of a multivariate normal distribution whenever $(\boldsymbol{\Sigma}^*)^{-1}$ has zero structure with respect to some $I \subseteq \tilde{I}$. Dempster has shown that $\boldsymbol{\Sigma}^*(\hat{\boldsymbol{\Sigma}}^*)$ can be derived uniquely from the unrestricted matrix $\boldsymbol{\Sigma}(\mathbf{S})$ such that $\sigma^*_{ij} = \sigma_{ij}$ $(\hat{\sigma}^*_{ij} = s_{ij})$ for all $(i, j) \notin I$ and for $i = j$.

Proposition 3: For every reducible $I \subseteq \tilde{I}$ and every pair $((\mathbf{A}, \mathbf{T}), \boldsymbol{\Sigma}^*)$, the following two statements are equivalent:

1. $\mathbf{A}$ has zero structure with respect to I, and
2. $(\boldsymbol{\Sigma}^*)^{-1}$ has zero structure with respect to I.

If $\mathbf{A}$ and $\boldsymbol{\Sigma}^{*-1}$ have zero structure with respect to a reducible I, then the regression coefficients and their least squares estimates in the ith equation are identical to the elements in the ith row of the triangular decomposition of $\boldsymbol{\Sigma}^*$ and $\mathbf{S}^*$, respectively, obtained by successively orthogonalizing $\boldsymbol{\Sigma}^*$ and $\mathbf{S}^*$ in the order $p, p-1, \ldots, 1$.

To see that (1) implies (2), note from (2.2) that we can define $\boldsymbol{\Sigma}^{*-1} = \mathbf{A}^T\mathbf{T}^{-1}\mathbf{A}$, which is of the same form as (2.8). Thus if $\mathbf{A}$ has a reducible zero pattern, then we get for each (i, j) with $a_{ij} = 0$ that $\sigma^{*ij} = 0$.

Conversely, we know from Proposition 2 that the triangular decomposition $\mathbf{B}^*$ of $\boldsymbol{\Sigma}^*$ has the same zero pattern as $(\boldsymbol{\Sigma}^*)^{-1}$ if I is reducible. Thus for a proof in the direction (2) to (1), we only need to show that the negative values of b_{ij}^* are identical to the regression coefficients in the regression of Z_i on Z_j for $(i, j) \notin I$. The inner product of row ℓ of $\boldsymbol{\Sigma}_{(1,\ldots,i)}{}^{*-1}$ with row j of $\boldsymbol{\Sigma}_{(1,\ldots,i)}{}^*$ is known to be zero for all $j \neq \ell$. By using (2.5) we can write these inner products as

$$\sigma_{i\ell}{}^* = -\sum_{k=i+1}^{p} b_{ik}{}^*\sigma_{k\ell}{}^* = -\sum_{(i,k)\notin I} b_{ik}{}^*\sigma_{k\ell}{}^* \quad \text{for} \quad \ell = i+1, \ldots, p . \qquad (2.11)$$

The second equality follows from Proposition 2, since $b_{ik}{}^* = 0$ for all $(i, k) \in I$, if I is reducible. Further, since $\sigma_{ij}{}^* = \sigma_{ij}$ for $(i, j) \notin I$ and since by the reducibility of I we know that $(i, \ell) \notin I$ and $(i, k) \notin I$ implies $(k, \ell) \notin I$, for $i < k < \ell$, we get

$$\sigma_{i\ell} = -\sum_{(i,k)\notin I} b_{ik}{}^*\sigma_{k\ell} \quad \text{for} \quad (i, \ell) \notin I . \qquad (2.12)$$

But (2.12) is just the parameter equivalent of the normal equations satisfied by the negative values of least squares coefficients in regressing Z_i on Z_j for $(i, j) \notin I$; we have established the claimed identity if I is reducible. The same argument applies if $\boldsymbol{\Sigma}^*$ is replaced by its sample equivalent $\hat{\boldsymbol{\Sigma}}^*$, hence the proof is complete.

3. COVARIANCE SELECTION

We know from the previous section that a covariance selection model with reducible zero pattern in the concentration matrix can equivalently be described by a linear recursive system $(\mathbf{A}, \mathbf{T})$ that has the same zero pattern for the regression coefficients, and vice versa. A concentration is known to relate to the partial correlation coefficient of a variable pair given all the remaining variables as

$$\rho_{ij.1,\ldots,i-1,i+1,\ldots j-1,j+1,\ldots,p} = -\sigma^{ij}/(\sigma^{ii}\sigma^{jj})^{\frac{1}{2}} . \qquad (3.1)$$

Therefore, a zero concentration for a pdc $\boldsymbol{\Sigma}$ is equivalent to conditional independence of this pair given the $p-2$ other variables, which has been termed *zero partial association* (ZPA), elsewhere (Birch 1964; Wermuth 1976a, b).

We do not know yet which subclass of covariance selection models can be characterized by a reducible zero pattern in the concentrations. Therefore, we want to show the following in this section:

1. Every incomplete system $(\mathbf{A}, \mathbf{T})$ with reducible zero pattern can equivalently be described by a de-

composable or multiplicative covariance selection model, in which decomposability means that the joint probability density function factors into the product of certain marginal probability density functions.

2. Every decomposable covariance selection model can, after a proper reordering of the variables, be described by an incomplete system (**A**, **T**) with reducible zero pattern (Propositions 4 and 5).
3. The decomposition rule can easily be derived from a given reducible zero pattern (Proposition 6).
4. A reducible zero pattern facilitates computation of the MLE's of the parameters in a covariance selection model in the sense that they become closed form and can be derived from least squares estimates in the corresponding incomplete system (Proposition 7).

3.1 Multiplicative Models

Multiplicative or decomposable models have been studied extensively for contingency tables (Goodman 1970; Bishop 1971; Haberman 1974; Sundberg 1975; Darroch, Lauritzen and Speed, 1980) and to some extent for covariance selection models (Wermuth 1976a, b, Speed 1976, 1978). We restate Sundberg's criterion for decomposability as the definition of a multiplicative model.

Let $N_0 = \{1, \ldots, p\}$ and let $\{N_1, \ldots, N_T\}$ be a class of subsets of N_0 such that no member N_t of the class is a subset of any other member of the class. $\{N_t\}$ denotes indices of those subgroups of $\mathbf{Z} = (Z_1, \ldots, Z_p)$ that jointly generate the distribution of $\mathbf{Z}$.

Definition 3: A *model is decomposable* if and only if the variables can be ordered in such a way that each N_t contained in the class $\{N_t\}$ is composed of one set of elements V_t, which are missing in all N_s for $s > t$ and one set of elements $W_t = N_t \cap (\cup_{s>t} N_s)$ that is contained in some N_r for $r > t$ (Sundberg 1975).

For instance, $\{N_t\} = \{\{1, 4\}, \{1, 3, 5, 6\}, \{2, 3, 5\}\}$ is the generating class of a multiplicative model, since we can find $V_1 = \{4\}$, $W_1 = \{1\}$, $V_2 = \{1, 6\}$, $W_2 = \{3, 5\}$, $V_3 = \{2, 3, 5\}$, but $\{N_t\} = \{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$ does not satisfy Sundberg's criterion.

We make use of the fact that there are dual characterizations of each multiplicative model, one in terms of its generating class $\{N_t\}$, the other in terms of a list $I \subseteq \tilde{I}$ with $(i, j) \in I$ if and only if $\{ij\} \nsubseteq N_t$ for all N_t, where I denotes those pairs in the given model that have ZPA. A simple rule has been given previously (Wermuth 1976b) to derive $\{N_t\}$ from any $I \subseteq \tilde{I}$ or to decide that I does not characterize a multiplicative model.

Proposition 4: If $I \subseteq \tilde{I}$ is the list of ZPAs of a covariance selection model, then the following statements are equivalent:

1. The covariance selection model is a multiplicative model;
2. There is no subset of $n \geq 4$ indices such that exactly n pairs are not contained in I, and these pairs can be written as $(1, 2), (2, 3), \ldots, (i-1, i), \ldots, (n-1, n), (1, n)$;
3. There is no subset of $n \geq 4$ indices such that exactly $n(n-3)/2$ pairs are contained in I and such that each ordering of the indices leads to $(1, k), (k-1, \ell)$, for $k = 3, \ldots, n-1$ and $\ell = k+1, \ldots, n$; and
4. There is an ordering of the variables such that I is reducible.

Before we prove this proposition, we give an example. With the purpose of simplification, we use the index groups in $\{N_t\}$ separated by dashes as the notation of a multiplicative model. Then, model 15/24/345 has as generating class $N_t = \{\{1, 5\}, \{2, 4\}, \{3, 4, 5\}\}$. The variables are ordered to satisfy Sundberg's criterion. The index pairs that do not appear jointly in any of the index groups of the model notation give the list of ZPAs: $I = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 5)\}$. I is reducible. For covariance selection, 15/24/345 specifies a model in which the density $f(Z_1, Z_2, Z_3, Z_4, Z_5)$ of $\mathbf{Z}$ is generated from the marginal joint densities of (Z_1, Z_5), (Z_2, Z_4), and (Z_3, Z_4, Z_5) as

$$\begin{aligned} f(Z_1, Z_2, Z_3, Z_4, Z_5) &= \frac{f_{1,5}(Z_1, Z_5)}{f_5(Z_5)} \frac{f_{2,4}(Z_2, Z_4)}{f_4(Z_4)} f_{3,4,5}(Z_3, Z_4, Z_5) \; ; \\ &= f_1(Z_1|Z_5) \quad f_2(Z_2|Z_4) \quad f_{3,4,5}(Z_3, Z_4, Z_5) \; . \end{aligned} \tag{3.2}$$

From the factorization rule for densities (e.g., Wilks 1962),

$$f(Z_1, \ldots, Z_p) = \Big(\prod_{i=1}^{p-1} f_i(Z_i|Z_{i+1}, \ldots, Z_p)\Big) f_p(Z_p) \; ; \tag{3.3}$$

it is known that the right-hand side of (3.2) in fact generates a five-dimensional distribution, if $f_1(Z_1|Z_5) = f_1(Z_1|Z_2, Z_3, Z_4, Z_5)$ and $f_2(Z_2|Z_4) = f_2(Z_2|Z_3, Z_4, Z_5)$. From the proof of Proposition 3, we know this holds for model 15/24/345, since its concentration matrix has zero structure with respect to the reducible I, specified previously as list of pairs with ZPA. The system of linear recursive equations corresponding to the given ordering of the variables is

$$\begin{aligned} Z_1 + \qquad\qquad\quad a_{15}Z_5 &= U_1 \\ Z_2 + \quad a_{24}Z_4 \qquad\quad &= U_2 \\ Z_3 + a_{34}Z_4 + a_{35}Z_5 &= U_3. \end{aligned} \tag{3.5}$$

For the proof of Proposition 4, note that the pairs listed in statement (2) have been called a closed loop by Bishop, Fienberg, and Holland (1975) and an n-cycle by Lauritzen, Speed, and Vijayan (1976); a proof of the equivalence of (1) and (2) can be found in the latter. Statement (3) describes the pairs in the complementary set to the subset of $\binom{n}{2}$ pairs that satisfy (2) so that (2) and (3) are equivalent. Whenever there is a subset as

described in (3) contained in I, then each ordering of the indices gives three pairs: $(2, n) \in I$, but $(1, 2) \notin I$ and $(1, n) \notin I$ and hence I cannot be reducible, which proves that (4) implies (3). To show that (1) implies (4), we again use Sundberg's criterion (Definition 3). A different first proof for this last assertion is due to Speed (1978).

Let $\{N_t\}$ be ordered to satisfy Sundberg's criterion, then we can rewrite the list I of ZPAs as

$$I = \{(i, j) \mid i \in V_t, \quad j \in C_t\}, \quad \text{with} \quad C_t = (N_0 - (\cup_{s \le t} V_s)) - W_t. \tag{3.6}$$

For $h \in V_t$ and $k \in V_r$, $r > t$, we need to prove that $(k, j) \in I$ implies $(h, k) \in I$ and/or $(h, j) \in I$. Note that by definition $V_t \cap V_r = \emptyset$, and either $W_t \cap N_r = \emptyset$ or $(W_t \cap N_r) \subseteq N_r$. If this intersection is empty, then it must be true that $k \in C_t$, and hence $(k, j) \in I \Rightarrow (h, k) \in I$. If the intersection $W_t \cap N_r$ is a subset of N_r, then it follows that $C_r \subseteq C_t$, and hence $(k, j) \in I \Rightarrow (h, j) \in I$. Thus if the indices for the variables have been reordered so that Sundberg's criterion is satisfied, then the concentration matrix has a reducible zero pattern. This completes the proof, and we are prepared to relate linear recursive systems $(\mathbf{A}, \mathbf{T})$ as defined in (2.2) to multiplicative models.

Proposition 5: A covariance selection model is multiplicative (decomposable) if and only if the variables can be reordered to imply an incomplete system of recursive equations in which the regression coefficients have the same reducible zero pattern as the concentration matrix.

The proof follows by using the equivalence of (1) and (4) in Proposition 5 and by applying Proposition 3.

Further, from any reducible set $I \subseteq \bar{I}$, we can derive the corresponding incomplete recursive equations as well as the generating class of the corresponding multiplicative model by using Proposition 6. Let k' be such that $I_{(1,\ldots,k')} = 0$ for all $i \ge k'$.

Proposition 6: For every reducible $I \subseteq \bar{I}$ such that $\mathbf{A}$ (respectively $(\mathbf{\Sigma}^*)^{-1}$) has zero structure with respect to I

1. The index set of the variables in the ith equation of the system $(\mathbf{A}, \mathbf{T})$ is

$$M_i = \{i\} \cup \{j \mid j > i \text{ and } (i, j) \notin I\}, \quad \text{for } i < k'; \qquad = \{i, i+1, \ldots, p\}, \quad \text{for } i \ge k' \tag{3.7}$$

2. The generating class $\{N_t\}$ of the multiplicative model with covariance matrix $\mathbf{\Sigma}^*$ is obtained by deleting all members from the class $\{M_1, \ldots, M_{k'}\}$ that are subsets of another member of this class.

For a proof, note that M_i just lists indices of variable pairs from $(Z_i, \ldots, Z_p)$ that are not affected by the zero restrictions, and the class derived in (2) implies the correct list I of ZPAs and satisfies Sundberg's criterion for decomposability.

3.2 Closed-Form MLE for Decomposable Covariance Matrices

In multiplicative models for contingency tables, the MLE's have been given in closed form (Goodman 1970). It has been asserted (Sundberg 1975) that the same can be done for all members of the exponential family. In the case of covariance selection, MLE's have been stated explicitly for all multiplicative models for four variables (Wermuth 1976a). Now, we can give the following result for an arbitrary number of variables.

Definition 4: A *covariance matrix is decomposable* if it gives the covariance structure of a multiplicative model in covariance selection.

Proposition 7: If the variables of a decomposable covariance matrix $\mathbf{\Sigma}$ have been reordered so that $\mathbf{\Sigma}^{-1}$ has zero structure with respect to a reducible I, then

1. The MLE $\hat{\mathbf{\Sigma}}$ can be built up in the order $p, p-1, \ldots, 1$ from

$$\hat{\sigma}_{ij} = \mathbf{z}_i^T \mathbf{z}_j / n \quad \text{for } (i, j) \notin I \text{ and } i = j; \qquad = -\sum_{(i,k) \notin I} \hat{a}_{ik} \hat{\sigma}_{jk}, \quad \text{for } (i, j) \in I. \tag{3.8}$$

2. If in addition no member of $\{M_1, \ldots, M_{k'}\}$ is a subset of another member in this class and $(\cup_{i \ne j} (M_i \cap M_j)) \subseteq M_{k'}$, we get

$$n\hat{\sigma}_{ij} = \mathbf{z}_i^T \mathbf{z}_j \quad \text{for } (i, j) \notin I \text{ and } i = j; \qquad = -\sum_{(i,k) \notin I} \hat{a}_{ik} \mathbf{z}_i^T \mathbf{z}_j \quad \text{for } (i, j) \in I \text{ and } j \ge k'; \qquad = \sum_{(i,k) \notin I, (j,\ell) \notin I} \hat{a}_{ik} \hat{a}_{j\ell} \mathbf{z}_k^T \mathbf{z}_\ell \quad \text{for } (i, j) \in I \text{ and } j < k'; \tag{3.9}$$

where M_i and k' have been defined for the corresponding incomplete system $(\mathbf{A}, \mathbf{T})$ for proposition 6 and $\hat{a}_{ik}$ are the least squares estimates in the ith equation of this system.

For the proof, note from Dempster's results that $\hat{\sigma}^{ij} = 0$ for $(i, j) \in I$ and $\hat{\sigma}_{ij} = \mathbf{z}_i^T \mathbf{z}_j / n$ for $(i, j) \notin I$ and $i = j$ define the unique MLE $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$. But we know from the proof of Proposition 3, especially from (2.11), that $\hat{\sigma}_{ij}$ as defined in (3.8) implies $\hat{\sigma}^{ij} = 0$ for $(i, j) \in I$, hence the proof of (1) is complete. Further, if the conditions in (2) are satisfied we know for $(i, k) \notin I$ and $j \ge k'$ that $(k, j) \notin I$ so that the elements in position (k, j) of $n\hat{\mathbf{\Sigma}}$ have to coincide with the observed inner products $\mathbf{z}_k^T \mathbf{z}_j$. Similarly, for $j < k'$ we know that $(i, k) \notin I$ and $(j, \ell) \notin I$ define for a reducible I the pairs (k, ℓ) with the property $\{k, \ell\} \subset (M_i \cap M_j) \subset M_{k'}$, and hence by the definition of $M_{k'}$, (k, ℓ) is not contained in I. This completes the proof.

In order to illustrate our results, we list in Table 1 all 26 multiplicative models for five variables that differ in the constellation of their ZPAs. Thereby, we used the following reordering rule to achieve reducibility of I:

1. Multiplicative Models for Five Variables

Model Number	Number of ZPA's[a]	Model Notation
1	1	1345/2345
2	2	145/2345
3	3	15/2345
4	3	145/245/345
5	3	145/235/345
6	4	1/2345
7	4	15/245/345
8	4	15/234/345
9	5	1/245/345
10	5	15/25/345
11	5	15/24/345
12	6	1/25/345
13	6	15/25/35/45
14	6	15/25/34/45
15	7	1/2/345
16	7	1/25/35/45
17	7	1/25/34/45
18	8	1/2/35/45
19	9	1/2/3/45
20	10	1/2/3/4/5
21	3	125/345
22	4	12/25/345
23	5	13/24/35/45
24	5	12/345
25	6	12/35/45
26	7	1/23/45

[a] Zero partial association.

Reordering rule: For any list $L = I \subseteq \tilde{I}$ of ZPAs of a decomposable covariance matrix, repeat the following two steps in the order $l = 1, 2, \ldots, p$:

1. Label with "l" the variable that is contained most frequently in L. If there are ties, make an arbitrary choice.
2. Define a new set L by deleting from the old set L all pairs (i, j) with $i = l$ or $j = l$.

All models in Table 1 have a reducible list of ZPAs. The first 20 models all satisfy the conditions of statement (2) in Proposition 7. Thus each term in the model notation N_t can also be viewed as a list M_i of variables in the ith equation of the corresponding incomplete system, and in addition all pairwise intersection of the terms in the model notation are contained in the last term of the model notation. These conditions will be satisfied if for any two pairs $(i, j) \notin I$ and $(k, l) \notin I$ for $1 \leq i < j < k < l \leq p$, it never happens that the following four pairs are all contained in I: (i, k), (i, l), (j, k), and (j, l). Since this does not hold in any one of the last six models in Table 1, the MLE for their decomposable covariance matrix can only be computed by using the recursive definition (3.8).

So far, iterative methods have been proposed (Dempster 1973, Wermuth and Scheidt 1977) for computing the MLE of the covariance structure in a covariance selection model. Now, we have given a recursive closed form (3.8) for the large subclass of decomposable covariance matrices and a nonrecursive closed form (3.9) for decomposable covariance matrices that satisfy some additional restrictions. In the next section we show how these are related to the path analysis rules for computing so-called implied correlations.

4. PATH ANALYSIS

The geneticist Sewell Wright (1923, 1934) proposed path analysis as a "method for working out the logical consequences of a hypothesis as to the causal relations in a system of correlated variables" (Wright 1923, p. 254). Some of these systems can be presented as a diagram with p points, each of which denotes an observable variable.

Definition 5: A path diagram for a recursive system with uncorrelated errors consists of p points such that each pair of points is either unconnected or connected by a two-headed arrow or by a one-headed arrow.

The linear dependency of variable Z_i from variable Z_j is called a direct path and is drawn as a one-headed arrow coming from j and pointing at i. An unanalyzed correlation between Z_i and Z_j is shown as a two-headed arrow between i and j, and a missing path between two points is to characterize an indirect dependency.

For a given system of k linear recursive equations (2.1), a path diagram can easily be drawn: For each pair $(i, j) \in \tilde{I}$, exactly one of the following three possibilities holds: (a) $i > k$, implying a two-headed arrow between i and j; (b) $a_{ij} \neq 0$, implying a one-headed arrow from j pointing at i; (c) $a_{ij} = 0$, implying no direct connecting line between i and j.

Proposition 8: A path diagram for $\mathbf{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ implies a complete system of $k < p$ linear recursive equations (2.1) if the following two statements are satisfied:

1. There is a set $\{k + 1, \ldots, p\}$ of $p - k$ points such that each pair (i, j) from this set with $i \neq j$ is connected by a two-headed arrow;
2. There are k points at which one-headed arrows are directed, and these points can be ordered such that for each $i \in \{1, \ldots, k\}$ the $p - i$ arrows pointing at i begin at j for $j = i + 1, \ldots, p$.

Proposition 8 follows from Definitions 5 and (2.1) and from the fact (cf. proof of Proposition 1) that the defined system has uncorrelated errors. As a consequence, our results in Sections 2 and 3 apply to this type of path analysis models. In particular, they imply the definition of a reducible zero pattern for path analysis.

Definition 6: A *path diagram has a reducible zero pattern* if (1) replacing all missing paths by one-headed arrows leads to a complete recursive system (2.1) and (2) the pairs with missing paths (labeled in reference to a complete recursive system) define a reducible set I (Definition 1).

Wright presented rules for computing so-called path coefficients $\hat{\alpha}_{ij}$ and implied correlations r_{ij}^*, from observed correlation coefficients r_{ij}. Furthermore, he had suggested concluding that the model assumptions are plausible if the deviations between observed and im-

plied correlations are small. We state the rules in our notation:

$$r_{ij} = \sum_k \hat{\alpha}_{ik} r_{kj} \tag{4.1}$$

$$\text{with } \hat{\alpha}_{ik} \text{ as given from (4.1)}, \; r_{ij}^* = \sum_k \hat{\alpha}_{ik} r_{kj}^* \tag{4.2}$$

where k denotes all points in the corresponding path diagram from which one-headed arrows are leading to point i. Equation (4.1) is known (Tukey 1954) to give the normal equations satisfied by least squares estimates of standardized regression coefficients α_{ij}. A standardized regression coefficient is defined as a usual regression coefficient for standardized variables $Z_i^* = Z_i/\sigma_{ii}^{\frac{1}{2}}$, and the covariance of Z_i^* and Z_j^* is just the correlation coefficient ρ_{ij}. Equation (4.2) has been called "an informal device with intuitive appeal" (Land 1973, p. 46). We have given an example (2.6) in which (4.2) in fact does not define the MLE of the correlation matrix, since for instance for the element in position (1, 4) we obtained from (4.2) and (4.1): $r_{14}^* = \sum_{k=2}^4 \hat{\alpha}_{1k} r_{k4} = r_{14}$, but the MLE $\hat{\rho}_{14} = \hat{\sigma}_{14}^*/(\hat{\sigma}_{11}^* s_{44})^{\frac{1}{2}} \neq r_{14}$. But, we also know from (3.8) a condition under which (4.2) always gives the MLE of a correlation matrix and has a theoretical justification from the theory of covariance selection.

Proposition 9: A sufficient condition that Wright's rule for computing implied correlations defines the MLE of a correlation matrix is that the path diagram has a reducible zero pattern and that the order for applying the rule is $i = p - 1, p - 2, \ldots, 1$. (The MLE then deviates from the observed correlation matrix only in positions corresponding to missing paths.)

For the proof, note first that correlations and standardized regression coefficients can be obtained by the following one-to-one transformations from the covariance matrix $\mathbf{\Sigma}$, the concentration matrix $\mathbf{\Sigma}^{-1}$, and the matrix $\mathbf{A}$ of unstandardized coefficients:

$$\rho_{ij} = \sigma_{ij}(\sigma_{ii}\sigma_{jj})^{-\frac{1}{2}}, \quad \rho^{ij} = \sigma^{ij}(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}};$$
$$\alpha_{ij} = a_{ij}\sigma_{jj}^{\frac{1}{2}}\sigma_{ii}^{-\frac{1}{2}}, \quad \text{for} \quad 1 \leq i < j \leq p; \tag{4.3}$$

and similarly for observed correlations, since $r_{ij} = (\mathbf{z}_i^T\mathbf{z}_j)((\mathbf{z}_i^T\mathbf{z}_i)(\mathbf{z}_j^T\mathbf{z}_j))^{-\frac{1}{2}}$. Then the proof follows from Definition 6, (3.8), and the invariance property of MLE under one-to-one transformations (Anderson 1958).

Furthermore, if the path diagram has a reducible zero pattern, then the following likelihood-ratio (LR) test statistic for the hypothesis of τ zero concentrations (Dempster 1972) is a measure for the deviations between implied and observed correlations. Let $\mathbf{R}$ denote the observed correlation matrix and $\hat{\mathbf{P}}$ the MLE under the assumption that $\mathbf{P}^{-1}$ has zero structure with respect to the reducible set $I \subseteq \bar{I}$ containing τ elements, then

$$\text{LR} - \chi^2 = n(\log \det\hat{\mathbf{P}} - \log \det\mathbf{R}) \tag{4.4}$$

can—for large sample sizes n—be regarded as roughly χ^2 distributed on τ degrees of freedom, and $\hat{\mathbf{P}}$ and $\mathbf{R}$ are identical except for the elements in positions $(i, j) \in I$; those may be computed from (4.2).

5. A SEARCH METHOD FOR FITTING RECURSIVE EQUATIONS

Suppose that a complete system of $k < p$ variables (2.1) is given. Then one can ask which of the regression coefficients can simultaneously be set to zero such that the obtained incomplete system of recursive equations is still compatible with the observations? After a small modification, model search among multiplicative models (Wermuth 1976b) can provide an answer to this question: The variables in the set $\{Z_k, \ldots, Z_p\}$ have to be specified as those variable whose correlations are not to be accounted for by other relations. The effect of such a specification is that only variable pairs (Z_i, Z_j) for $i \in \{1, \ldots, k'\}$ with $k' = k - 1$ and for $j \in \{i + 1, \ldots, p\}$ can have ZPA (cf. Section 3), whereas the estimated correlations for all variable pairs from $\{Z_k, \ldots, Z_p\}$ have to coincide with the observed correlation coefficients (cf. Propositions 7 and 9). Then we know from Proposition 6 how to obtain the incomplete recursive equations that correspond to any well-fitting multiplicative model with reducible list of ZPAs.

Model search among multiplicative models is used to find a set of variable pairs that can be assumed to have ZPA. In covariance selection an appropriate measure for ZPA is the partial correlation coefficient given all $p - 2$ remaining variables (3.1). Many pairs with ZPA are considered to be desirable because they reduce the number of interdependent variables and therefore tend to simplify the interpretation of a multivariate correlation structure. The search procedure has been programmed (Wermuth, Wehner, and Gönner 1976) and applied to subject matter problems in medicine (Wermuth 1978) and in psychology (Wermuth, Hodapp, and Weyer 1976), Likelihood-ratio statistics as defined in (4.4) are used to judge the fit of the model assumptions. A theoretical justification for the search procedure is contained in Sundberg's (1975) results about the partitioning of tests.

We illustrate the modified search with a set of sociological data (Goldberg 1971). Goldberg's "analysis is concerned with making inferences about the pattern of causal relationships among six variables: father's sociological characteristics (FSC); father's party identification (FPI); respondent's sociological characteristics (RSC); respondent's party identification (RPI); respondent's partisan attitudes (RPA); and respondent's vote for President in 1956 (RV)" (Goldberg 1971, pp. 35–36). We have numbered the variables in reverse order and display in Table 2 the observed simple correlations r_{ij}, as well as the observed measures for ZPA: $r_{ij.klrs} = -r^{ij}/(r^{ii}r^{jj})^{\frac{1}{2}}$. We are only interested in equations for the first four variables, therefore we specify

2. Observed Correlations Among Variables on Voting Behavior

Variables	1	2	3	4	5	6
1. RV	—	.742	.722	.271	.466	.282
2. RPA	.469	—	.710	.289	.453	.318
3. RPI	.365	.328	—	.411	.603	.400
4. RSC	−.021	−.053	.144	—	.420	.808
5. FPI	.054	.003	.388	.024	—	.454
6. FSC	−.019	.081	−.026	.755	.173	—

NOTE: Upper half: r_{ij}; lower half: $r_{ij \cdot klrs}$; $n = 645$.

Z_4 (= RSC) Z_5 (= FPI), and Z_6 = (FSC) as variables whose correlations are to stay fixed. It should be noted that the assumption of multivariate normality is at best crudely approximated by these data.

The search procedure applied to the $\binom{6}{2} - \binom{3}{2}$ pairs leads to the results displayed in Table 3. At step r of the search, a statistic with 1 *df* is computed for each pair available for selection at this step. These are test statistics for the hypotheses that pair (Z_i, Z_j) has ZPA given that $r - 1$ pairs—selected at previous steps—already have ZPA. The pair with the smallest statistic is selected, then the likelihood-ratio statistic (4.4) for the test of r ZPAs is simply the sum of the r-selected statistics on 1 *df*. In analogy to tests of hypotheses, the fit of a model is judged to be poor if the fractile value of a statistic falls below .05 or .01. Thus, for the given set of data, the fit starts to be poor at step 8, whereas the model with $I = \{(1, 4), (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 6)\}$ is judged to be compatible with the observations. This set $I \subseteq \bar{I}$ is reducible so that we can use Proposition 6 to get model 123/345/456 as a well-fitting multiplicative model and to obtain as well-fitting incomplete recursive system:

$$\begin{aligned} Z_1 + a_{12}Z_2 + a_{13}Z_3 &= U_1 \\ Z_2 + a_{23}Z_3 &= U_2 \\ Z_3 + a_{34}Z_4 + a_{35}Z_5 &= U_3 \\ Z_4 + a_{45}Z_5 + a_{46}Z_6 &= U_4 \,. \end{aligned} \tag{5.1}$$

The interpretation of the multiplicative model is the following: Variables 1 and 2 jointly are conditionally independent of variables 4, 5, and 6 jointly, given variable 3; and variable 3 is conditionally independent of variable 6, given variables 4 and 5. This is to mean the following:

1. For voter groups that are homogeneous with respect to party identification (variable 3), knowledge of the sociological characteristics RSC, FPI, and PSC (variables 4, 5, 6) does not help to predict the vote or the partisan attitude prior to the election (variables 1 and 2).
2. For voter groups that are homogeneous with respect to the respondent's sociological characteristics and the father's party identification (variables 4 and 5), the sociological characteristics of the father (variable 6) do not help to predict the respondent's party identification (variable 3).

This correlation structure can equivalently be described by the system of incomplete recursive equations given in (5.1).

Although we discussed the step from interdependencies of associations to dependencies and vice versa, we have not touched the problems connected with the step from dependencies to causation. Instead, we refer the reader to Cochran's (1965) discussion of difficulties connected with observational studies on human populations.

3. Model Search Results

Step r	Selected Pair (i,j)	Test Statistic with 1 df[a]	Fractile value	Test Statistic with r df	Fractile value
1	(2,5)	.005	.94	.005	.94
2	(2,4)	1.802	.18	1.807	.41
3	(1,4)	1.775	.18	3.582	.31
4	(1,5)	2.427	.12	6.009	.20
5	(1,6)	1.049	.30	7.058	.22
6	(2,6)	1.792	.18	8.850	.18
7	(3,6)	.206	.65	9.056	.25
8	(3,4)	31.370	.00	40.426	.00
9	(2,3)	98.156	.00	138.582	.00
10	(3,5)	291.503	.00	430.085	.00
11	(1,3)	475.138	.00	905.223	.00
12	(1,2)	515.846	.00	1421.069	.00

[a] Degrees of freedom.

APPENDIX—PROOFS FOR (2.6) AND (2.7)

Since

$$f_{1,2,3,4}(Z_1, Z_2, Z_3, Z_4) = f_1(Z_1 | Z_2, Z_3, Z_4) f_{2,3,4}(Z_2, Z_3, Z_4) \,,$$

the maximization of the likelihood function for the left-hand side can be split into two independent maximizations, as long as restrictions do not affect both functions on the right-hand side simultaneously. Applying the notation of (2.2) to (2.9), we can write the resulting $\hat{\Sigma}^{-1}$ as

$$\hat{\Sigma}^{-1} = \begin{bmatrix} \hat{\vartheta}^{11} & \hat{a}_{12}\hat{\vartheta}^{11} & \hat{a}_{13}\hat{\vartheta}^{11} & \hat{a}_{14}\hat{\vartheta}^{11} \\ & \hat{\vartheta}^{22.1} + \hat{a}_{12}^2\hat{\vartheta}^{11} & \hat{\vartheta}^{23.1} + \hat{a}_{12}\hat{a}_{13}\hat{\vartheta}^{11} & \hat{\vartheta}^{24.1} + \hat{a}_{12}\hat{a}_{14}\hat{\vartheta}^{11} \\ & & \hat{\vartheta}^{33.1} + \hat{a}_{13}^2\hat{\vartheta}^{11} & \hat{\vartheta}^{34.1} + \hat{a}_{13}\hat{a}_{14}\hat{\vartheta}^{11} \\ & & & \hat{\vartheta}^{44.1} + \hat{a}_{14}^2\hat{\vartheta}^{11} \end{bmatrix} . \tag{A.1}$$

From (A.1) we obtain with the help of the sweep operator (SWP):

$$\mathrm{SWP}(1)\hat{\Sigma}^{-1} = \begin{bmatrix} -1/\hat{\vartheta}^{11} & \hat{a}_{12}\,\hat{a}_{13}\,\hat{a}_{14} \\ & (\hat{\Sigma}_{(1)})^{-1} \end{bmatrix}, \tag{A.2}$$

and $-\hat{\Sigma} = \mathrm{SWP}(1, 2, 3, 4)\,\hat{\Sigma}^{-1}$ as

$$-\hat{\Sigma} = \begin{bmatrix} -1/\hat{\vartheta}^{11} - (\hat{a}_{12}\,\hat{a}_{13}\,\hat{a}_{14})\hat{\Sigma}_{(1)}(\hat{a}_{12}\,\hat{a}_{13}\,\hat{a}_{14})^T & (\hat{a}_{12}\,\hat{a}_{13}\,\hat{a}_{14})\hat{\Sigma}_{(1)} \\ & -\hat{\Sigma}_{(1)} \end{bmatrix}, \tag{A.3}$$

and from these we can immediately write down the MLE for the case with no restrictions (a), one restriction $a_{23} = 0$ (b), and two restrictions $a_{12} = a_{23} = 0$ (c).

Case a: If there are no restrictions at all, it is known that

$$(\hat{\alpha}_{12}\ \hat{\alpha}_{13}\ \hat{\alpha}_{14}) = -(s_{12}\ s_{13}\ s_{14})(\mathbf{S}_{(1)})^{-1},$$

$$-(s_{12}\ s_{13}\ s_{14})(\mathbf{S}_{(1)})^{-1}(s_{12}\ s_{13}\ s_{14})^T \quad \text{and} \quad \hat{\mathbf{\Sigma}}_{(1)} = \mathbf{S}_{(1)},$$

so that (A.3) gives $\hat{\mathbf{\Sigma}} = \mathbf{S}$ as it should.

Case b: If the only restriction is $a_{23} = \sigma_{23.4} = 0$, then $(\hat{\alpha}_{12}\ \hat{\alpha}_{13}\ \hat{\alpha}_{14})$ and $1/\hat{\sigma}^{11}$ are unrestricted as under Case a, but

$$\hat{\mathbf{\Sigma}}_{(1)} = \begin{bmatrix} s_{22} & s_{24}s_{34}/s_{44} & s_{24} \\ & s_{33} & s_{34} \\ & & s_{44} \end{bmatrix}$$

$$(\hat{\mathbf{\Sigma}}_{(1)})^{-1} = \begin{bmatrix} 1/s_{22.4} & 0 & -s_{14}/s_{44}s_{22.4} \\ & 1/s_{33.4} & -s_{34}/s_{44}s_{33.4} \\ & & 1/s_{44.3} - s_{24}^2 s_{44}^2/s_{22.4} \end{bmatrix}$$

as is known from covariance selection (cf. Wermuth 1976a). Hence the MLE given in (2.6) follows from substitution in (A.3).

Case c: If there are two restrictions and these are $a_{23} = \sigma_{23.4} = 0$ and $a_{12} = \sigma_{12.34} = 0$, then $\hat{\mathbf{\Sigma}}_{(1)}$ is as given in Case b, but $\hat{a}_{12} = 0$ and

$$(\hat{a}_{13}\ \hat{a}_{14}) = -(s_{13}\ s_{14})(\mathbf{S}_{(1,2)})^{-1},$$

$$1/\hat{\sigma}^{11} = s_{11} - (s_{13}\ s_{14})(\mathbf{S}_{(1,2)})^{-1}(s_{13}\ s_{14})^T$$

Substituting these in (A.3) gives the MLE in (2.7).

[*Received January 1978. Revised January 1980.*]

REFERENCES

Anderson, Theodore W. (1958), *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley & Sons.

Beaton, Albert E. (1964), "The Use of Special Matrix Operators in Statistical Calculus," unpublished EdD dissertation Harvard University, Dept. of Education. Reprinted Princeton, N.J.: Educational Testing Service Research Bulletin 64–51.

Birch, W.M. (1964), "The Detection of Partial Association, I. The 2 × 2 Case," *Journal of the Royal Statistical Society*, Ser. B, 26, 313–324.

Bishop, Yvonne M.M. (1971), "Effects of Collapsing Multi-dimensional Contingency Tables," *Biometrics*, 27, 545–562.

Bishop, Yvonne M.M., Fienberg, Stephen E., and Holland, Paul. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, Mass.: MIT Press.

Blalock, Hubert M., Jr. (ed.) (1971), *Causal Models in the Social Sciences*, Chicago: Aldine-Atherton.

Cochran, William G. (1965), "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society*, Ser. A, 128, 234–265.

Darroch, John N., Lauritzen, Steffen L., and Speed, Terry (1980), "Markov Fields and Log-linear Interaction Models for Contingency Tables," *Annals of Statistics*, 8, 522–539.

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

——— (1972), "Covariance Selection," *Biometrics*, 28, 157–175.

Duncan, Otis D. (1975), *Introduction to Structural Equation Models*, New York: Academic Press.

Goldberg, Arthur S. (1971), "Discerning a Causal Pattern Among Data on Voting Behaviour," in *Causal Models in the Social Sciences*, ed. H.M. Blalock, Jr., Chicago: Aldine-Atherton.

Goldberger, Arthur S. (1964), *Econometric Theory*, New York: John Wiley & Sons.

Goldberger, Arthur S., and Duncan, Otis D. (eds) (1973), *Structural Equation Models in the Social Sciences*, New York: Seminar Press.

Goodman, Leo A. (1970), "The Multivariate Analysis of Qualitative Data: Interaction Among Multiple Classifications," *Journal of the American Statistical Association*, 65, 226–256.

Haberman, Shelby J. (1974), *The Analysis of Frequency Data*, Chicago: The University of Chicago Press.

Heise, David R. (1975), *Causal Analysis*, New York: John Wiley & Sons.

Hodapp, Volker (1978), "Causal Inference From Nonexperimental Research on Anxiety and Educational Achievement," in *Achievement, Stress, and Anxiety*, eds. Heinz H. Krone and Lothar Laux, Washington D.C.: Hemisphere, in press.

Land, Kenneth C. (1973), "Identification, Parameter Estimation and Hypothesis Testing in Recursive Sociological Models," in *Structural Equation Models in the Social Sciences*, eds. A.S. Goldberger and O.D. Duncan, New York: Seminar Press, 19–49.

Lauritzen, Steffen L., Speed, Terry, and Vijayan, K. (1976), "Decomposable Graphs and Hypergraphs," *Proceedings of the Royal Statistical Society*, in press.

Li, Ching C. (1975), *Path Analysis: A Primer*, Pacific Grove, Calif.: The Boxwood Press.

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

Speed, Terry P. (1976), "Interaction," unpublished manuscript.

——— (1978), "Graph-theoretic Methods for the Analysis of Data," unpublished lecture notes.

Sundberg, Rolf (1975), "Some Results About Decomposable (or Markov-Type) Models for Multidimensional Contingency Tables: Distribution of Marginals and Partitioning of Tests," *Scandinavian Journal of Statistics*, 2, 71–79.

Tukey, John W. (1954), "Causation, Regression, and Path Analysis," in *Statistics and Mathematics in Biology*, eds. O. Kempthorne, T.A. Bancroft, J.W. Gowen, and J.L. Lush, Ames: The Iowa State College Press, 35–66.

Wermuth, Nanny (1976a), "Analogies Between Multiplicative Models in Contingency Tables and Covariance Selection," *Biometrics*, 32, 95–108.

——— (1976b), "Model Search Among Multiplicative Models," *Biometrics*, 32, 253–263.

——— (1978), *Zusammenhangsanalysen Medizinischer Daten*, Band 5, Lecture Notes in Medizinischer Informatik und Statistik, Berlin: Springer Verlag.

Wermuth, Nanny, Hodapp, Volker, and Weyer, Geerd (1976), "Die Methode der Kovarianzselektion als Alternative zur Faktorenanalyse, untersucht an Persönlichkeitsmerkmalen," *Zeitschrift für Experimentelle und Angewandte Psychologie*, 23, 320–338.

Wermuth, Nanny, and Scheidt, Eberhard (1977), "Fitting a Covariance Selection Model to a Matrix," Algorithm AS 105, *Journal of the Royal Statistical Society, Ser. C, Applied Statistics*, 26, 88–92.

Wermuth, Nanny, Wehner, Theo, and Gönner, Herbert (1976), "Finding Condensed Descriptions for Multidimensional Data," *Computer Programs in Biomedicine*, 6, 23–38.

Wilks, Samuel S. (1962), *Mathematical Statistics*, New York: John Wiley & Sons.

Wold, Herman O.A. (1954), "Causality and Econometrics," *Econometrica*, 22, 162–177.

——— (1960), "A Generalization of Causal Chain Models," *Econometrica*, 28, 443–463.

Wright, Sewall (1923), "The Theory of Path Coefficients: A Reply to Niles' Criticism," *Genetics*, 8, 239–255.

——— (1934), "The Method of Path Coefficients," *The Annals of Mathematical Statistics*, 5, 161–215.