Using Path Diagrams as a Structural Equation Modelling Tool

by Peter Spirtes, Thomas Richardson, Chris Meek, Richard Scheines, and Clark Glymour

1. Introduction

Linear structural equation models (SEMs) are widely used in sociology, econometrics, biology, and other sciences. A SEM (without free parameters) has two parts: a probability distribution (in the Normal case specified by a set of linear structural equations and a covariance matrix among the "error" or "disturbance" terms), and an associated path diagram corresponding to the functional composition of variables specified by the structural equations and the correlations among the error terms. It is often thought that the path diagram is nothing more than a heuristic device for illustrating the assumptions of the model. However, in this paper, we will show how path diagrams can be used to solve a number of important problems in structural equation modelling.

There are a number of problems associated with structural equation modeling. These problems include:

- How much do sample data underdetermine the correct model specification? Of course, one must decide how much credence to give alternative explanations that afford different fits to any particular data set. There are a variety of techniques for that purpose, including Bayesian updating, and a variety of fit measures with well understood large sample properties . But what about two or more alternative models that fit a specific data set equally well, or, subject to certain restrictions, fit any data set meeting the restrictions equally well? The number of such equivalents for a given linear structural equation model may be very large. Even if there are sources of knowledge about structure from outside the data set, the number of equivalent models all meeting those knowledge constraints may be considerable, and the structures they postulate may have importantly different implications for policy. Unless we characterize such equivalencies, selection of a particular model can only involve an element of arbitrary choice.

- Given that there are equivalent models, is it possible to extract the features common to those models? Under some circumstances, every member of a set of equivalent models may share some of the same linear coefficients or correlated errors. If that is the case, then it is possible that even though the data may not help us choose between the different models, the data may provide evidence for features common to all of the best models.

• When a modeler draws conclusions about coefficients in an unknown underlying structural equation model from a multivariate regression, precisely what assumptions are being made about the structural equation model? For example, when does a non-zero partial regression coefficient correspond to a non-zero coefficient in a structural equation?

These questions have been addressed many times, though usually only for models with special structures, and usually relying on linear algebra, the mathematics that seems most natural for a study of linear models. The aim of this paper is to explain how the path diagram provides much more than heuristics for special cases; the theory of path diagrams helps to clarify several of the issues just noted, issues that have been the focus of intelligent--if, in our judgment, ultimately too sweeping-- criticism of the use of structural equation models. What follows is a report that describes some of what has been learned about these issues by following a different set of mathematical ideas that exploit the graphical structure implicit in structural equation models.

In particular, we will present answers to these questions that depend upon an understanding of the relationship between the path diagram used to represent a structural equation model, and the zero partial correlations entailed by that path diagram (entailed in the sense that every structural equation model that shares the path diagram has a zero partial correlation). We will describe a graphical relation, the Pearl-Geiger-Verma d-separation criterion, among a pair of variables X and Y, and a set of variables **Z**, that is a necessary and sufficient condition for a structural equation model to entail a zero partial correlation. Such necessary and sufficient conditions have been known for path diagrams without correlated errors, but we will extend the conditions to path diagrams with correlated errors.

In section 2 we will motivate interest in the d-separation relation by describing the problems that it helps to solve in more detail. Then in section 3 we will show how the zero partial correlations entailed by a structural equation model can be read off from its path diagram, and in section 4 use the machinery developed in section 3 to provide some solutions to problems described in section 2. In section 6 we prove the main theorem, hitherto unpublished, which justifies the use of d-separation in path diagrams representing correlated errors (represented by edges of the form $\leftrightarrow$, which we call double-headed arrows).

2. Problems in SEM Modeling

In order to describe the problems listed in section 1 in more detail, we will first review how path diagrams are used to represent structural equation models without free parameters. The path diagram contains a directed edge from B to A if and only if there is a non-zero coefficient for B in the equation for A; and there is a double-headed arrow between A and B if and only if the error term for A and the error term for B have a non-zero correlation.[1] The path diagram associated with a SEM may contain directed cycles (representing feedback), and double-headed arrows (representing correlated errors.) We will call a path diagram which contains no double-headed arrows a **directed graph**. (We place sets of variables and defined terms in boldface.) In a SEM M, we will denote the correlation matrix among the non-error variables by **Σ(M)**, and the corresponding path diagram by **G(M)**. We will now review the problems mentioned in section 1 in more detail.

2.1. Covariance Equivalence

Consider the following example. The graph in Figure 1(a) is the path diagram of a SEM M proposed by Aberle (Blalock, 1961) as a model for evolutionary culture in American Indian tribes, where W is matridominant division of labor, X is matrilocal residence, Y is matricentered land tenure, and Z is matrilinear system of descent.

Suppose for the moment that there is a SEM with the path diagram in Figure 1(a) and the p(χ^2), the AIC (Aikake Information Criterion), and the BIC (Bayes Information Criterion) score for this SEM are all high[2] (See Raftery 1995 for a discussion of the BIC score.) In order to evaluate how well the data supports this model, it is still necessary to know whether or not there are other models compatible with background knowledge that fit the data equally well. (Lee and Hershberger 1990, Stelzl, 1986). In this case, for each of the path diagrams in Figure 1, and for *any* data set D, there is a SEM with that path diagram that fits D as well as M does (in the sense that each SEM has the same p(χ^2) and the same BIC and AIC scores.) If **O** represent the set of measured variables in path diagrams G_1 and G_2, then G_1 and G_2 are **covariance equivalent over O**[3] if and only if for every SEM M

---

[1] This is slightly different than the usual convention in which if ε_A and ε_B are correlated, then they are explicitly included in the graph, there is a directed edge from ε_A to A, a directed edge from ε_B, and the double-headed arrow is placed between ε_A and ε_B. However, the convention adopted here will simplify later theorems and proofs.

[2] In counting degrees of freedom, we will assume that a SEM with free parameters (and no latents) associates a linear coefficient parameter with each directed edge (i.e. $\rightarrow$) in its path diagram, a correlation parameter with each double-headed arrow (i.e. $\leftrightarrow$) in its path diagram, and a variance parameter with each vertex. We also assume that no extra constraints (such as equality constraints among parameters) are imposed.

[3] For technical reasons, a more formal definition requires a slight complication. G is a **sub-path diagram** of G' when G and G' have the same vertices, and G has a subset of the edges in G'. G_1 and G_2 are **covariance equivalent over O** if for every SEM M such that G(M) = G_1, there is a SEM M' with path diagram G(M') that is a sub-path diagram of G_2, and the marginal over **O** of Σ(M') equals the marginal over

such that $G(M) = G_1$, there is a SEM M' with path diagram $G(M') = G_2$, and the marginal of $\Sigma(M')$ over **O** equals the the marginal of $\Sigma(M)$ over **O**, and vice-versa. (Informally, any covariance matrix over **O** generated by a parameterization of path diagram G_1 can be generated by a parameterization of path diagram G_2, and vice-versa.) If G_1 and G_2 have no latent variables, (i.e all of the variables in their path diagrams are in **O**), then we will simply say that G_1 and G_2 are **covariance equivalent**. If two covariance equivalent models are equally compatible with background knowledge, and have the same degrees of freedom, the data does not help distinguish them, so it is important to be able to find the complete set of path diagrams that are covariance equivalent to a given path diagram. (Every SEM that contains a path diagram in Figure 1 has the same number of degrees of freedom.)

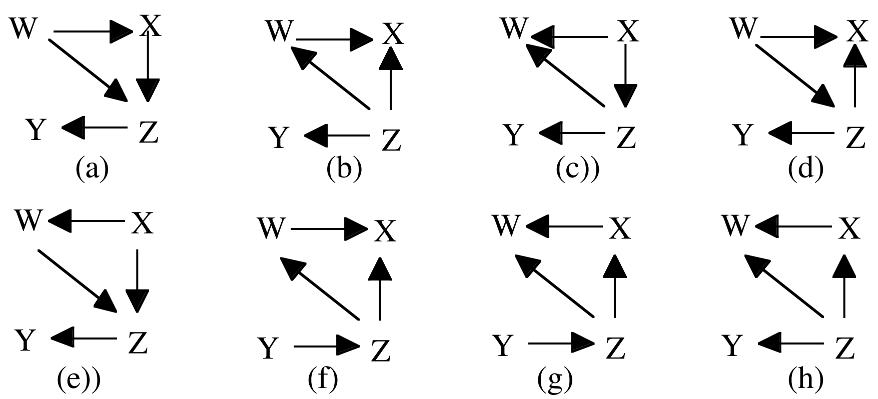


Figure 1

It is often far from obvious what constitutes a complete set of path diagrams covariance equivalent to a given path diagram. We will call such a complete set a **covariance equivalence class over O**. (Again, if we consider only SEMs without latent variables, we will call such a complete set a **covariance equivalence class**.) If it is a complete set of path diagrams without correlated errors or directed cycles, i.e. directed acyclic graphs, that are covariance equivalent we will call it a **simple covariance equivalence class over O**.) As shown in section 4, the path diagrams in Figure 1 are a simple covariance equivalence class.

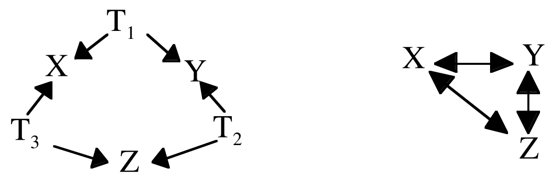
Another example of a case where it is not obvious whether or not two path diagrams are covariance equivalent over **O** is shown below. It is often thought that the two path diagrams in Figure 2 (each of which is part of a just-identified SEM) are covariance equivalent over **O** = {X,Y,Z}. However, as shown in Spirtes et al. (1996), there is a SEM with path diagram in Figure 2(b) with the covariance matrix Σ over X, Y, and Z, but there

---

O of $\Sigma(M)$, and for every SEM M' such that $G(M') = G_2$, there is a SEM M with path diagram G(M) that is a sub-path diagram of G_1, and the marginal over **O** of $\Sigma(M)$ equals the marginal over **O** of $\Sigma(M')$.

is no SEM that contains path diagram in Figure 2 (a) with marginal covariance matrix Σ (where T_1, T_2, and T_3 are latent variables).

$$\Sigma = \begin{pmatrix} 1.0 & 0.99 & 0.99 \\ 0.99 & 1.0 & 0.99 \\ 0.99 & 0.99 & 1.0 \end{pmatrix}$$



(a) (b)

Figure 2

In section 4, we will describe how to efficiently test when two path diagrams without correlated errors or directed cycles are covariance equivalent. We will also give informative necessary conditions for two path diagrams with correlated errors, cycles, or latent variables to be covariance equivalent over **O**. For related theorems see also Pearl (1997).

2.2. Features Common to a Covariance Equivalence Class

A second important question that arises with respect to covariance equivalence classes is whether it is possible to extract the features that the set of covariance equivalent path diagrams have in common. For example, every path diagram in Figure 1 has the same adjacencies, but the path diagrams do not have any edge with the same orientation in every member of the equivalence class (e.g. both $W \rightarrow X$, and $W \leftarrow X$ occur in path diagrams in Figure 1).

However, there are other sets of covariance equivalent path diagrams in which a given edge always occurs with the same orientation in every member of the equivalence class. For example, Figure 3 shows another simple covariance equivalence class of path diagrams in which the orientation $X \rightarrow Z$ occurs in every member of the equivalence class.

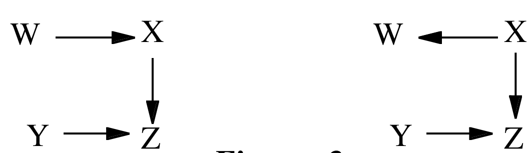


Figure 3

This is informative because even though the data does not help choose between members of the equivalence class, insofar as the data is evidence for the disjunction of the members in the equivalence class, it is evidence for the orientation $X \rightarrow Z$.

In section 4 we will show how to extract all of the features common to a simple covariance equivalence class of path diagrams, and briefly indicate that it is possible to extract some features common to a covariance equivalence class of path diagrams with correlated errors, cycles, or latent variables.

2.3. Regression Coefficients and Structural Equation Coefficients

It is common knowledge among practising social scientists that for the coefficient of X in the regression of Y upon X to be interpretable as the effect of X on Y there should be no "confounding" variable Z which is a cause of both X and Y:

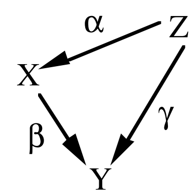


Figure 4

Simple calculations confirm this conclusion (using the notation in Figure 4)[4]:

$$\mathrm{Cov}(X,Y) = \beta V(X) + \alpha\gamma V(Z)$$

Hence

$$\frac{\mathrm{Cov}(X,Y)}{V(X)} = \frac{\beta V(X) + \alpha\gamma V(Z)}{V(X)} \neq \beta.$$

Thus the coefficient from the regression of Y on X alone will be a consistent estimator only if either α or γ is equal to zero. Further, observe that the bias term $\alpha\gamma V(Z)/V(X)$ my be either positive or negative, and of arbitrary magnitude.

However, $\mathrm{Cov}(X,Z) = \alpha V(Z)$ and $\mathrm{Cov}(Y,Z) = (\alpha\beta + \gamma)V(Z)$, and hence

$$\begin{aligned}\mathrm{Cov}(X,Y \mid Z) &\equiv \mathrm{Cov}(X,Y) - \frac{\mathrm{Cov}(X,Z)\mathrm{Cov}(Y,Z)}{V(Z)} \\ &= \beta V(X) + \alpha\gamma V(Z) - \alpha V(Z)(\alpha\beta + \gamma) \\ &= \beta(V(X) - \alpha^2 V(Z))\end{aligned}$$

and

$$V(X \mid Z) \equiv V(X) - \frac{\mathrm{Cov}(X,Z)^2}{V(Z)} = V(X) - \alpha^2 V(Z),$$

so the coefficient of X in the regression of Y on X and Z is a consistent estimator of β since $\mathrm{Cov}(X,Y|Z)/V(X|Z) = \beta$.

---

[4] Section 6 contains a simple rule for calculating covariances from a path diagram.

The danger presented by failing to include confounding variables is well understood by social scientists. Indeed, it is often used as the justification for considering a long “laundry list” of “potential confounders” for inclusion in a given regression equation.

What is perhaps less well understood is that including a variable which is not a confounder can also lead to biased estimates of the structural coefficient. We now consider a number of simple cases demonstrating this:

$$X \xrightarrow{\beta} Y \xrightarrow{\eta} Z$$

Figure 5

In the SEM with the path diagram depicted in Figure 5, Cov(X,Y) = βV(X), hence the coefficient of X in the regression of Y upon X is a consistent estimator of β. However, Cov(Y,Z) = ηV(Y), and Cov(X,Z) = $\beta\eta$V(X), so that

$$\frac{\mathrm{Cov}(X,Y|Z)}{V(X|Z)} = \beta\frac{V(Z) - \eta^2 V(Y)}{V(Z) - \beta^2\eta^2 V(X)} = \beta\left(\frac{V(\varepsilon_Z)}{V(\varepsilon_Z) + \eta^2 V(\varepsilon_Y)}\right)$$

Hence the coefficient of X in the regression of Y on X and Z is an inconsistent estimator of β. The estimate will have the same sign as β, but will have smaller absolute magnitude. Note that Cov(X,Y|Z)/V(X|Z) = 0 if and only if β = 0.

It might be objected that this type of error is unlikely to arise in practise since often information about time order would rule out Z as a potential unmeasured confounder. In the next example this response is not available since Z may temporally precede both X and Y. Let ε_X, ε_Y, and ε_Z be the error variables in Figure 6(a), and ε'_X, ε'_Y, and ε'_Z be the error variables in Figure 6 (b).

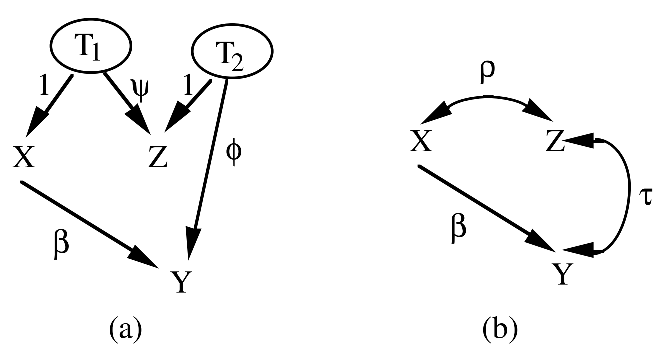


Figure 6

In the path diagram depicted in Figure 6(a) there are two unmeasured confounders T_1 and T_2, which are uncorrelated with one another. Any SEM with this path diagram may be converted into a SEM with the path diagram depicted in Figure 6(b), letting ρ = Cov(X,Z)

$= \psi V(T_1)$, $\tau = \phi V(T_2)$, $V(\varepsilon_X^*) = V(\varepsilon_X) + V(T_1)$, $V(\varepsilon_Y^*) = V(\varepsilon_Y) + \psi^2 V(T_1) + V(T_2)$, and $V(\varepsilon_Z^*) = V(\varepsilon_Z) + \phi^2 V(T_2)$.

Note however, that the reverse is not in general true: not every model containing correlated errors ($X \leftrightarrow Y$) can be converted into a SEM model with latent variables but without correlated errors by introducing a latent T that is a parent of X and Y ($X \leftarrow T \rightarrow Y$), as pointed out in section 2.1. (It is however always possible to convert a model with correlated errors into *some* latent variable model without correlated errors, because every normal distribution is a linear transformation of a set of independent normal variables.)

Returning to the path diagram in Figure 6(b) note that the regression of Y on X yields a consistent estimate of β since $\mathrm{Cov}(X,Y) = \beta V(X)$. However,

$$\begin{aligned}\frac{\mathrm{Cov}(X,Y|Z)}{V(X|Z)} &= \frac{\mathrm{Cov}(X,Y)V(Z) - \mathrm{Cov}(X,Z)\mathrm{Cov}(Y,Z)}{V(X)V(Z) - \mathrm{Cov}(X,Z)^2} \\ &= \frac{\beta V(X)V(Z) - \rho(\rho\beta + \tau)}{V(X)V(Z) - \rho^2} \\ &= \beta - \frac{\rho\tau}{V(X)V(Z) - \rho^2}\end{aligned}$$

Hence the coefficient of X in the regression of Y on X and Z is not a consistent estimate of β, (unless $\rho = 0$ or $\tau = 0$), and may even have a completely different sign. In the case where $\beta = 0$, the coefficient of X in the regression of Y on X typically will not be significantly different from zero, but will become so once Z is included.

Analyses often appear to suggest that it is better to include rather than exclude a variable from a regression. This notion is perhaps given support by reference to “controlling for Z”, the implication being that controlling for Z eliminates a source of bias. The conclusion to be drawn from these examples is that there is no sense in which one is “playing safe” by including rather than excluding “potential' confounders”; if they turn out not to be potential confounders then this could change a consistent estimate into an inconsistent estimate.

The situation is also made somewhat worse by the use of misleading definitions of 'confounder': sometimes a confounder is said to be a variable that is strongly correlated with both X and Y, or even a variable whose inclusion changes the coefficient of X in the regression. Since, for sufficiently large τ and ρ, Z in Figure 6 would qualify as a confounder under either of these definitions, it follows that under either definition including confounding variables in a regression may make a higherto consistent estimator inconsistent.

Finally, it is worth reiterating the well-known fact that in certain circumstances there may be no regression which will estimate the parameter of interest, (although some other consistent estimator may exist):

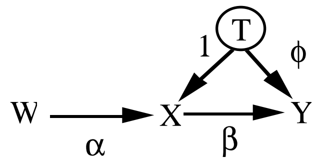


Figure 7

In the SEM shown in Figure 7, Cov(X,Y) = βV(X) + ϕV(T); hence the coefficient of X in the regression of Y on X is not a consistent estimator of β. Further

$$\frac{\mathrm{Cov}(X,Y|W)}{V(X|W)} = \beta + \frac{\phi V(T)}{V(X) - \alpha^2 V(W)} = \beta + \frac{\phi V(T)}{V(T) + V(\varepsilon_X)}$$

hence including W in the regression does not help matters. However, a consistent estimator exists, the so-called Instrumental Variable estimator:

$$\frac{\mathrm{Cov}(Y,W)}{\mathrm{Cov}(X,W)} = \frac{\alpha\beta V(W)}{\alpha V(W)} = \beta$$

In this discussion we have highlighted a number of problems that arise when estimating structural coefficients via regression. These examples raise the following general questions:

(a) If Y is regressed on a set of variables **W**, including X, in which SEMs will the partial regression coefficient of X be a consistent estimate of the structural coefficient β associated with the $X \to Y$ edge?

(b) If Y is regressed on the set **W**, which includes X, in which SEMs will the partial regression coefficient of X be zero if the structural coefficient associated with the $X \to Y$ edge is zero?

(c) Given a particular SEM in which there is an edge $X \to Y$ with coefficient β, is it possible to find a subset **W** of observed variables (including X), such that when Y is regressed on the set **W**, the coefficient of X in the regression is a consistent estimate of β?

(d) Given a particular SEM and a structural coefficient β, is it possible to find a function h(**S**) (where **S** is the sample covariance matrix) that is a consistent estimator of β?

We shall answer questions (a), (b) and (c), by applying the graphical criterion of d-separation. One advantage of a graphical criterion is that it can be applied simply by visual inspection of the path diagram, and does not require lengthy algebraic manipulations which become increasingly arduous when more variables are involved in the calculation.

We do not know the answer to (d), which is one form of the well-known "identification problem"; it is possible that extensions of the graphical criteria we present may hold the key. For related theorems, see Pearl(1997).

2.4. Other Applications

In addition to the uses described above, there are a number of other applications that we do not have the space to describe here. The d-separation relation has proved useful in automated search for causal structure from data and background knowledge (Spirtes and Glymour, 1991, Spirtes, Glymour and Scheines, 1993, Pearl and Verma, 1991, Cooper, 1992), in calculating the effects of interventions on causal systems (Spirtes, Glymour and Scheines, 1993, and Pearl, 1995), and has shed light on a number of issues in statistics ranging from Simpson's Paradox to experimental design (Spirtes, Glymour and Scheines, 1993). See also the applications in Pearl(1997).

3. Linear Structural Equation Models and d-separation

In a linear SEM the random variables are divided into two disjoint sets, the substantive variables and the error variables. Corresponding to each substantive random variable V is a unique error term ε_V.[5] A linear SEM contains a set of linear equations in which each substantive random variable V is written as a linear function of other substantive random variables together with ε_V, and a correlation matrix among the error terms. Initially, we will assume that the error variables are multi-variate Gaussian. However, many of the results that we will prove are about partial correlations, which do not depend upon the distribution of the error terms, but depend only upon the linear equations and the correlations among the error terms.

Since we have no interest in first moments, without loss of generality each variable can be expressed as a deviation from its mean.

For example, the following is a linear SEM M, ε_A, ε_B, ε_C, ε_D, and ε_E are Gaussian "error terms", and A, B, C, D, and E are substantive random variables:

$$A = \varepsilon_A$$
$$B = \varepsilon_B$$
$$C = .2B + .8D + \varepsilon_C$$
$$D = -.5C + .1E + \varepsilon_D$$
$$E = \varepsilon_E$$

[5] There is an equivalent definition of a linear SEM in which parent-less or 'exogenous' substantive variables have no associated error variables.

Correlation Matrix Among Error Terms

$$\begin{array}{c} \\ \varepsilon_A \\ \varepsilon_B \\ \varepsilon_C \\ \varepsilon_D \\ \varepsilon_E \end{array} \begin{array}{c} \begin{array}{ccccc} \varepsilon_A & \varepsilon_B & \varepsilon_D & \varepsilon_D & \varepsilon_E \end{array} \\ \begin{pmatrix} 1.0 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.5 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix} \end{array}$$

If the coefficients in the linear equations are such that the substantive variables are a unique linear function of the error variables alone, the set of equations is said to have a **reduced form**. A linear SEM with a reduced form also determines a joint distribution over the substantive variables. We will consider only linear SEMs which have coefficients for which there is a reduced form, all variances and partial variances among the substantive variables are finite and positive, and all partial correlations among the substantive variables are well defined (e.g. not infinite).

The path diagram of a linear SEM with uncorrelated errors is written with the conventions that it contains an edge A → B if and only if the coefficient for A in the structural equation for B is non-zero, and there is a double-headed arrow between two variables A and B if and only if the correlation between ε_A and ε_B is non-zero. Thus the path diagram for M is shown in Figure 8.

In order to define the d-separation relation, we need to introduce the following path diagram terminology. The concepts defined here are illustrated in Figure 8. A path diagram consists of two parts, a set of vertices **V** and a set of edges **E**. Each edge in **E** is between two distinct vertices in **V**. There are two kinds of edges in **E**, directed edges A → B or A ← B, and double-headed edges A ↔ B; in either case A and B are **endpoints** of the edge; further, A and B are said to be **adjacent**. There may be multiple edges between vertices. In Figure 8 the set of vertices is {A,B,C,D,E} and the set of edges is {A ↔ B, B → C, C → D, D → C, E → D}. For a directed edge A → B, A is the **tail** of the edge and B is the **head** of the edge, A is a **parent** of B, and B is a **child** of A.

An **undirected path** U between X_1 and X_n is a sequence of edges $<E_1,...,E_m>$ such that one endpoint of E_1 is X_1, one endpoint of E_m is X_n, and for each pair of consecutive edges E_i, E_{i+1} in the sequence, $E_i \neq E_{i+1}$, and one endpoint of E_i equals one endpoint of E_{i+1}. In Figure 8, A ↔ B → C ← D is an example of an undirected path between A and D. A **directed path** P between X_1 and X_n is a sequence of directed edges $<E_1,...,E_m>$ such that the tail of E_1 is X_1, the head of E_m is X_n, and for each pair of edges E_i, E_{i+1} adjacent in the sequence, $E_i \neq E_{i+1}$, and the head of E_i is the tail of E_{i+1}. For example, B → C → D is a directed path. A **vertex occurs on a path** if it is an endpoint of one of the edges in the path. The set of vertices on A ↔ B → C ← D is {A, B, C, D}. A path is **acyclic** if no vertex occurs more than once on the path. C → D → C is a cyclic directed path. The following is a list of all the acyclic directed paths in Figure 8: B → C, C → D, E → D, D → C, B → C → D, E → D → C.

A vertex A is an **ancestor** of B (and B is a **descendant** of A) if and only if either there is a directed path from A to B or A = B. Thus the ancestor relation is the transitive, reflexive closure of the parent relation. The following table lists the child, parent, descendant and ancestor relations in Figure 8.

Vertex	Children	Parents	Descendants	Ancestors
A	∅	∅	{A}	{A}
B	{C}	∅	{B,C,D}	{B}
C	{D}	{B,D}	{C,D}	{B,C.D,E}
D	{C}	{C,E}	{C,D}	{B,C,D,E}
E	{D}	∅	{C,D,E}	{E}

A vertex X is a **collider** on undirected path U if and only if U contains a subpath Y ↔ X ↔ Z, or Y → X ↔ Z, or Y → X ← Z, or Y ↔ X ← Z; otherwise if X is on U it is a **non-collider** on U. For example, C is a collider on B → C ← D but a non-collider on B → C → D. X is an **ancestor of a set** of vertices **Z** if X is an ancestor of some member of **Z**.

For disjoint sets of vertices, **X**, **Y**, and **Z**, **X** is **d-connected** to **Y** given **Z** if and only if there is an acyclic undirected path U between some member X of **X**, and some member Y of **Y**, such that every collider on U is an ancestor of **Z**, and every non-collider on U is not in **Z**. For disjoint sets of vertices, **X**, **Y**, and **Z**, **X** is **d-separated** from **Y** given **Z** if and only if **X** is not d-connected to **Y** given **Z**.

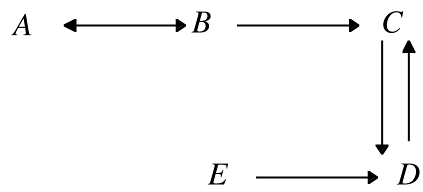


Figure 8

For example, the path E → D → C d-connects E and C given ∅; it also d-connects E and C given {A}, {B}, or {A,B}. E → D ← C d-connects E and C given {D}, given {D,B}, {D,A}, or {D,A,B}. The following is a list of all the pairwise d-separation relations in Figure 8 (where each pair is followed by a list of all of the sets that d-separate them):

{A} and {C} are d-separated given: {B}, {B,D}, {B,E}, {B,D,E}

{A} and {D} are d-separated given: {B}, {B,C}, {B,E}, {B,C,E}

{A} and {E} are d-separated given: ∅, {B}, {B,C}, {B,D}, {B,C,D}, {C,D}

{B} and {E} are d-separated given: ∅, {C,D}

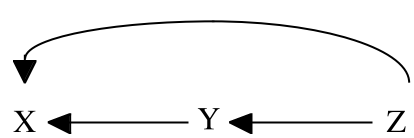
The first theorem states that d-separation in a path diagram G is a sufficient condition for G to entail that $\rho(X,Y.\mathbf{Z}) = 0$ (i.e. in every SEM with path diagram G, the partial correlation of X and Y given **Z** equals 0.)

Theorem 1: If M is a SEM, and {X} and {Y} are d-separated given **Z** in G(M), then $\rho(X,Y.\mathbf{Z}) = 0$ in $\Sigma(M)$.

The second theorem states that d-separation is a necessary condition for a path diagram to entail a zero partial correlation.

Theorem 2: If $\{X_i\}$ and $\{X_j\}$ are not d-separated given **Z** in path diagram G, then there is a SEM M such that G(M) = G, and $\rho(X_i,X_j.\mathbf{Z}) \neq 0$ in $\Sigma(M)$.

Theorem 2 does *not* say that there might not be an individual SEM M with “extra” zero partial correlations among variables that are not d-separated in G(M), as the following example shows.



$$X = .3\,Y + .6\,Z + \varepsilon_X$$

$$Y = -2\,Z + \varepsilon_Y$$
$$Z = \varepsilon_Z$$

Figure 9

(The errors are uncorrelated because there are no double-headed arrows in the path diagram.) In this case X and Y are independent, i.e. $\rho(X,Y) = 0$, even though $\{X\}$ and $\{Y\}$ are not d-separated given $\emptyset$. However, this zero correlation holds because of the particular linear coefficients. Thus, according to Theorem 2 there is some other SEM M such with the same path diagram in which $\rho(X,Y) \neq 0$. It has been shown (Spirtes et. al 1993) that the set of parameters which produce conditional independence relations among variables which are not d-separated in G has zero Lebesgue measure over the parameter space.

4. Applications

4.1. Covariance Equivalence for Path diagrams Without Correlated Errors or Directed Cycles

If for SEM M, there is another SEM M' with a different path diagram but the same number of degrees of freedom, and the same marginal distribution over the measured variables in M, then the $p(\chi^2)$ for M' equals $p(X^2)$ for M, and they have the same BIC scores and AIC scores. Such SEMs are guaranteed to exist if there are SEMs that have the same number of degrees of freedom and contain path diagrams which are covariance equivalent to each other. Stelzl(1986) and Lee and Hershberger(1990) discuss sufficient conditions for covariance equivalence (which they call simply equivalence). Theorem 3 states necessary, as well as sufficient conditions for covariance equivalence in path diagrams without corrrelated errors or directed cycles.

G_1 and G_2 are **d-separation equivalent** if for each disjoint $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in G_1 if and only if $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in G_2.

Theorem 3: If G_1 and G_2 are directed acyclic graphs, G_1 and G_2 are covariance equivalent if and only if G_1 and G_2 are d-separation equivalent.

The test for covariance equivalence of two path diagrams described in Lee and Hershberger(1990) requires determining whether there is a series of edge replacements or reversals preserving equivalence that lead from one path diagram to the other. Because they do not specify an ordering in which the tests are to be done, this could be a very slow process. The following theorem, due to Pearl and Verma (1991) shows how d-separation equivalence can be calculated in $O(E^2)$ time, where E is the number of edges in a path diagram. X is an **unshielded collider** in directed acyclic graph G if and only if G contains edges $A \rightarrow X \leftarrow B$, and A is not adjacent to B in G.

Theorem 4: Two directed acyclic graphs are d-separation equivalent if and only if they contain the same vertices, the same adjacencies, and the same unshielded colliders.

It is apparent from Theorem 4 that any two SEMs with covariance equivalent directed acyclic graphs have the same degrees of freedom.

4.2. Covariance Equivalence for Path Diagrams with Correlated Errors or Directed Cycles

Necessary conditions for covariance equivalence for path diagrams with correlated errors or cycles, and for path diagrams with latent variables follow from Theorem 1 and Theorem 2. If **O** is a subset of the vertices in G_1 and a subset of the vertices in G_2, then G_1 and G_2 are **d-separation equivalent over O** if for each disjoint **X**, **Y**, and **Z** included in **O**, **X** is d-separated from **Y** given **Z** in G_1 if and only if **X** is d-separated from **Y** given **Z** in G_2.

Theorem 5: If G_1 and G_2 are path diagrams that are covariance equivalent over **O**, then G_1 and G_2 are d-separation equivalent over **O**.

The converse is not generally true because while d-separation equivalence guarantees that the conditional independence constraints imposed by two path diagrams are the same, there are other, non-conditional independence constraints, that can be imposed by one path diagram but not the other. The path diagrams in Figure 2 are examples of path diagrams that are d-separation equivalent, but not covariance equivalent over **O = {X,Y,Z}**.

If V is the maximum of the number of variables in G_1 or G_2, and M is the number of variables in **O**, Spirtes and Richardson 1996 presents an $O(M^3 \times V^2)$ algorithm for checking whether two acyclic path diagrams G_1 and G_2 (which may contain latent variables and correlated errors) are d-separation equivalent over **O**. Richardson (1996) presents an $O(V^7)$ algorithm for determining when two cyclic path diagrams without latent variables are d-separation equivalent.

4.3. Extracting Features Common to a Covariance Equivalence Class

Theorem 4 is also the basis of a simple representation (called a pattern in Verma and Pearl 1990) of the entire set of path diagrams without correlated errors or cycles covariance equivalent to a given path diagram without correlated errors or cycles. The pattern for each path diagram in Figure 1 is shown in Figure 10(a), and the pattern for each path diagram in Figure 3 is shown in Figure 10(b).

(a) (b)

Figure 10

A pattern has the same adjacencies as the path diagrams in the covariance equivalence class that it represents. In addition, an edge is oriented as X $\rightarrow$ Z in the pattern if and only if it is oriented as X $\rightarrow$ Z in every path diagram in the simple covariance equivalence class. Meek 1995, Andersson et al. 1995, and Chickering 1995 show how to generate a pattern from an acyclic graph in O(E) time (where E is the number of edges.)

In the case of acyclic path diagrams which may also contain latent variables, and the case of cyclic path diagrams which do not contain latent variables, there is an object analogous to a pattern called a Partial Ancestral Graph (PAG), which represents some of the features common to the members of a covariance equivalence class over **O**. Spirtes and Verma (1992) shows how to create a PAG[6] from an acyclic path diagram in $O(V^5)$ time (where V is the number of vertices in the path diagram). Richardson(1996c) presents an $O(V^7)$ algorithm for constructing a PAG from a (possibly cyclic) graph.

4.4. Solutions to the questions on regression

In this section we apply d-separation in order to answer three questions about the use of regression to estimate structural coefficients that we raised earlier. We introduce the following notation first: Given an SEM with path diagram G, we define **G\{X→Y}** as the path diagram in which the X $\rightarrow$ Y edge is removed.

(a) If Y is regressed on a set of variables W, including X, in which SEMs will the partial regression coefficient of X be a consistent estimate of the structural coefficient β associated with the X $\rightarrow$ Y edge?

The coefficient of X is a consistent estimator of β if **W** does not contain any descendant of Y in G, and X is d-separated from Y given **W** in G\{X→Y}.[7] If this condition does not

[6] The algorithm given by Spirtes and Verma was designed to output an object called a partially oriented inducing path graph (POIPG); however, it has subsequently been shown that the output can be re-interpreted as a PAG.

[7] Note this criterion is similar to Pearl's back door criterion (Pearl, 1993), except that the back-door criterion was proposed as a means of estimating the *total* effect of X on Y.

hold, then for almost all instantiations of the parameters in the SEM, the coefficient of X will fail to be a consistent estimator of β.

It follows directly from this that (almost surely) β cannot be estimated consistently via any regression equation if either there is an edge X ↔ Y (i.e. ε_X and ε_Y are correlated) or if X is a descendant of Y (so that the path diagram is cyclic). The result itself follows from the fact that under the conditions stated,

$$\mathrm{Cov}(X,\varepsilon_Y \mid \mathbf{W}\backslash\{X\}) = \mathrm{Cov}(X,Q \mid \mathbf{W}\backslash\{X\}) = 0$$

for each Q ∈ **Parents**(Y,G)\{X}. (**Parents**(Y,G) is the set of parents of Y in G.) It follows that

$$\mathrm{Cov}(X,Y|\mathbf{W} \setminus \{X\}) = \mathrm{Cov}(X,\beta X + \sum_{T_i} a_i T_i + \varepsilon_Y|\mathbf{W} \setminus \{X\}) = \beta V(X|\mathbf{W} \setminus \{X\})$$

and hence $\dfrac{\mathrm{Cov}(X,Y|\mathbf{W} \setminus \{X\})}{V(X|\mathbf{W} \setminus \{X\})} = \beta$.

(b) If Y is regressed on the set **W**, including X, in which SEMs will the partial regression coefficient of X be zero if there is no edge between X and Y?

The coefficient of X will be zero if X and Y are d-separated given **W**\{X}. (See Scheines 1994 and Glymour 1994). This follows directly from the fact that the coefficient of X in the regression equation is proportional to ρ(X,Y,**W**\{X}), which in turn will be zero if {X} is d-separated from {Y} given **W**\{X}. As before, if {X} and {Y} are not d-separated given **W**\{X}, then, even if there is no edge between X and Y, for almost all assignments of values to the model parameters the coefficient of X will be non-zero.

(c) Given a particular SEM, with path diagram G, in which there is an edge X → Y, with coefficient β, is it possible to find a subset **W** of observed variables, (including X), such that when Y is regressed on the set **W**, the coefficient of X in the regression is a consistent estimate of β?

From (a), we know that if there is a subset **W** of the observed variables which contains no descendant of Y, but which d-separates X from Y in G\{X→Y}, then the regression coefficient of X in the regression of Y on **W** will be a consistent estimate of β.

5. Conclusion

D-separation is a widely studied graphical relation, which has proved useful in solving many problems. We have illustrated some of its applications, such as finding covariance equivalence classes, finding features common to all members of covariance equivalence classes, and explaining relationships between regression coefficients and structural equation coefficients. Other applications to problems in structural equation modelling are described in Spirtes et al. (1993) and Pearl(1997). Instructions for using a program, TETRAD II, that uses d-separation as a basis for searching for automated searching for models given data and background knowledge can be found on the world wide web at http://hss.cmu.edu/html/departments/philosophy/TETRAD/tetrad.html. By extending the relation to path diagrams, as well as graphs, we have shown how theorems proved about d-separation can be applied to a wider class of structural equation models.

6. Proofs of Main Results

We will prove Theorem 1 and Theorem 2 in two steps. First we will prove them for the case where G(M) contains no double-headed arrows; then we will prove it for the case where G(M) does contain double-headed arrows.

A probability measure P over **V** satisfies the **global directed Markov property** for path diagram G if and only if for any three disjoint sets of variables **X**, **Y**, and **Z** included in **V**, if **X** is d-separated from **Y** given **Z**, then **X** is independent of **Y** given **Z** in P.

The following lemma relates the global directed Markov property to factorizations of a density function. Denote a density function over **V** by f(**V**), where for any subset **X** of **V**, f(**X**) denotes the marginal of f(**V**). Let **An(X)** be the set of ancestors of members of **X**. If f(**V**) is the density function for a probability measure over a set of variables **V**, say that f(**V**) **factors according to directed graph** G with vertices **V** if and only if for every subset **X** of **V**,

$$f(\mathbf{An}(\mathbf{X})) = \prod_{V \in \mathbf{An}(X)} g_V(V, \mathbf{Parents}(\mathbf{V}))$$

where g_V is a non-negative function.

Lemma 1 was proved in Lauritzen et al. (1990) for the acyclic case, and the proof carries over essentially unchanged for the cyclic case.

Lemma 1: If **V** is a set of random variables with a probability measure P that has a density function f(**V**) and f(**V**) factors according to directed graph G, then P satisfies the global directed Markov property for G.

Lemma 2 was proved in Spirtes(1995) and Koster(1995).

Lemma 2: If M is a SEM, and {X} and {Y} are d-separated given **Z** in directed graph G(M), then $\rho(X,Y.\mathbf{Z}) = 0$ in $\Sigma(M)$.

Lemma 3 was proved in Spirtes(1995).

Lemma 3: For any directed graph G, if {X} and {Y} are not d-separated given **Z** in G(M), there is a SEM M, G = G(M) and $\rho(X,Y.\mathbf{Z}) \neq 0$ in $\Sigma(M)$.

We will now show that Theorem 1 and Theorem 2 hold even when G contains double-headed arrows. Let the set of vertices in G be **V**. For a given triple X, Y, and **Z**, if {X} is d-separated from {Y} given **Z** in G(M) and G(M) contains double-headed arrows, the strategy is to convert M into another SEM M'(M,X,Y,**Z**) such that G(M'(M,X,Y,**Z**)) has additional latent variables, but no double-headed arrows, the marginal over **V** of Σ(M'(M,X,Y,**Z**)) is equal to $\Sigma(M)$, and {X} and {Y} are d-separated given **Z** in G(M'(M,X,Y,**Z**)). (We write M'(M,X,Y,**Z**) in order to emphasize that the SEM M' constructed from M is a function of the path diagram of M, and the vertices X, Y, and **Z** in the d-separation relation being considered.) It will then follow from Lemma 2 that $\rho(X,Y.\mathbf{Z}) = 0$ in $\Sigma(M)$.

If {X} is d-separated from {Y} given **Z** in G(M), the graph G(M'(M,X,Y,**Z**)) is constructed by the following algorithm, where a **trek** between X_i and X_j is an undirected path between X_i and X_j that contains no colliders. (We will illustrate the application of the algorithm to the path diagram in Figure 11.)

Algorithm: Construct Latent Directed Graph

Inputs - Path Diagram G with vertex set **V**, Vertices X, Y, **Z**;

Output - Directed Graph $G_{Construct}$(G,X,Y,**Z**), with vertex set $\mathbf{V} \cup \mathbf{T}$;

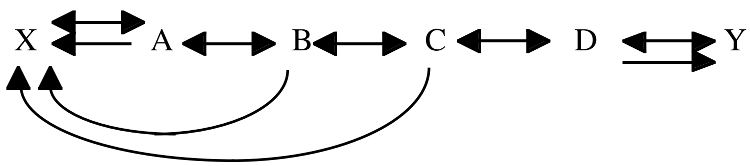
1. Order the variables so that X is first, Y is second, followed by each variable with a descendant in **Z**, followed by any remaining variables that have X or Y as descendants in G(M), followed by the rest of the variables. Given this ordering, we will now refer to the variables as $X_1,...,X_n$, where for all i, X_i is the i^{th} variable in the ordering.

2. For each variable X_i, add to the existing graph G, a variable T_i, and edges from T_i to X_j, for each $j \geq i$. Call the resulting graph, which has vertex set $(X_1,\ldots,X_n, T_1,\ldots,T_n)$ $G_{Constsruct(0)}$.

3. Let $G_{Construct(i)}$ be the the graph constructed after the i^{th} iteration of the following step, starting with i = 1: If r > i, and there is no trek between X_r and X_i in $G_{Construct(i-1)}$ containing a variable T_j, where j < i, and ε_i and ε_j are uncorrelated in Σ, then remove the $T_i \rightarrow X_r$ edge.

For inputs G, X, Y, and **Z**, we will refer to the output of this algorithm as $G_{Construct}(G,X,Y,\mathbf{Z})$. Note that it follows from step 2 of the construction algorithm that if there is a trek $X_i \leftarrow T_j \rightarrow X_k$, then $j \leq \min(i,k)$.

Suppose for the graph in Figure 11 we are interested in whether $\rho(X,Y) = 0$ (i.e. $\mathbf{Z} = \varnothing$).



G

Figure 11

Applying the first step of Algorithm Construct Latent Directed Graph to G in Figure 11 with vertex inputs X, Y, ∅, results in the naming of the vertices shown in Figure 12.

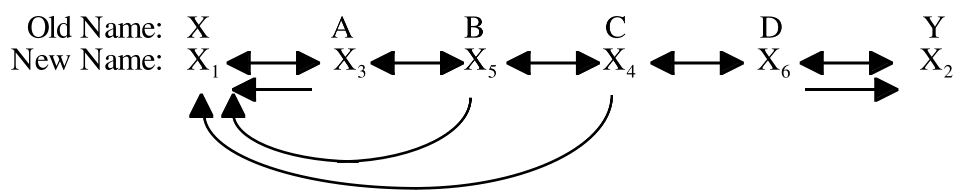


Figure 12: G with vertices renamed

Applying steps 2 and 3 of Algorithm Construct Latent Directed Graph results in the directed graph shown in Figure 13.

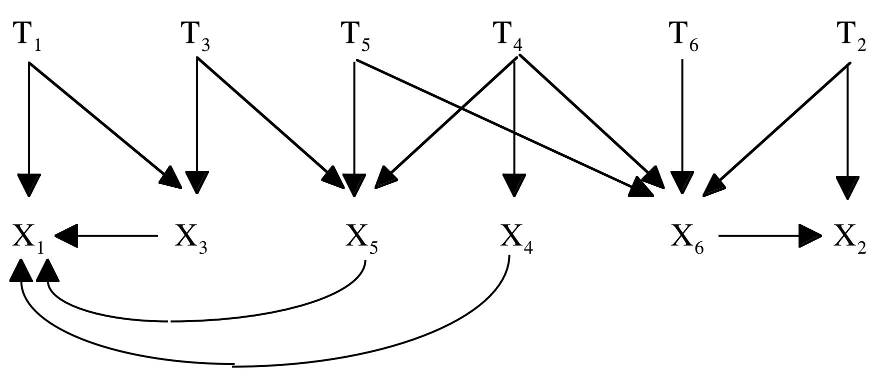


Figure 13: $G_{Construct}(X,Y, \varnothing)$

As an example of an application of step 3, the edge from T_3 to X_4 is removed because in $G_{Construct(2)}$ there is no trek between X_3 and X_4 that contains T_1 or T_2, and there is no double-headed arrow between X_3 and X_4 in G.

The next series of lemmas shows how to construct a SEM M'(M,X,Y,**Z**) with measured variables **V** and latent variables **T** , so that the marginal over **V** of Σ(M'(M,X,Y,**Z**)) = Σ(M), and G(M'(M,X,Y,**Z**)) = $G_{Construct}$(G(M),X,Y,**Z**).

Lemma 4: If Σ is a positive definite matrix, then there exists a positive definite matrix $\Sigma' = \Sigma - \delta I$, where δ is a real positive number.

Proof. Suppose that Σ is a positive definite matrix. It follows then that for all solutions of $\det(\Sigma - \lambda I) = 0$, λ is positive. Let the smallest solution of $\det(\Sigma - \lambda I) = 0$ be λ_1. Let δ be less than λ_1 and greater than 0. Let $\Sigma' = \Sigma - \delta I$. We will now show that all of the solutions of $\det(\Sigma' - \lambda' I) = 0$ are positive. $\Sigma' - \lambda' I = \Sigma - \delta I - \lambda' I = \Sigma - (\lambda' + \delta)I$. If we set $\lambda' = \lambda - \delta$, then for each solution of $\det(\Sigma - \lambda I) = 0$, there is a solution of $\det(\Sigma - (\lambda' + \delta)I) = 0$. Since $\lambda' = \lambda - \delta$, and δ is less than λ_1, the smallest solution of $\det(\Sigma' - \lambda' I) = 0$ is greater than 0. ∴

A linear transformation of a set of random variables is **lower triangular** if and only if there is an ordering of the variables such that the matrix representing the transformation is zero for all entries a_{ij}, when $j > i$.

Lemma 5: If $X_1, \ldots, X_n$ have a joint normal distribution $N(0,\Sigma)$, where Σ is positive definite, then there is a set of n mutually independent standard normal variables $T_1, \ldots, T_n$, such that $X_1, \ldots, X_n$ are a lower triangular linear transformation of $T_1, \ldots, T_n$ and for each i, the coefficient of T_i in the equation for X_i is not equal to zero.

Proof. For every positive definite correlation matrix Σ, there is a SEM M with correlation matrix $\Sigma(M) = \Sigma$, and directed acyclic graph G(M) that has each pair of vertices in G(M) adjacent (Spirtes et al. 1993). The reduced form of a complete directed acyclic graph is a lower triangular transformation of independent error variables (in this case the T variables) that is non-zero on the diagonal, because Σ is positive definite. ∴

There is a simple rule for calculating Cov(X,Y) from a path diagram with no directed cycles that is used in the following lemmas. There is an edge directed into a vertex A on a path P if and only if P contains an edge $A \leftarrow B$ or $A \leftrightarrow B$. The **source** of a trek is the vertex on the trek with no edges directed into it, if there such a vertex. (For example B is the source of the trek $A \leftarrow B \rightarrow C$, A is the source of $A \rightarrow B \rightarrow C$, and the trek $A \leftarrow B \leftrightarrow C \rightarrow D$ has no source.) Associated with each edge $A \rightarrow B$ in a graph is a label that corresponds to the coefficient of A in the equation for B, and associated with each edge $A \leftrightarrow B$ is a label that corresponds to the correlation of the error terms for A and B. Cov(X,Y)is equal to the sum over all of the treks, of the product of the edge labels on the trek, times the variance of the source of the trek (if there is one). For example, in Figure 4, $\mathrm{Cov}(Y,Z) = (\alpha\beta + \gamma)V(Z)$. For a proof of the case without correlated errors, see Glymour et al. 1987; the case with correlated errors is a simple modification of the latter proof.

Lemma 6: There is a SEM M'(M,X,Y,**Z**) with measured variables **V** and latent variables **T**, such that G(M'(M,X,Y,**Z**)) = $G_{Construct}$(G(M),X,Y,**Z**), and the marginal over **V** of Σ(M'(M,X,Y,**Z**)) is equal to Σ(M).

Proof. Let the correlation matrix among the error terms of M be Σ. If the equations in M are:

$$X_i = \sum_{j \neq i} b_{ij} X_j + \varepsilon_i \tag{1}$$

(where some of the b_{ij} may equal zero, and some of the ε_i may be correlated) we will construct equations in M'(M,X,Y,**Z**) that are:

$$X_i = \sum_{j \neq i} b_{ij} X_j + \sum_{j \leq i} a_{ij} T_j + \varepsilon''_i \tag{2}$$

by showing that there is a latent variable model of Σ of the form

$$\varepsilon_i = \sum_{j \leq i} a_{ij} T_j + \varepsilon''_i \tag{3}$$

where each of the T_i and ε''_i are uncorrelated.

By hypothesis, Σ is a positive definite matrix. By Lemma 4 there is a set of variables $\varepsilon'_1,\ldots,\varepsilon'_n$ with positive definite matrix Σ' = Σ - δI, where δ > 0. So we can write

$$\varepsilon_i = \varepsilon_i' + \varepsilon_i'' \tag{4}$$

where the ε_i'' are uncorrelated with each other and the ε_i' variables, each ε_i'' is normally distributed with mean zero and variance δ. The ε'' variables will serve as the uncorrelated error terms in the new model that we construct; the ε' variables are used only in intermediate stages of constuction, and have the same covariance matrix as the ε variables, except that the variances of the variables have been decreased by a small amount δ, i.e. Σ' = Σ - δI. As a first step to constructing a latent variable model of **V**, we will construct a latent variable model of ε'.

By Lemma 5, there is a set of variables **T**={T_1, ..., T_n} such that ε'_1, ..., ε'_n with correlation matrix Σ' are a lower triangular linear transformation of $T_1,\ldots, T_n$ and for each i, the coefficient of T_i in the equation for ε'_i is not equal to zero. That is

$$\varepsilon'_i = \sum_{j<i} a_{ij} T_j \tag{5}$$

where $a_{ii} \neq 0$.

There is a directed graph H that represents this latent variable model of the ε'_i variables, in which there is an edge from T_j to ε'_i only if $j \leq i$. From the construction of H, there are no edges from T_j to ε'_1 unless $j = 1$. Hence, for every $j \neq 1$, in H every trek between ε'_1 and ε'_j contains T_1. It follows that there is at most one trek between ε'_1 and ε'_j. The edge from T_1 to ε'_1 is not zero. Hence it follows from the trek rule for calculating covariances from a path diagram, that if ε_1 and ε_j are not correlated in Σ (i.e. there is no double-headed arrow between X_1 and X_j in G(M)) then the edge from T_1 to ε_j is zero. (In the example from Figure 12, $a_{12} = a_{14} = a_{15} = a_{16} = 0$.)

Applying this strategy to each of the T_i variables in turn, we can now show that for each i and $r > i$, if there is no trek between ε'_r and ε'_i containing a variable T_j, where $j < i$, and ε'_i and ε'_j are uncorrelated in Σ, then there is no $T_i \rightarrow \varepsilon'_r$ edge in H. Suppose on the contrary that in H there is no trek between ε'_r and ε'_i containing a variable T_j, where $j < i$, and ε_i and ε_j are uncorrelated in M, but the $T_i \rightarrow \varepsilon'_r$ edge is in H. By the construction of H, if $k > i$, then there is no edge from T_k to ε'_i. It follows that if in H there is no trek between ε'_r and ε'_i containing a variable T_j, where $j < i$, then every trek between ε'_i and any other variable contains the edge from T_i to ε'_i, which is in H since $a_{ii} \neq 0$. The $T_i \rightarrow \varepsilon'_r$ edge exists by hypothesis, so there is exactly one trek between ε'_i and ε'_r in H. Hence, in every SEM L with vertices $\{\varepsilon'_1,\ldots,\varepsilon'_n\}$ and directed graph G(L) = H, ε'_i and ε'_r are correlated in Σ(L). (Note that this could not be claimed if there were more than one trek between ε'_i and ε'_r since in that case the treks might cancel each other.) Since the covariances between distinct ε' variables are equal to the correlations between the corresponding ε variables, it follows that ε_i and ε_r are correlated in Σ, and hence there is a double-headed arrow between ε'_i and ε'_j in G(M). This is a contradiction.

The graph H for the path diagram in Figure 12 is shown in Figure 14.

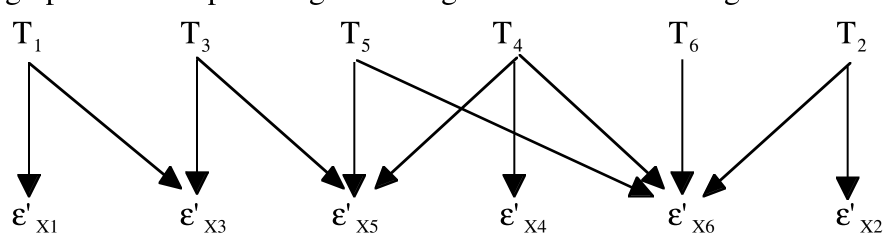


Figure 14: H

From the latent variable model of the ε' variables, we can now form a model M'(M,X,Y,**Z**) with measured variables **V** and latent variables $T_1,\ldots,T_n$, but without correlated errors.

$$X_i = \sum_{j \neq i} b_{ij} X_j + \sum_{j<i} a_{ij} T_j + \varepsilon''_i$$

It follows from equations (1), (4), and (5) that the marginal distribution of $\mathbf{V}=\{X_1,\ldots X_n\}$ in M'(M,X,Y,**Z**) is the same as the distribution of **V** in M.

We will now show that G(M'(M,X,Y,**Z**)) = $G_{Construct}$(G(M),X,Y,**Z**). For variables A and B in **V**, by the construction of M', there is an edge between A and B in G(M'(M,X,Y,**Z**)) if and only if there is an edge between A and B in G, and hence an edge between A and B in $G_{Construct}$(G(M),X,Y,**Z**). (Hence the ancestor relations among the substantive variables in G(M'(M,X,Y,**Z**)) are the same as the ancestor relations among the corresponding variables in G(M).) There is an edge between a variable T in **T** and a variable A in **V** in G(M'(M,X,Y,**Z**)) if and only if there is an edge between T and ε'_A in H. We have already shown that for each i and r > i, if there is no trek between ε'_r and ε'_i containing a variable T_j, where j < i, and ε'_i and ε'_j are uncorrelated in Σ, then there is no $T_i \rightarrow \varepsilon'_r$ edge in H. It follows that for each i and r > i, if there is no trek between X_r and X_i containing a variable T_j, where j < i, and ε_i and ε_j are uncorrelated in Σ, then there is no $T_i \rightarrow X_r$ edge in G(M'(M,X,Y,**Z**)). (This latter property is the property obtaining in $G_{Construct}$(G(M),X,Y,**Z**) by application of steps 2 and 3.) Hence $G_{Construct}$(G(M),X,Y,**Z**) = G(M'(M,X,Y,**Z**)) $\therefore$

The next series of lemmas show that if X_1 and X_2 are d-separated given **Z** in G(M), then X_1 and X_2 are d-separated given **Z** in G(M'(M,X,Y,**Z**)).

We will call a trek $X_j \leftarrow T_m \rightarrow X_i$ that contains a T variable a **latent trek** in $G_{Construct}$(G(M),X,Y,**Z**). In G(M), a **correlated error trek sequence** is a sequence of vertices $<X_i, .., X_k>$ such that no pair of vertices adjacent in the sequence are identical, and for each consecutive pair of vertices X_r and X_s in the sequence, there is an edge $X_r \leftrightarrow X_s$. For example in Figure 11, the sequence of vertices <X,A,B,C,D,Y> is a correlated error trek sequence between X and Y.

Lemma 7: If there is a latent trek between X_i and X_j in $G_{Construct}$(G(M),X,Y,**Z**) that contains a variable T_r, i.e. $X_i \leftarrow T_r \rightarrow X_j$, then in G(M) there is a correlated error trek sequence between X_i and X_j, such that every variable in the correlated error trek sequence, with the possible exception of the endpoints, X_i and X_j,, has index (i.e. subscript) less than or equal to r (henceforth referred to as the correlated error trek sequence in G(M) corresponding to the latent trek between X_i and X_j in $G_{Construct}$(G(M),X,Y,**Z**).)

Proof. The proof is by induction on r. Suppose first that r = 1. From the construction algorithm for $G_{Construct}$(G(M),X,Y,**Z**), if there is a latent trek between X_i and X_j in $G_{Construct}$(G(M),X,Y,**Z**) that contains T_1 then there are edges $X_i \leftrightarrow X_1$ and and $X_j \leftrightarrow X_1$ in G(M). The concatenation of these two edges forms a correlated error trek sequence in which (trivially) every variable in the sequence except for the endpoints has an index less than or equal to 1. The induction hypothesis is that for all $r \leq n$, if there is a latent trek between X_i and X_j in $G_{Construct}(G(M),X,Y,\mathbf{Z})$ that contains T_r, then in G(M) there is a correlated error trek sequence between X_i and X_j, such that every variable in the sequence, with the possible exception of the endpoints has an index less than r. Suppose now that in $G_{Construct}(G(M),X,Y,\mathbf{Z})$ there is a latent trek between X_i and X_j such that the trek contains T_{n+1}, where $i, j \geq n+1$. Since the edge between T_{n+1} and X_i exists in $G_{Construct}(G(M),X,Y,\mathbf{Z})$, it follows from the construction algorithm for $G_{Construct}(G(M),X,Y,\mathbf{Z})$ that either there is a latent trek between X_i and X_{n+1} in $G_{Construct}(G(M),X,Y,\mathbf{Z})$ that contains some T_r, $r < n+1$, or there is a double-headed arrow between X_{n+1} and X_i in G(M). In the former case, by the induction hypothesis there is a correlated error trek sequence between X_i and X_{n+1} that, except for the endpoints, contains only vertices whose indices are less than or equal to r, and hence less than or equal to n+1. In the latter case, $\langle X_i,X_{n+1}\rangle$ is a correlated error trek sequence between X_i and X_{n+1}. Similarly, there is a correlated error trek sequence between X_{n+1} and X_j that, except for the endpoints, contains only vertices whose indices are less than or equal to n+1. These two correlated error trek sequences can be concatenated to form a correlated error trek sequence between X_i and X_j that, except for the endpoints, contains only vertices whose indices are less than or equal to n+1. $\therefore$

For $G_{Construct}(G(M),X,Y,\mathbf{Z})$ shown in Figure 13, there is a latent trek between $X_5 \leftarrow T_5 \rightarrow X_6$, and a corresponding correlated error trek sequence $\langle X_5,X_4,X_6\rangle$ in the graph G in Figure 12.

We will make use of the following Lemma which is a simple extension to path diagrams with directed cycles of Lemma 3.3.1 in Spirtes *et al.* (1993). This Lemma allows us to concatenate 'small' d-connecting paths to form a larger d-connecting path. We say a path is **into** endpoint X if the path contains some edge $X \leftrightarrow Y$ or $X \leftarrow Y$.

Lemma 8: In a path diagram G over a set of vertices **V**, if:

(a) **Q** is a sequence of vertices in **V** from A to B, $\mathbf{Q} \equiv \langle A \equiv X_0, \ldots X_{n+1} \equiv B\rangle$, such that $\forall i, 0 \leq i \leq n, X_i \neq X_{i+1}$ (the X_i are only *pairwise distinct* , i.e. not necessarily distinct),

(b) $\mathbf{Z} \subseteq \mathbf{V} \setminus \{A,B\}$,

(c) **P** is a set of undirected paths such that

(i) for each pair of consecutive vertices in **Q**, X_i and X_{i+1}, there is a unique undirected path in **P** that d-connects X_i and X_{i+1} given $\mathbf{Z} \setminus \{X_i , X_{i+1}\}$,

(ii) if some vertex X_k in **Q**, is in **Z**, then the paths in **P** that contain X_k as an endpoint collide at X_k, (i.e. all such paths are directed into X_k)

(iii) if for three vertices X_{k-1}, X_k, X_{k+1} occurring in **Q** the d-connecting paths in **P** between X_{k-1} and X_k, and X_k and X_{k+1}, collide at X_k then X_k has a descendant in **Z**,

then there is a path U in G that d-connects $A \equiv X_0$ and $B \equiv X_{n+1}$ given **Z**.

Note that we do not require that a vertex occur only once in **Q**. Hence one occurrence of a vertex in **Q** may be a collider, and another occurrence of the same vertex in **Q** may be a non-collider. (We say that Y_k is a collider (non-collider) in **Q** if the pair of consecutive paths in **P** that contain Y_k as an endpoint collide (do not collide) at Y_k.)

Lemma 9: If $X_1 \equiv X$ and $X_2 \equiv Y$ are d-connected given **Z** in the directed graph $G_{Construct}$(G(M),X,Y,**Z**), then X and Y are d-connected given **Z** in the path diagram G(M). (G(M) has vertex set **V**, $G_{Construct}$(G(M),X,Y,**Z**) has vertex set $\mathbf{V} \cup T$, and $\{X,Y\} \cup Z \subseteq \mathbf{V}$.)

Proof. Suppose that there is an undirected path U that d-connects X_1 and X_2 given **Z** in $G_{Construct}$(G(M),X,Y,**Z**). We will prove that X and Y are d-connected given **Z** in G(M) by constructing a sequence of vertices **Q** and a set **P** of paths in G between pairs of consecutive vertices in **Q** satisfying the conditions of Lemma 8.

Our first step will be to use U to construct a sequence **Q'** and a set of paths **P'** in $G_{Construct}$(G(M),X,Y,**Z**) from which we will then construct **P** and **Q**. Intuitively, we form **Q'** and **P'** by breaking U into pieces, such that each latent trek occurs as a separate piece. More formally, form a sequence **Q'** of vertices and an associated sequence **P'** of paths in $G_{Construct}$(G(M),X,Y,**Z**) with the following properties: (i) every vertex in **Q'** is in **V** and occurs on U; (ii) no vertex occurs in **Q'** more than once; (iii) if X_i occurs before X_j in **Q'**, then X_i occurs before X_j on U; (iv) if the subpath of U between X_i and X_j is a latent trek, $X_i \leftarrow T_r \rightarrow X_j$, then X_i and X_j both occur in that order in **Q'**. The path in **P'** associated with a pair X_i and X_j of consecutive vertices in **Q'** is the subpath of U between X_i and X_j. In the example in Figure 13, in $G_{Construct}$(G(M),X,Y,**Z**) the d-connecting path between X_1 and X_2 given $\mathbf{Z} = \varnothing$ is $X_1 \leftarrow X_5 \leftarrow T_4 \rightarrow X_6 \rightarrow X_2$, $\mathbf{Q'} = \langle X_1,X_5,X_6,X_2 \rangle$, and $\mathbf{P'} = \langle X_1 \leftarrow X_5, X_5 \leftarrow T_4 \rightarrow X_6, X_6 \rightarrow X_2 \rangle$. In this example, there are no colliders in **Q'**.

Because U is a path that d-connects X_1 and X_2 given **Z** in $G_{Construct}$(G(M),X,Y,**Z**), it is clear that the paths in **P'** have the following properties in $G_{Construct}$(G(M),X,Y,**Z**): (i) Each path in **P'** d-connects its endpoints X_i and X_j given $\mathbf{Z} \backslash \{X_i,X_j\}$; (ii) if paths in **P'** collide at X_i then X_i has a descendant in **Z**; and (iii) if X_i is in **Z** then the paths in **P'** collide at X_i.

We will now show how to construct a sequence of vertices **Q** and a set **P** of paths in G(M) between pairs of consecutive vertices in **Q** satisfying the conditions of Lemma 8; it follows then that X and Y are d-connected given **Z** in G.

We will create **Q** by several modifications of **Q'**. Step (1) in creating **Q** is to replace each subsequence $<X_r,X_s>$ of **Q'** such that X_r and X_s are the endpoints of a latent trek in **P'**, with the corresponding correlated error trek sequence $<X_r, \ldots, X_s>$ in G(M). Then replace the latent trek in **P'** with the corresponding correlated error trek sequence in **P'**. Note that each occurrence of X_k between $<X_r, \ldots,X_s>$ is a collider in **Q**. In the example, after the first step **Q** = $<X_1,X_5,X_4,X_6,X_2>$ and **P** = $<X_1 \leftarrow X_5, X_5 \leftrightarrow X_4, X_4 \leftrightarrow X_6, X_6 \rightarrow X_2>$, i.e. we replaced the subsequence $<X_5,X_6>$ in **Q'** by $<X_5,X_4,X_6>$, and the latent trek $X_5 \leftarrow T_4 \rightarrow X_6$ by $X_5 \leftrightarrow X_4$ and $X_4 \leftrightarrow X_6$ in **Q'**.

Recall that the ancestor relations among the variables in **V** (which includes the variables in **Z**) in $G_{Construct}$(G(M),X,Y,**Z**) are the same as the ancestor relations among the variables in G(M). After stage (1) in creating **Q**, if X_k is not an ancestor of **Z** in G(M) (or in $G_{Construct}$(G(M),X,Y,**Z**)), but has an occurrence in **Q** that is a collider, it follows that X_k was added to **Q** by replacing a subsequence $<X_r,X_s>$ of **Q'** by a corresponding correlated error trek sequence $<X_r, \ldots,X_s>$ in G(M). Hence any such X_k lies between some pair of vertices X_r and X_s that are adjacent in **Q'**. Because every vertex in $<X_r, \ldots, X_s>$ in **Q** (except for X_r and X_s) has an index less than r and s, and X_k is not an ancestor of **Z** in $G_{Construct}$(G(M),X,Y,**Z**), it follows from the ordering of the variables that we chose, that X_r and X_s are not ancestors of **Z** in $G_{Construct}$(G(M),X,Y,**Z**). If a path U d-connects X_1 and X_2 given **Z**, then every vertex on U is an ancestor of X_1 or X_2 or **Z**. Because X_r and X_s are on U, but not ancestors of **Z** in $G_{Construct}$(G(M),X,Y,**Z**), and U d-connects X_1 and X_2 given **Z**, X_r and X_s are both ancestors of $\{X_1,X_2\}$. Because in $G_{Construct}$(G(M),X,Y,**Z**), X_r and X_s are both ancestors of $\{X_1,X_2\}$, and k < r and s, it follows from the ordering of the variables that X_k is also an ancestor of $\{X_1,X_2\}$ in $G_{Construct}$(G(M),X,Y,**Z**). Hence X_k is an ancestor of $\{X_1,X_2\}$ in G(M). In the example, in G(M), X_4 is not an ancestor of the empty set but is an ancestor of X_1, and it is between two vertices X_5 and X_6 which also are not ancestors of the empty set but are ancestors of X_1 or X_2.

Thus, if there is some vertex X_k in **Q** that is not an ancestor of **Z**, but occurs in **Q** as a collider then X_k is an ancestor of X_1 or X_2. Let X_a be the last occurrence of a collider in **Q** that is an ancestor of X_1 but not of **Z**, if there is one, otherwise let $X_a = X_1$. Step (2) in forming **Q** and **P** is to replace the subsequence $<X_1,\ldots,X_a>$ by $<X_1,X_a>$ if $X_a \neq X_1$, and replacing the corresponding paths in **P** by a directed path from X_a to X_1 if $X_a \neq X_1$. (Such a directed path exists if $X_a \neq X_1$ because X_a is an ancestor of X_1.) This removes all occurrences of vertices between X_1 and X_a that are not ancestors of **Z**, but are colliders in **Q**. In the example, $X_a = X_4$, and after step 2, **Q** = $<X_1,X_4,X_6,X_2>$ and **P** = $<X_1 \leftarrow X_4, X_4 \leftrightarrow X_6, X_6 \rightarrow X_2>$.

By definition, every vertex that occurs as a collider between X_a and X_2 in **Q** is an ancestor of **Z** or of X_2. Let X_b be the first vertex after X_a in **Q** that is an ancestor of X_2 but not of **Z**, if there is one, otherwise let $X_b = X_2$. Step (3) in forming **Q** and **P** is to replace the subsequence $<X_b, \ldots, X_2>$ by $<X_b, X_2>$ if $X_b \neq X_2$, and replacing the corresponding paths in **P** by a directed path from X_b to X_2 if $X_b \neq X_2$. This removes all occurrences of colliders between X_b and X_2 that are not ancestors of **Z**. Note that all occurrences of colliders that are left are between X_a and X_b, and every occurrence of a collider between X_a and X_b is an ancestor of **Z** by construction. In the example, $X_b = X_2$, and after step (3), **Q** and **P** are unchanged.

We will now show that every path between a pair of variables X_u and X_v in **P** d-connects X_u and X_v given $\mathbf{Z}\backslash\{X_u,X_v\}$. If the path between X_u and X_v is also in **P'**, then it d-connects X_u and X_v given $\mathbf{Z}\backslash\{X_u,X_v\}$ because every path in **P'** has this property. If the path between X_u and X_v is not in **P'**, but was added in step (1) of the formation of **P**, then the path between X_u and X_v is a correlated error trek $X_u \leftrightarrow X_v$, which clearly d-connects X_u and X_v given $\mathbf{Z}\backslash\{X_u,X_v\}$. If the path between X_u and X_v is not in **P'**, but was added in step (2) of the formation of **P'**, then $X_u = X_1$, $X_v = X_a$, and the path between X_u and X_v is a directed path from X_a to X_1 that does not contain any member of **Z**. Hence the path d-connects X_u and X_u given **Z**. Similarly, if the path between path between X_u and X_v is not in **P'**, but was added in step (3) of the formation of **P**, then $X_u = X_b$, $X_v = X_2$, and the path between X_u and X_v is a directed path from X_b to X_2 that does not contain any member of **Z**. Hence the path d-connects X_u and X_v given **Z**.

We will now show that every vertex that occurs as a collider in **Q** has a descendant in **Z**, and every vertex that occurs as a non-collider in **Q** is not in **Z**. Every vertex that occurs as a collider in **Q** is an ancestor of **Z**, because steps (2) and (3) in the formation of **Q** removed all occurrences of colliders that were not ancestors of **Z**. Every vertex that occurs as a non-collider in **Q'** and as a non-collider in **Q** is not in **Z**, because every vertex that occurs as a non-collider in **Q** is not in **Z**. The only vertices that may occur as non-colliders in **Q** but not in **Q'** are X_a and X_b. X_a is not in **Z**, because either it is equal to X_1 or X_2, neither of which is in **Z**, or it is not an ancestor of **Z** by construction. Similarly, X_b is not in **Z**.

Hence **Q** is a sequence of paths that satisfy properties (i), (ii), and (iii) of Lemma 8. It follows from Lemma 8 that $X_1 \equiv X$ and $X_2 \equiv Y$ are d-connected given **Z** in G(M). ∴

Theorem 1: If M is a SEM, and {X} and {Y} are d-separated given **Z** in G(M), then $\rho(X,Y.\mathbf{Z}) = 0$ in $\Sigma(M)$.

Proof. By Lemma 6 and Lemma 9 there is a SEM M'(M,X,Y,**Z**) with the marginal of Σ(M'(M,X,Y,**Z**)) = Σ(M), and {X} and {Y} d-separated given **Z** in G(M'(M,X,Y,**Z**)) = $G_{Construct}$(G(M),X,Y,**Z**). Because $G_{Construct}$(G(M),X,Y,**Z**) is the directed graph of a latent variable model M'(M,X,Y,**Z**) with correlation matrix that has marginal Σ(M), no correlated errors, and X and Y are d-separated given **Z** in $G_{Construct}$(G(M),X,Y,**Z**), it follows from Lemma 2 that ρ(X,Y.**Z**) = 0 in Σ. ∴

Theorem 2: If M is a SEM, and $\{X_i\}$ and $\{X_j\}$ are d-connected given **Z** in G(M), then $\rho(X_i,X_j.\mathbf{Z}) \neq 0$ in Σ(M).

Proof. Suppose that $\{X_i\}$ and $\{X_j\}$ are d-connected given **Z** in G, and the set of vertices in G is **V**. Form a graph Transform(G) with vertices $\mathbf{T} \cup \mathbf{V}$ in the following way. For a pair of vertices X_k and X_m in **V**, there is a directed edge $X_k \rightarrow X_m$ in Transform(G) if and only if there is a directed edge $X_k \rightarrow X_m$ in G. For vertices X_k and X_m in **V**, there is a vertex $T(X_k,X_m)$ in **T**,and edges $X_m \leftarrow T(X_k,X_m) \rightarrow X_k$ if and if and only if there is a double-headed arrow $X_k \leftrightarrow X_m$ in G. (For convenience in writing equations, for each latent variable $T(X_k,X_m)$ in Transform(G), we will also refer to it as $T(X_m,X_k)$.)

For $\{X_i,X_j\} \cup \mathbf{Z} \subseteq \mathbf{V}$, if $\{X_i\}$ and $\{X_j\}$ are d-connected given **Z** in G, then they are d-connected given **Z** in Transform(G). By Lemma 3 there is a SEM M', with G(M') = Transform(G(M)), and $\rho(X_i,X_j.\mathbf{Z}) \neq 0$.

Let **Double**(X_i) be the set of vertices X_m in G such that there is an edge $X_m \leftrightarrow X_j$ in G. In M',

$$X_i = \sum_{X_m \in \mathbf{Parents}(X_i)} a_{im} X_m + \sum_{X_m \in \mathbf{Double}(X_i)} b_{ij} T(X_i, X_m) + \varepsilon'_i$$

Now define

$$\varepsilon_i = \sum_{X_m \in \mathbf{Double}(X_i)} b_{ij} T(X_i, X_m) + \varepsilon'_i$$

It follows then that

$$X_i = \sum_{X_m \in \mathbf{Parents}(X_i)} a_{ij} X_m + \varepsilon_i$$

is a SEM M, with G(M) = G, and $\rho(X_i,X_j.\mathbf{Z}) \neq 0$. ∴

We will prove Theorem 5 before Theorem 3 because we will use Theorem 5 in the proof of Theorem 3.

Theorem 5: If G_1 and G_2 are path diagrams that are covariance equivalent over **O**, then G_1 and G_2 are d-separation equivalent over **O**.

Proof. Suppose that G_1 and G_2 are not d-separation equivalent over **O**. Suppose without loss of generality that there is some {X}, {Y} and **Z** included in **O**, such that {X} and {Y} are d-connected given **Z** in G_1, but not in G_2. By Theorem 2, there is some SEM M with $G(M) = G_1$ such that $\rho(X,Y.\mathbf{Z}) \neq 0$. By Theorem 1, there is no SEM M' with $G(M') = G_2$, in which $\rho(X,Y.\mathbf{Z}) \neq 0$. Hence G_1 and G_2 are not covariance equivalent over **O**. ∴

Let **Ancestors***(X,G) be the set of ancestors of X, exccluding X, in directed graph G, and **Descendants***(X,G) be the set of descendants of X excluding X in G.

Lemma 10: In a directed acyclic graph G, if X and Y are not adjacent, Y is not an ancestor of X, **Ancestors***(Y,G)\{X} $\subseteq$ **Z**, and **Z** $\cap$ **Descendants***(Y,G) = $\emptyset$, then {X} and {Y} are d-separated given **Z**.

Proof. Suppose that X and Y are not adjacent, but there is a path U that d-connects {X} and {Y} given **Z**. Suppose that U contains an edge $A \rightarrow Y$ and $A \in \mathbf{Z}$. Since A is a non-collider on U ($A \neq X$), it follows that U does not d-connect {X} and {Y} given **Z**, contrary to hypothesis. Suppose then that U contains an edge $A \leftarrow Y$. It follows that U contains a collider, because by hypothesis, Y is not an ancestor of X. Let C be the collider on U closest to Y. C is a descendant of Y, so **Descendants**(C,G)$\subseteq$ **Descendants***(Y,G). Hence **Descendants**(C,G) $\cap$ **Z** = $\emptyset$, so again in this case U does not d-connect {X} and {Y} given **Z**. ∴

Theorem 3: If G_1 and G_2 are directed acyclic graphs, G_1 and G_2 are covariance equivalent if and only if G_1 and G_2 are d-separation equivalent

Proof. By Theorem 5, if G_1 and G_2 are covariance equivalent G_1 and G_2 are d-separation equivalent.

Suppose that G_1 and G_2 are d-separation equivalent, and M is a SEM with directed acyclic graph $G(M) = G_1$. We can form a SEM M'' where G(M'') is a subgraph of G_2[8] and $\Sigma(M'') = \Sigma(M)$ in the following way.

Order the variables in G_2 so that X comes before Y in the ordering only if X is not a descendant of Y. Form a directed acyclic graph G_2' that has G_2 as a subgraph by putting an edge between X and Y if and only if X precedes Y in the ordering. In any SEM M' with graph $G(M') = G_2$', the error term for a variable Y is independent of the parents of Y. Hence if X is a parent of Y in G(M'), the regression coefficient of X when Y is regressed on its parents in G(M') using $\Sigma(M)$ is equal to the linear coefficient of X in the equation for Y. Remove an edge from X to Y G(M') if and only if the linear coefficient of X in the equation for Y is 0. The result is a SEM M'', where $\Sigma(M'') = \Sigma(M)$, and G(M'') is a subgraph of G_2'.

---

[8] Note this includes the possibility that $G(S'') = G_2$.

Suppose X and Y are not adjacent in G_2. In G_2, either X is not an ancestor of Y or Y is not an ancestor of X; suppose without loss of generality that Y is not an ancestor of X. Then X and Y are d-separated given **Parents**(Y,G_2) in G_2.By Lemma 10, in G_2, {X} and {Y} are d-separated given **Parents**(Y,G_2'), because **Parents**(Y,G_2') contains all of the ancestors of Y in G_2, and no descendants of Y in G_2. Because G_1 and G_2 are d-separation equivalent, {X} and {Y} are d-separated given **Parents**(Y,G_2') in G_1. By Theorem 1, ρ(X,Y.**Parents**(Y,G_2')) = 0 in Σ(M). Hence the regression coefficient of X when Y is regressed on **Parents**(Y,G_2') using Σ(M), is equal to 0. It follows that there is no edge between X and Y in G(M''). Hence Σ(M'') = Σ(M), and G(M'') is a subgraph of G_2. ∴

Bibliography

Andersson, S., Madigan, D., and Perlman, M. (1995) A Characterization of Markov Equivalence Classes for Acyclic Digraphs, Technical Report 287, Department of Statistics, University of Washington.

Blalock, H., 1961, Causal Inferences in Nonexperimental Research, (W. W. Norton and Co., New York).

Bollen, K., 1989, Structural Equations with Latent Variables. (Wiley, New York).

Chickering, D. (1995) A Transformational Characterization of Equivalent Bayesian Network Structures, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Philippe Besnard and Steve Hanks (Eds.), Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Frydenberg, M., 1990, The chain graph Markov property,*Scandinvaian Journal of Statistics,* **17**, 333-353.

Geiger, D., and Pearl, J., 1988, Logical and Algorithmic properties of Conditional Independence. Technical Report R-97, Cognitive Systems Laboratory, University of California, Los Angeles.

C. Glymour, R. Scheines, P. Spirtes, and K. Kelly (1987) Discovering Causal Structure: Artificial Intelligence, Philosophy and Statistical Modeling, Academic Press, San Diego, CA.

C. Glymour, P. Spirtes, and R. Scheines (1994), In Place of Regression (in Patrick Suppes: Scientific Philosopher, Paul Humphreys (editor), Vol. 1, Kluwer Academic Publishers, Dordrecht, Holland.

Goldberger, A., Duncan, O. (eds.), 1973, Structural Equation Models in the Social Sciences (Seminar Press, New York).

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

Kiiveri, H. and Speed, T.,1982, Structural analysis of multivariate data: A review, *Sociological Methodology*, Leinhardt, S. (ed.). Jossey-Bass, San Francisco.

Kiiveri, H., Speed, T., and Carlin, J., 1984, Recursive causal models, *Journal of the Australian Mathematical Society*, **36**, 30-52.

Koster, J., (1995) Markov Properties of Non-Recursive Causal Models, Annals of Statistics, November 1995.

Lee, S., and Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. Multivariate Behavioral Research, 25, 313-334.

Lauritzen, S., Dawid, A., Larsen, B., Leimer, H.,1990, Independence properties of directed Markov fields, *Networks*, **20**, 491-505.

Meek, C. (1995) Causal inference and causal explanation with background knowledge, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Philippe Besnard and Steve Hanks (Eds.), Morgan Kaufmann Publishers, Inc., San Mateo, CA, pp. 403-410.

Pearl, J., (1986) Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29**, 241-88.

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

Pearl, J. and Verma, T. (1991). A theory of inferred causation, in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference* (Morgan Kaufmann, San Mateo, CA).

Pearl, J. (1997). Graphs, Causality and Structural Equation Models, Technical Report R-253, Cognitive Science Laboratory, UCLA.

Raftery, A. (1995), Baysian Model Selection in social research, in Marsden, ed. Sociological Methodology, Blackwells, Cambridge MA.

Richardson, T. (1994) Properties of Cyclic Graphical Models, Master's Thesis, Carnegie Mellon University.

Richardson, T. (1996a). A Polynomial Algorithm for Deciding Equivalence in Directed Cyclic Graphical Models. Technical Report PHIL-63, Department of Philosophy, Carnegie Mellon University.

Richardson, T. (1996b) A Discovery Algorithm for Dirtected Cyclic Graphs. In *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Conference* (F.Jensen and E.Horvitz, eds.), 462-469, Morgan Kaufmann, San Francisco.

Richardson, T. (1996c) Feedback Models: Interpretation and Discovery. Ph.D. Thesis, Dept. of Philosophy, Carnegie-Mellon University.

Scheines, R. (1994) Causation, Indistinguishability, and Regression, in Sofstat '93: Advances in Statistial Software 4, Frank Faulbaum (editor), Gustav Fischer Verlag, pp. 89-98.

Spirtes, P. and Glymour, C., 1990, Causal Structure Among Measured Variables Preserved with Unmeasured Variables. Technical Report CMU-LCL-90-5, Laboratory for Computational Linguistics, Carnegie Mellon University.

Spirtes, P., and Glymour, C.(1991) An algorithm for fast recovery of sparse causal graphs, *Social Science Computer Review,* **9**, 62-72.

Spirtes, P., Verma, T. (1992) Equivalence of Causal Models with Latent Variables. Technical Report CMU-PHIL-33, Department of Philosophy, Carnegie Mellon University, October, 1992.

Spirtes, P., Glymour, C., and Scheines, R.(1993) *Causation, Prediction, and Search***,** (Springer-Verlag Lecture Notes in Statistics 81, New York).

Spirtes, P., (1995) Directed Cyclic Graphical Representation of Feedback Models, in Proceedings *of the Eleventh Conference on Uncertainty in Artificial Intelligence*, ed. by Philippe Besnard and Steve Hanks, Morgan Kaufmann Publishers, Inc., San Mateo.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C., (1996). Using D-separation to Calculate Zero Partial Correlations in Linear Models with Correlated Errors, Technical Report CMU-72-Phil.

Spirtes, P., and Richardson, T. (1996), A Polynomial Time Algorithm For Determining DAG Equivalence in the Presence of Latent Variables and Selection Bias, Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics.

Stelzl, I. (1986) Changing causal relations withou changing the fit: Some rules for generating equivalent LISREL-models. Multivariate Behavioral Research, 21, 309-331.

Whittaker, J.,1990, *Graphical Models in Applied Multivariate Statistics* (Wiley, New York).

Wright, S. (1934). The method of path coefficients, *Annals of Mathematical Statistics* **5**, 161-215.

Appendix

d-separation equivalence holds between two path diagrams without correlated errors if and only if they have the same adjacencies and the same unshielded colliders. The following examples show that this simple rule does not work when there are correlated errors. Consider the following path diagrams.

Path Diagram G **Path Diagram G'**

Figure 15

Both of these path diagrams contain the same adjacencies and same unshielded colliders. However, in Path Diagram G', X is d-separated from W given the empty set; in Path Diagram G it is not.