A Generalization of the Tetrad Representation Theorem

Glenn Shafer*, Alexander Kogan*†, and Peter Spirtes‡

Abstract

The tetrad representation theorem, due to Spirtes, Glymour, and Scheines (1993), gives a graphical condition necessary and sufficient for the vanishing of an individual tetrad difference in a linear structural equation model. In this paper, we generalize their result from individual tetrad differences to sets of tetrad differences of a certain form, and we simplify their proof. The generalization allows tighter constraints to be placed on the set of linear models compatible with given data and thereby facilitates the search for parsimonious models for large data sets.

1 Introduction

A *linear structural model* for a set of variables V consists of a set of linear equations, one for each variable $X_i \in V$. The linear equation for X_i expresses X_i as a linear combination of certain of the other variables and an error term E_i. Associated with the model is a directed graph, in which there is an edge from X_i to X_j if and only if X_i appears in the equation for X_j. The model is said to be *recursive* if the directed graph is acyclic (has no directed cycles), or equivalently, if the variables can be ordered so that each variable X_i in the equation for X_j satisfies $i < j$. Such models are widely studied in the social sciences (see, e.g., Bollen 1989).

Recursive linear structural models involving only measured variables generally permit some of the error terms to be correlated (Kang and Seneta 1980). This means that the exogenous variables (the variables with no parents in the graph) are correlated. If we allow latent (unmeasured) variables, however, we can always enlarge the model by introducing latent common parents for any variables with correlated errors, so as to arrive at a model in which all error terms are uncorrelated. This paper is concerned with models of this type.

A *tetrad difference* in a linear structural model is a quantity of the form

$$\rho(X,Y)\rho(Z,W) - \rho(X,Z)\rho(Y,W),$$

where $\rho(X,Y)$ denotes the correlation between X and Y (Spearman 1928). The tetrad representation theorem, due to Spirtes, Glymour, and Scheines (1993), gives a condition on the directed acyclic graph that is necessary and sufficient for such a difference to vanish structurally – i.e., for any values of the coefficients in the linear equations. This condition is concerned with "common

*Accounting and Information Systems, Faculty of Management, Rutgers University, Newark, NJ 07102
E-mail: gshafer@andromeda.rutgers.edu
†RUTCOR, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903-5062
E-mail: kogan@rutcor.rutgers.edu
‡Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213
E-mail: ps7z+@andrew.cmu.edu

causes of $\{X, W\}$ and $\{Y, Z\}$" – variables that are ancestors of either X or W and are also ancestors of either Y or Z. The condition is that either (1) there exists a variable U such that any directed path from a common cause of $\{X, W\}$ and $\{Y, Z\}$ to X or W must go through U, or else (2) there exists a variable U such that any directed path from such a common cause to Y or Z must go through U. Such a variable U is called a *choke point*. The theorem depends on the assumptions mentioned above: the model is recursive, and the errors and hence the exogenous variables are uncorrelated.

In this paper, we generalize the tretrad representation theorem from individual tetrad differences to sets of tetrad differences of a certain form, and we greatly simplify its proof.

For the sake of mathematical clarity, we begin, in Section 2, by studying choke points from a purely graph-theoretical point of view. In Section 3, we take a step from the graph theory towards the statistical interpretation that motivates us by studying polynomials formed from symbols attached to the edges and exogenous nodes in a directed acyclic graph. Then, in Section 4, we obtain the tetrad representation theorem and our generalization of it by interpreting the symbols attached to edges as regression coefficients and the symbols attached to exogenous variables as variances.

In Section 5, we very briefly discuss the enterprise of model search to which the tetrad representation theorem has been applied by Spirtes, Glymour, and Scheines (1993). It is beyond the scope of this paper to review this application in detail or to explore how it can exploit our generalization. The mathematical clarification provided by this paper should, however, provide an underpinning for more effective model search algorithms.

For a full account of these ideas, with proofs, see Shafer, Kogan, and Spirtes (1993).

2 Treks and Choke Points in a Directed Acyclic Graph

We assume that the reader is familiar with the most basic definitions of graph theory. We assume that we are working with finite directed graphs. We call a node *exogenous* if it has no parents, *endogenous* if it does have parents, and *barren* if it has no children. A *path* is a sequence of nodes connected by edges. We allow a sequence consisting of a single node to qualify as a path. If the first node in a path is I, and the last is J, then we say that the path is a *path from I to J*. If it is a path from I to J or a path from J to I, then we say that it is a *path between I and J*.

A path $< X_1 X_2 \ldots X_k >$ is *directed* if either (1) the edge between X_i and X_{i+1} has its arrow pointing to X_{i+1}, for $i = 1, 2, \ldots, k-1$ (in this case, we say that the path is *directed from X_1 to X_k*), or else (2) the edge between X_{i-1} and X_i has its arrow pointing to X_{i-1}, for $i = 2, 3, \ldots, k$ (in this case, we say that the path is *directed from X_k to X_1*, even though it is a *path from X_1 to X_k*).

A path in which the first and last nodes are equal is called a *cycle*. A directed graph in which there are no directed paths that are cycles (no cycles following the arrows) is *acyclic*; it is a *directed acyclic graph*. We henceforth assume that the directed graph with which we are working is a directed acyclic graph containing at least one node.

A node on a path (or more precisely, an occurrence of a node on a path) is a *collider on the path* if (1) it has two neighbors in the sequence (it is not at the beginning or the end), and (2) it has arrows directed to it from both these neighbors.

A *trek between I and J* is a path between I and J that does not contain any colliders. Since there cannot be any arrows pointing towards each other in a trek, there are only a few possibilities

for how the directions of the arrows can change as we move along the trek. First of all, there might not be any arrows at all; if I and J are identical, then $< I >$, the path consisting of I alone, qualifies as a trek between I and J. Second, all the arrows might go from I to J. Third, all the arrows might go from J to I. Fourth, the arrows might change direction once, at a third node Q. The last three possibilities are shown in Figure 1. (The treks shown in this figure are simple, though this is not required by the definition. It should be noted that Spirtes, Glymour, and Scheines (1993), from whom we borrow the name "trek", do require that a trek be simple.)

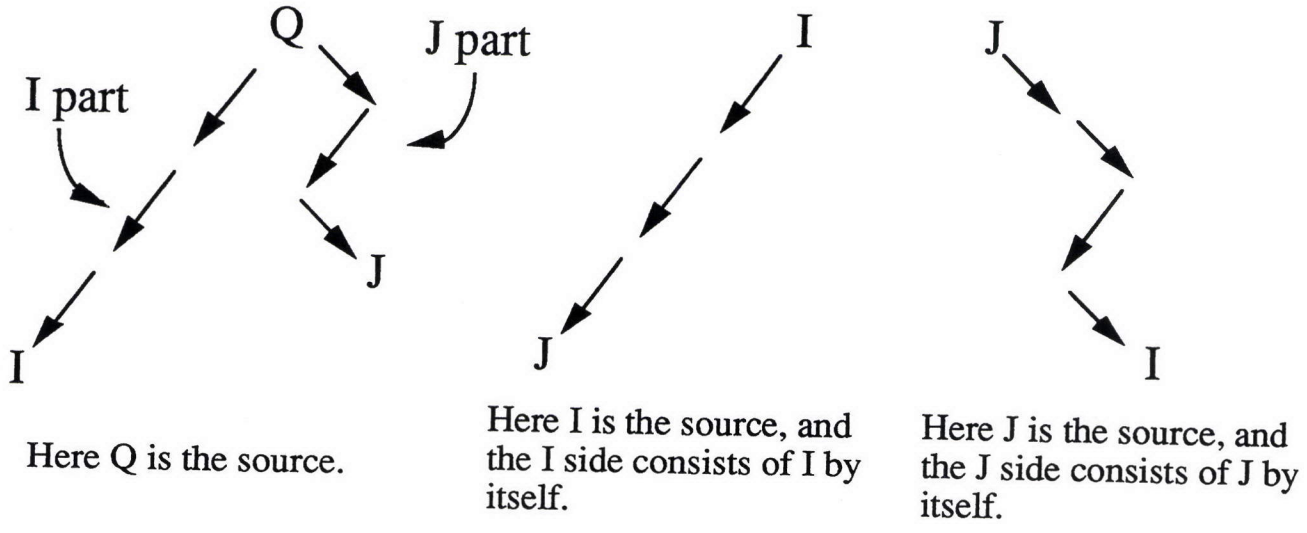


Figure 1: Three types of treks.

Every trek has a unique node to which no arrows are directed; this is called its *source*. If the trek is a path directed from I, then I is its source. If it is composed of a pair of paths directed from Q, then Q is its source. If it consists of a single node I, then I is its source.

Every trek between I and J also has an I *side* and a J *side*. The I side is the subpath directed from the source to I; the J side is the subpath directed from the source to J. The two sides may have edges in common.

Consider two sets of nodes, **I** and **J**. We say that a trek is a *trek between* **I** *and* **J** if it is a trek between some element I of **I** and some element J of **J**. If X is a node in such a trek τ, then we say that X is *on the* **I** *side of* τ if X is in τ's I side, and we say that X is *on the* **J** *side of* τ if X is in τ's J side. If X is the source of τ, then it is on both the **I** side and the **J** side. If τ is simple, its source is the only node that is on both sides. Notice also that if one or both of **I** and **J** are empty, then there are no treks between them.

The definitions in the preceding paragraph apply even if the sets **I** and **J** overlap. If they do overlap, then a trek consisting of a single node that is in their intersection qualifies as a trek between them. A trek between two distinct nodes that are both in both **I** and **J** is also a trek between **I** and **J**, but when we speak of it as such, we must arbitrarily specify one side as the **I** side and the other as the **J** side. The definitions even apply in the case where **I** and **J**, as sets, are identical. In this case, we still think of **I** and **J** as two distinct labels, and we still label one of the sides of the trek as the **I** side and the other as the **J** side.

We say that a node X is a *choke point between* **I** *and* **J** if two conditions are met:

1. every trek between **I** and **J** (if there are any) goes through X, and

2. either (a) X is on the **I** side of every such trek, or (b) X is on the **J** side of every such trek.

If condition 2a is satisfied, then we say that X is an **I**-*side choke point*. If condition 2b is satisfied, then we say that X is a **J**-*side choke point*. If condition 1 is satisfied (whether or not condition 2 is satisfied), then we say that X is a *weak choke point between* **I** *and* **J**. Figures 2 and 3 illustrate these definitions.

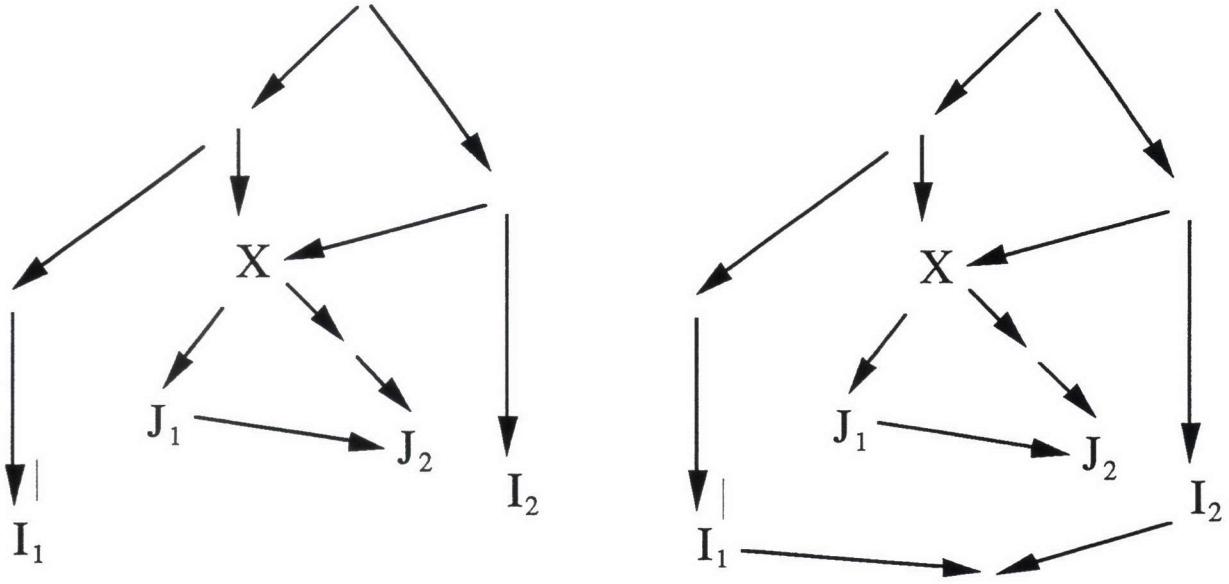


Figure 2: In both these graphs, X is a choke point between $\{I_1, I_2\}$ and $\{J_1, J_2\}$ on the $\{J_1, J_2\}$ side.

Figure 3: There are no choke points between $\{I_1, I_2\}$ and $\{J_1, J_2\}$ in either of these graphs, though there is a weak choke point in both cases: X on the left and I_2 on the right. On the left, X is not on the $\{I_1, I_2\}$ side of $< I_1 X J_2 >$, and not on the $\{J_1, J_2\}$ side of $< I_2 X J_1 >$. On the right, I_2 is not on the $\{I_1, I_2\}$ side of $< I_1 I_2 J_2 >$, and not on the $\{J_1, J_2\}$ side of $< I_2 J_1 >$.

In the case where **I** and **J** each contain exactly two nodes, our definition of choke point is essentially equivalent to the definition given by Spirtes, Glymour, and Scheines (1993, p. 196). It is simpler than their definition, however, and this simplification is basic to the contributions of this paper.

The next lemma clarifies the structure of the choke points between two sets **I** and **J**.

Lemma 2.1 *Let* **C** *designate the set of weak choke points between* **I** *and* **J**. *Then every trek from* **I** *to* **J** *goes through the nodes in* **C** *in the same order.*

Lemma 2.1 tells us in particular that any trek between **I** and **J** goes through all the choke points in the same order. So if there are choke points between **I** and **J**, we can talk about the one nearest **I** and the one nearest **J**. Similarly, if there are **I**-side choke points, then we can talk about the **I**-side choke point nearest the sources of the treks between **I** and **J**; this is the same choke point for all such treks. The source of a trek from **I** to **J** always lies between the last **I**-side choke point and before the first **J**-side choke point, except that in some cases it may be equal to one or the other or both.

The next lemma will help us prove Theorem 2.3, which explains what happens when a choke point does not exist.

Lemma 2.2 *Consider sets* $\mathbf{I_1}, \mathbf{I_2}, , \ldots, \mathbf{I_k}$. *Suppose that for each* i, $1 \leq i \leq k$, *there is at least one choke point between* $\mathbf{I_i}$ *and* **J**. *Let* $\mathbf{C_i}$ *designate the set consisting of all the choke points between* $\mathbf{I_i}$ *and* **J**. *Set* $\mathbf{C} = \bigcup_{i=1}^{k} \mathbf{C_i}$ *and* $\mathbf{I} = \bigcup_{i=1}^{k} \mathbf{I_i}$. *Then the following statements hold.*

1. *Every trek from* $\bigcap_{i=1}^{k} \mathbf{I_i}$ *to* **J** *(if there are any) goes through all the nodes in* **C** *and does so in the same order.*
2. *Suppose there does exist a trek from* $\bigcap_{i=1}^{k} \mathbf{I_i}$ *to* **J**. *(This means, in particular, that* $\bigcap_{i=1}^{k} \mathbf{I_i}$ *is non-empty.) Then the node in* **C** *nearest* **J** *is a choke point between* **I** *and* **J**.

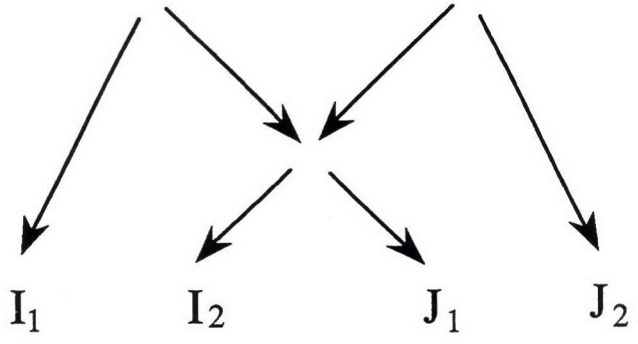


Figure 4: There cannot be a choke point between $\{I_1, I_2\}$ and $\{J_1, J_2\}$, because the $\{I_1, I_2\}$ side of π is disjoint from τ, and the $\{J_1, J_2\}$ side of τ is disjoint from π.

It is evident from the definition of choke point that there is no choke point between **I** and **J** if there are two non-intersecting treks between **I** and **J**, or even if there are two treks between **I** and **J** that intersect only on the **I** side of the first and the **J** side of the second, as in Figure 4. The following theorem tells us that the converse is true as well: if there is no choke point, then there exist treks π and τ such that the **I** side of π does not intersect τ and the **J** side of τ does not intersect π, though possibly the **J** side of π and the **I** side of τ may intersect one or more times.

Theorem 2.3 *If there is no choke point between* **I** *and* **J**, *then there exist treks* π *and* τ *between* **I** *and* **J** *such that the* **I** *side of* π *is disjoint from* τ *and the* **J** *side of* τ *is disjoint from* π.

In summary, there is no choke point between **I** and **J** if and only if there exist treks π and τ such that the **I** side of π does not intersect τ and the **J** side of τ does not intersect π.

Corollary 2.4 *There is a choke point between* **I** *and* **J** *if and only if there is a choke point between* $\{I_1, I_2\}$ *and* $\{J_1, J_2\}$ *whenever* $\{I_1, I_2\} \subseteq \mathbf{I}$ *and* $\{J_1, J_2\} \subseteq \mathbf{J}$.

3 Correlation and Covariance Structures

In this section we study polynomials formed from symbols attached to the edges and exogenous variables of a directed acyclic graph. Anticipating the statistical interpretation in the next section, we call the structures obtained by attaching these symbols to the directed acyclic graph *correlation* and *covariance structures*.

A *correlation structure* is a directed acyclic graph in which we have attached distinct symbols to the edges. We are interested in polynomials formed by multiplying the symbols along treks and then adding the products.

Given a trek π in a correlation structure, we write $*\pi$ for the product of the edge symbols along the trek, and we call $*\pi$ the *edge product over* π. If π consists of a single node, then by convention, $*\pi$ is equal to 1. Given any two nodes I and J, we set

$$*(I,J) = \sum\{*\pi | \pi \in \sigma(I,J)\},$$

where $\sigma(I,J)$ is the set of all simple treks from I to J. We call $*(I,J)$ the *simple trek sum between* I *and* J. Notice that since $< I >$ is the only simple trek between I and itself, $*(I,I)$ is equal to 1.

By convention, $*(I,J)$ is equal to 0 if there are no simple treks (i.e., no treks at all) between I and J. The converse is also true, of course. Though mathematically trivial, this observation is sufficiently significant for the application considered in the next section that we call it a theorem:

Theorem 3.1 *If I and J are nodes in a correlation structure, then the following statements are equivalent:*

1. $*(I,J) = 0$.
2. *There is no trek between I and J.*
3. *There is no node X such that there exists a directed path from X to I and a directed path from X to J.*

Given a trek π and a node X, we write $\#_X(\pi)$ for the difference between the number of arrows on π that come into X and the number that go out of X. It is easily seen that if the endpoints of π are distinct, then

$$\#_X(\pi) = \begin{cases} 1, & X \text{ is an endpoint and not the source,} \\ 0, & X \text{ is neither the source nor endpoint,} \\ -1, & X \text{ is both the source and an endpoint,} \\ -2, & X \text{ is the source and not an endpoint.} \end{cases}$$

Given two treks, π and τ, we write $\#_X(\pi,\tau)$ for the sum $\#_X(\pi) + \#_X(\tau)$.

The following lemma will help us understand the significance of individual terms in the product of two simple trek sums.

Lemma 3.2 *Suppose the four endpoints of the treks π and τ are distinct. Then $\#_X(\pi,\tau)$ is negative if and only if X is the source of one (or both) of π and τ.*

This lemma is useful when we are given a product $*\pi * \tau$ but do not have the treks π and τ themselves. All we can see by looking at $*\pi * \tau$ is the edges involved, with their multiplicities (some edges may be in both treks or on both sides of one of the treks). But from this information, we can calculate $\#_X(\pi, \tau)$ for every node X, and hence we can identify the sources involved in the two treks.

The following theorem is analogous to Theorem 3.1, inasmuch as it shows how a fact about the graph can be represented by a fact about a polynomial in the edge symbols.

Theorem 3.3 *If I_1, I_2, J_1, and J_2 are distinct nodes in a correlation structure, then the following statements are equivalent.*

1. *There is a choke point between $\{I_1, I_2\}$ and $\{J_1, J_2\}$.*
2. *The polynomial $*(I_1, J_1) * (I_2, J_2)$ is equal to the polynomial $*(I_1, J_2) * (I_2, J_1)$.*

We now develop an analogue of Theorem 3.3 for the case where we assume that each endogenous node is the only child of at least one exogenous parent, we attach symbols to the exogenous nodes as well as to all the edges, and in the place of simple treks, we consider treks that have simple sides and have exogenous nodes as sources.

A *covariance structure* is a directed acyclic graph in which each endogenous node is the only child of at least one exogenous parent, and we have attached distinct symbols to each exogenous node as well as to each edge.

We call a trek an *ultratrek* if its source is exogenous. Given an ultratrek α between I and J, we write $X(\alpha)$ for the first node starting from I (or, equivalently, the first node starting from J) where α's I and J sides intersect. We call $X(\alpha)$ the *base* of α. We write $\downarrow \alpha$ for the subtrek that follows α from I to $X(\alpha)$ and then directly to J, and we write $\uparrow \alpha$ for the trek that follows α from $X(\alpha)$ to the source and then back to $X(\alpha)$. (See Figure 5.) Notice that $\downarrow \alpha$ is a simple trek from I to J, in fact it is α's only simple subtrek from I to J. On the other hand, $\uparrow \alpha$ is an ultratrek.

Given an ultratrek in a covariance structure, we write $\Diamond\pi$ for the product of all the symbols along the ultratrek (the edge symbols together with the symbol attached to the source, which is exogenous), and we call $\Diamond\pi$ the *symbol product over* π. If π consists of a single node, its source, then $\Diamond\pi$ is simply the symbol attached to the source. Given any two nodes I and J, we set

$$\Diamond(I, J) = \sum\{\Diamond\pi | \pi \in \upsilon(I, J)\},$$

where $\upsilon(I, J)$ is the set of all ultratreks from I to J. We call $\Diamond(I, J)$ the *ultratrek sum between I and J*. By convention, the ultratrek sum is equal to 0 if there are no ultratreks (i.e., no treks) between I and J. If I is exogenous, then $< I >$ is the unique ultratrek between I and I, and hence $\Diamond(I, I)$ is the symbol attached to I.

Theorem 3.4 *If I_1, I_2, J_1, and J_2 are distinct nodes in a covariance structure, then the following statements are equivalent.*

1. *There is a choke point between $\{I_1, I_2\}$ and $\{J_1, J_2\}$.*
2. *The polynomial $\Diamond(I_1, J_1)\Diamond(I_2, J_2)$ is equal to the polynomial $\Diamond(I_1, J_2)\Diamond(I_2, J_1)$.*

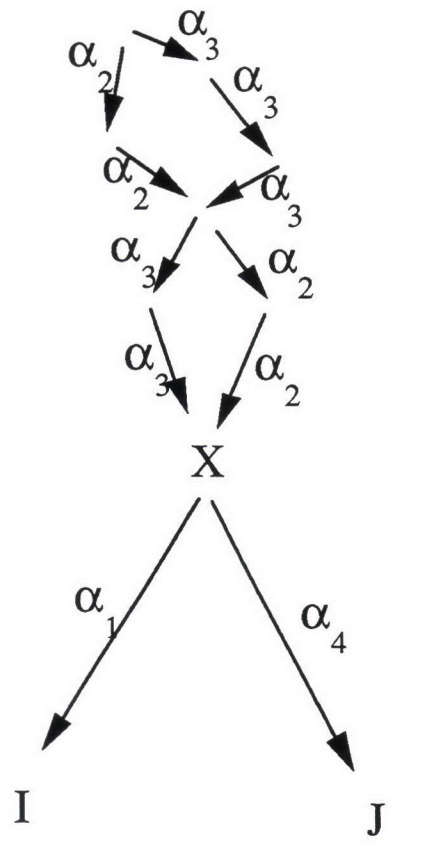


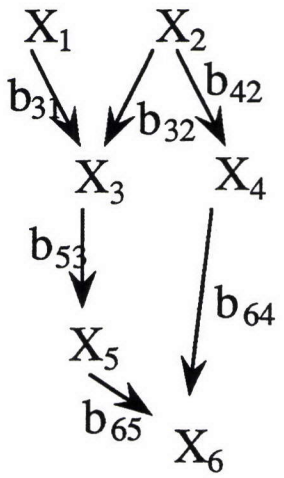
Figure 5: $\downarrow \alpha = \alpha_1\alpha_4$ and $\uparrow \alpha = \alpha_2\alpha_3$.

4 Application to Statistical Inference

Let us now interpret correlation structures by taking the nodes to represent real-valued random variables, with a joint probability distribution in which each variable has zero partial correlations, given its parents, with its non-descendants (Pearl 1988). This implies in particular that the exogenous variables are all uncorrelated with each other. We interpret the symbols on the edges pointing into an endogenous variable as the regression coefficients in the linear regression of that variable on its parents, as in Figure 6.

We interpret covariance structures in a similar way: we take the nodes to represent real-valued random variables, such that the exogenous variables are uncorrelated, the symbols on the exogenous variables represent their variances, and each endogenous variable is a linear combination of its parents, with the symbols on the edges representing the coefficients.

Since the errors in the regression equations for the endogenous variables in a correlation structure have zero correlations with each other and with the exogenous variables in the structure (this follows from the assumption that each variable has zero partial correlations with all its non-descendants given its parents), we can expand the correlation structure to a covariance structure by adding to each endogenous variable a parent representing the error in its regression equation, as in Figure 7. Notice that when we add the error, we put a new symbol on it, representing its variance, and also a symbol on the new edge. The symbol on the edge replaces the unit coefficient for the error



$$X_3 = b_{31}X_1 + b_{32}X_2 + E_3$$

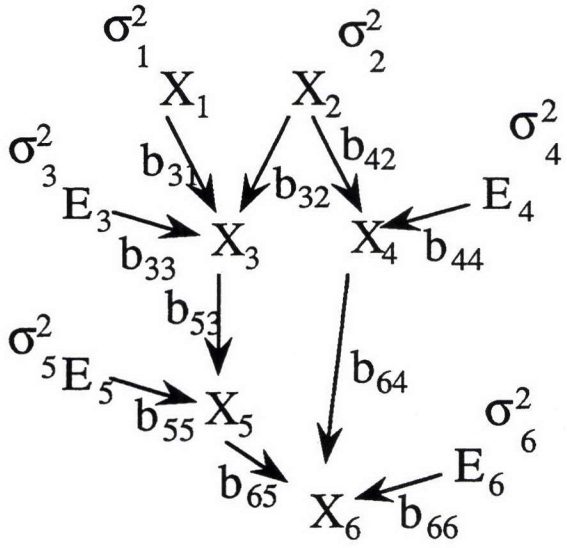
$$X_4 = b_{42}X_2 + E_4$$

$$X_5 = b_{53}X_3 + E_5$$

$$X_6 = b_{64}X_4 + b_{65}X_5 + E_6$$

Figure 6: The graph on the left is a correlation structure associated with the recursive linear regression equations on the right.

in the regression equation, so that the regression equation becomes the equation representing the endogenous variable as a linear combination of its parents in the covariance structure.



$$X_3 = b_{31}X_1 + b_{32}X_2 + b_{33}E_3$$

$$X_4 = b_{42}X_2 + b_{44}E_4$$

$$X_5 = b_{53}X_3 + b_{55}E_5$$

$$X_6 = b_{64}X_4 + b_{65}X_5 + b_{66}E_6$$

Figure 7: The covariance structure corresponding to the correlation structure of Figure 6.

Let us write $\rho(I, J)$ and $\mathrm{Cov}(I, J)$, respectively, for the correlation and covariance of any pair of random variables I and J. The following theorem, which is easily proven by induction on the number of variables in the directed acyclic graph, shows the substantive significance of the trek sums.

Theorem 4.1

1. *If every variable in a correlation structure has variance one, then* $\rho(I, J) = *(I, J)$ *for every pair of variables* I *and* J *in the structure.*

2. $\mathrm{Cov}(I, J) = \diamondsuit(I, J)$ *for every pair of variables* I *and* J *in a covariance structure.*

Statement 1 is a special case of Sewall Wright's (1934) rule for decomposing correlation by path analysis. Wright's rule reduces to this statement when the errors in the path model are uncorrelated.

We are interested in constraints on correlations or covariances that are equivalent to the vanishing of polynomials in the symbols in a covariance structure. Examples include the constraint that a particular correlation, say $\rho(I, J)$, should equal zero, which is equivalent to

$$\diamondsuit(I, J) = 0, \tag{1}$$

or the constraint that a particular "tetrad difference", say

$$\rho(I_1, J_1)\rho(I_2, J_2) - \rho(I_1, J_2)\rho(I_2, J_1),$$

should vanish, which is equivalent to

$$\diamondsuit(I_1, J_1)\diamondsuit(I_2, J_2) - \diamondsuit(I_1, J_2)\diamondsuit(I_2, J_1) = 0. \tag{2}$$

We call such a constraint on correlations *structural* if the polynomial is identically equal to zero – i.e., if the constraint holds for every possible choice of the exogenous variances and endogenous coefficients. We call it *accidental* otherwise – i.e., if it holds only for particular variances and coefficients. It is reasonable to call such constraints accidental, for they would not be expected if the variances and correlations were themselves chosen at random from some continuous joint probability distribution. If we specify a finite class of such constraints (e.g., all possible vanishing correlations, partial correlations, and tetrad differences for a set of variables) before examining a body of data extensive enough to test them, then it will be reasonable for us to conjecture that those constraints that do hold are structural, and this will give us information about the correlation structure.

The next theorem, the tetrad representation theorem, is an important tool in this program of statistical inference. The ideas involved in this theorem go back to Spearman (1928), but the theorem was formulated and proven only recently, by Spirtes, Glymour, and Scheines (1993).

Theorem 4.2 *Suppose I_1, I_2, J_1, and J_2 are distinct variables. Then*

$$\rho(I_1, J_1)\rho(I_2, J_2) - \rho(I_1, J_2)\rho(I_2, J_1) = 0$$

is a structural constraint if and only if there is a choke point between $\{I_1, I_2\}$ and $\{J_1, J_2\}$.

This theorem follows immediately from Theorems 3.4 and 4.1.

Corollary 2.4 yields the following generalization of the tetrad representation theorem.

Theorem 4.3 *Suppose* **I** *and* **J** *are disjoint set of variables in a correlation structure. Then*

$$\rho(I_1, J_1)\rho(I_2, J_2) - \rho(I_1, J_2)\rho(I_2, J_1) = 0$$

is a structural constraint for every subset $\{I_1, I_2\}$ of **I** *and every subset $\{J_1, J_2\}$ of* **J** *if and only if there is a choke point between* **I** *and* **J**.

5 Conclusion

The assumption that given tetrad differences among measured variables are structurally zero will obviously impose constraints on a structural model that includes those variables. How can these constraints be exploited in searching for or constructing such models?

Answers to this question were first explored by Spearman and by Kelley (1928), who used vanishing tetrad and pentad differences to search for psychological models of intelligence. Their strategy can be generalized in the following way:

1. Perform a statistical test that each tetrad difference is equal to zero, and form the set of tetrad differences that pass the test. (Wishart (1928) described a statistical test for vanishing tetrad differences drawn from a joint normal distribution, and Bollen (1990) has described an asymptotically distribution free test.)

2. Identify for an initial model the set of tetrad differences that are equal to zero for all values of the linear coefficients and all positive values of the variances.

3. Modify the model so that the set of tetrad differences entailed to vanish is closer (relative to a suitable metric) to the set of tetrad differences that pass the test. (Several methods for this step have been described by Spirtes, Glymour, and Scheines (1993) and Spirtes, Scheines, and Glymour (1990).)

In a variety of simulation tests reported in the preceding references, variations of this strategy (implemented in the TETRAD II program) have been shown to be reliable for large ($n \approx 2000$) samples from Gaussian models.

The results of this paper are used in Step 2. The evaluation of their usefulness will depend, therefore, on an eventual overall assessment of the three-step strategy. We expect that the strategy will be most successful when it is extended so as to use additional constraints and other sources of evidence in addition to vanishing tetrad differences.

Acknowledgements

Glenn Shafer's research for this paper was partially supported by the National Science Foundation (Grant SBE9213674), Alexander Kogan's research for this paper was partially supported by the Office of Naval Research (Grants N00014-92-J1375 and N00014-92-J4083) and AFOSR (Grant F49620-93-1-0041), and Peter Spirtes' research for this paper was partially supported by the Office of Naval Research (Grant N00014-91-J-1361).

References

[1] Bollen, K. A. (1989) *Structural Equations with Latent Variables*. John Wiley & Sons, New York.

[2] Bollen, K. A. (1990). "Outlier screening and a distribution-free test for vanishing tetrads". *Sociological Methods and Research* **19**, 80-92.

[3] Kang, K. M. and E. Seneta (1980). "Path Analysis: An Exposition". In *Developments in Statistics*, Vol. 3, ed. Paruchuri R. Krishnaiah, 217-246. Academic Press, New York.

[4] Kelley, T. (1928). *Crossroads in the Mind of Man.* Stanford University Press, Stanford, CA.

[5] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

[6] Spearman, C. (1904). "General intelligence objectively determined and measured". *American Journal of Psychology* **15**, 201-293.

[7] Spearman, C. (1928). "Pearson's contribution to the theory of two factors". *British Journal of Psychology* **19**, 95-101.

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

[9] Spirtes, P., R. Scheines, and C. Glymour (1990). "Simulation studies of the reliability of computer aided specification using the TETRAD II, EQS, and LISREL Programs". *Sociological Methods and Research* **19**, 3-66.

[10] Wishart, J. (1928). "Sampling errors in the theory of two factors". *British Journal of Psychology* **19**, 180-187.

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