A Generalization of the Tetrad Representation Theorem

Glenn Shafer, Alex Kogan, and Peter Spirtes

The tetrad representation theorem, due to Spirtes, Glymour, and Scheines (1993), gives a graphical condition necessary and sufficient for the vanishing of tetrad differences in a linear correlation structure. This note simplifies their proof and generalizes the theorem. In order to make the ideas as accessible as possible to mathematicians who might develop them further, we begin with a thorough exposition of their purely graph-theoretical aspects.

Part I. Treks and Choke Points in a Directed Acyclic Graph

We assume that the reader is familiar with the most basic definitions of graph theory. Recall that a graph is an object consisting of nodes and edges between them. It is directed if its edges are directed (marked with arrows). We assume that we are working with a finite directed graph, in which edges are always between distinct nodes and there is at most one edge between any pair of distinct nodes. We use the usual definitions of parent, child, descendant, and ancestor; if there is an edge between X and Y with its arrowing pointing from X to Y, we say that X is a parent of Y and Y is a child of X. We call a node exogenous if it has no parents, endogeneous 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 $<X_1X_2...X_k>$ 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,...,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,...,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. It is easy to see that a directed acyclic graph always has at least one barren node, and it remains a directed acyclic graph if we delete that node and any edges to it.

We call any subsequence of a path $<X_1X_2...X_k>$ that is also a path a subpath. Notice that any subpath of a directed path is also a directed path. We call a subpath of the form $<X_iX_{i+1}...X_j>$, where $1 \leq i \leq j \leq k$, a chunk. If $1<i_1<\cdots<i_r<k$, then we say that $<X_1X_2...X_k>$ is composed of the r+1 chunks $<X_1X_2...X_{i_1}>$, $<X_{i_1}X_{i_1+1}...X_{i_2}>$, ... $<X_{i_r}X_{i_r+1}...X_k>$. (Each chunk begins with the node with which the preceding chunk ends.)

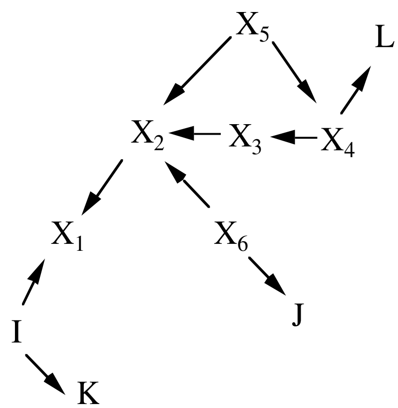


Figure 1 The path <$IX_1X_2X_3X_4X_5X_2X_6J$> has the simple subpath <$IX_1X_2X_6J$>.

If the nodes in a path are distinct (the path does not intersect itself), then we say that the path is <u>simple</u>. We leave the proof of the following lemma to the reader.

> **Lemma 1** Any path from I to J has at least one subpath that is a simple path from I to J.

Figure 1 illustrates the lemma.

A node on a path (or more precisely, an occurrence of a node on a path) is a <u>collider on the path</u> 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. In Figure 1, for example, there are two colliders on the path <$IX_1X_2X_3X_4X_5X_2X_6J$>: X_1 and the second occurrence of X_2. We will leave it to the reader to prove the following lemma.

> **Lemma 2** A path <$X_1X_2...X_k$> has a collider if and only if there exist integers i and j such that $1<i\leq j<k$ such that the arrow between X_{i-1} and X_i points to X_i, and the arrow between X_j and X_{j+1} points to X_j.

This lemma says that if there are arrows pointing towards each other on a path (as in Figure 2), then there must be a collider somewhere between them.

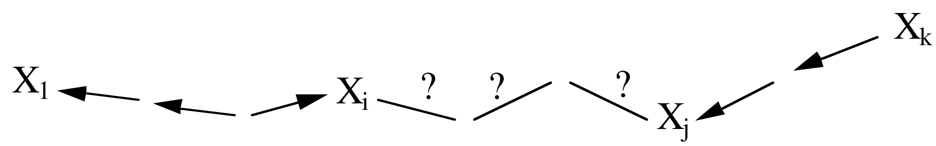
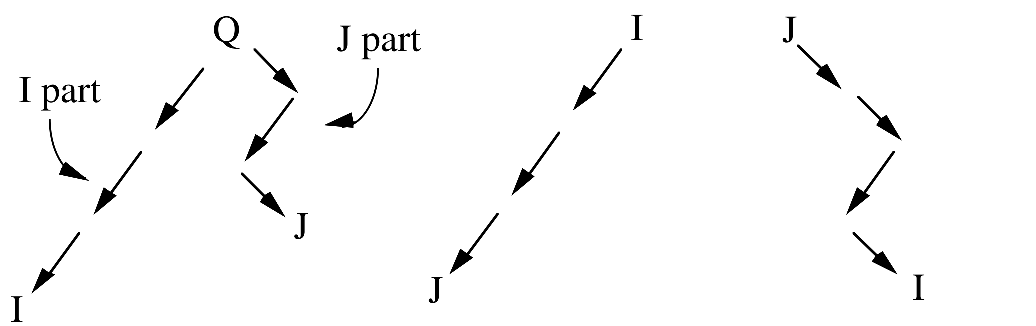


Figure 2 We do not know the directions of the arrows on the edges X_i between X_j, but no matter what their directions, there must be a collider between X_i and X_j.

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 3. (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.)



Here Q is the source.

Here I is the source, and the I side consists of I by itself.

Here J is the source, and the J side consists of J by itself.

Figure 3 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. If the trek is a path directed from I to J, then the I side consists of I by itself. If it is a path directed from J to I, then the J side consists of J by itself. If it consists of a single node, this node is both the I side and the J side.

Any subpath of a trek is also a trek—we call it, naturally, a subtrek. If π is a subtrek of τ, and both go from I to J, then the I side of π is a subpath of the I side of τ, and the J side of π is a subpath of the J side of τ. By Lemma 1, every trek between I and J has at least one subtrek that is simple trek between I and J. Notice also that if $<X_1X_2...X_r>$ and $<X_rX_{r+1}...X_k>$ are treks, then the composition $<X_1X_2...X_rX_{r+1}...X_k>$ is a trek if and only if the edges $X_{r-1}X_r$ and X_rX_{r+1} do not both have their arrows pointing towards X_r.

Though we usually think of a trek visually, as a set of edges, or as two paths directed from the source, we will sometimes need to insist on the formal definition, according to which a trek, like any path, is a sequence of nodes. We will have occasion, for example, to speak of a trek from I to J. This refers to a sequence of nodes beginning with I and ending with J that forms a trek between I and J; the arrows need not be directed from I to J.

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 <u>choke point between **I** and **J**</u> 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 <u>**I**-side choke point</u>. If condition 2b is satisfied, then we say that X is a <u>**J**-side choke point</u>. If condition 1 is satisfied (whether or not condition 2 is satisfied), we that X is a <u>weak choke point between **I** and **J**</u>. Figures 4 and 4a illustrate these definitions.

We may, if we wish, require that the treks in the definition of choke point be simple. Every trek between **I** and **J** goes through X if and only if every simple trek between **I** and **J** goes through X, and every trek between **I** and **J** goes through X on the **I** side if and only if every simple trek between **I** and **J** goes through X on the **I** side.

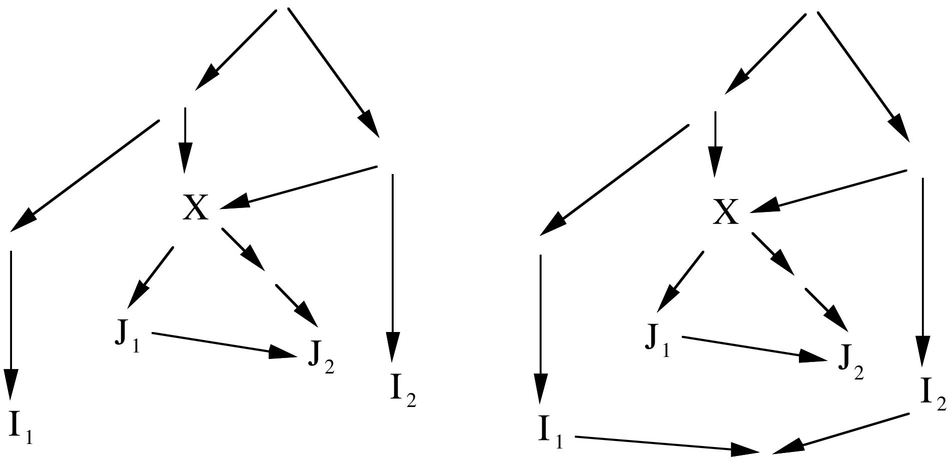


Figure 4 In both these graphs, X is a choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$ on the $\{J_1,J_2\}$ side.

Figure 5 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 $<I_1XJ_2>$, and not on the $\{J_1,J_2\}$ side of $<I_2XJ_1>$. On the right, I_2 is not on the $\{I_1,I_2\}$ side of $<I_1I_2J_2>$, and not on the $\{J_1,J_2\}$ side of $<I_2J_1>$.

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 lists some obvious consequences of the definition of choke point.

Lemma 3

(1) If there is no trek between **I** and **J**, then every node in the graph qualifies as a choke point between **I** and **J**. This happens, for example, when one or both of the sets are empty.

(2) If **I** contains only one node, then this node is a choke point between **I** and **J**.

(3) If X is a choke point between **I** and **J**, and $\mathbf{I'} \subseteq \mathbf{I}$, then X is a choke point between **I'** and **J**.

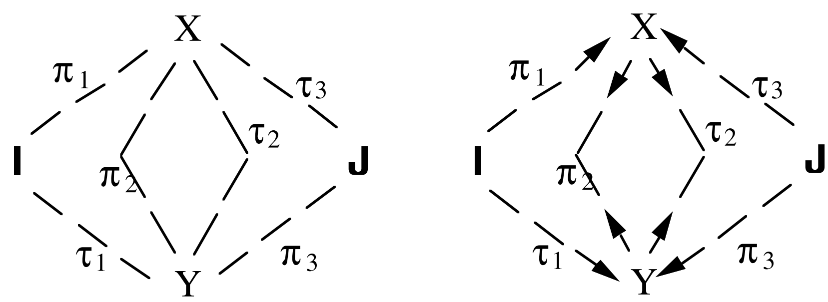
(4) If I is in both **I** and **J**, and there is a choke point X between **I** and **J**, then X=I.

(5) If **I** and **J** have more than one node in common, then they do not have a choke point between them.

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

Lemma 4 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.

Proof We prove the lemma by contradiction. Suppose X and Y are distinct nodes in **C**, π is a trek from **I** to **J** that goes through X first, and τ is a trek from **I** to **J** that goes through Y first. Decompose π and τ into chunks as in Panel i of Figure 6; $\pi=\pi_1\pi_2\pi_3$, where π_1 goes from a node in **I** to X, π_2 goes from X to Y, and π_3 goes from Y to a node in **J**; and $\tau=\tau_1\tau_2\tau_3$, where τ_1 goes from a node in **I** to Y, τ_2 goes from Y to X, and τ_3 goes from X to a node in **J**. Use these chunks to form two new paths from **I** to **J**: $\lambda_1=\pi_1\tau_3$ and $\lambda_2=\tau_1\pi_3$. Neither of these new paths are treks; λ_1 cannot be a trek because it avoids the weak choke point Y, and λ_2 cannot be a trek because it avoids the weak choke point X. So both must contain colliders. Since there are no colliders in the chunks, the colliders must occur where the chunks are joined; X must be a collider on λ_1, and Y must be a collider on λ_2, as indicated by the arrows into X and Y in Panel ii. In order to avoid X or Y being colliders on π or τ, we must then also have the arrows out of X and Y shown there. But even this does not avoid colliders on π and τ, for by Lemma 2, the arrows out of X and Y imply that there must be colliders on the chunks π_2 and τ_2. This contradicts our assumption that π and τ are treks. **Q.E.D.**



i. The treks π and τ. ii. Implied arrows.

Figure 6

Lemma 4 tells us in particular if 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 1, which explains what happens when a choke point does not exist.

Lemma 5 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 $\mathbf{C}=\cup\mathbf{C}_i$ and $\mathbf{I}=\cup\mathbf{I}_i$. Then the following statements hold.

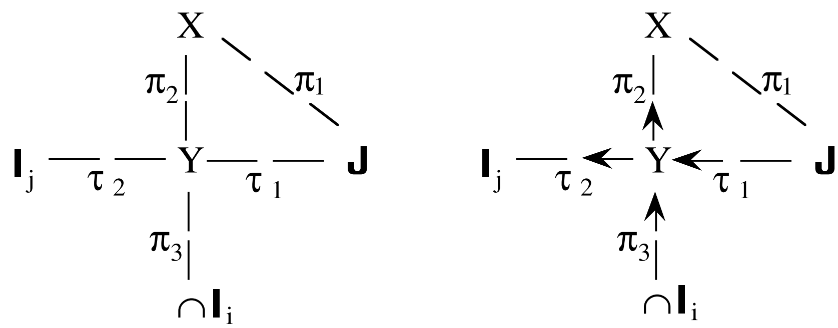
(1) Every trek from $\cap\mathbf{I}_i$ to $\mathbf{J}$ (if there are any) goes through all the nodes in $\mathbf{C}$ and does so in the same order.

(2) Suppose there does exist a trek from $\cap\mathbf{I}_i$ to $\mathbf{J}$. (This means, in particular, that $\cap\mathbf{I}_i$ is non-empty.) Then the node in $\mathbf{C}$ nearest $\mathbf{J}$ is a choke point between $\mathbf{I}$ and $\mathbf{J}$.

Proof The truth of Statement 1 is obvious: all the nodes in $\mathbf{C}$ are choke points between $\cap\mathbf{I}_i$ and $\mathbf{J}$ (by Statement 3 of Lemma 3) and every trek from $\cap\mathbf{I}_i$ to $\mathbf{J}$ goes through the choke points between $\cap\mathbf{I}_i$ and $\mathbf{J}$ in the same order (by Lemma 4).

We prove Statement 2 by contradiction. Choose a trek π from $\mathbf{J}$ to $\cap\mathbf{I}_i$. Let X designate the node in $\mathbf{C}$ nearest $\mathbf{J}$, and suppose X is not a choke point between $\mathbf{I}$ and $\mathbf{J}$. We consider two cases: X is a weak choke point between $\mathbf{I}$ and $\mathbf{J}$, or it is not.

First, suppose X is not a weak choke point between $\mathbf{I}$ and $\mathbf{J}$. Then we can choose a trek τ from $\mathbf{J}$ to $\mathbf{I}$ that does not go through X at all. The last node in τ is in one of the $\mathbf{I}_i$, say $\mathbf{I}_j$. Both π and τ are treks from $\mathbf{J}$ to $\mathbf{I}_j$. Let Y designate a choke point between $\mathbf{I}_j$ and $\mathbf{J}$. Then π goes through X and then Y, and τ goes through Y but not X. Decompose π and τ into chunks; $\pi=\pi_1\pi_2\pi_3$, where π_1 goes from $\mathbf{J}$ to X, π_2 goes from X to Y, and π_3 goes from Y to $\cap\mathbf{I}_i$; and $\tau=\tau_1\tau_2$, where τ_1 goes from $\mathbf{J}$ to Y, and τ_2 goes from Y to $\mathbf{I}_j$. See Panel i of Figure 7. (In interpreting Panel i, the reader should remember that we are not making any assumptions about the number of edges in the chunks. Some chunks might have no edges at all. For example, we might have $\tau_2=<Y>$.) Since it does not go through X, the path $\tau_1\pi_3$ is not a trek. So Y must be a collider on it. This gives us the arrows into Y in Panel ii. Since Y is not a collider on the treks π and τ, we must also have the arrows out from Y in that panel. The resulting picture implies that Y is on the **J** side (and not on the $\mathbf{I}_j$ side) of the trek π but on the $\mathbf{I}_j$ side (and not on the **J** side) of the trek τ, and this contradicts the assumption that Y is a choke point between $\mathbf{I}_j$ and **J**.



i. The treks π and τ. ii. Implied arrows.

Figure 7

Now suppose X is a weak choke point between **I** and **J**. Then any trek τ from **J** to **I** must go through X, and it must do so before it goes through Y (for this is the order in which π, which is also a trek from **J** to **I**, goes through them), but we can choose τ so that it and π do not go through X on the same side. Again, we choose j so that the last node in τ is in $\mathbf{I}_j$, and we choose a choke point Y between $\mathbf{I}_j$ and **J**. The graphs on the left of Figure 8 illustrate the two possibilities. In each case, the graph on the right shows the additional arrows that are implied, and these arrows imply that $\tau_1\tau_2\pi_3$ is a trek from **J** to $\cap\mathbf{I}_i$ that goes through X on a different side than the trek π does, thus contradicting the assumption that X is a choke point between **J** and $\cap\mathbf{I}_i$. (Most of the arrows on the right in Figure 8 are implied by the requirement that once one edge points towards an endpoint of a trek, all the edges between it and the endpoint must also point towards that endpoint. The arrow into Y on π_2 in the top graph and the arrow into Y on τ_2 in the bottom graph are needed to prevent a cycle between X and Y.) **Q.E.D.**

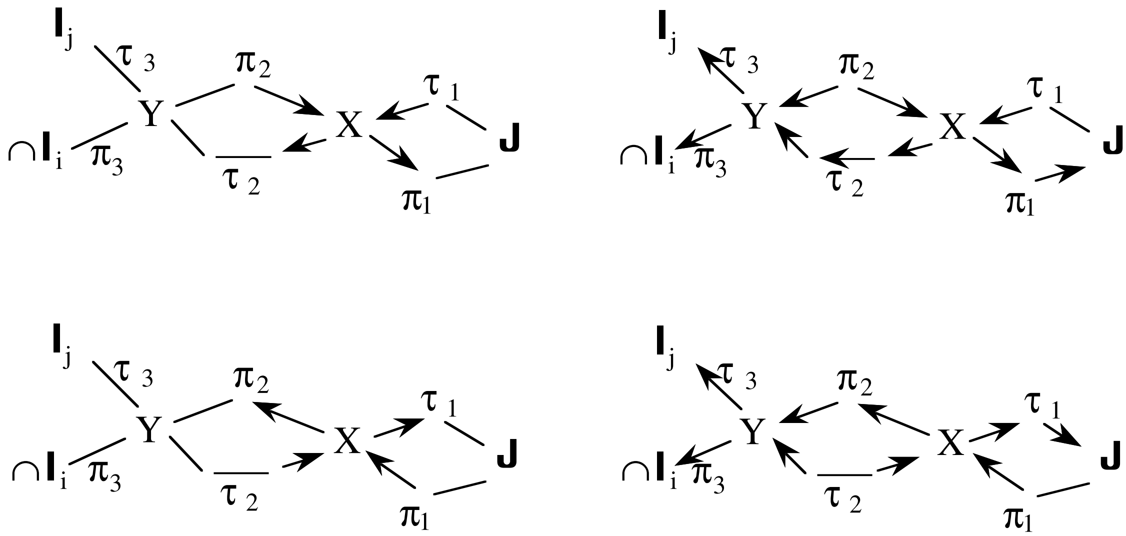


Figure 8 The graphs on the left show the two ways in which the treks π and τ from **J** to $\mathbf{I}_j$ might go through X on different sides. The graphs on the right show the further arrows that are implied.

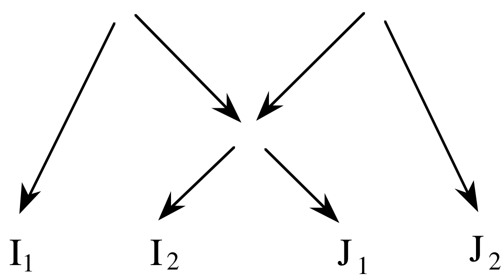


Figure 9 There cannot be a choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$, because the $\{I_1,I_2\}$ side π is disjoint from τ, and the $\{J_1,J_2\}$ side of τ is disjoint from π.

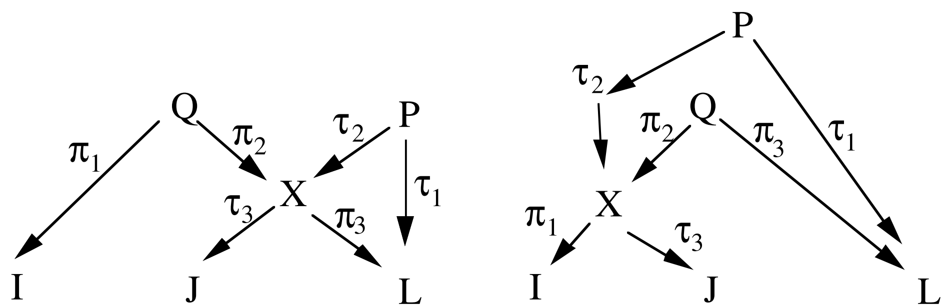
It is evident from the definition of choke point that there is no choke point between **I** and **J** if there are two non-intersecting treks between **I** and **J**, or even if there are two treks between **I** and **J** that intersect only on the **I** side of the first and the **J** side of the second, as in Figure 9. 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 τ, though possibly the **J** side of π and the **I** side of τ may intersect one or more times.

Theorem 1 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.

Proof We will prove the theorem by induction on the number of nodes in the directed acyclic graph. If there is only one node in the graph, then the theorem holds because its hypothesis cannot be satisfied; as we noted in Lemma 3, there is necessarily a choke point between **I** and **J** if **I** is empty or contains only one node. To complete the proof, we assume that the theorem holds for directed acyclic graphs with n or fewer nodes, and we show that it holds for any directed acyclic graph with n+1 nodes.

Suppose, then, that **I** and **J** are sets of nodes in a directed acyclic graph with n+1 nodes, and that there is no choke point between them. We may assume that every node in the graph is either an element of **I** or **J** or an ancestor of an element of **I** or **J**. For only nodes in **I** and **J** and their ancestors can be involved in treks between **I** and **J**, and if there were other nodes, we could delete them from the graph, thus obtaining a smaller directed acyclic graph containing **I** and **J** and all the treks between them, and thus obtaining the conclusion by the inductive hypothesis. Under this assumption, at least one element of **I** or **J**, say L, is barren.

Consider first the case where L is in both **I** and **J**. In this case, <L> is a trek between **I** and **J**. Since L is not a choke point between **I** and **J**, either there exists a trek π between **I** and **J** that does not contain L, in which case <L> and π are treks that do not intersect at all and hence satisfy the conclusion of the theorem, or else there exists a trek π from **I** to **J** that contains <L> only on its **J** side and another trek τ from **I** to **J** that contains <L> only on its **I** side. Since L is barren, it must be the last node in π and the first node in τ. Suppose π and τ do not satisfy the conclusion of the theorem, say because the **J** side of τ intersects π. Then we can combine the part of the **J** side of τ that comes after the intersection with the part of π that comes before the intersection to construct a trek λ between **I** and **J** that does not contain L, so that λ and <L> satisfy the conclusion of the theorem. The graph on the left in Figure 10 illustrates the case where the intersection is on π's **J** side; in this case, λ is $\pi_1\pi_2\tau_3$. The graph on the right illustrates the case where the intersection is on π's **I** side; in this case, λ is $\pi_1\tau_3$.



i. On π's **J** side. ii. On π's **I** side.

Figure 10 Two ways the **J** side of τ can intersect π.

Consider next the case where L is in only one of the sets, say in **I** but not in **J**, and L has no parents. In this case, L is an isolated node. Since L is not in **J**, no trek between **I** and **J** contains L, and hence the hypothesis that there is no choke point between **I** and **J** implies that there is no choke point between **I**\{L} and **J**. Hence we can obtain the conclusion by applying the inductive hypothesis to the smaller graph obtained by deleting L.

Finally, consider the case where L is in **I** but not in **J**, and L has at least one parent. Designate L's parents by $I_1, I_2, \ldots, I_k$. Set $\mathbf{I}_i$ = $\{I_i\} \cup (\mathbf{I} \setminus \{L\})$. Then $\cap \mathbf{I}_i = \mathbf{I} \setminus \{L\}$. Since L is not a choke point between **I** and **J**, there exists a trek between $\cap \mathbf{I}_i$ and **J**. So Lemma 5 tells us that either (1) for some j, there is no choke point between $\{I_j\}\cup(\mathbf{I}\backslash\{L\})$ and $\mathbf{J}$, or (2) there is a choke point between $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ and $\mathbf{J}$.

Suppose there is no choke point between $\{I_j\}\cup(\mathbf{I}\backslash\{L\})$ and $\mathbf{J}$. Then we use the inductive hypothesis to obtain treks π and τ from $\{I_j\}\cup(\mathbf{I}\backslash\{L\})$ to $\mathbf{J}$ such that the $\mathbf{I}$ side of π is disjoint from τ and the $\mathbf{J}$ side of τ is disjoint from π. Since the $\mathbf{I}$ sides of the two treks do not intersect, at most one of the two can begin with I_j on the $\mathbf{I}$ side. If neither begins with I_j on the $\mathbf{I}$ side, then they are both treks from $\mathbf{I}$ to $\mathbf{J}$ and hence satisfy the conclusion of the theorem. If one of them begins with I_j on the $\mathbf{I}$ side, then the second is already a trek from $\mathbf{I}$ to $\mathbf{J}$, and adding L at the beginning of the first makes it into a trek from $\mathbf{I}$ to $\mathbf{J}$, without disrupting the disjointness of π's $\mathbf{I}$ side and τ or the disjointness of τ's $\mathbf{J}$ side and π (since it is barren and is not an endpoint of either trek, L is disjoint from both treks), thus giving us the desired treks.

Finally, suppose there is a choke point, between $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ and $\mathbf{J}$. Designate this choke point by X, and suppose, first, that it is a choke point on the $\mathbf{J}$ side. We will show that every trek π from $\mathbf{I}$ to $\mathbf{J}$ goes through X on the $\mathbf{J}$ side, contradicting our assumption that there is no choke point between $\mathbf{I}$ and $\mathbf{J}$. Indeed, if π does not go through L, then it is a trek from $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ to $\mathbf{J}$ and hence goes through X on the $\mathbf{J}$ side. And if π does go through L, then, since L is barren and is not in $\mathbf{J}$, π must start from L on the $\mathbf{I}$ side and immediately go to one of the I_i. Dropping L from π gives a trek that goes from $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ to $\mathbf{J}$ and hence goes through X on the $\mathbf{J}$ side. Thus π goes through X on the $\mathbf{J}$ side. The argument is the same if X is a choke point between $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ and $\mathbf{J}$ on the $\{I_1,I_2,...,I_k\}\cup(\mathbf{I}\backslash\{L\})$ side, in which case the conclusion is that every trek π from **I** to **J** goes through X on the **I** side. **Q.E.D.**

We can summarize Theorem 1 and its converse by saying that there is a choke point between **I** and **J** if and only if for every pair of treks, say π and τ, between **I** and **J**, either the **I** side of π intersects τ or the **J** side of π intersects τ.

We leave it to the reader to verify the following corollary of Theorem 1.

Corollary 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 $\{J_1,J_2\}\subseteq\mathbf{J}$.

Part II. Linear Correlation and Covariance Structures

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

$$*(I,J) = \Sigma\{*\pi \mid \pi\in\sigma(I,J)\},$$

where σ(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 2 If I and J are nodes in a linear 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

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

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 6 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.

Proof It suffices to notice that the distinctness of the endpoints makes it impossible that $\#_X(\pi)$ should equal 1 and $\#_X(\tau)$ should equal -1, or vice versa. **Q.E.D.**

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 2, 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 If I_1, I_2, J_1, and J_2 are distinct nodes in a linear 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 $*(I_1,J_2)*(I_2,J_1)$.

Proof First we show that Statement 1 implies Statement 2. Suppose X is a choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$. Given a simple trek π from I_1 to J_1, and a simple trek τ from I_2 to J_2, we may decompose the two treks as in Figure 11. Let $F(\pi,\tau)$ be the pair $(\pi_1\tau_2,\tau_1\pi_2)$. It is easy to verify that $\pi_1\tau_2$ is a simple trek from I_1 to J_2, and $\tau_1\pi_2$ is a simple trek from I_2 to J_1. Moreover, the mapping F is its own inverse, and hence it is a one-to-one mapping between the set of (I_1-J_1 trek,I_2-J_2 trek) pairs and the set of (I_1-J_2 trek,I_2-J_1 trek) pairs. Since the same edges (with the same multiplicities) are involved in the pair $(\pi_1\tau_2,\tau_1\pi_2)$ as in the pair (π,τ), the products $*\pi*\tau$ and $*(\pi_1\tau_2)*(\tau_1\pi_2)$ are equal. This establishes that the polynomials $*(I_1,J_1)*(I_2,J_2)$ and $*(I_1,J_2)*(I_2,J_1)$ have the same terms and hence are equal.

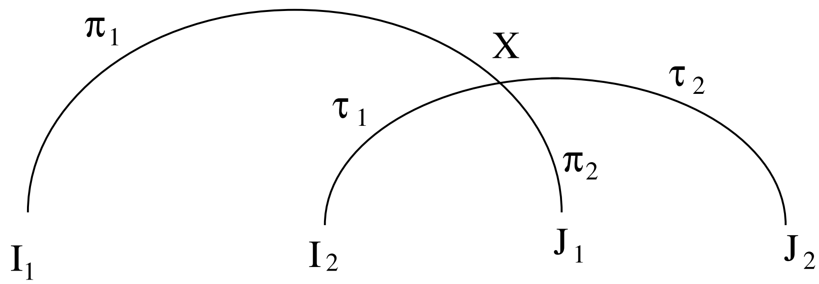


Figure 11

Now we show that if Statement 1 is false, then Statement 2 is false. Suppose there is no choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$. By Theorem 1, we know that there exist treks π and τ between $\{I_1,I_2\}$ and $\{J_1,J_2\}$ such that 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. These treks must have distinct sources, as in Figure 12.

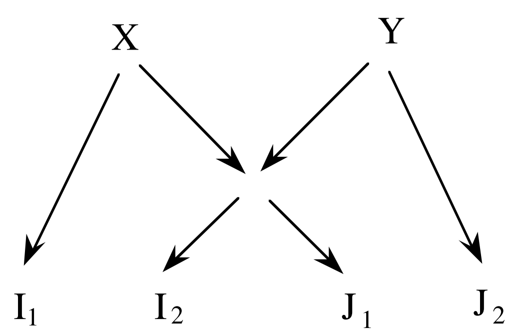


Figure 12

It is evident from Figure 12 that the edges in these two treks (with whatever multiplicities they may have because of edges shared by τ's I_2 side and π's J_1 side) cannot be rearranged into treks from I_1 to J_2 and I_2 to J_1. Indeed, since π's I_1 side is disjoint from the rest of π and all of τ, any trek from I_1 to J_2 that uses only the edges in π and τ must go up π's I_1 side to X and start down π's J_1 side. Similarly, a trek from J_2 to I_1 using only these edges must go up τ's J_2 side to Y and start back down τ's I_2 side. These two parts cannot be extended so as to form into a trek between I_1 and J_2, since they have arrows pointing towards each other. It follows that the term $*\pi*\tau$, which appears in $*(I_1,J_1)*(I_2,J_2)$, does not appear in $*(I_1,J_2)*(I_2,J_1)$; the two polynomials are not equal. **Q.E.D.**

We now develop an analogue of Theorem 3 for the case where we assume that each endogeneous node is the only child of at least one exogeneous parent, we attach symbols to the exogeneous nodes as well as to all the edges, and in the place of simple treks, we consider treks that have simple sides and have exogeneous nodes as sources.

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

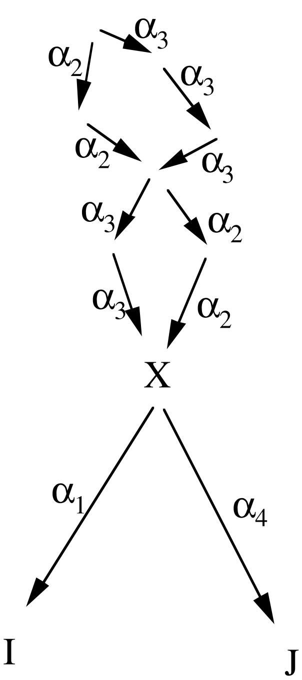


Figure 13 $\downarrow\alpha = \alpha_1\alpha_4$ and $\uparrow\alpha = \alpha_2\alpha_3$.

We call a trek an ultratrek if its source is exogenous and both its sides are simple. 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 13.) 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 linear covariance structure, we write $\blacklozenge\pi$ 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 $\blacklozenge\pi$ the symbol product over π. If π consists of a single node, its source, then $\blacklozenge\pi$ is simply the symbol attached to the source. Given any two nodes I and J, we set

$$\blacklozenge(I,J) = \Sigma\{\blacklozenge\pi \mid \pi \in \upsilon(I,J)\},$$

where $\upsilon(I,J)$ is the set of all ultratreks from I to J. We call $\blacklozenge(I,J)$ 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 $\blacklozenge(I,I)$ is the symbol attached to I.

Theorem 4 If I_1, I_2, J_1, and J_2 are distinct nodes in a linear 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 $\blacklozenge(I_1,J_1)\blacklozenge(I_2,J_2)$ is equal to the polynomial $\blacklozenge(I_1,J_2)\blacklozenge(I_2,J_1)$.

Proof The proof is completely analogous to the proof of Theorem 3.

First we show that Statement 1 implies Statement 2. Suppose X is a choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$; we assume without loss of generality that it is a choke point on the $\{J_1,J_2\}$ side. Given an ultratrek π from I_1 to J_1, and an ultratrek τ from I_2 to J_2, we may decompose the two treks as in Figure 11. Here π_2 is the part of π that goes from X to J_1, and τ_2 is the part of τ that goes from X to J_2. Since X is a choke point on the $\{J_1,J_2\}$ side, the bases of two treks lie on the other parts, π_1 and τ_1. Let

$F(\pi,\tau)$ be the pair $(\pi_1\tau_2,\tau_1\pi_2)$. It is obvious that $\pi_1\tau_2$ is an ultratrek from I_1 to J_2, and $\tau_1\pi_2$ is an ultratrek from I_2 to J_1. Moreover, the mapping F is its own inverse, and hence it is a one-to-one mapping between the set of (I_1-J_1 ultratrek,I_2-J_2 ultratrek) pairs and the set of (I_1-J_2 ultratrek,I_2-J_1 ultratrek) pairs. Since the same edges (with the same multiplicities) are involved in the pair $(\pi_1\tau_2,\tau_1\pi_2)$ as in the pair (π,τ), the products $\blacklozenge\pi\blacklozenge\tau$ and $\blacklozenge(\pi_1\tau_2)\blacklozenge(\tau_1\pi_2)$ are equal. This establishes that the polynomials $\blacklozenge(I_1,J_1)\blacklozenge(I_2,J_2)$ and $\blacklozenge(I_1,J_2)\blacklozenge(I_2,J_1)$ have the same terms and hence are equal.

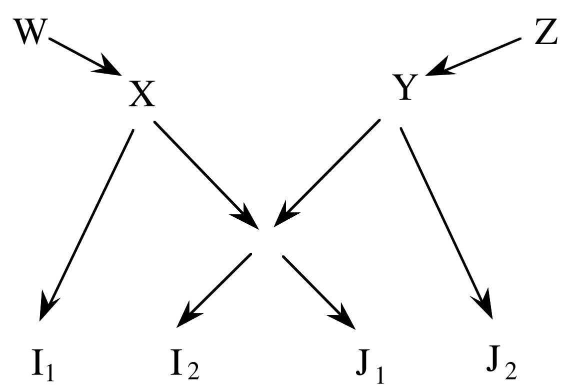
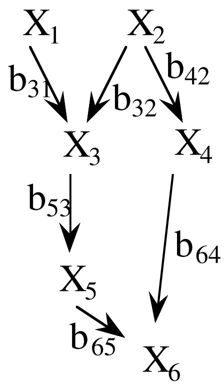


Figure 14

Now we show that if Statement 1 is false, then Statement 2 is false. Suppose there is no choke point between $\{I_1,I_2\}$ and $\{J_1,J_2\}$. By Theorem 1, we know that there exist treks π_0 and τ_0 between $\{I_1,I_2\}$ and $\{J_1,J_2\}$ such that the $\{I_1,I_2\}$ side of π_0 is disjoint from τ_0 and the $\{J_1,J_2\}$ side of τ_0 is disjoint from π_0. These treks must have distinct sources, as in Figure 12, where we have assumed, without loss of generality, that π_0 is a trek from I_1 to J_1 and τ_0 is a trek from I_2 to J_2. If π_0 is not an ultratrek—i.e., if X is endogeneous, then we extend it to an ultratrek π from I_1 to J_1 by adding an excursion from X to an endogeneous parent for which X is the only child. We designate this parent by W, as it Figure 14; notice that W must be distinct from Y (since Y has other children) as well as from all the other nodes on τ_0 (since they are endogeneous). Similarly, we extend τ_0, if necessary to an ultratrek τ from I_2 to J_2. Thus we obtain ultratreks π from I_1 to J_1 and τ from I_2 to J_2 such that the I_1 side of π is disjoint from τ and the J_2 side of τ is disjoint from π. It is evident from Figure 14 that the edges in these two ultratreks (with whatever multiplicities they may have) cannot be rearranged into treks from I_1 to J_2 and I_2 to J_1. Indeed, since π's I_1 side is disjoint from the rest of π and all of τ, any trek from I_1 to J_2 that uses only the edges in π and τ must go up π's I_1 side to W, go back down to X, and then start on down toward J_1. Similarly, any trek from J_2 to I_1 using only these edges must go up τ's J_2 side to Z, go back down to Y, and then start on down toward I_2. These two parts cannot be extended so as to form into a trek between I_1 and J_2, since they have arrows pointing towards each other. It follows that the term $\blacklozenge\pi\blacklozenge\tau$, which appears in $\blacklozenge(I_1,J_1)\blacklozenge(I_2,J_2)$, does not appear in $\blacklozenge(I_1,J_2)\blacklozenge(I_2,J_1)$; the two polynomials are not equal. **Q.E.D.**

Part III. Application to Statistical Inference

Let us now interpret linear 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 endogeneous variable as the regression coefficients in the linear regression of that variable on its parents, as in Figure 15.



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

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

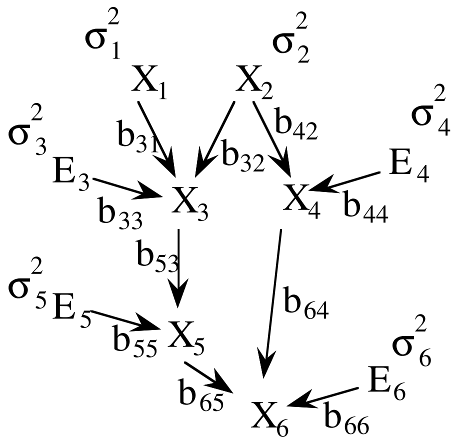
$$X_5 = b_{53}X_3 + E_5$$

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

Figure 15 The graph on the left is a linear correlation structure associated with the recursive linear regression equations on the right.

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

Since the errors in the regressions equations for the endogeneous variables in a linear correlation structure have zero correlations with each other and with the exogeneous 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 linear correlation structure to a linear covariance structure by adding to each endogeneous variable a parent representing the error in its regression equation, as in Figure 16. 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 in the regression equation, so that the regression equation becomes the equation representing the endogeneous variable as a linear combination of its parents in the linear 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 16 The linear covariance structure corresponding to the linear correlation structure of Figure 15.

Let us write $\rho(I,J)$ and 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 5

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

(2) Cov(I,J) = ♦(I,J) for every pair of variables I and J in a linear covariance structure.

Statement 3 is due to Sewall Wright (1934); it is the centerpiece of his method of path analysis.

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

$$♦(I,J) = 0, \qquad \text{(III.1)}$$

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

$$\blacklozenge(I_1,J_1)\blacklozenge(I_2,J_2) - \blacklozenge(I_1,J_2)\blacklozenge(I_2,J_1) = 0. \qquad \text{(III.2)}$$

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 endogeneous 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 infer that those constraints that do hold are structural, and this will give us information about the linear 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 6 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 4 and 5.

The corollary to Theorem 4 yields the following generalization of the tetrad representation theorem.

Theorem 7 Suppose **I** and **J** are disjoint set of variables in a linear 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 $\{I_1,I_1\}$ of **I** and every subset $\{J_1,J_2\}$ of **J** if and only if there is a choke point between **I** and **J**.

Acknowledgements

Glenn Shafer’s research for this paper was partially supported by National Science Foundation grant SBE9213674 .

References

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann. San Mateo, California.

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

Spirtes, Peter, Clark Glymour, and Richard Scheines (1993). *Causation, Prediction, and Search*. Lecture Notes in Statistics, No. 81. Springer-Verlag. New York.

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