TECHNICAL REPORT
R-432
July 2015

Graphical Tools for Linear Structural Equation Modeling

Bryant Chen and Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
(310) 825-3243

This paper surveys graphical tools developed in the past three decades that are applicable to linear structural equation models (SEMs). These tools permit researchers to answer key research questions by simple path-tracing rules, even for highly complex models. They include parameter identification, causal effect identification, regressor selection, selecting instrumental variables, finding testable implications of a given model, identifying equivalent models and estimating counterfactual relationships.

Keywords: causal effects, counterfactuals, equivalent models, goodness of fit, graphical models, identification, linear regression, misspecification test

Introduction

Structural Equation Models (SEMs) are the dominant research paradigm in the quantitative, data-intensive behavioral sciences. These models permit a researcher to express theoretical assumptions meaningfully, using equations, derive their consequences and test their statistical implications against data. The result is a powerful symbiosis between theory and data which underlies much of current research in causal analysis, especially in therapy evaluation (Shrout et al., 2010), education testing and management (Muthén and Muthén, 2010), and personality research (Lee, 2012).

While advances in graphical models have had a transformative impact on causal analysis and machine learning, only a meager portion of these developments have found their way to mainstream SEM literature which, by and large, prefers algebraic over graphical representations (Joreskog and Sorbom, 1982; Bollen, 1989; Mulaik, 2009; Hoyle, 2012). One of the reasons for this disparity rests on the fact that graphical techniques were developed for non-parametric analysis, while much of SEM research is conducted within the confines of Gaussian linear models, to which matrix algebra and powerful statistical tests are applicable. Among the tasks facilitated by graphical models are: model testing, identification, policy analysis, bias control, mediation, external validity, and the analysis of counterfactuals and missing data (Pearl, 2014a).

The purpose of this paper is to introduce psychometric researchers to modern tools of graphical models and to describe some of the benefits, as well as new insights that graphical models can provide. We will begin by introducing basic definitions and tools used in graphical modeling, including graph construction, definitions of causal effects, and Wright's path tracing rules. We then introduce more advanced notions of graph separation, which were developed for non-parametric analysis, but have simple and meaningful interpretation in linear models. These tools provide the basis for model testing and identification criteria, discussed in subsequent sections. We then cover advanced applications of path diagrams including equivalent regressor sets, minimal regressor sets, and variance minimizing for causal effect estimation. Lastly, we discuss counterfactuals and their computation in linear SEMs before showing how the tools presented in this paper provide simple solutions to five examples representing non-trivial problems in SEM research.

With the exception of the "Causal Effects among Latent Variables" section, we focus on models where all variables are observable (often called path analysis models), allowing for error terms to be correlated. As graphical techniques were originally developed for non-parametric models, they have not traditionally addressed the identification of effects among latent variables, which is impossible without parametric assumptions. Instead, the presence of latent variables was taken into account through the correlations they induce on the error terms. We will demonstrate how latent variables can be summarized using error terms and briefly discuss how the results in this paper, while not directly addressing causal effects among latent variables, can nevertheless be applied to their analysis.

Path Diagrams and Graphs

Path diagrams or graphs[1] are graphical representations of the model structure. They were introduced by Sewell Wright (1921), who aimed to estimate causal influences from statistical data on animal breeding. Today, SEM is generally

[1] We use both terms interchangeably.

implemented in software[2], and, as a result, when users experience unexpected behavior (due to unidentified parameters, for example) they are often at a loss as to the source of the problem[3]. For the remainder of this section, we will review the basics of path diagrams and provide users with simple, intuitive tools that will be used to resolve questions of identification, goodness of fit, and more using graphical methods.

We introduce path diagrams by way of example. Suppose we wish to estimate the effect of attending an elite college on future earnings. Clearly, simply regressing earnings on college rating will not give an unbiased estimate of the target effect. This is because elite colleges are highly selective, so students attending them are likely to have qualifications for high-earning jobs prior to attending the school. This background knowledge can be expressed in the following SEM specification. Throughout the paper, we will use lowercase letters and the Greek letter α to represent model parameters.

Model 1.

$$\begin{aligned} Q_1 &= U_1 \\ C &= a \cdot Q_1 + U_2 \\ Q_2 &= c \cdot C + d \cdot Q_1 + U_3 \\ S &= b \cdot C + e \cdot Q_2 + U_4, \end{aligned}$$

where Q_1 represents the individual's qualifications prior to college, Q_2 represents qualifications after college, C contains attributes representing the quality of the college attended, and S the individual's salary.

Figure 1a is a causal graph that represents this model specification. Each variable in the model has a corresponding node or vertex in the graph. Additionally, for each equation, arrows are drawn from the independent variables to the dependent variables. These arrows reflect the direction of causation. In some cases, we may label the arrow with its corresponding structural coefficient as in Figure 1a. Error terms are typically not displayed in the graph, unless they are correlated.

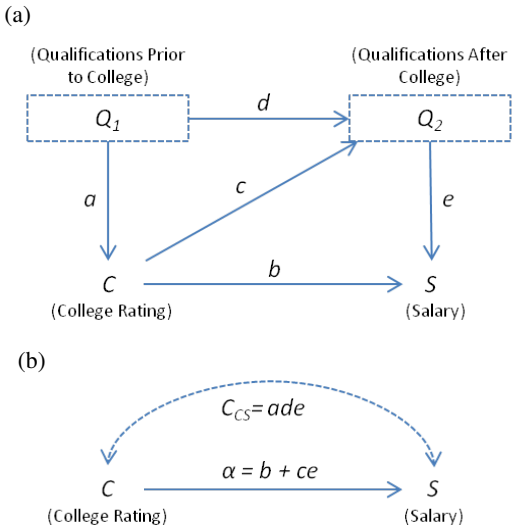


Figure 1. (a) Model with latent variables (Q_1 and Q_2) shown explicitly (b) Same model with latent variables summarized

The variables Q_1 and Q_2 represent quantities that are not directly measurable. As a result, they are *latent variables*. In this paper, we distinguish latent variables from observable variables in the graph by surrounding the former with a dashed box. As we mentioned in the Introduction, the presence of latent variables is taken into account by the correlations they induce on the error terms[4]. For example, the effect of the latent variables in Figures 1a is summarized by Figure 1b. We see that the effect of College on Salary in Figure 1a is now summarized by the coefficient α in Figure 1b. Similarly, the bidirected arc between C and S (representing the correlation of the error terms of C and S) in Figure 1b summarizes the correlation between C and S due to the path $C \leftarrow Q_1 \rightarrow Q_2 \rightarrow S$. The corresponding model is as follows:

Model 2.

$$\begin{aligned} C &= U_C \\ S &= \alpha C + U_S \end{aligned}$$

The background information specified by Model 1 implies that the error term of S, U_S, is not correlated with U_C, and this correlation is depicted in Figure 1a by the bidirected arc between C and S.

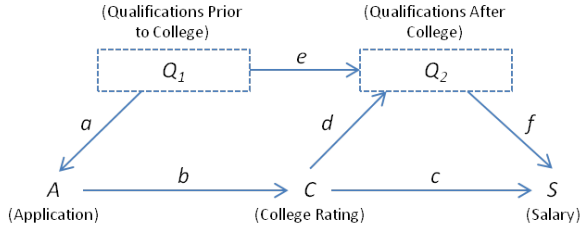
In order to estimate α, the causal effect of attending an elite college on future earnings, the coefficients must have a unique solution in terms of the covariance matrix or probability distribution over the observable variables, C and S. The task of finding this solution is known as *identification* and is discussed in a later section. In some cases, one or more coefficients may not be identifiable, meaning that no matter the size of the dataset, it is impossible to obtain point estimates for their values. Indeed, we will see that the coefficients in Model 1 are not identified if Q_1 and Q_2 are latent. However, if we include the strength of an individual's college application, A, as shown in Figure 2a, we obtain the following model:

[2]Common software packages include AMOS (Arbuckle, 2005), EQS (Bentler, 1989), LISREL (Jöreskog and Sörbom, 1989), and MPlus (Muthén and Muthén, 2010) among others.

[3]Kenny and Milan (2012) write, "Identification is perhaps the most difficult concept for SEM researchers to understand. We have seen SEM experts baffled and bewildered by issues of identification."

[4]While we do not directly address the identification of causal effects among latent variables, the results in this paper are nevertheless applicable to this problem. See section Identification.

(a)



(b)

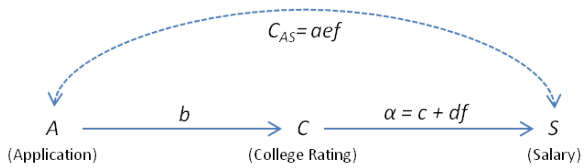


Figure 2. Graphs associated with Model 3 in the text (a) with latent variables shown explicitly (b) with latent variables summarized

Model 3.

$$\begin{aligned} Q_1 &= U_1 \\ A &= a \cdot Q_1 + U_2 \\ C &= b \cdot A + U_3 \\ Q_2 &= e \cdot Q_1 + d \cdot C + U_4 \\ S &= c \cdot C + f \cdot Q_2 + U_5. \end{aligned}$$

By removing the latent variables from the model specification we obtain:

Model 4.

$$\begin{aligned} A &= a \cdot Q_1 + U_A \\ C &= b \cdot A + U_C \\ S &= \alpha \cdot C + U_S. \end{aligned}$$

The corresponding path diagram is displayed in Figure 2b. The coefficients in this model are combinations of coefficients of the original model and each of these combinations is identifiable, as we will show.

The ability to determine identifiability directly from the model specification is a valuable feature of graphical models. For example, it would be a waste of resources to specify the structure in Model 2 and gather data only to find that the parameter of interest is not identified. The tools provided in subsequent sections will allow modelers to determine immediately from the path diagram that the effect of attending an elite college on future salary, α, is not identified using Model 2 but is identified (and equal to the coefficient of C in the regression of S on C and A, denoted $\beta_{SC.A}$) using Model 4. This conclusion is a consequence of the model specification and $\alpha = \beta_{SC.A}$ holds only if the specification accurately reflects reality (see section "Causal Effects among Latent Variables"). The ability to derive testable implications and test the model specification is another valuable feature of graphical models. For example, we will see that Model 3 implies that the partial correlation between S and A given C and Q_1, $\rho_{SA.CQ_1}$, is equal to zero. If this constraint does not hold in the data, then we have evidence that the model is missing either an arrow or a bidirected arc between A and S. Most importantly, these tools will be applicable to far more complex models where questions of identifiability and testable implications are near impossible to determine by hand or even by standard software.

In summary, the causal graph is constructed from the model equations in the following way: Each variable in the model has a corresponding vertex or node in the graph. For each equation, arrows are drawn in the graph from the nodes representing dependent variables to the node representing the independent variable. Each arrow, therefore, is associated with a coefficient in the SEM, which we will call its structural coefficient. Finally, if the error terms of any two variables are correlated, then a bidirected arc is drawn between the two variables. Conversely, the lack of a bidirected arc indicates that the error terms are independent.

Before continuing, we review some basic graph terminology. An *edge* is defined to be either an arrow or a bidirected arc. If an arrow exists from X to Y, we say that X is a *parent* of Y. If there exists a sequence of arrows all of which are directed from X to Y we say that X is an *ancestor* of Y. If X is an ancestor of Y then Y is a *descendant* of X. Finally, the set of nodes connected to Y by a bidirected arc are called the *siblings* of Y.

A *path* between X and Y is a sequence of edges, connecting the two vertices[5]. A path may go either along or against the direction of the arrows. A *directed path* from X to Y is a path consisting only of arrows pointed towards Y. A *back-door path* from X to Y is a path begins with an arrow pointing to X and ends with an arrow pointing to Y. For example, in Figure 4, $C \leftarrow B \rightarrow E$, $C \rightarrow D \rightarrow E$, $C \leftarrow B \rightarrow D \rightarrow E$, and $C \rightarrow D \leftarrow B \rightarrow E$ are all paths between C and E. However, only $C \rightarrow D \rightarrow E$ is a directed path, and only $C \leftarrow D \rightarrow E$ and $C \leftarrow B \rightarrow D \rightarrow E$ are back-door paths. The significance of directed paths stems from the fact that they convey the flow of causality, while the significance of back-door paths stems from their association with confounding.

A node in which two arrowheads meet is called a *collider*. For example, Z in Figure 5a and C in Figure 4 are colliders. The significance of colliders stems from the fact that they block the flow of information along a path (see section D-separation).

A graph or model is *acyclic* if it does not contain any *cy-*

[5] We emphasize that, in this paper, we refer to paths as sequences of arrows and/or bidirected arcs, not single arrows.

cles, that is a directed path that begins and ends with the same node. A model or graph is *cyclic* if it contains a cycle. An acyclic model without correlated error terms is called *Markovian*. Models with correlated error terms are called *non-Markovian* while acyclic non-Markovian models are additionally called *semi-Markovian*. For example, Figure 4 is both acyclic and Markovian. If we were to add a bidirected arc between any two variables, then it would no longer be Markovian and would instead be semi-Markovian. If we were to instead reverse the edge from B to E, then we would create a cycle and the model would be non-Markovian and cyclic.

Lastly, we note that, for simplicity, we will assume without loss of generality that all variables have been standardized to mean 0 and variance 1.

Causal Effects

Let $\Pi = \{\pi_1, \pi_2, ..., \pi_k\}$ be the set of directed paths from X to Y and p_i be the product of the structural coefficients along path π_i. The *total effect* or *average causal effect* (ACE) of X on Y is often defined as the $\sum_i p_i$ (Bollen, 1989). For example, in Figure 2a, the total effect of C on S is $c + d \cdot f$ and that of A on S is $b(c + df)$.

The rational for this additive formula and its extension to non-linear systems can best be seen if we define the causal effect of X on Y as the expected change in Y when X is assigned to different values by intervention, as in a randomized experiment. The act of assigning a variable X to the value x is represented by removing the structural equation for X and replacing it with the equality $X = x$. This replacement dislodges X from its prior causes and ensures that covariation between X and Y reflects causal paths from X to Y only.

The expected value of a variable, Y, after X is assigned the value x by intervention is denoted $E[Y|do(X = x)]$, and the ACE of X on Y is defined as

$$ACE = E[Y|do(X = x + 1)] - E[Y|do(X = x)], \qquad (1)$$

where x is some reference point (Pearl, 2009, ch. 5)[6]. In non-linear systems, the effect will depend on the reference point but in the linear case, x will play no role and we can replace (1), with the derivative,

$$ACE = \frac{\partial}{\partial x} E[Y|do(X = x)]. \qquad (2)$$

Consider again Model 3 with C a binary variable taking value 1 for elite colleges and 0 for non-elite colleges. To estimate the total effect of attending an elite college on salary, we would hypothetically assign each member of the population to an elite college and observe the average salary, $E[S|do(C = 1)]$. Then we would rewind time and assign each member to a non-elite college, observing the new average salary, $E[S|do(C = 0)]$. Intuitively, the causal effect of attending an elite college is the difference in average salary,

$$E[S|do(C = 1)] - E[S|do(C = 0)].$$

The above operation provides a mathematical procedure that mimics this hypothetical (and impossible) experiment using a SEM.

In linear systems, this "interventional" definition of causal effect coincides with the aforementioned "path-tracing" definition as we will demonstrate by computing $E[S|do(C = 1)] - E[S|do(C = 0)]$ in Model 3. (For clarity, we will consider Q_1 and Q_2 to be observable and not latent variables for the remainder of this section.)

The intervention, $do(C = c_0)$, modifies the equations in the following way:

Model 5.

$$\begin{aligned} Q_1 &= U_1 \\ A &= a \cdot Q_1 + U_2 \\ C &= c_0 \\ Q_2 &= e \cdot Q_1 + d \cdot C + U_4 \\ S &= c \cdot C + f \cdot Q_2 + U_5. \end{aligned}$$

The corresponding path diagram is displayed in Figure 3a. Notice that back-door paths, due to common causes, between C and S have been cut, and as a result, all unblocked paths between C and S now reflect the causal effect of C on S only.

Recalling that we assume model variables have been standardized to mean 0 and variance 1 implying that $E[U_i] = 0$ for all i, we see that setting C to c_0 gives the following expectation for S:

$$\begin{aligned} E[S|do(C = c_0)] &= E[c \cdot C + f \cdot Q_2 + U_5] \\ &= c \cdot E[C] + f \cdot E[Q_2] + E[U_5] \\ &= c \cdot c_0 + fE[e \cdot Q_1 + d \cdot C + U_4] \\ &= c \cdot c_0 + f \cdot e \cdot E[U_1] + f \cdot d \cdot c_0 + f \cdot E[U_4] \\ &= c \cdot c_0 + f \cdot d \cdot c_0 \end{aligned}$$

As a result,

$$E[S|do(C = c_0 + 1)] - E[S|do(C = c_0)] = c + fd \qquad (3)$$

for all c_0, aligning the two definitions[7].

[6]Holland (2001) defines causal effects in counterfactual terminology (also known as potential outcomes (Rubin, 1974)), which will be discussed in section Counterfactuals in Linear Models. The logical equivalence between these two notational systems is shown in (Pearl, 2009, ch. 7.4). SEMs provide a semantics for the potential outcomes framework, which is based on scientific knowledge as opposed to experimental design.

[7]Moreover, this equality holds even when the parameters, c, d, and f, are not identified (e.g. if the U terms are correlated). Causal effects are defined in terms of hypothetical interventions, and the parameters determine the impact of these interventions. Identification is only the means to obtain causal effects from statistical data and has nothing to do with the definition.

In many cases, we may be interested in the *direct effect* of C on S. The term "direct effect" is meant to quantify an effect that is not mediated by other variables in the model or, more accurately, the sensitivity of S to changes in C while all other factors in the analysis are held fixed (Pearl, 2009, ch. 4.5). In Model 3, the direct effect of C on S represents the effects on salary due to factors other than the superior qualifications obtained by attending an elite college. For example, it could represent the value that employers place on the reputation of the school.

"Holding all other factors fixed" can be simulated by intervening on all variables other than C and S and assigning them an arbitrary set of reference values[8]. (Like the total effect, in linear systems, the direct effect does not change with respect to the reference values.) Doing so severs all causal links in the model other than those leading into S. As a result, all links from C to S other than the direct link will be severed. For example, Figure 3b shows the path diagram of Model 3 after intervention on all variables other than C and S.

Now, the direct effect of C on S can be defined as

$$E[S|do(C = c_0 + 1, T = t)] - E[S|do(C = c_0, T = t)],$$

where T is a set containing all model variables other than C and S and $\{c_0 \cup t\}$ a set of reference values. This causally defined notion of direct effect differs fundamentally from the traditional definition which is based on *conditioning* on intermediate variables (Baron and Kenny, 1986). The former is valid in the presence of correlated errors and permits us to extend this notion to non-linear models (Pearl, 2014b).

Notice that the direct effect of C on S in Figure 2b is equal to the total effect of C on S in Figure 2a. Direct effects depend on the set of variables that we decide to include in the model.

Lastly, in linear models, the effect of C on S *mediated* by Q_2 is equal to the sum of the product of coefficients associated with directed paths from C to S that go through Q_2 (i.e. the effect on salary due to the knowledge and skills obtained from attending an elite college). In Figure 2a, we see that this effect is equal to df. For a non-linear and non-parametric extension of this definition, see *indirect effect* in (Pearl, 2014b).

Wright's Path Tracing Rules

The earliest usage of graphs in causal analysis can be found in Sewell Wright's 1921 paper, "Correlation and Causation". This seminal paper gives a method by which the covariance of any two variables in an acyclic, standardized model can be expressed as a polynomial over a subset of the model coefficients.

Wright's method consists of equating the covariance, σ_{YX}, between any pair of variables, X and Y, to the sum of products of structural coefficients and error covariances along certain paths between X and Y. Let $\Pi = \{\pi_1, \pi_2, ..., \pi_k\}$ denote the paths between X and Y that do not trace a collider, and

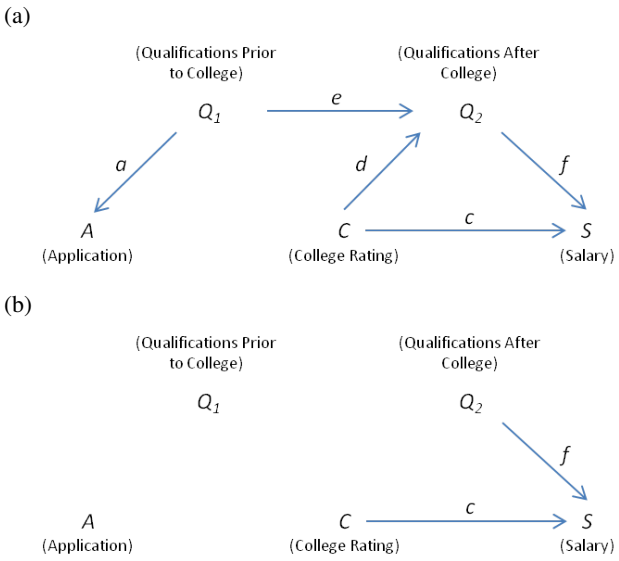


Figure 3. Models depicting interventions (a) After intervening on C (c) After intervening on C, A, Q_1, and Q_2

let p_i be the product of structural coefficients along path π_i. Then the covariance between variables X and Y is $\sum_i p_i$. For example, we can calculate the covariance between C and S in Figure 2b in the following way: First, we note that there are two paths between C and S and neither trace a collider, $\pi_1 = C \rightarrow S$ and $\pi_2 = C \leftarrow A \leftrightarrow S$. The product of the coefficients along these paths are $p_1 = \alpha$ and $p_2 = b \cdot C_{AS}$. Summing these products together we obtain the covariance between C and S, $\sigma_{CS} = \alpha + b \cdot C_{AS}$.

Consider the more complicated example of calculating σ_{CE} in Figure 4. The paths between C and E that do not trace a collider are $C \leftarrow F \rightarrow A \rightarrow E$, $C \leftarrow A \rightarrow E$, and $C \rightarrow D \rightarrow E$. (Note that we do not include $C \rightarrow D \leftarrow B \rightarrow E$ because it traces a collider, D.) Summing the products of coefficients along these paths gives $\sigma_{CE} = b \cdot a \cdot g + c \cdot g + d \cdot h$.

To express the partial covariance, $\sigma_{YX.Z}$, partial correlation, $\rho_{YX.Z}$ or regression coefficient, $\beta_{YX.Z}$, of Y on X given Z

[8] In footnote 15 we give an example demonstrating that "holding all other factors fixed" cannot be simulated using conditioning but instead must invoke intervention.

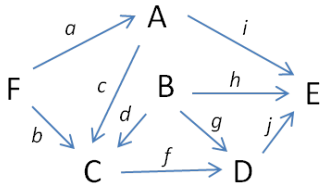


Figure 4. Model illustrating Wright's path tracing rules and d-separation

in terms of structural coefficients we can first apply the following reductions given by Crámer (1946), before utilizing Wright's rules. When Z is a singleton, these reductions are:

$$\rho_{YX.Z} = \frac{\rho_{YX} - \rho_{YZ}\rho_{XZ}}{[(1-\rho_{YZ}^2)(1-\rho_{XZ}^2)]^{\frac{1}{2}}} \tag{4}$$

$$\sigma_{YX.Z} = \sigma_{YX} - \frac{\sigma_{YZ}\sigma_{ZX}}{\sigma_Z^2} \tag{5}$$

$$\beta_{YX.Z} = \frac{\sigma_Y}{\sigma_X}\frac{\rho_{YX} - \rho_{YZ}\rho_{ZX}}{1-\rho_{XZ}^2} \tag{6}$$

When Z is a singleton and S a set, we can reduce $\rho_{YX.ZS}$, $\sigma_{YX.ZS}$, or $\beta_{YX.ZS}$ as follows:

$$\rho_{YX.ZS} = \frac{\rho_{YX.S} - \rho_{YZ.S}\rho_{XZ.S}}{[(1-\rho_{YZ.S}^2)(1-\rho_{XZ.S}^2)]^{\frac{1}{2}}} \tag{7}$$

$$\sigma_{YX.ZS} = \sigma_{YX.S} - \frac{\sigma_{YZ.S}\sigma_{ZX.S}}{\sigma_{Z.S}^2} \tag{8}$$

$$\beta_{YX.ZS} = \frac{\sigma_{Y.S}}{\sigma_{X.S}}\frac{\rho_{YX.S} - \rho_{YZ.S}\rho_{ZX.S}}{1-\rho_{XZ.S}^2} \tag{9}$$

We see that $\rho_{YX.ZS}$, $\sigma_{YX.ZS}$, or $\beta_{YX.ZS}$ can be expressed in terms of pair-wise coefficients by recursively applying the above formulas for each element of S. Then, using Equations 4-9, we can express the reduced pairwise covariances / correlations in terms of the structural coefficients. For example, reducing $\beta_{CS.A}$ for Figure 2b can be done as follows:

$$\beta_{CS.A} = \frac{\sigma_C}{\sigma_S}\frac{\rho_{CS} - \rho_{CA}\rho_{AS}}{1-\rho_{SA}^2} \tag{10}$$

$$= \frac{1}{1}\frac{(\alpha + bC_{AS}) - (b\alpha + C_{AS})(b)}{1-b^2} \tag{11}$$

$$= \frac{\alpha + bC_{AS} - b^2\alpha - bC_{AS}}{1-b^2} \tag{12}$$

$$= \frac{\alpha - b^2\alpha}{1-b^2} \tag{13}$$

$$= \alpha \tag{14}$$

D-Separation

When the conditioning set becomes large, applying the recursive formula of Equations 7-9 can become complex. Vanishing partial correlations, however, can be readily identified from the path diagram using a criterion called *d-separation* (Pearl, 1988)[9]. In other words, d-separation allows us to determine whether correlated variables become uncorrelated when conditioning on a given set of variables. Not only will this criterion allow us to use these zero partial correlations for model testing, but it will also be utilized extensively in the analysis of identification that follows.

The idea of d-separation is to associate "correlation" with "connectedness" in the graph, and independence with "separation". The only twist on this simple idea is to define what we mean by "connected path", since we are dealing with a system of directed arrows in which some nodes (those residing in the conditioning set, Z) correspond to variables whose values are given. To account for the orientations of the arrows we use the terms "d-separated" and "d-connected" (d denotes "directional").

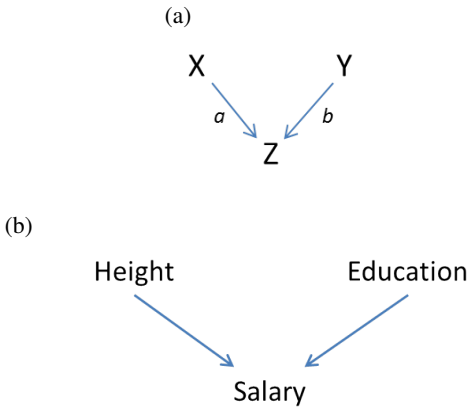


Figure 5. Examples illustrating conditioning on a collider

Rule 1: X and Y are d-separated if there is no active path between them.

By "active path", we mean a path that can be traced without traversing a collider. If no active path exists between X and Y then we say that X and Y are d-separated. As we can see from Wright's rules, $\rho_{XY} = 0$ when X and Y are d-separated.

When we measure a set Z of variables, and take their values as given, the partial covariances of the remaining variables changes character; some correlated variables become uncorrelated, and some uncorrelated variables become correlated. To represent this dynamic in the graph, we need the notion of "partial d-connectedness" or more concretely, "d-connectedness conditioned on a set Z of measurements".

Rule 2: X and Y are d-connected, conditioned on a set of Z nodes, if there is a collider-free path between X and Y that traverses no member of Z. If no such path exists, we say that X and Y are d-separated by Z or we say that every path between X and Y is "blocked" by Z.

A common example used to show that correlation does not imply causation is the fact that ice cream sales are correlated with drowning deaths. When the weather gets warm people

[9] See also Hayduk et al. (2003) and Mulaik (2009) for an introduction to d-separation tailored to SEM practitioners.

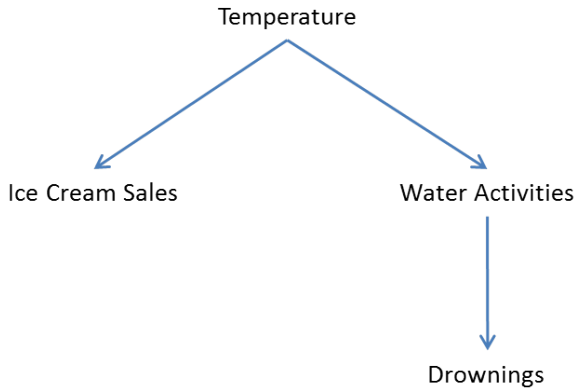


Figure 6. Diagram illustrating why Ice Cream Sales and Drowning are uncorrelated given Temperature and/or Water Activities

tend to both buy ice cream and play in the water, resulting in both increased ice cream sales and drowning deaths. This causal structure is depicted in Figure 6. Here, we see that Ice Cream Sales and Drownings are d-separated given either Temperature or Water Activities. As a result, if we only consider days with the same temperature and/or the same number of people engaging in water activities then the correlation between Ice Cream Sales and Drownings will vanish.

Rule 3: If a collider is a member of the conditioning set Z, or has a descendant in Z, then the collider no longer blocks any path that traces it.

According to Rule 3, conditioning can unblock a blocked path from X to Y. This is due to the fact that conditioning on a collider or its descendant opens the flow of information between the parents of the collider. For example, X and Y are uncorrelated in Figure 5a. However, conditioning on the collider, Z, correlates X and Y giving $\rho_{XY.Z} \neq 0$. This phenomenon is known Berkson's paradox or "explaining away". To illustrate, consider the example depicted in Figure 5b. It is well known that higher education often affords one a greater salary. Additionally, studies have shown that height also has a positive impact on one's salary. Let us assume that there are no other determinants of salary and that Height and Education are uncorrelated. If we observe an individual with a high salary that is also short, our belief that the individual is highly educated increases. As a result, we see that observing Salary correlates Education and Height. Similarly, observing an effect or indicator of salary, say the individual's Ferrari, also correlates Education and Height.

The fact that $\sigma_{YX.Z} \neq 0$ when $\sigma_{YX} = 0$ and Z a common child of X and Y can also be illustrated using Wright's path tracing rules. Consider Figure 5a where Z is a common effect of X and Y. We have $\sigma_{YX} = 0$ and, using Equation 5,

$$\begin{aligned}\sigma_{YX.Z} &= \sigma_{YX} - \frac{\sigma_{YZ}\sigma_{ZX}}{\sigma_Z^2} \\ &= 0 - \frac{ab}{1} \\ &= -ab.\end{aligned}$$

When a and b are non-zero we have an algebraic confirmation of our intuition from the salary example that X and Y are uncorrelated marginally, but becoming correlated when we condition on Z.

Berkson's paradox implies that paths containing colliders can be unblocked by conditioning on colliders or their descendants. Let π' be a path from X to Y that traces a collider. If for each collider on the path π', either the collider or a descendant of the collider is in the conditioning set Z then π' is unblocked given Z. The exception to this rule is if Z also contains a non-collider along the path π' in which case X and Y are still blocked given Z. For example, in Figure 4 the path $F \to C \leftarrow A \to E$ is unblocked given C or D. However, it is blocked given $\{A, C\}$ or $\{A, D\}$.

The above three rules can be used to determine if X and Y are d-separated given a set Z while the following theorem makes explicit the relationship between partial correlation and d-separation.

Theorem 1. *Let G be the path diagram for a SEM over a set of variables V. If $X \in V$ and $Y \in V$ are d-separated given a set $Z \subset V$ in the path diagram, G, then $\sigma_{XY.Z} = \rho_{XY.Z} = \beta_{XY.Z} = \beta_{YX.Z} = 0$.*

If X and Y are d-connected given Z then $\sigma_{XY.Z}$ is generally not equal to zero but may equal zero for particular parameterizations. For example, it is possible that the values of the coefficients are such that the unblocked paths between X and Y perfectly cancel one another.

We use the diagram depicted in Figure 4 as an example to illustrate the rules of d-separation. In this example, F is d-separated from E by A and C. However, C is not d-separated from E by A and D since conditioning on D opens the collider $C \to D \leftarrow B$. Finally C is d-separated from E by conditioning on A, D, and B.

D-separation formalizes the intuition that paths carry associational information between variables and that this flow of information can be blocked by conditioning. This intuition drives many of the results in identification, model testing, and other problems that will be discussed in subsequent sections, making d-separation an essential component of graphical modeling.

We conclude this section by noting that d-separation implies vanishing partial correlation in both acyclic and cyclic linear models (Spirtes, 1995). Further, all vanishing partial correlations implied by a SEM can be obtained using

d-separation (Pearl, 2009, ch. 1.2.3). Finally, in models with independent error terms, these vanishing partial correlations represent all of the model's testable implications (Pearl, 2009, ch. 5.2.3).

Identification

A model parameter is *identified* if it can be uniquely determined from the probability distribution over the model variables. If a parameter is not identified, then it cannot be estimated from data because there are many (often infinite) values for the parameter compatible with a given dataset.

If every parameter in the model is identified then the model is said to be identified. If there is at least one unidentified parameter than the model is *not identified* or *unidentified*[10].

In SEMs, a parameter can be identified by expressing it uniquely in terms of the covariance matrix. For example, consider the model represented by Figure 2b. In the previous section (see Equations 10-14), we used Wright's rules to show that the parameter α, which is equivalent to the causal effect of C on S, is identified and equal to $\beta_{SC.A} = \frac{\sigma_A^2\sigma_{CS} - \sigma_{CA}\sigma_{AS}}{\sigma_S^2\sigma_A^2 - \sigma_{SA}^2}$.

In contrast, α is not identified in Figure 1b, whose standardized covariance matrix is:

$$\begin{pmatrix} 1 & \sigma_{SC} \\ \sigma_{SC} & 1 \end{pmatrix}$$

Using Wright's rules we obtain a single equation: $\alpha + C_{CS} = \sigma_{SC}$. Since there are infinite values for α and C_{CS} that satisfy this equation, neither parameter is identified and the model is not identified[11].

Many SEM researchers determine the identifiability of the model by submitting the specification and data to software, which attempts to estimate the coefficients by minimizing a fitting function[12]. If the model is not identified, then the program will be unable to complete the estimation and warns that the model may not be identified. While convenient, there are disadvantages to using typical SEM software to determine model identifiability (Kenny and Milan, 2012). If poor starting values are chosen, the program could mistakenly conclude the model is not identified when in fact it may be identified. When the model is not identified, the program is not helpful in indicating which parameters are not identified nor are they able to provide estimates for identifiable coefficients[13]. Most importantly, the program only gives an answer after the researcher has taken the time to collect data.

Rather than determining the identifiability of parameters by fitting the model, the tools described in this paper enable us to detect identifiability directly from the model specification and express identified parameters in terms of the population covariance matrix. As a result, the modeler can estimate their values from the sample covariance matrix, usually invoking only a few variables, and the resulting estimates will be consistent, as long as the model accurately reflects the data generating mechanism. (In the next section, Model Testing, we will give graphical criteria for testing whether this is indeed the case.) Futher, we avoid issues of poor starting values, are able to identify individual parameters when the model as a whole is not identified, and can determine the identifiability of parameters prior to collecting data. For example, in the previous section, we demonstrated, without data, that α was not identified in Figure 1b, but was identified in Figure 2b. As a result, the researcher knows when designing the study that if α is the effect of interest, she must collect data on A, in addition to C and S.

In this section, we give graphical criteria that allow the modeler to determine whether a given parameter is identified by inspecting the path diagram. While these methods are not complete in the sense that they may not be able to identify *every* coefficient that is identifiable, they subsume the identifiability rules in the existing SEM literature, including the recursive and null rules (Bollen, 1989) and the regression rule (Kenny and Milan, 2012).

Selecting Regressors

It is well known that the coefficients of a structural equation, $Y = \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_k X_k + U_Y$, are identified and can be estimated using regression if the error term, U, is independent of $\mathbf{X} = \{X_1, X_2, ...X_k\}$. However, in some cases, α_i can be estimated using regression even when $\mathbf{X}$ is correlated with U. For example, we showed that α in Figure 2b is equal to $\beta_{CS.A}$ (see Equations 10-14), even though in the corresponding structural equation, $S = \alpha C + U_S$, permits C and U_S to be correlated. As a result, α can be estimated using the regression $S = \beta_1 C + \beta_2 A + \epsilon_S$, which yields $\alpha = \beta_1$[14].

[10] Many authors also use the term "under-identified". This term can be confusing because it suggests models that are not identifiable have no testable implications. This is not the case.

[11] While α is is not identified in Figure 1b, the causal effect of C on S is still well-defined. It is equal to $E[S|do(C = 1)] - E[S|do(C = 0)] = \alpha$. The fact that this quantity is not identified simply means that we cannot estimate it from data on C and S alone.

[12] Determining identifiability by fitting a model has become so commonplace in the SEM community that it is often forgotten that identification and model testing are separate concepts. Identifiability is a property of the model specification only, and remains independent of the data actually observed. "Fitting", on the other hand, is a relationship between the model specification and the data observed. Models can be "fitted" or "tested for fitting" regardless of whether they are identified, although most available software today require identification for "fitting" to produce meaningful results.

[13] According to Kenny and Milan (2012), AMOS is the only program that attempts to estimate parameters when the model is not identified.

[14] We distinguish between structural equations, in which the parameters, $\alpha_1, \alpha_2, ..., \alpha_k$, represent causal effects, and regression

Adding a set of variables, Z, to a regression to estimate a parameter is often called *adjusting for* Z. The question arises, how can we, in general, determine whether a set of variables is adequate for adjustment when attempting to identify a given structural coefficient α? Put another way, how can we determine whether a set Z of variables, when added to the regression of Y on X would render the slope (of Y on X) equal to the desired structural coefficient, α? The following criterion, called *single-door*, allows the modeler to answer this question by inspection of the path diagram.

Theorem 2. *(Pearl, 2009, ch. 5.3.1) (Single-door Criterion) Let G be any acyclic causal graph in which α is the coefficient associated with arrow $X \to Y$, and let G_α denote the diagram that results when $X \to Y$ is deleted from G. The coefficient α is identifiable if there exists a set of variables Z such that (i) Z contains no descendant of Y and (ii) Z d-separates X from Y in G_α. If Z satisfies these two conditions, then α is equal to the regression coefficient $\beta_{YX.Z}$. Conversely, if Z does not satisfy these conditions, then $\beta_{YX.Z}$ is not a consistent estimand of α (except in rare instances of measure zero).*

In Figure 7a, we see that W blocks the spurious path $X \leftarrow Z \to W \to Y$ and X is d-separated from Y by W in Figure 7b. Therefore, α is identified and equal to $\beta_{YX.W}$. This is to be expected since X is independent of U_Y in the structural equation, $Y = \alpha X + cW + U_Y$. Theorem 2 tells us, however, that Z can also be used for adjustment since Z also d-separates X from Y in Figure 7b, and we obtain $\alpha = \beta_{YX.W}$. (We will see in a subsequent section, however, that the choice of W is superior to that of Z in terms of estimation power.)

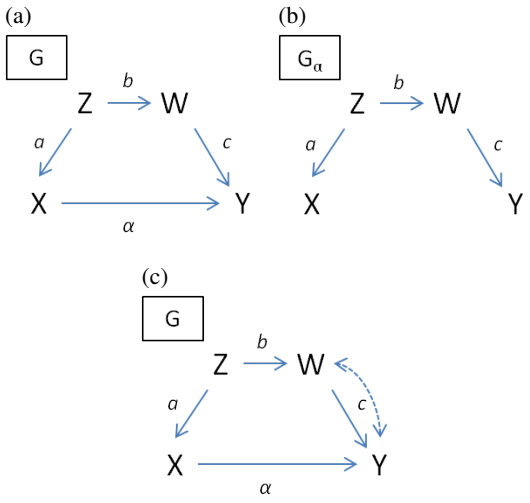


Figure 7. Diagrams illustrating identification by the single-door criterion (a) α is identified by adjusting for Z or W (b) The graph G_α used in the identification of α (c) α is identified by adjusting for Z (or Z and W) but not W alone

Consider, however, Figure 7c. Z satisfies the single-door criterion but W does not. Being a collider, W unblocks the spurious path, $X \leftarrow Z \to W \leftrightarrow Y$, in violation of Theorem 2, leading to bias if adjusted for[15]. In conclusion, α is equal to $\beta_{YX.Z}$ in Figures 7a and 7c. However, α is equal to $\beta_{YX.W}$ in Figure 7a only.

Returning to Figure 2a, we see that S is d-separated from C given A when we remove the edge from S to C, confirming that $\beta_{SC.A} = \alpha$.

The intuition for the requirement that Z not be a descendant of Y is depicted in Figures 8a and 8b. We typically do not display the error terms, which can be understood as latent causes. In Figure 8b, we show the error terms explicitly. It should now be clear that Y is a collider and conditioning on Z will create spurious correlation between X, U_Y, and Y leading to bias if adjusted for. This means that α can be estimated by the regression slope of Y on X, but adding Z to the regression equation would distort this slope, and yield a biased result.

Notice that any coefficient, say from X to Y, is identifiable in a Markovian model using the single-door criterion, since $Pa(Y) \setminus \{X\}$ is a single-door admissible set. As a result, the structural equation for Y, which consists of X and the other parents of Y, can be converted into a regression equation that gives an unbiased estimate of each coefficient in the equation. For example, in Figure 4, f is identifiable because the other parents of E, D and A, represent a single-door admissible

equations, in which the coefficients, $\beta_1, \beta_2, ..., \beta_k$ represent regression slopes. The equation, $S = \beta_1 C + \beta_2 A + \epsilon_S$, is a regression equation, where $\beta_1 = \frac{\partial}{\partial C} E[S|C, A]$, $\beta_2 = \frac{\partial}{\partial A} E[S|C, A]$, and $\epsilon_S = S - \beta_1 C - \beta_2 A$ the residual term. The equation is not structural since β_2 does not equal the direct effect of A on S, $\frac{\partial}{\partial A} E[S|do(C, A)]$, which equals 0. It is for this reason that we refrain from referring to $S = \beta_1 C + \beta_2 A + \epsilon_S$ as a regression "model". It is merely a specification for running a least square routine on the data and estimating the slopes β_1 and β_2.

[15] It is for this reason that the direct effect cannot be defined by conditioning on a mediator but must instead invoke intervention (Pearl, 2014b,c), as we did earlier.

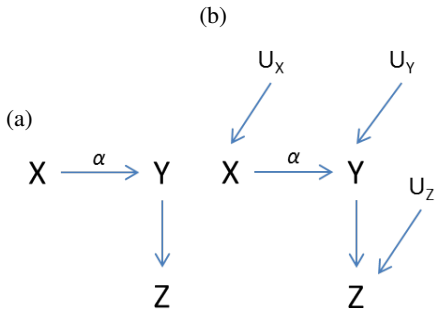


Figure 8. Example showing that adjusting for a descendant of Y induces bias in the estimation of α

set. The same is true for the other coefficients, g and h. As a result, each of these coefficients is identified and can be estimated using the regression, $E = \beta_1 A + \beta_2 B + \beta_3 D + \epsilon$. Using this method, we have the following lemma:

Lemma 1. *Any Markovian (acyclic without correlated error terms) model can be identified equation by equation using regression.*

No matter how complex the model, the single-door theorem gives us a quick and reliable criterion for identification of a structural parameter using regression. It allows us to choose a variety of regressors using considerations of estimation power, sample variability, cost of measurement and more. Further, it is an important tool that plays a role in the identification of parameters in more elaborate models.

Instrumental Variables

In Figure 9a, no single-door admissible set exists for α and it cannot be estimated using regression. However, using Wright's equations we see that $\sigma_{YZ} = \gamma\alpha$ and $\sigma_{XZ} = \gamma$. As a result, $\alpha = \frac{\sigma_{YZ}}{\sigma_{XZ}}$. In this case, we were able to identify α using an auxiliary variable Z called an *instrumental variable (IV)*.

In this subsection, we will provide a graphical method that allows modelers to quickly determine whether a given variable is an IV by inspecting the path diagram. Additionally, we will introduce *conditional instrumental variables* and *instrumental sets*, which will significantly increase the identification power of the instrumental variable method.

The usage of IVs to identify causal effects in the presence of confounding can be traced back to Sewall Wright (1925) and his father Philip Wright (1928), and the following is a standard definition adapted from Bollen (2012):

Definition 1. *For a structural equation, $Y = \alpha_1 X_1 + ... + \alpha_k X_k + U_Y$, Z_i is an* instrumental variable *if*

(i) Z_i is correlated with $X = \{X_1, ..., X_k\}$ and

(ii) Z_i is uncorrelated with U_Y.

Implicit in this definition is that Z_i has no effect on Y except through X. According to Bollen (2012), a necessary condition for the identification of $\alpha_1, ..., \alpha_k$ is that there exists at least k IVs satisfying (i) and (ii), but this condition is not sufficient.

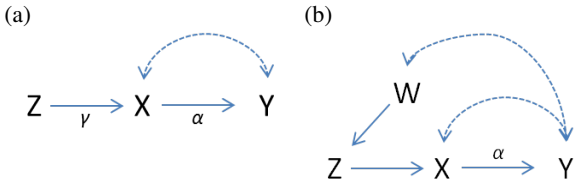


Figure 9. (a) Z qualifies as an instrumental variable (b) Z is an instrumental variable given W

As is typical in the SEM literature, the above definition defines an IV relative to an equation. However, by defining an IV relative to a specific parameter, we will be able to greatly expand the power of IVs. First, this will allow the identification of parameters of interest, even when the equation, as a whole, is not identifiable. For example, in Figure 10a, $\alpha = \frac{\beta_{YZ_1}}{\beta_{X_1 Z_1}}$ and is identified but γ is not. Second, we refine the conditions under which the equation, as a whole, is identified using IVs. For example, in Figure 10b, we have two instruments for Y satisfying the condition of Definition 1, yet (as we shall see later) γ remains unidentified. A sufficient condition for the identification of an equation with k coefficients is the existence of at least one IV for each coefficient. Finally, thinking about IVs as pertaining to individual parameters will also allows us to generalize them and develop new tools like instrumental sets.

Economists have always recognized the benefit of defining IVs relative to parameters rather than equations (Wright, 1928; Bowden and Turkington, 1984) but have had difficulties articulating the conditions that would qualify a variable Z as an instrument in a system of multiple equations. For example, the following requirements for an instrument are offered by Angrist and Pischke (2014):

(i) The instrument has a causal effect on the variable whose effects we're trying to capture...

(ii) Z is randomly assigned or "as good as randomly assigned," in the sense of being unrelated to the omitted variables that we would like to control for...

(iii) Finally, IV logic requires an exclusion restriction. The exclusion restriction describes a single channel through which the instrument affects outcomes.

As we shall see from the graphical criterion of Definition 2, condition (i) is overly restrictive; a proxy of an instrument could also qualify as an instrument. Condition (ii) leaves ambiguous the choice of those "omitted variables", and condition (iii) wrongly excludes multiple channels between Z and X, as well as between X and Y.

The following graphical characterization rectifies such ambiguities and allows us to determine through quick inspection of the path diagram whether a given variable is an instrument for a given parameter. Moreover, it provides a necessary and sufficient condition for when α_i in the equation $Y = \alpha_1 X_1 + ... + \alpha_k X_k + U_Y$ is identified by $\frac{\beta_{YZ_i}}{\beta_{X_i Z_i}}$.

Definition 2. *(Pearl, 2009, p. 248) A variable Z qualifies as an instrumental variable for coefficient α from X to Y if*

(i) Z is d-separated from Y in the subgraph G_α obtained by removing edge $X \to Y$ from G and

(ii) Z is not d-separated from X in G_α.

In Figure 9a, Z is d-separated from Y when we remove the edge associated with α. As a result, Z is an instrumental variable for α and we have $\alpha = \frac{\beta_{YZ}}{\beta_{XZ}}$.

Now, consider Figure 9b. In this diagram, Z is not an instrument for α because it is d-connected to Y through the path $Z \leftarrow W \leftrightarrow Y$, even when we remove the edge associated with α. However, if we condition on W, this path is blocked. Thus, we see that some variables may become instruments by conditioning on covariates.

While this fact is known (in general terms) in the econometric literature (Angrist and Pischke, 2014; Imbens, 2014), finding an appropriate set W in a system of equations, unaided by the graph, is an intractable task. The following definition allows researchers to determine which variables W would allow the identification of a given coefficient using conditional IVs.

Definition 3. *(Brito and Pearl, 2002a) A variable Z is a* conditional instrumental variable *given a set W for coefficient α (from X to Y) if*

(i) W contains only non-descendants of Y

(ii) W d-separates Z from Y in the subgraph G_α obtained by removing edge $X \rightarrow Y$ from G

(iii) W does not d-separate Z from X in G_α

Moreover, if (i)-(iii) are satisfied, then $\alpha = \frac{\beta_{YZ.W}}{\beta_{XZ.W}}$. To demonstrate the power of Definition 3, consider the models in Figure 11.

In Figure 11a, Z is an instrument for α given W because Z is d-separated from Y given W in G_α. However, in Figure 11b, Z is not an instrument given W because conditioning on W opens the paths $Z \rightarrow X \leftrightarrow Y$ (W is a descendant of the collider, X) and $Z \rightarrow W \leftarrow X \leftrightarrow Y$ (W is a collider). Finally, in Figure 11c, Z is again an instrument given W since W is not a descendant of X and the path $Z \rightarrow W \leftrightarrow X \leftrightarrow Y$ is blocked by the collider, X.

Finally, it may be possible to use several variables in order to identify a set of parameters when, individually, none of the variables qualifies as an instrument. In Figure 12a, neither Z_1

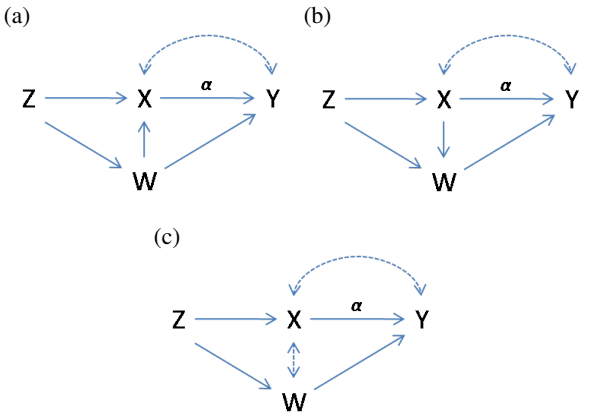


Figure 11. (a) Z is an instrument for α given W (b)

nor Z_2 are instruments for the identification of γ or α. However, using them simultaneously allows the identification of both coefficients. Using Wright's equations, as we did in the single instrumental variable case, we have:

$$\sigma_{Z_1Y} = \sigma_{Z_1X_1}\gamma + \sigma_{Z_1X_2}\alpha$$
$$\sigma_{Z_2Y} = \sigma_{Z_2X_1}\gamma + \sigma_{Z_2X_2}\alpha$$

Solving these two linearly independent equations for γ and α identifies the two parameters. We call a set of variables that enables a solution in this manner an *instrumental set* and characterize them in Definition 4[16].

Note that Z_1 and Z_2 in Figure 10b qualify as IVs according to Definition 1, but do not enable the identification of α and

[16]It can be shown that the well-known rank and order rules (Bollen, 1989; Kline, 2011), which are necessary and sufficient for models that satisfy specific structural properties, are subsumed by instrumental sets. For the class of models that the rank and order rules are applicable to (of all the graphs given in this paper, they can be applied only to Figures 9a, 9b, and 12a), the rank and order rules for the equation, $Y_i = \Lambda_{1i}Y_1 + \Lambda_{2i}Y_2 + ... + \Lambda_{ni}Y_n + U_i$, are satisfied if and only if there exists an instrumental set for the coefficients, $\Lambda_{1i}, \Lambda_{2i}, ..., \Lambda_{ni}$.

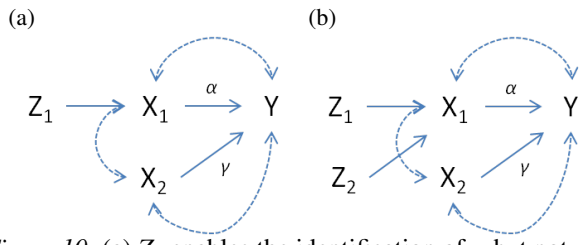


Figure 10. (a) Z_1 enables the identification of α but not γ (b) Adding Z_2 does not enable the identification of γ

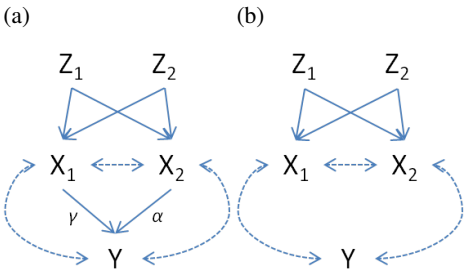


Figure 12. Diagrams illustrating instrumental sets

γ. Likewise, Z_1 and Z_2 in Figure 13b qualify as IVs according to Definition 1, but do not enable the identification of α and γ. Definition 4, adapted from Brito and Pearl (2002a), correctly disqualifies $\{Z_1, Z_2\}$ as an instrumental set in both scenarios.

Definition 4 (Instrumental Set)**.** *For a path π_h that passes through nodes V_i and V_j, let $\pi_h[V_i...V_j]$ denote the "sub-path" that begins with V_i, ends with V_j, and follows the same sequence of edges and nodes as π_h does from V_i to V_j. Then $\{Z_1, Z_2, ..., Z_k\}$ is an* instrumental set *for the coefficients $\alpha_1, ..., \alpha_k$ associated with edges $X_1 \to Y, ..., X_k \to Y$ if the following conditions are satisfied.*

(i) Let $\bar{G}$ be the graph obtained from G by deleting edges $X_1 \to Y, ..., X_k \to Y$. Then, Z_i is d-separated from Y in $\bar{G}$ for all $i \in \{1, 2, ..., k\}$.

(ii) There exists paths $\pi_1, \pi_2, ..., \pi_k$ such that π_i is a path from Z_i to Y that includes edge $X_i \to Y$ and if paths π_i and π_j have a common variable V, then either

(a) both $\pi_i[Z_i...V]$ and $\pi_j[V...Y]$ point to V or

(b) both $\pi_j[Z_j...V]$ and $\pi_i[V...Y]$ point to V.

for all $i, j \in \{1, 2, ..., k\}$ and $i \neq j$.

The following theorem, adapted from (Brito and Pearl, 2002a), explains how instrumental sets can be used to obtain closed form solutions for the relevant coefficients.

Theorem 3. *Let $\{Z_1, Z_2, ..., Z_n\}$ be an instrumental set for the coefficients $\alpha_1, ..., \alpha_n$ associated with edges*

$$X_1 \to Y, ..., X_n \to Y.$$

Then the linear equations,

$$\begin{aligned}
\sigma_{Z_1Y} &= \sigma_{Z_1X_1}\alpha_1 + \sigma_{Z_1X_2}\alpha_2 + ... + \sigma_{Z_1X_n}\alpha_n \\
\sigma_{Z_2Y} &= \sigma_{Z_2X_1}\alpha_1 + \sigma_{Z_2X_2}\alpha_2 + ... + \sigma_{Z_2X_n}\alpha_n \\
&\vdots \\
\sigma_{Z_nY} &= \sigma_{Z_nX_1}\alpha_1 + \sigma_{Z_nX_2}\alpha_2 + ... + \sigma_{Z_nX_n}\alpha_n,
\end{aligned}$$

are linearly independent for almost all parameterizations of the model and can be solved to obtain expressions for $\alpha_1, ..., \alpha_n$ in terms of the covariance matrix.

The second condition in Definition 4 can be understood as requiring that two paths π_i and π_j cannot be broken at a common variable V and have their pieces swapped and rearranged to form two unblocked paths. One of the rearranged paths must contain a collider. For example, in Figure 12a, $\pi_1 = Z_1 \to Z_2 \to X_1 \to Y$ and $\pi_2 = Z_2 \leftrightarrow X_2 \to Y$ satisfy the second condition of Definition 4 because in π_1, the arrow associated with coefficient, a, *entering* the shared node, Z_2, is pointing at Z_2 while in π_2, the arrow associated

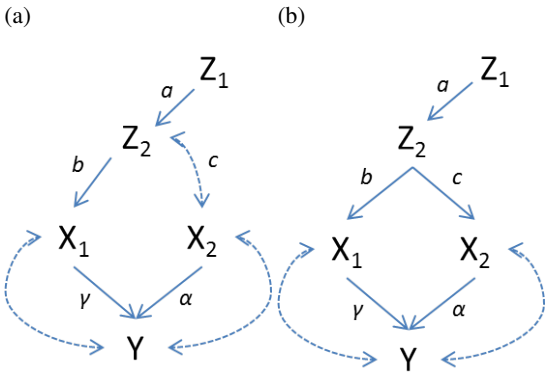


Figure 13. (a) Z_1 and Z_2 qualify as an instrumental set (b) Z_1 and Z_2 do not qualify as an instrumental set

with parameter, c, *leaving* Z_2 is also pointing at the shared node, Z_2. As a result, if the paths π_1 and π_2 are broken at the common variable, Z_2, and their pieces swapped and rearranged, π_1 will become a blocked path due to the collider at Z_2. Algebraically, this means that σ_{Z_1Y} lacks the influence of the path $Z_2 \leftrightarrow X_2 \to Y$ and, therefore, does not contain the term $ac\alpha$. σ_{Z_2Y}, on the other hand, contains the term $c\alpha$ associated with the path. It is in this way that condition (ii) of Definition 4 allows π_i and π_j to share a node, while still ensuring linear independence of the covariance equations and, therefore, identification. To see this, we use Wright's rules to obtain,

$$\begin{aligned}
\sigma_{Z_1Y} &= ab\gamma = \sigma_{Z_1X_1}\gamma + 0 \cdot \alpha = \sigma_{Z_1X_1}\gamma + \sigma_{Z_1X_2}\alpha \text{ and} \\
\sigma_{Z_2Y} &= b\gamma + c\alpha = \sigma_{Z_2X_1}\gamma + \sigma_{Z_2X_2}\alpha,
\end{aligned}$$

which are linearly independent. Solving the equations identifies α and γ giving:

$$\begin{aligned}
\gamma &= \frac{\sigma_{Z_1Y}}{\sigma_{Z_1X_1}} \\
\alpha &= \frac{\sigma_{Z_2Y}}{\sigma_{ZX_2}} - \frac{\sigma_{Z_2X_1}\sigma_{Z_1Y}}{\sigma_{Z_2X_2}\sigma_{Z_1X_1}}
\end{aligned}$$

In contrast, consider Figure 13b. Here, Z_1 and Z_2 are not an instrumental set for α and γ. Every path from Z_2 to Y is a "sub-path" of a path from Z_1 to Y, which, using Wright's rules, implies that the equation for σ_{Z_1Y} is not linearly independent of σ_{Z_1Y} with respect to Y's coefficients:

$$\begin{aligned}
\sigma_{Z_1Y} &= b\gamma + c\alpha \\
\sigma_{Z_2Y} &= ab\gamma + ac\alpha = a(b\gamma + c\alpha) = a\sigma_{Z_1Y}
\end{aligned}$$

In some cases, condition (i) of Definition 4 can be satisfied by conditioning on a set W. Brito and Pearl (2002a) show how conditioning can be used to obtain a *conditional instrumental set*. Due to the more complex nature of applying Wright's rules over partial correlations, we do not cover

conditional instrumental sets in this paper and instead refer the reader to Brito and Pearl (2002a).

C-Component Decomposition

In this subsection, we show that the question of coefficient identification can be addressed using smaller and simpler sub-graphs of the original causal graph. Further, in some cases, the coefficient is not identified using any methods considered thus far on the original graph but is identified using those methods on the sub-graph.

A *c-component* in a causal graph is a maximal set of nodes such that all nodes are connected to one another by paths consisting of bidirected arcs. For example, the graph in Figure 13b consists of three c-components, $\{X_1, X_2, Y\}$, $\{Z_2\}$, and $\{Z_1\}$, while the graph depicted in Figure 15 consists of a single c-component. Tian (2005) showed that a coefficient is identified if and only if it is identified in the sub-graph consisting of its c-component and the parents of the c-component.

More formally, a coefficient from X to Y is identified if and only if it is identified in the sub-model constructed in the following way:

(i) The sub-model variables consist of the c-component to which Y belongs, C_Y, union the parents of all variables in that c-component.

(ii) The structural equations for the variables in C_Y are the same as their structural equations in the original model.

(iii) The structural equations for the parents simply equate each parent with its error term.

(iv) If the error terms of any two variables in the sub-model were uncorrelated in the original model then they are uncorrelated in the sub-model.

For example, the sub-model for the coefficient α from X to Y in Figure 14a consists of the following equations:

$$
\begin{aligned}
Z &= U_Z \\
X &= aX + U_X \\
W &= bW + U_W \\
V &= U_V \\
Y &= \alpha X + dV + U_Y
\end{aligned}
$$

Additionally, $\rho_{U_X U_Y}$ and $\rho_{U_W U_Y}$ are unrestricted in their values. All other error terms are uncorrelated.

It is not clear how to identify the coefficient α depicted in Figure 14a using any of the methods considered thus far. However, the sub-graph for the c-component, $\{W, X, Y\}$, depicted in Figure 14b, shows that α is identified using Z as an instrument. Therefore, α is identified in the original model.

It is important to note that the covariances in the sub-model are not necessarily the same as the covariances in the original model. As a result, the identified expressions obtained from the sub-model may not apply to the original model. For example, Figure 14b shows that $\alpha = \frac{\beta_{ZY}}{\beta_{ZX}}$. However, this is clearly not the case in Figure 14a. The above method simply tells us that α is identified. It does not give us the identified expression for α. Tian (2005) shows how the covariance matrix for the sub-model can be obtained from the original covariance matrix thus enabling us to obtain the identified expression for the parameter in the original model. However, we do not cover it here.

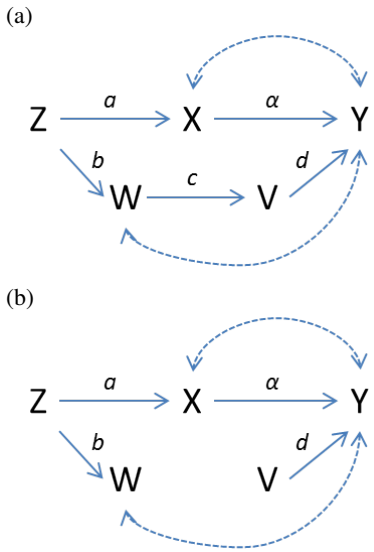


Figure 14. (a) Example illustrating c-component decomposition (b) Sub-graph consisting of c-component, $\{W, X, Y\}$, and its parents, Z and V.

A Simple Criterion for Model Identification

The previous criteria allow researchers to determine whether a given coefficient is identifiable and provide closed form expressions for the coefficients in terms of the covari-

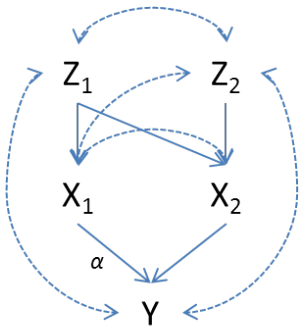


Figure 15. A bow-free graph; the absence of a 'bow' pattern assures identification

ance matrix. As a result, they provide an alternative to using system-wide ML methods (e.g. Full Information Maximum Likelihood) that is unbiased in small samples, can be used when the model is not identified, and do not require data. Should modelers choose to identify and estimate models using software incorporating system-wide ML methods and the estimation fail, it can be useful to know whether the failure is due to non-identification or other issues.

In order to determine identifiability of the model using the single-door criterion or instrumental variables, the modeler must check the identifiability of each structural coefficient. In large and complex models, this process can be tedious. In this section, we give a simple, sufficient criterion that allows the modeler to determine immediately whether an acyclic model is identified called the *bow-free rule* (Brito and Pearl, 2002b; Brito, 2004). We will see that even a model as complicated as Figure 15 can be immediately determined to be identified using this rule.

A *bow-arc* is a pair of variables, one of which is a direct function of the other, whose error terms are correlated. This is depicted in the path diagram as a parent-child pair that are also siblings and looks like a bow-arc. In Figure 7c, the variables W and Y create a bow-arc.

Theorem 4. *(Brito and Pearl, 2002b) (Bow-free Rule) Every acyclic model whose path diagram lacks bow-arcs is identified*[17].

The bow-free rule is able to identify models that the single-door criterion is not. In Figure 15, for example, the coefficient α is not identified using the single-door criterion. Attempting to block the back-door path, $X_1 \leftrightarrow X_2 \rightarrow Y$, by conditioning on X_2 opens the path $X_1 \leftrightarrow Z_2 \leftrightarrow Y$ because X_2 is a descendant of the collider, Z_2. However, because Figure 15 does not contain any bow-arcs it is identified according to Theorem 4. Finally, since the single-door criterion is unable to identify any model that contain bow-arcs[18], the bow-free rule subsumes the single-door criterion when applied to model identification. (Note that the single-door criterion may be able to identify some coefficients even when the model as a whole is not identified. In contrast, the bow-free rule only addresses the question of model identifiability, not the identifiability of individual coefficients in unidentified models.)

Advanced Identification Algorithms

In this subsection, we survey advanced algorithms that utilize the path diagram to identify model parameters. The details of these algorithms are beyond the scope of this paper, and we instead refer the reader to the relevant literature for more information.

Instrumental variables and sets demonstrate that algebraic properties of linear independence translate to graphical properties in the path diagram that can be used to identify model coefficients. The G-Criterion algorithm (Brito, 2004; Brito and Pearl, 2006) expands this notion in order to give a method for systematically identifying the coefficients of an acyclic SEM.

This algorithm was generalized by Foygel et al. (2012) to determine identifiability of a greater set of graphs[19]. Additionally their criterion, called the half-trek criterion, applies to both acyclic and cyclic models. The half-trek algorithm was further generalized by Chen et al. (2014) to identify more coefficients in unidentified models.

The aforementioned algorithms of Brito (2004), Foygel et al. (2012), and Chen et al. (2014) identify coefficients by searching for graphical patterns in the diagram that correspond to linear independence between Wright's equations. Tian (2005), Tian (2007), and Tian (2009) approach the problem differently and give algorithms that identify parameters by converting the structural equations into orthogonal partial regression equations.

Finally, do-calculus (Pearl, 2009) and non-parametric algorithms for identifying causal effects (Tian and Pearl, 2002a; Tian, 2002; Shpitser and Pearl, 2006; Huang and Valtorta, 2006) may also be applied to parameter identification in linear models. These methods have been shown to be complete for non-parametric models (Shpitser and Pearl, 2006; Huang and Valtorta, 2006) and, if theoretically possible, are able to identify any expectations of the form $E(Y|do(X = x, Z = z)$, where Z represents any susbet of variables in the model other than X and Y. As mentioned in the preliminaries, a coefficient from X to Y equals $\frac{\partial}{\partial x}E[Y|do(X = x, S = s)$, where S represents all variables in the model other than X and Y.

Total Effects

When the model is not identifiable, modelers typically consider research with SEMs "impossible" (Kenny and Milan, 2012) without imposing additional constraints or collecting additional data. However, as should be clear from the single-door criterion (and is acknowledged by Kenny and Milan (2012)), it is often possible to identify some of the model coefficients even when the model as a whole is not

[17] Note that the equations in such models are not regression equations as suggested by Kenny and Milan (2012). The independent variable may be correlated with the error term of the dependent variable as in X_1 and Y in Figure 15. X_1 is correlated with the error term of Y through the path $X_1 \leftarrow Z_1 \leftrightarrow Y$. Another way of defining a bow-free model is a model where error terms of every parent-child pair are not correlated.

[18] To prove this statement, consider any model that contains a bow-arc from X to Y. There is no way to block the path $X \leftrightarrow Y$ and identify the coefficient from X to Y using the single-door criterion.

[19] Foygel et al. (2012) also released an R package implementing their algorithm called SEMID, which determines whether the entire model is identifiable given its causal graph.

identifiable. Further, we show in this section that it is often not necessary to identify all coefficients along a causal path in order to identify the causal effect of interest[20]. For example, in Figure 13b, the total effect or ACE of Z on Y, $\frac{\partial}{\partial z}E[Y|do(Z = z)]$, is identified and equal to β_{ZX} even though γ and α are not identified. The back-door criterion, given below, is a sufficient condition for the identification of a total effect.

Theorem 5. *(Pearl, 2009, ch. 3.3.1) (Back-door Criterion) For any two variables X and Y in a causal diagram G, the total of effect of X on Y is identifiable if there exists a set of variables Z such that*

(i) no member of Z is a descendant of X; and

(ii) Z d-separates X from Y in the subgraph $G_{\underline{X}}$ formed by deleting from G all arrows emanating from X.

Moreover, if the two conditions are satisfied, then the total effect of X on Y is given by $\beta_{YX.Z}$.

Returning to the example in Figure 13b we see that the total of effect of Z on Y, $\frac{\partial}{\partial z}E[Y|do(Z = z)]$, is β_{ZX}.

Do-calculus and the aforementioned non-parametric algorithms (Tian and Pearl, 2002a; Tian, 2002; Shpitser and Pearl, 2006; Huang and Valtorta, 2006) can also be used to identify total effects in linear models.

Causal Effects among Latent Variables

The graphical methods described above do not explicitly address the identification of causal effects among latent variables (e.g. the effect of a latent variable on another latent variable, the effect of an observed variable on a latent, or the the effect of latent variable on an observed variable). They are, nevertheless, applicable to the identification of such effects. With respect to non-identification, if we assume that all latent variables are observed and are still unable to identify the effect of interest then it clearly cannot be identified when one or more of the variables are latent. With respect to identification, if a latent variable has three or more observed indicators without any edges between them (see Figure 16a) then we can consider that latent variable to be observed and apply the above methods (Bollen, 1989). In certain cases, only two indicators per latent variable may be enough as in Figure 16b (Bollen, 1989) and Figure 16c (Kuroki and Pearl, 2014). In the former, the four indicators are enough to ensure identification of the coefficients from the latents to their indicators and the coefficient from L_1 to L_2, which then allows identification of the covariance between the latents and any observed variables in the model. In the latter, X and Y together act as a third indicator, which also allows identification of the coefficients from L to its indicators.

In general, we can apply the above graphical methods to the identification of coefficients in latent variable models in the following way. First, consider any latent variables that exhibit the patterns in Figures 16a, 16b, and 16c to be observed variables. Any remaining latent variables are summarized using the method described earlier. We are now left with no explicit latent variables (other than the error terms) and can apply the methods described above. If we find that a coefficient is identified in this augmented model then we know it is also identified in the original latent variable model.

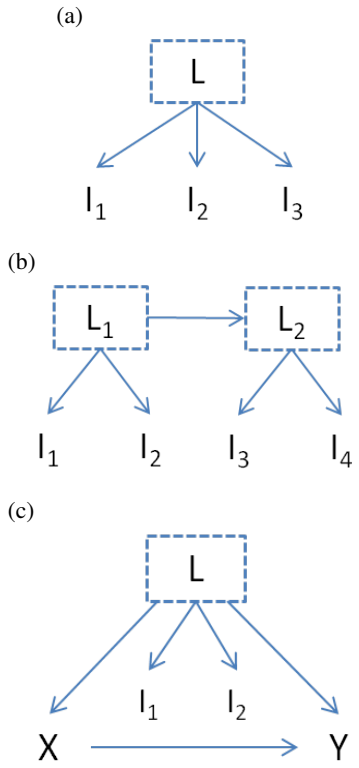


Figure 16. Graphical patterns that allow latent variables to be considered observed for purposes of identification.

More recently, researchers have begun using the power of graphical representations to identify the coefficients between latent variables and their indicators in linear SEMs. For example, Cai and Kuroki (2008) and Leung et al. (2015) give sufficient graphical identifiability conditions for models that contain a single latent variable.

Model Testing

A crucial step of structural equation modeling is to test the structural and causal assumptions of the model, ensuring to the best of our ability that the model specification is correct. A given model often imposes certain constraints on the probability distribution or covariance matrix and checking

[20] This fact was noted by Marschak (1942) and was dubbed "Marschak's Maxim" by Heckman (2000).

whether these constraints hold in the data provides a means of testing the model. For example, we showed, in the section on d-separation, that a model may imply that certain partial correlations are equal to zero. If these constraints do not hold in the data, then we have reason to doubt the validity of our model.

The most common method of testing a linear SEM is a likelihood ratio or chi-square test that compares the covariance matrix implied by the model to that of the sample covariance matrix (Bollen, 1989; Shipley, 2000). While this test simultaneously tests all of the restrictions implied by the model, it relies critically on our ability to identify the model. Moreover, bad fit does not provide the modeler with information about which aspect of the model needs to be revised[21]. Finally, if the model is very large and complex, it is possible that a global chi-square test will not reject the model even when a crucial testable implication is violated. Global tests represent summaries of the overall model-data fit and, as a result, violation of specific testable implications may be masked (Tomarken and Waller, 2003). In contrast, if the testable implications are enumerated and tested individually, the model can be tested even when unidentified, the power of each test is greater than that of a global test (Bollen and Pearl, 2013; McDonald, 2002), and, in the case of failure, the researcher knows exactly which constraint was violated.

Vanishing Correlation Constraints

D-separation allows modelers to predict vanishing partial correlations simply by inspecting the graph, and in the case of Markovian models, these vanishing partial correlations represent *all* of the constraints implied by the model (Geiger and Pearl, 1993)[22]. For the example depicted in Figure 17a, we obtain the following vanishing partial correlations: $\rho_{V_2V_3.V_1} = 0$, $\rho_{V_1V_4.V_2V_3} = 0$, $\rho_{V_2V_5.V_4} = 0$, and $\rho_{V_3V_5.V_4} = 0$. If a constraint, say $\rho_{V_2V_3.V_1} = 0$ does not hold in the dataset, we have reason to believe that the model specification is incorrect and should reconsider the lack of edge between V_2 and V_3.

In large and complex graphs, it may be infeasible to list all conditional independence constraints by inspection. Additionally, some constraints obtained using d-separation may be redundant. Kang and Tian (2009) gave an algorithm that utilizes the graph to enumerate a set (not necessarily minimal) of vanishing partial correlations that imply all others for semi-Markovian models.

Lastly, we note that d-separation implies vanishing partial correlation even in non-linear models.

Equivalent Models

Since vanishing partial correlations represent all of the constraints that Markovian SEMs impose on the data, two Markovian models are observationally indistinguishable if they share the same set of vanishing partial correlations. In other words, Markovian models that share the same set of vanishing partial correlations cannot be distinguished using data. In this case, we say that the models are *covariance equivalent* since every covariance matrix generated by one model (through some choice of parameters) can also be generated by the other. The *skeleton* of a graph, used in the following theorem, is the undirected graph obtained by replacing all arrows with undirected edges. For example, the skeleton for Figure 17a is Figure 17b.

Theorem 6. *(Verma and Pearl, 1990) Two Markovian linear-normal models are covariance equivalent if and only if they entail the same sets of zero partial correlations. Moreover, two such models are covariance equivalent if and only if their corresponding graphs have the same skeletons and the same sets of v-structures, that is, two converging arrows whose tails are not connected by an arrow.*

The first part of Theorem 6 defines the testable implications of linear Markovian models. It states that, in non-experimental studies, Markovian SEMs cannot be tested for any feature other than those vanishing partial correlations that the d-separation test imposes. It also provides a simple test for equivalence that requires merely a comparison of corresponding edges and their directionalities (Pearl, 2009, ch. 5.2).

The graphs in Figures 18a, 18b, and 18c are equivalent because they share the same skeleton and v-structures. Note

[21] While modification indices can be used, they also require the model to be identified.

[22] These constraints may induce non-conditional independence constraints when projected onto a subset of variables. For example, suppose that L_1 and L_2 in Figure 16b are observed and the model is, therefore, Markovian. While this model implies a vanishing tetrad constraint, $\sigma_{I_1I_4}\sigma_{I_2I_3} = \sigma_{I_1I_3}\sigma_{I_2I_4}$ (Spearman, 1904), this constraint can, in fact, be derived from the vanishing partial correlations among the I variables given the L variables.

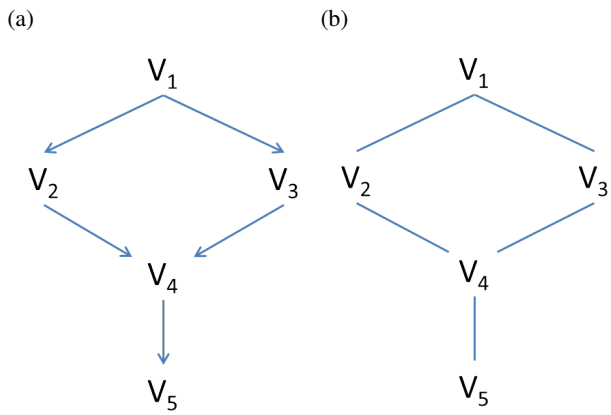


Figure 17. (a) Example illustrating vanishing partial correlation (b) The skeleton of the model in (a)

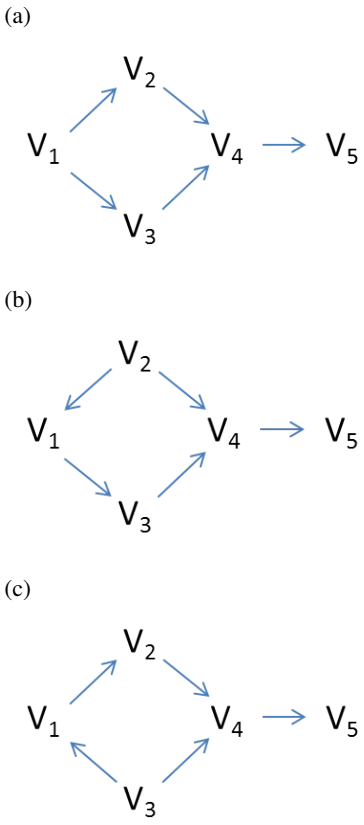


Figure 18. Models (a), (b), and (c) are equivalent.

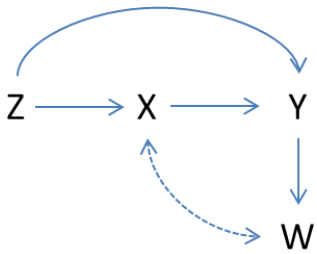


Figure 19. Counterexample to the standard Replacement Rule; The arrow $X \rightarrow Y$ cannot be replaced.

that we cannot reverse the edge from V_4 to V_5 since doing so would generate a new v-structure, $V_2 \rightarrow V_4 \leftarrow V_5$.

The graphical criterion given in Theorem 6 is necessary and sufficient for equivalence between Markovian models. It is a necessary condition for equivalence between non-Markovian models since d-separation in the graph implies vanishing partial correlation in the covariance matrix. In contrast, the more prevalent replacement criterion (Lee and Hershberger, 1990) is not always valid[23]. Pearl (2012) gave the following example depicted in Figure 19. According to the replacement criterion, we can replace the arrow $X \rightarrow Y$ with a bidirected edge $X \leftrightarrow Y$ and obtain a covariance equivalent model when all predictors (Z) of the effect variable (Y) are the same as those for the source variable (X). Unfortunately, the post-replacement model imposes the constraint, $\rho_{WZ.Y} = 0$, which is not imposed by the original model. This can be seen from the fact that, conditioned on Y, the path $Z \rightarrow Y \leftarrow X \leftrightarrow W$ is unblocked and becomes blocked if replaced by $Z \rightarrow Y \leftrightarrow X \leftrightarrow W$. The same applies to path $Z \rightarrow X \leftrightarrow W$, since Y would cease to be a descendant of X.

Testable Implications in Non-Markovian Models

In the case of non-Markovian models, additional testable implications may be present, which are not revealed by d-separation. In the non-parametric literature, these constraints are often called *Verma constraints* (Verma and Pearl, 1990) and impose invariance rather than conditional independence restrictions. In Figure 20, for example, one can show that the quantity $\sum_{V_2} P(V_4|V_3, V_2, V_1)P(V_2|V_1)$ is not a function of V_1. Algorithms that enumerate certain types of Verma constraints for semi-Markovian, non-parameteric SEMs are given by Tian and Pearl (2002b) and Shpitser and Pearl (2008).

Testable implications in non-Markovian models can also be obtained by *overidentifying* model parameters. In some cases, these constraints will be vanishing partial correlations, while in other cases they are not. For example, in Figure 20, b can be identified by using the single-door criterion, yielding $b = \beta_{V_3V_2}$, and by using V_1 as an IV, yielding $b = \frac{\beta_{V_3V_1}}{\beta_{V_2V_1}}$. Equating the two expressions and rearranging terms gives the constraint, $\beta_{V_3V_1} - \beta_{V_2V_1}\beta_{V_3V_2} = 0$, which is equivalent to the constraint, $\rho_{V_3V_1.V_2} = 0$ (see Equations 4-6), as seen in the graph. Similarly, c can be overidentified by using the single-door criterion and V_1 as an IV, yielding two estimands, $\beta_{V_4V_3.V_2}$ and $\frac{\beta_{41}}{\beta_{21}\beta_{32}}$. Equating the two expressions gives the constraint $\beta_{V_4V_3.V_2} = \frac{\beta_{41}}{\beta_{21}\beta_{32}}$, which is not a vanishing partial correlation. It is equivalent to the Verma constraint obtained using the non-parametric methods of Tian and Pearl (2002b) and Shpitser and Pearl (2008).

Parameters are often described as overidentified when they have "more than one solution" (MacCallum, 1995) or are "determined from [the covariance matrix] in different ways" (Jöreskog et al., 1979). Overidentified parameters are presumed to be more firmly collaborated by data than "just

[23]The replacement rule violates the transitivity of equivalence (Hershberger and Marcoulides, 2006), yet it is still used in most of the SEM literature (Mulaik, 2009; Williams, 2012, pp. 247-260).

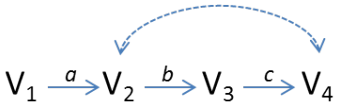


Figure 20. A graph illustrating a Verma constraint

identified" parameters (i.e. those that have only one solution). However, simply finding two distinct expressions for a parameter in terms of the covariance matrix does not necessarily give another independent testable implication. For example, in Figure 20, we have that $\sigma_{V_1V_3.V_2} = 0$. By adding $\sigma_{V_1V_3.V_2}$ to any expression for a parameter, we will have two different expressions for that parameter, implying that every parameter in the model is overidentified. Clearly, this notion of overidentification does not turn a parameter "overidentified". Pearl (2004) gives a formal definition of overidentification that ensures multiple independent tests, and Chen et al. (2014) give an algorithm that utilizes advanced identification methods to systematically discover overidentifying constraints given a non-Markovian model.

We conclude this section by noting that testable implications in semi-Markovian and non-Markovian models have not been fully characterized, and subsequently, we do not have a necessary and sufficient condition for equivalence between semi-Markovian or non-Markovian models.

Learning Structure from Data

The question naturally arises whether one can learn the structure of the data generating model from its data. In other words, rather than specify the structural equation model and use the data to test it, can one use the data to discover aspects of the model's structure? There are a number of algorithms that search the data for vanishing partial correlations to accomplish this goal for acyclic models. See Cooper (1999), (Pearl, 2009, ch. 2), and (Spirtes et al., 2000, chs. 5 and 6)[24] for examples. For cyclic models, Hoover and Phiromswad (2013) make use of overidentifying constraints obtained using instrumental variables in addition to vanishing partial correlations to uncover aspects of the model's structure.

Additional Applications of Graphical Models

Equivalent Regressor Sets and Minimal Regressor Sets

In some cases, we may wish to know whether two sets, when used for adjustment, have the same asymptotic bias. For example, an investigator may wish to assess, prior to taking any measurement, whether two candidate sets of covariates, differing substantially in dimensionality, measurement error, cost or sample variability are equally valuable in their bias-reduction potential (Pearl and Paz, 2014). This problem pertains to prediction tasks as well. A researcher wishing to predict the value of some variable given a set of observations may wonder whether another set of observations is a valid substitute.

In the linear case, the problem can be stated in the following way. Under what conditions would replacing $Z = \{Z_1, ..., Z_n\}$ with $W = \{W_1, ..., W_n\}$ yield the same value for α in the regression $Y = \alpha X + \beta_1 Z_1 + ... + \beta_n Z_n + \epsilon_n$, or equivalently, when does $\beta_{YX.Z} = \beta_{YX.W}$?

Here we adapt Theorem 3 in (Pearl and Paz, 2014) for linear SEMs.

Theorem 7. *(Pearl and Paz, 2014) Let Z and W be two sets of variables in G containing no descendants of X. $\beta_{YX.Z} = \beta_{YX.W}$ if and only if one of the following holds:*

(i) Z and W satisfy the back-door criterion for the total effect of X on Y

(ii) $Z \cap W$ separates X from all other elements of Z and W

If $\beta_{YX.Z} = \beta_{YX.W}$ then we say that Z and W are *confounding equivalent*, or c-equivalent for short.

Consider the graph depicted in Figure 21. Let $Z = \{V_1, W_2\}$ and $W = \{W_1, V_2\}$. Since both Z and W satisfy the back-door criterion they are c-equivalent and $\beta_{YX.Z} = \beta_{YX.W}$. Now consider $Z = \{V_1\}$ and $W = \{V_1, V_2\}$. Z and W no longer satisfy the back-door criterion. However, since $Z \cap W = \{V_1\}$ separates X from $(Z \cup W) \setminus Z \cap W = \{V_2\}$, Z and W are c-equivalent and $\beta_{YX.Z} = \beta_{YX.W}$.

C-equivalence can also be used to find a minimal subset of regressors needed for estimating a given partial regression coefficient. Consider a regression equation, $Y = \alpha X + \beta_1 Z_1 + ... + \beta_n Z_n$. What is the smallest subset of $Z = \{Z_1, ..., Z_n\}$ that yields the same value for the regression coefficient, α? This subset is unique and can be found simply by removing elements from Z one at a time such that every removed node is d-separated from X given the remaining elements of Z.

Some economists (e.g. Lu and White (2014)) regard the resilience and stability of regression coefficients to additional regressors to be a sign of model correctness. The idea is to assess whether any additional variables, hypothesized to have no influence on the equation tested, leave intact the regression slopes that were estimated. Using Theorem 7, we see that this sort of robustness is expected to hold when the back-door criterion is satisfied before and after the addition of regressors or when the added regressors are separated from X given the existing regressors. When these conditions are

[24] Software implementing these algorithms is available from the TETRAD Project (http://www.phil.cmu.edu/projects/tetrad/).

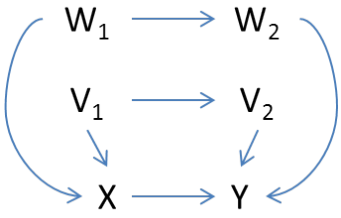


Figure 21. $\{V_1, W_2\}$ and $\{V_2, W_1\}$ are c-equivalent but not $\{W_1\}$ and $\{W_2\}$

satisfied, a shift in the regression coefficient of X is indicative of model misspecification, while stability of the coefficient is indicative of model correctness. Robustness tests are discussed in more detail in Chen and Pearl (2015).

Variance Minimization

In some cases, there may be multiple sets that satisfy the back-door criterion when identifying a total effect. While each set provides an unbiased estimate of the causal effect, the estimates may differ in their asymptotic variance. As a result, some sets may be preferable to others. The following theorem is adapted from Theorem 5 of (Kuroki and Miyakawa, 2003):

Theorem 8. *Suppose that sets* $\{W, Z_1\}$ *and* $\{W, Z_2\}$ *satisfy the back-door criterion relative to* (X, Y) *in a linear SEM in which the error terms are normally distributed. If* $\{W, Z_1\}$ *d-separates* X *from* Z_2 *and* $\{X, Z_2, W\}$ *d-separates* Y *from* Z_1*, then* $Var[\beta_{YX.WZ_2}] \leq Var[\beta_{YX.WZ_1}]$*. In other words, the asymptotic variance of the effect estimated when controlling for* $\{W, Z_2\}$ *is less than or equal to the one estimated by controlling for* $\{W, Z_1\}$*.*

For the model depicted by Figure 22, both $\{W, Z_1\}$ and $\{W, Z_2\}$ are back-door admissible sets for estimating the total effect of X on Y. However, $\{W, Z_2\}$ is preferable since $\{W, Z_1\}$ d-separates X from Z_2 while $\{X, Z_2, W\}$ d-separates Y from Z_1. The intuition here is that Z_2 is 'closer' to Y hence more effective in reducing variations in Y due to uncontrolled factors. Similar results were derived without graphs by Hahn (2004).

Counterfactuals in Linear Models

We have seen in the subsection on causal effects how a SEM can be used to predict the effect of actions and policies that have never been implemented before. The action of setting a variable, X, to value x, is simulated by replacing the structural equation for X with the equation $X = x$. In this section, we show further that SEMs can be used to answer counterfactual queries. A counterfactual query asks, "Given that we observe $E = e$ for a given individual, what would we expect the value of B for that individual to be if A had been a?" For example, given that Joe's salary is s, what would his salary be had he had five more years of education? This expectation is denoted $E[B_{A=a}|E = e]$. The $E = e$ after the conditioning bar represents the observed evidence while the subscript $A = a$ represents a hypothetical condition specified by the counterfactual sentence. Structural equation models are able to answer counterfactual queries because each equation represents an invariant mechanism by which a variable obtains its values. If we identify these mechanisms we should also be able to predict what values would be obtained had circumstances been different. As a result, it is natural to view counterfactuals to be derived properties of structural equations and not the other way around. This is in contrast to the Neyman-Rubin potential outcomes framework where counterfactuals are taken as primitives (Rubin, 1974; Holland, 2001).

The following model, depicted in Figure 23a, represents an "encouragement design" (Holland, 1988; Pearl, 2014c) where X represents the amount of time a student spends in an after-school remedial program, H the amount of homework a student does, and Y a student's score on the exam. The value of each variable is given as the number of standard deviations above the mean so that the model is standardized to mean 0 and variance 1. For example, if $Y = 1$ then the student scored 1 standard deviation above the mean on his or her exam.

Model 6.

$$
\begin{aligned}
X &= U_X \\
H &= a \cdot X + U_H \\
Y &= b \cdot X + c \cdot H + U_Y \\
\sigma_{U_i U_j} &= 0 \text{ for all } i, j \in \{X, H, Y\}
\end{aligned}
$$

We also give the values for the coefficients (which can be

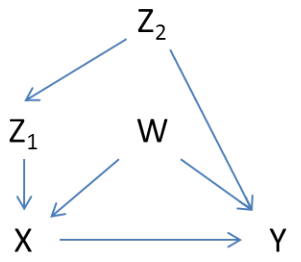


Figure 22. Graph illustrating preference to Z_1 over Z_2; $Var[\beta_{YX.WZ_2}] \leq Var[\beta_{YX.WZ_1}]$

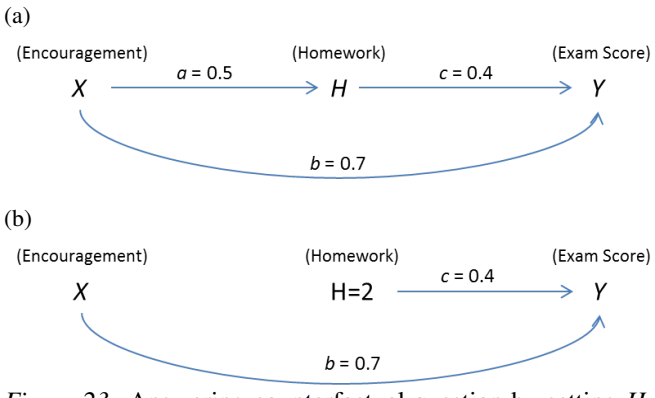


Figure 23. Answering counterfactual question by setting H equal to 2

estimated from population data):

$$a = 0.5$$
$$b = 0.7$$
$$c = 0.4$$

Let us consider a student named Joe, for whom we measure $X = 0.5, H = 1, Y = 1.5$. Suppose we wish to answer the following query: What would Joe's score have been had he doubled his study time?

In a linear SEM, the value of each variable in the model is determined by the coefficients and U variables, and the latter accounts for all variations among individuals. As a result, we can use the evidence $X = 0.5, H = 1, Y = 1.5$ to determine the values of the U variables associated with Joe. These values are invariant to external variations, such as those which might cause Joe to double his homework.

In this case, we are able to obtain the specific characteristics of Joe from the evidence:

$$\begin{aligned} U_X &= 0.5, \\ U_H &= 1 - 0.5 \cdot 0.5 = 0.75, \text{ and} \\ U_Y &= 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75. \end{aligned}$$

Next, we simulate the action of doubling Joe's study time by replacing the structural equation for H with the constant $H = 2$. The modified model is depicted in Figure 23b. Finally, we compute the value of Y in our modified model using the updated U values giving:

$$\begin{aligned} Y_{H=2}(U_X = 0.5, U_H = 0.75, U_Y = 0.75) \\ &= 0.5 \cdot 0.7 + 2.0 \cdot 0.4 + 0.75 \\ &= 1.90 \end{aligned}$$

We thus conclude that Joe's new score, predicated on doubling his homework, would have been 1.9 instead of 1.5.

In summary, we first applied the evidence $X = 0.5, H = 1, Y = 1.5$ to update the values for the U variables or their probabilities. We then simulate an external intervention to force the condition $H = 2$ by replacing the structural equation $H = aX + U_H$ with the equation $H = 2$. Finally, we computed the value of Y given the structural equations and the updated U values.

The following three steps generalize the above procedure for non-linear systems and arbitrary counterfactuals of the form, $E[B_{A=a}|E = e]$ (Pearl, 2009, ch. 7.1):

(i) **Abduction** - Update $P[U]$ by the evidence to obtain $P[U|E = e]$

(ii) **Action** - Modify the model, M, by removing the structural equations for the variables in A and replacing them with the appropriate equalities to obtain the modified model, M_A.

(iii) **Prediction** - Use the modified model, M_A, and the updated probabilities over the U variables, $P[U|E = e]$, to compute the expectation of B, the consequence of the counterfactual.

Notice that the above procedure applies not only to retrospective counterfactual queries (queries of the form "What would have been the value of Y had X been x?") but also prospective counterfactual queries (queries of the form "What will the value of Y be if X is set to x by intervention?"). For example, suppose we wish to estimate the effect on test score provided by a school policy that sends students who are lazy on their homework ($S \leq -1$) to attend the afterschool program for $X = 1$. The expected value of this quantity is denoted $E[Y_{X=1}|S \leq -1]$ and can, in principle, be computed using the above three step method. Counterfactual reasoning and the above procedure are necessary for estimating the effect of actions and policies on subsets of the population characterized by features that, in themselves, are policy dependent (e.g. $S \leq -1$).

In non-parametric models, counterfactual quantities of the form $E[Y_{X=x}|E = e]$ may not be identifiable, even if we have the luxury of running experiments (Pearl, 2009, ch. 9). In linear models, however, any counterfactual quantity is identifiable whenever the model parameters are identifiable (**?**). Moreover, even when some parameters are not identified, a counterfactual of the form, $E[Y_X = x|E = e]$ is identified whenever $E[Y|do(X = x)]$ is identified (Cai and Kuroki, 2005; Pearl, 2009, p. 389). The relation between the two is summarized in Theorem 9.

Theorem 9. *(Pearl, 2009, p. 389) Let T be the slope of the total effect of X on Y, $\frac{\partial}{\partial x}E[Y|do(X = x)]$, then $E[Y_{X=x}|E = e] = E[Y|E = e] + T(x - E[X|E = e])$.*

This provides an intuitive interpretation of counterfactuals in linear models: $E[Y_{X=x}|E = e]$ can be computed by first calculating the best estimate of Y conditioned on the evidence e, $E[Y|e]$, and then adding to it whatever change is expected in Y when X is shifted from its current best estimate, $E[X|E = e]$, to its hypothetical value, x.

Methodologically, the importance of Theorem 9 lies in enabling researchers to answer hypothetical questions about individuals (or set of individuals) from population data. The ramifications of this feature in legal contexts and political science are explored, respectively, in (Pearl, 2009, ch. 9) and Yamamoto (2012).

Example Problems

In this section, we apply graphical tools to solve non-trivial problems that SEM researchers are likely to encounter.

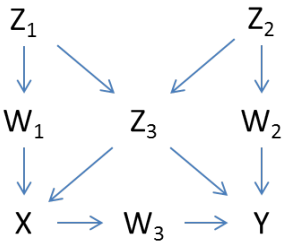


Figure 24. Graph corresponding to Model 7 in text

Model 7.

$$Y = aW_3 + bZ_3 + cW_2 + U \qquad X = t_1W_1 + t_2Z_3 + U'$$
$$W_3 = c_3X + U'_3 \qquad W_1 = a'_1Z_1 + U'_1$$
$$Z_3 = a_3Z_1 + b_3Z_2 + U_3 \qquad Z_1 = U_1$$
$$W_2 = c_2Z_2 + U'_2 \qquad Z_2 = U_2$$

Given the model depicted above, we pose the following questions:

(i) Identify three testable implications of this model

(ii) Identify a testable implication assuming that only X, Y, W_3, and Z_3 are observed

(iii) Suppose X, Y, and W_3 are the only variables observed. Which parameters can be identified from the data?

(iv) If we regress Z_1 on all other variables in the model, which regression coefficient will be zero?

(v) The model in Figure 24 implies that certain regression coefficients will remain invariant when an additional variable is added as a regressor. Identify five such coefficients with their added regressors.

Solutions:

(i) Figure 24 shows that $\{W_1, Z_3, W_2, W_3\}$ d-separates X and Y. Therefore, $\sigma_{XY.W_1Z_3W_2W_3} = 0$. Likewise, $\{W_1, Z_3\}$ blocks all paths between X and Z_1 and $\{Z_3, W_2\}$ blocks all paths between Y and Z_2. As a result, $\sigma_{XZ_1.W_1Z_3} = 0$ and $\sigma_{YZ_2.Z_3W_2} = 0$.

(ii) When X, Y, W_3, and Z_3 are latent variables, Model 7 is equivalent to the graph in Figure 25. We see that W_3 is d-separated from Z_3 by X. Therefore, $\sigma_{W_3Z_3.X} = 0$.

(iii) c_3 is identified using the single-door criterion. When we remove the edge $X \to W_3$, X is d-separated from W_3. Likewise, a can be identified using the single-door criterion. When we remove the edge $W_3 \to Y$, W_3 is d-separated from Y by X. Therefore, $c_3 = \beta_{W_3X}$ and $a = \beta_{YW_3.X}$.

(iv) The coefficients for X, W_3, W_2, and Y will be zero since they are d-separated from Z_1 by $\{W_1, Z_3, Z_2\}$. The coefficient for Z_2 may not be zero since Z_3 is a collider.

(v) (a) $\beta_{YX.W_1Z_3} = \beta_{YX.W_1,Z_3Z_1}$ since both $\{W_1, Z_3\}$ and $\{W_1, Z_3, Z_1\}$ satisfy the back-door criterion for the total effect of X on Y.

(b) $\beta_{YW_3.X} = \beta_{YW_3.XW_1}$ since $\{X\}$ and $\{X, W_1\}$ satisfy the back-door criterion for the total effect of W_3 on Y.

(c) $\beta_{Z_2Z_1} = \beta_{Z_2Z_1.W_1}$ since Z_2 is d-separated from Z_1 by $\emptyset$ and W_1. As a result, both regression coefficients vanish.

(d) $\beta_{YW_2.Z_2} = \beta_{YW_2.Z_2Z_3Z_1}$ since both $\{Z_2\}$ and $\{Z_2, Z_3, Z_1\}$ satisfy the back-door criterion for the total effect of W_2 on Y.

(e) $\beta_{W_1Z_1} = \beta_{W_1Z_1.Z_3}$ since both $\emptyset$ and $\{Z_3\}$ satisfy the back-door criterion for the total effect of Z_1 on W_1.

Conclusion

The benefit of graphs are usually attributed to their ability to represent theoretical assumptions visibly and transparently, by abstracting away unnecessary algebraic details. What is not generally recognized is graphs' ability to serve as efficient computational engines for tasks that would otherwise be intractable. This paper demonstrates how graphs can compute the testable implications of modeling assumptions, combine those assumption with data, and generate quantitative answers to both statistical and causal questions about populations and individuals.

We showed that a few basic principles of reading vanishing partial correlations from graphs can give rise to new methods of model testing and identification that substantially enrich traditional methods of SEM. The construction of equivalent models and characterization of instrumental variables follow directly from these principles. Auxiliary techniques of counterfactual analysis further permit researchers to quantify individual behavior from population data and to reason backward into alternative courses of action.

Graphical representations have become an indispensable second language in the health sciences (Glymour and Green-

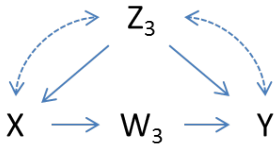


Figure 25. Graph representing Model 7 when Z_1, W_1, Z_2, and W_2 are unobserved

land, 2008; Lange et al., 2012) and are making their way towards the social and behavioral sciences (Chalak and White, 2011; Lee, 2012; Morgan and Winship, 2007). It is hoped that this survey unveils the potentials of these tools to quantitative methodologists engaged in psychometric research.

References

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'Metrics: The Path from Cause to Effect*. Princeton University Press.

Arbuckle, J. (2005). *Amos 6.0 user's guide*. Marketing Department, SPSS Incorporated.

Baron, R. and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* **51** 1173–1182.

Bentler, P. M. (1989). Eqs structural equations program manual .

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

Bollen, K. and Pearl, J. (2013). Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research* (S. Morgan, ed.). Springer, 301–328.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology* **38** 37–72.

Bowden, R. and Turkington, D. (1984). *Instrumental Variables*. Cambridge University Press, Cambridge, England.

Brito, C. (2004). *Graphical methods for identification in structural equation models*. Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.
URL {$<$http://ftp.cs.ucla.edu/pub/stat_ser/r314.pdf$>$}

Brito, C. and Pearl, J. (2002a). Generalized instrumental variables. In *Uncertainty in Artificial Intelligence, Proceedings of the Eighteenth Conference* (A. Darwiche and N. Friedman, eds.). Morgan Kaufmann, San Francisco, 85–93.

Brito, C. and Pearl, J. (2002b). A new identification condition for recursive models with correlated errors. *Journal Structural Equation Modeling* **9** 459–474.

Brito, C. and Pearl, J. (2006). Graphical condition for identification in recursive SEM. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Corvallis, OR, 47–54.

Cai, Z. and Kuroki, M. (2005). Counterfactual reasoning in linear structural equation models. In *Proceedings of the 21st conference on Uncertainty in Artificial Intelligence*. AUAI Press.

Cai, Z. and Kuroki, M. (2008). On identifying total effects in the presence of latent variables and selection bias. In *Uncertainty in Artificial Intelligence, Proceedings of the Twenty-Fourth Conference* (???, ed.). AUAI, Arlington, VA, 62–69.

Chalak, K. and White, H. (2011). Viewpoint: An extended class of instrumental variables for the estimation of causal effects. *Canadian Journal of Economics/Revue canadienne d'économique* **44** 1–51.

Chen, B. and Pearl, J. (2015). Exogeneity and robustness. Tech. Rep. R-449, <http://ftp.cs.ucla.edu/pub/stat_ser/r449.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Submitted.

Chen, B., Tian, J. and Pearl, J. (2014). Testable implications of linear structual equation models. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence* (C. E. Brodley and P. Stone, eds.). AAAI Press, Palo, CA. <http://ftp.cs.ucla.edu/pub/stat_ser/r428-reprint.pdf>.

Cooper, G. (1999). An overview of the representation and discovery of causal relationships using bayesian networks. *Glymour and Cooper [36]* 4–62.

Crámer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ.

Foygel, R., Draisma, J. and Drton, M. (2012). Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics* **40** 1682–1713.

Geiger, D. and Pearl, J. (1993). Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics* 2001–2021.

Glymour, M. and Greenland, S. (2008). Causal diagrams. In *Modern Epidemiology* (K. Rothman, S. Greenland and T. Lash, eds.), 3rd ed. Lippincott Williams & Wilkins, Philadelphia, PA, 183–209.

Hahn, J. (2004). Functional restriction and efficiency in causal inference. *Review of Economics and Statistics* **86** 73–76.

Hayduk, L., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., Gillespie, M., Pazderka-Robinson, H. and Boadu, K. (2003). Pearl's d-separation: One more step into causal thinking. *Structural Equation Modeling* **10** 289–311.

HECKMAN, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* **115** 45–97.

HERSHBERGER, S. L. and MARCOULIDES, G. A. (2006). The problem of equivalent structural models. In *Structural equation modeling: A second course* (G. R. Hancock and R. O. Mueller, eds.). Information Age Publishing, Greenwich, CT, 3–37.

HOLLAND, P. (1988). Causal inference, path analysis, and recursive structural equations models. In *Sociological Methodology* (C. Clogg, ed.). American Sociological Association, Washington, D.C., 449–484.

HOLLAND, P. W. (2001). *Stochastic causality*, chap. The causal interpretation of regression coefficients. CLSI Publications, Stanford, CA, 173–177.

HOOVER, K. D. and PHIROMSWAD, P. (2013). Selecting instrumental variables: a graph-theoretic approach. *Available at SSRN 2318552* .

HOYLE, R. H. (2012). *Handbook of structural equation modeling*. Guilford Press.

HUANG, Y. and VALTORTA, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 217–224.

IMBENS, G. (2014). Instrumental variables: An econometrician's perspective. Tech. rep., National Bureau of Economic Research.

JORESKOG, K. and SORBOM, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research* **XIX** 404–416.

JÖRESKOG, K. G. and SÖRBOM, D. (1989). *LISREL 7: A guide to the program and applications*, vol. 2. Spss Chicago.

JÖRESKOG, K. G., SÖRBOM, D., MAGIDSON, J. and COOLEY, W. W. (1979). *Advances in factor analysis and structural equation models*. Abt Books Cambridge, MA.

KANG, C. and TIAN, J. (2009). Markov properties for linear causal models with correlated errors. *The Journal of Machine Learning Research* **10** 41–70.

KENNY, D. A. and MILAN, S. (2012). Identification: A nontechnical discussion of a technical issue. In *Handbook of Structural Equation Modeling* (G. M. S. W. R. Hoyle, D. Kaplan, ed.). Guilford Press, New York, 145–163.

KLINE, R. (2011). *Principles and Practice of Structural Equation Modeling*. Methodology in the social sciences, Guilford Press.
URL `http://books.google.com/books?id=-MDPILyu3DAC`

KUROKI, M. and MIYAKAWA, M. (2003). Covariate selection for estimating the causal effect of control plans by using causal diagrams. *Journal of the Japanese Royal Statistical Society, Series B* **65** 209–222.

KUROKI, M. and PEARL, J. (2014). Measurement bias and effect restoration in causal inference. Tech. rep.

LANGE, T., VANSTEELANDT, S. and BEKAERT, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology* **176** 190–195.

LEE, J. J. (2012). Correlation and causation in the study of personality. *European Journal of Personality* **26** 372–390.

LEE, S. and HERSHBERGER, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research* **25** 313–334.

LEUNG, D., DRTON, M. and HARA, H. (2015). Identifiability of directed gaussian graphical models with one latent source. *arXiv preprint arXiv:1505.01583* .

LU, X. and WHITE, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics* **178** 194–206.

MACCALLUM, R. C. (1995). Model specification: Procedures, strategies, and related issues. In *Structural Equation Modeling* (R. Hoyle, ed.). Sage, Thousand Oaks, CA.

MARSCHAK, J. (1942). Studies in mathematical economics and econometrics – in memory of henry schultz. chap. Economic interdependence and statistical analysis. University of Chicago Press, Chicago, 135–150. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 427–439, 1995.

MCDONALD, R. (2002). What can we learn from the path equations?: Identifiability constraints, equivalence. *Psychometrika* **67** 225–249.

MORGAN, S. and WINSHIP, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press, New York, NY.

MULAIK, S. A. (2009). *Linear causal modeling with structural equations*. CRC Press.

MUTHÉN, L. K. and MUTHÉN, B. O. (2010). *Mplus: Statistical analysis with latent variables: User's guide*. Muthén & Muthén.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

PEARL, J. (2004). Robustness of causal claims. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press.

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

PEARL, J. (2012). The causal foundations of structural equation modeling. In *Handbook of Structural Equation Modeling* (R. Hoyle, ed.). Guilford Press, New York, 68–91.

PEARL, J. (2014a). The deductive approach to causal inference. *Journal of Causal Inference* **2** 115–129.

PEARL, J. (2014b). Interpretation and identification of causal mediation. *Psychological Methods* **19** 459–481.

PEARL, J. (2014c). Reply to commentary by Imai, Keele, Tingley, and Yamamoto concerning causal mediation analysis. *Psychological Methods* **19** 488–492.

PEARL, J. and PAZ, A. (2014). Confounding equivalence in causal inference. *Journal of Causal Inference J. Causal Infer.* **2** 75–93.

RUBIN, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.

SHIPLEY, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling* **7** 206–218.

SHPITSER, I. and PEARL, J. (2006). Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 437–444.

SHPITSER, I. and PEARL, J. (2008). Dormant independence. In *Proceedings of the Twenty-Third Conference on Artificial Intelligence*. AAAI Press, Menlo Park, CA, 1081–1087.

SHROUT, P., KEYES, K. and ORNSTEIN, K. (2010). *Causality and psychopathology: Finding the determinants of disorders and their cures*. Oxford University Press.

SPEARMAN, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology* **15** 201–292.

SPIRTES, P. (1995). Directed cyclic graphical representation of feedback. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (P.Besnard and S. Hanks, eds.). Morgan Kaufmann, San Mateo, CA, 491–498.

SPIRTES, P., GLYMOUR, C. N. and SCHEINES, R. (2000). *Causation, prediction, and search*, vol. 81. MIT press.

TIAN, J. (2002). *Studies in Causal Reasoning and Learning*. Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.

TIAN, J. (2005). Identifying direct causal effects in linear models. In *Proceedings of the National Conference on Artificial Intelligence*, vol. 20. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

TIAN, J. (2007). A criterion for parameter identification in structural equation models. In *Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07)*. AUAI Press, Corvallis, Oregon.

TIAN, J. (2009). Parameter identification in a class of linear structural equation models. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*.

TIAN, J. and PEARL, J. (2002a). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*. AAAI Press/The MIT Press, Menlo Park, CA, 567–573.

TIAN, J. and PEARL, J. (2002b). On the testable implications of causal models with hidden variables. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence* (A. Darwiche and N. Friedman, eds.). Morgan Kaufmann, San Francisco, CA, 519–527.

TOMARKEN, A. J. and WALLER, N. G. (2003). Potential problems with" well fitting" models. *Journal of abnormal psychology* **112** 578.

VERMA, T. and PEARL, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. Cambridge, MA. Also in P. Bonissone, M. Henrion, L.N. Kanal and J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, Elsevier Science Publishers, B.V., 255–268, 1991.

WILLIAMS, L. (2012). Equivalent models: Concepts, problems, alternatives. *Handbook of Structural Equation Modeling (R. Hoyle, ed.). Guilford Press, New York* 247–260.

WRIGHT, P. (1928). *The Tariff on Animal and Vegetable Oils*. The MacMillan Company, New York, NY.

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20** 557–585.

WRIGHT, S. (1925). Corn and hog correlations. Tech. Rep. 1300, U.S. Department of Agriculture.

YAMAMOTO, T. (2012). Understanding the past: Statistical analysis of causal attribution. *American Journal of Political Science* **56** 237–256.