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

Recent advances in graphical models have had a transformative impact on causal analysis and machine learning. 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). Only a meager portion of these developments have found their way to mainstream structural equation modeling (SEM) literature which, by and large, prefers algebraic over graphical representations (Joreskog and Sorbom, 1982; Bollen, 1989; Mulaik, 2009; Hoyle, 2012). The reason for this disparity rests primarily in the fact that graphical techniques were developed for non-parametric analysis, while most SEM research is conducted within the confines of Gaussian linear models, to which matrix algebra and powerful statistical tests are applicable.

The purpose of this paper is to introduce SEM researchers to modern tools of graphical models and to describe the benefits, as well as new insights that graphical models can provide. These include new methods of testing models, new ways of identifying structural parameters, new techniques of reducing confounding bias, and new paradigms for handling missing data.

To make this paper self contained, we will start with the basic definitions of regression analysis, linear structural equations models, path analysis, 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.

Preliminaries

Expected Value, Covariance, Regression, and Correlation

The expected value of a variable X, denoted $E[X]$, is defined as

$$E[X] = \int x \cdot P(x)dx \qquad (1)$$

and can be interpreted as the "weighted average" of X, where $P(x)$ stands for the probability density function of X.

The *variance* of X is defined as

$$\sigma_X^2 = E[X - E(X)]^2 \qquad (2)$$

and measures the degree to which X deviates from its mean $E[X]$.

The *covariance* of X and Y is defined as

$$\sigma_{XY} = E[(X - E[X])(Y - E[Y])] \qquad (3)$$

and measures the degree to which X and Y vary together.

The *covariance matrix* of a set of variables, $\{X_1, X_2, ..., X_n\}$, is the matrix, $[\sigma_{X_i X_j}]$, containing the covariance between each pair of variables and the variance of each variable along the diagonal.

The *correlation coefficient* of X and Y, ρ_{XY}, and *regression coefficient*, β_{YX} of Y on X are two other measures of association, which we define in terms of the covariance:

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **JUN 2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE **Graphical Tools for Linear Structural Equation Modeling**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Los Angeles,Department of Computer Sciences,Los Angeles,CA,90095-1596**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**23**	

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad (4)$$

$$\beta_{YX} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\sigma_Y}{\sigma_X} \rho_{XY} \qquad (5)$$

ρ_{XY} is a normalized measure of association and confined to the unit interval; $0 \leq \rho_{XY} \leq 1$. If the distribution is Gaussian (assumed for the remainder of the paper), the regression coefficient, β_{YX}, represents the slope of the least squares line in the prediction of Y given X: $\beta_{YX} = \frac{\partial}{\partial x} E[Y|X = x]$ for all x. Finally, notice that if the variables have been standardized (also assumed, without loss of generality, for the remainder of the paper) so that the mean and variance of each variable are equal to 0 and 1 respectively, we have $\sigma_{XY} = \rho_{XY} = \beta_{XY} = \beta_{YX}$.

Conditional Expectation, Partial Covariance, Partial Correlation, and Partial Regression

The conditional expectation of Y given $X = x$, denoted $E[Y|X = x]$, is the expected value of Y when $X = x$. More formally,

$$E[Y|X = x] = \int y \cdot P(y|X = x) dy \qquad (6)$$

We will also utilize partial covariances, correlations, and regressions, which measure the association between X and Y conditioned on a given set of variables, Z. For example, the *partial regression coefficient* of Y on X given $Z = z$ or *z-specific regression coefficient* of Y on X is given by:

$$\beta_{YX.z} = \frac{\partial}{\partial x} E[Y|X = x, Z = z] \qquad (7)$$

In words, $\beta_{YX.z}$ is the slope of the regression line of Y on X when we consider only cases for which $Z = z$. Since we are assuming a Gaussian distribution, $\beta_{YX.z}$, does not change for different values of z, which allows us to write $\beta_{YX.z} = \beta_{YX.Z}$, where $\beta_{YX.Z}$ is the coefficient for X when we regress Y on X and Z.

The partial correlation coefficient, $\rho_{XY.Z}$ can be defined by normalizing $\beta_{YX.Z}$:

$$\rho_{XY.Z} = \beta_{YX.Z} \frac{\sigma_{X.Z}}{\sigma_{Y.Z}} \qquad (8)$$

A well known result in regression analysis (Crámer, 1946) permits us to express $\rho_{YX.Z}$, $\sigma_{YX.Z}$, or $\beta_{YX.Z}$ recursively in terms of pair-wise correlation coefficients. When Z is a singleton, this reduction reads:

$$\rho_{YX.Z} = \frac{\rho_{YX} - \rho_{YZ}\rho_{XZ}}{[(1 - \rho_{YZ}^2)(1 - \rho_{XZ}^2)]^{\frac{1}{2}}} \qquad (9)$$

$$\sigma_{YX.Z} = \sigma_{YX} - \frac{\sigma_{YZ}\sigma_{ZX}}{\sigma_Z^2} \qquad (10)$$

$$\beta_{YX.Z} = \frac{\sigma_Y}{\sigma_X} \frac{\rho_{YX} - \rho_{YZ}\rho_{ZX}}{1 - \rho_{XZ}^2} \qquad (11)$$

If we wish to reduce $\rho_{YX.ZS}$, $\sigma_{YX.ZS}$, or $\beta_{YX.ZS}$ when Z is a singleton and S a set containing one or more variables, it can be done as follows:

$$\rho_{YX.ZS} = \frac{\rho_{YX.S} - \rho_{YZ.S}\rho_{XZ.S}}{[(1 - \rho_{YZ.S}^2)(1 - \rho_{XZ.S}^2)]^{\frac{1}{2}}} \qquad (12)$$

$$\sigma_{YX.ZS} = \sigma_{YX.S} - \frac{\sigma_{YZ.S}\sigma_{ZX.S}}{\sigma_{Z.S}^2} \qquad (13)$$

$$\beta_{YX.ZS} = \frac{\sigma_{Y.S}}{\sigma_{X.S}} \frac{\rho_{YX.S} - \rho_{YZ.S}\rho_{ZX.S}}{1 - \rho_{XZ.S}^2} \qquad (14)$$

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. When the conditioning set becomes large, this procedure can yield rather complicated expressions. However, if our aim is merely to identify vanishing partial correlations, which is the case in many applications, we can be spared the effort entailed by this recursion and use graphical criteria instead. These are reviewed in the subsection on d-separation.

Linear Structural Equation Models

Structural equation models (SEMs) use mathematical functions to describe the data generating mechanism for a set of variables. For example, the structural equation $Y = aX + U_Y$ describes a physical process by which Nature examines the values of X and U_Y and assigns the value $aX + U_Y$ to variable Y[1]. If the model specification accurately reflects the data generating mechanism, it is capable of answering all causally related queries regarding the model variables, including questions of prospective and introspective counterfactuals[2]. In this paper, we focus on linear SEMs.

Consider a set of observed variables, $y_1, y_2, ..., y_n$. A *linear SEM* consists of a set of equations of the form,

---

[1]The equal sign in structural equations is not symmetric. $Y = aX + U_Y$ does not imply the structural equation $X = \frac{1}{a}(Y - U_Y)$ because X may be assigned its value according to other variables in the model, not Y and U_Y.

[2]Prospective counterfactual queries are queries of the form, "What value would Y take if X were set to x?" Introspective counterfactual queries are queries of the form, "Given that Y currently takes the value of y, what would have been the value of Y if X had been x?" Counterfactuals will be discussed in more detail in the section on counterfactuals.
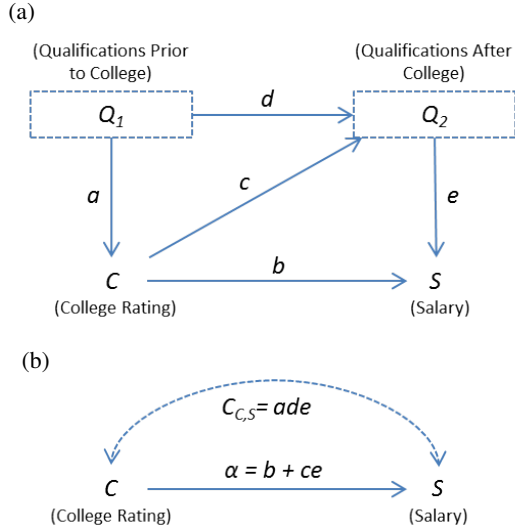
(a)

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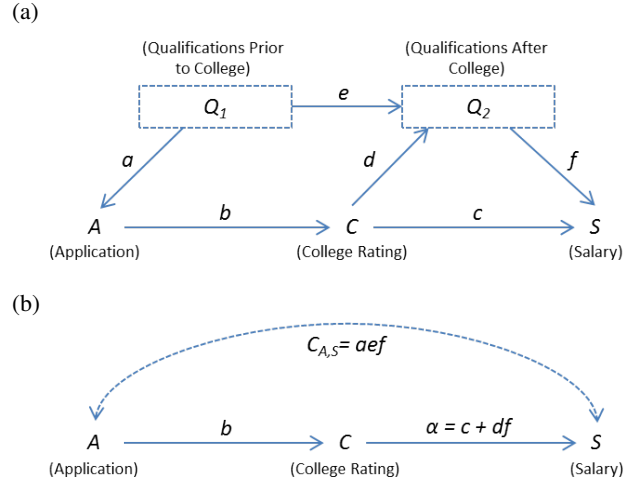
(a)

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

$$y_i = \Lambda_{1i}y_1 + \Lambda_{2i}y_2 + ... + \Lambda_{ni}y_n + u_i$$

where $y_1, ..., y_n$ are the model variables, Λ_{ji} is a coefficient that conveys the strength of the causal relationship from y_j to y_i, and u_i a random error term due to latent or omitted factors that is generally assumed to be normally distributed. If $\Lambda_{ji} = 0$, indicating zero direct influence of y_j on y_i then we usually omit the term $\Lambda_{ji}y_j$ from the equation. Note that the notion of direct effect depends on the set of variables we decide to include in the model, with the understanding that every coefficient Λ_{ji} from y_j to y_i may summarize a chain of many microprocesses whose precise nature remains outside the model. Throughout the paper we often use distinct letters (e.g. *a*, *b*, *c*) in place of Λ_{ji} for the coefficients. For example, in Model 1 shown below, the coefficient Λ_{CS} is given the label *b* while coefficient Λ_{Q_2S} is given the label *e*.

Typically, the modeler specifies the equations from domain knowledge and attempts to estimate the coefficients from data. For example, suppose we wish to estimate the effect of attending an elite college on future earnings (Figure 1a). Clearly, simply regressing earnings on college rating will not give an unbiased estimate of the target effect. Since elite colleges are highly selective, students attending them are likely to have qualifications for high-earning jobs prior to attending the school. This background knowledge is expressed in the following model specification:

Model 1.

$$Q_1 = U_1$$
$$C = a \cdot Q_1 + U_2$$
$$Q_2 = c \cdot C + d \cdot Q_1 + U_3$$
$$S = b \cdot C + e \cdot Q_2 + U_4,$$

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. Both Q_1 and Q_2 may be *latent variables*, meaning they are unobserved and, therefore, not present in the dataset. The path diagram for this model is depicted in Figure 1a.

In order to estimate the values of the coefficients *b*, *c*, and *e*, which convey the causal effect of attending an elite college on future earnings, the coefficients must have a unique solution in terms of the covariance matrix, $[\sigma_{ij}]$. 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 an unbiased estimate of these 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*, in the model, as shown in Figure 2a, we obtain the following structural equations:

Model 2.

$$Q_1 = U_1$$
$$A = a \cdot Q_1 + U_2$$
$$C = b \cdot A + U_3$$
$$Q_2 = e \cdot Q_1 + d \cdot C + U_4$$
$$S = c \cdot C + f \cdot Q_2 + U_5,$$

from which the causal effect of attending an elite college on future salary 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 1 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, $c + df$, is not identified using Model 1 but is identified using Model 2. Further, we will also be able to determine, again by inspection, that $c + df$ equals $\beta_{SC.A}$, and that the regression coefficient $\beta_{SA.CQ_1}$ vanishes, which can be used to test whether the specification of Model 2 is compatible with the data. 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.

Path Diagrams and Graphs

Path diagrams were introduced by Sewell Wright (1921), who aimed to estimate causal influences from statistical data on animal breeding. Today, SEM is generally implemented in software[3], and, as a result, when users experience unexpected behavior (due to unidentified parameters, for example) they are often at a loss as to how to proceed[4]. For the remainder of this section, we will review the basics of path diagrams and provide users with simple, intuitive tools to resolve questions of identification, goodness of fit, and more using graphical methods.

The path diagram or causal graph[5] of an SEM can be easily drawn from the equations in the model. The vertices represent model variables, and for each equation,

$$y_i = \Lambda_{1i}y_1 + \Lambda_{2i}y_2 + ... + \Lambda_{ni}y_n + u_i,$$

arrows are drawn from the variables in y_j to y_i whenever $\Lambda_{ji} \neq 0$. Each arrow, therefore, is associated with a coefficient in the SEM, which we will label as its path coefficient. The error terms, u_i, are usually not represented in the graph. However, a bidirected arc between two variables, y_i and y_j, indicates that their corresponding error terms, u_i and u_j, may be statistically dependent while the lack of a bidirected arc indicates that the error terms are independent. An *edge* is

defined to be either an arrow or a bidirected arc. Figure 1a is a path diagram for Model 1 while Figure 2a is a path diagram for Model 2.

If an arrow, called (X, Y), 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 to Y is a sequence of edges, connecting the two vertices. 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 graph is *acyclic* if it does not contain any cycles, a directed path that begins and ends with the same node. A graph is *cyclic* if it contains a cycle. A model in which the causal graph is acyclic is called *recursive* while models with cyclic graphs are called *non-recursive*.

For conciseness and clarity, latent variables will not be included in the model specification or depicted as nodes in the graph. Instead, their effect will be summarized by the correlation they induce on error variables as represented in the diagram by bidirected arcs. It was shown by Verma (1993) that one can always summarize a sequence of interconnected latent variables in the form of pairwise correlations. For example, the effect of the latent variables in Figures 1a and 2a are summarized by Figures 1b and 2b respectively. 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 $C \leftarrow Q_1 \rightarrow Q_2 \rightarrow S$.

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 path coefficients along path π_i. The *total effect* of X on Y is often defined as the $\sum_i p_i$ (Bollen, 1989). For example, in Figure 3a, the total effect of X on Y is $a \cdot b + e$.

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

---

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

[4]Kenny and Milan (2011) 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."
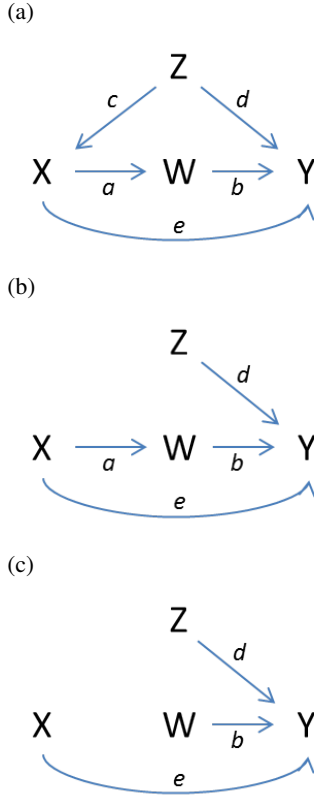
[5]We use both terms interchangeably.

(a)

(b)

(c)

Figure 3. Models depicting interventions (a) Original model (b) After intervening on X (c) After intervening on X, Z, and W

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 variation 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 causal effect of X on Y is defined as

$$E[Y|do(X = x + 1)] - E[Y|do(X = x)], \qquad (15)$$

where x is some reference point. 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 (15), with the derivative,

$$\frac{\partial}{\partial x} E[Y|do(X = x)]. \qquad (16)$$

Consider again Models 1 and 2 with C a binary variable taking value 1 for elite colleges and 0 for non-elite colleges. For defining 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 aver-

age 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 can be seen using Figure 3a and its corresponding structural equations:

Model 3.

$$
\begin{aligned}
Z &= U_Z \\
X &= cZ + U_X \\
W &= aX + U_W \\
Y &= dZ + bW + eX + U_Y
\end{aligned}
$$

Using the *do* operation we obtain the new set of equations:

$$
\begin{aligned}
Z &= U_Z \\
X &= x \\
W &= aX + U_W \\
Y &= dZ + bW + eX + U_Y
\end{aligned}
$$

The corresponding path diagram is displayed in Figure 3b. (Notice that paths between X and Y due to common causes, Z, have been cut, and as a result, all paths between X and Y now reflect the causal effect from X to Y only.)

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

$$
\begin{aligned}
E[Y|do(X = x)] &= E[d \cdot Z + b \cdot W + e \cdot X + U_Y] \\
&= d \cdot E[Z] + b \cdot E[W] + e \cdot x + E[U_Y] \\
&= d \cdot E[U_Z] + b \cdot E[a \cdot x + U_W] + e \cdot x + E[U_Y] \\
&= 0 + b \cdot a \cdot x + b \cdot E[U_W] + e \cdot x + 0 \\
&= b \cdot a \cdot x + 0 + e \cdot x + 0 \\
&= b \cdot a \cdot x + e \cdot x.
\end{aligned}
$$

As a result,

$$E[Y|do(X = x + 1)] - E[Y|do(X = x)] = ab + e \qquad (17)$$

for all x, aligning the two definitions. Note, moreover, that this equality holds even when the coefficients, a, b, and e, are not identified (e.g. if the U terms are correlated).

In many cases, we may be interested in the *direct effect* of X on Y. The term "direct effect" is meant to quantify an
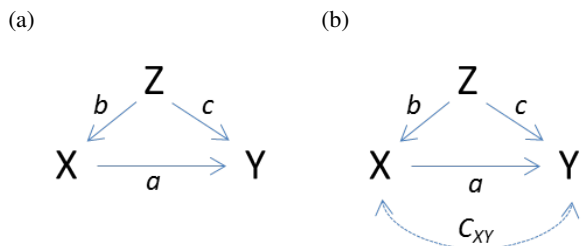
(a)   (b)

Figure 4. Models illustrating Wright's rules of path tracing

Figure 5. Model illustrating the rules of d-separation

effect that is not mediated by other variables in the model or, more accurately, the sensitivity of Y to changes in X while all other factors in the analysis are held fixed (Pearl, 2000).

"Holding all other factors fixed" can be simulated by assigning all variables other than X and Y by intervention to arbitrary set of reference values[6]. (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 Y. As a result, all links from X to Y other than the direct link will be severed. For example, Figure 3c shows the path diagram of Model 3 after intervention on all variables other than X and Y.

Now, the direct effect of X on Y can be defined as

$$E[Y|do(X = x + 1, S = s)] - E[Y|do(X = x, S = s)],$$

where S is a set containing all model variables other than X and Y and $\{x \cup s\}$ a set of reference values. It is not hard to show that, in linear models, the direct effect is equal to the path coefficient from X to Y. It provides, in fact, a formal interventional justification for associating the path coefficient with the notion of "direct effect", and permits us to extend this notion to non-linear models.

Finally, in linear models, the effect of X on Y *mediated* by W is equal to the sum of the product of coefficients associated with directed paths from X to Y that go through W. In Figure 3a, we see that this effect is equal to ab. For a non-linear and non-parametric extension of this definition, see *indirect effect* in Pearl (2000).

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 σ_{YX} of any two variables in a recursive model can be expressed as a polynomial over a subset of the model coefficients. In this way, Wright's equations characterize the relationship between the model coefficients and the covariance matrix and, subsequently, provide an algebraic representation of the identification problem. A coefficient is identified if and only if it can be solved in terms of the elements of the covariance matrix using Wright's equations[7].
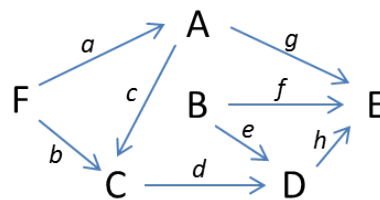
Wright's method consists of equating the standardized covariance $\sigma_{YX} = \rho_{YX}$ between any pair of variables to the sum of products of path coefficients and error covariances along certain paths between X and Y. Let $\Pi = \{\pi_1, \pi_2, ..., \pi_k\}$ denote the paths between X and Y that do not trace colliding arrowheads, i.e. $\rightarrow\leftarrow$, $\leftrightarrow\leftarrow$, $\rightarrow\leftrightarrow$, or $\leftrightarrow\leftrightarrow$, and let p_i be the product of path coefficients along path π_i. (We call nodes where colliding arrowheads meet a *collider*, e.g. Z in Figure 6a and C in Figure 5.) Then the covariance between variables X and Y is $\sum_i p_i$. For example, if we wish to express the covariance of C and E in Figure 5, we sum the product of the coefficients along paths $C \leftarrow F \rightarrow A \rightarrow E, C \leftarrow A \rightarrow E$, and $C \rightarrow D \rightarrow E$, giving $\sigma_{CE} = b \cdot a \cdot g + c \cdot g + d \cdot h$. However, we do not include the coefficients along $C \rightarrow D \leftarrow B \rightarrow E$ because it traces a collider. For models represented by the diagram in Figure 4a, we have $\sigma_{YX} = a + bc$ while $\sigma_{YX} = a + bc + C_{YX}$ for the diagram in Figure 4b.

To express partial covariances, correlations, or regression coefficients in terms of path coefficients we first apply Equations 9-14 and then use Wright's tracings rules for each covariance term. For example, reducing $\beta_{YX.Z}$ for the model represented by Figure 4a can be done as follows:

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

$$= \frac{1}{1} \frac{(a + bc) - (c + ab)(b)}{1 - b^2}$$

$$= \frac{a + bc - bc - ab^2}{1 - b^2}$$

$$= \frac{a(1 - b^2)}{1 - b^2}$$

$$= a$$

D-Separation

As mentioned previously, when the conditioning set becomes large, applying the recursive formula of Equations

---

[6]In footnote 12 we give an example demonstrating that "holding all other factors fixed" cannot be simulated using conditioning but instead must invoke intervention.

[7]However, as Wright's equations are non-linear, it can be very difficult to analyze the identification of large models by studying solutions for the system of equations.
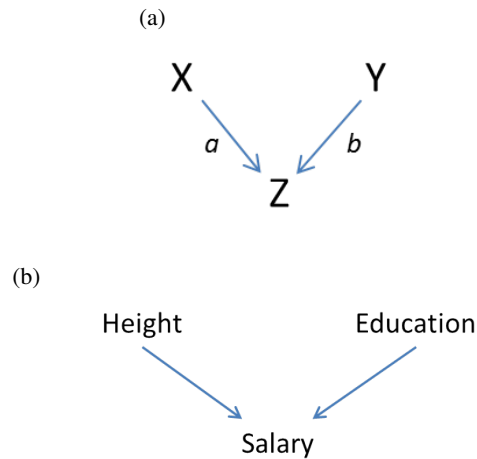
(a)

(b)

Figure 6. Examples illustrating conditioning on a collider

Figure 7. Diagram illustrating why Ice Cream Sales and Drowning are uncorrelated given Temperature and/or Water Activities

12-14 can become complex. Vanishing partial correlations, however, can be readily identified from the path diagram using a criterion called *d-separation* (Pearl, 1988)[8].

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", given that we are dealing with a system of directed arrows in which some vertices (those residing in the conditioning set, Z) correspond to variables, whose values are measured precisely. To account for the orientations of the arrows we use the terms "d-separated" and "d-connected" (d denotes "directional").

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, ρ_{XY} vanishes 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
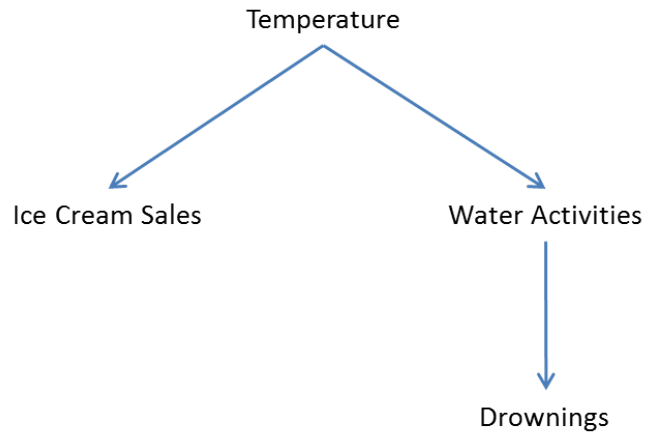
between X and Y is "blocked" by Z.

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 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 7. 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 6a. 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 6b. 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

---

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

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 6a where Z is a common effect of X and Y. We have $\sigma_{YX} = 0$ and, using Equation 10,

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

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 5 the path $F \rightarrow C \leftarrow A \rightarrow E$ is unblocked given C or D. However, it is blocked given $\{A, C\}$ or $\{A, D\}$.

The above three rules characterize d-separation 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 5 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 $C \rightarrow D \leftarrow B$. Finally C is d-separated from E by conditioning on A, D, and B.

To illustrate the power and applicability of d-separation, we pose and answer two questions regarding regression on the model depicted in Figure 5.

Figure 8. The fact that M d-separates X from Y implies $\beta_{YX.M} = 0$

(i) Suppose we regress B on all other variables,

$$B = \beta_E \cdot E + \beta_D \cdot D + \beta_A \cdot A + \beta_C \cdot C + \beta_F \cdot F + \epsilon_B,$$

which regression coefficients will be 0?

We might naively expect that all regression coefficients associated with variables that are not connected to B, β_A, β_C , and β_F, will vanish. However, D and E are colliders so regressing on them opens the path to A and C. Therefore, the only vanishing regression coefficient is β_F.

(ii) Suppose we regress E on A and B. Which variable can be added to the regression without changing the coefficient of B?

Since F and C are d-separated from B given A they can be added to the regression without changing the coefficient of B. However, adding D is liable to change the regression coefficient of B. Equivalent regressor sets will be discussed in more detail in a later section.

Some SEM researchers regard the resilience and stability of regression coefficients to additional regressors to be a sign of robustness[9]. As can be seen from the above example, sensitivity to adding regressors has little to do with misspecification; whether or not a regression coefficient changes when regressors are added is dependent on the structure of the data-generating model.

To emphasize this point, we demonstrate an extreme case of sensitivity in a well-specified model. Those familiar with the concept of *exogeneity* will recognize that, in Figure 8, X is uncorrelated with the error term of Y. As a result, simply

---

[9]According to Lu and White (2014), "A common exercise in empirical studies is a 'robustness check,' where the researcher examines how certain 'core' regression coefficient estimates behave when the regression specification is modified by adding or removing regressors." "Of the 98 papers published in *The American Economic Review* during 2009, 76 involve some data analysis. Of these, 23 perform a robustness check along the lines just described, using a variety of estimators." In a more recent survey of non-experimental empirical work, Oster (2013) finds that 75% of 2012 papers published in *The American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics* explored the sensitivity of results to varying control sets. Since this practice is conducted to help diagnose misspecfication, the answer to Question 5 is essential for discerning whether an altered coefficient indicates misspecification or not.

regressing Y on X will give an unbiased estimate of the total effect of X on Y. However, if we add the mediator, M, to the regression then the coefficient for X, $\beta_{YX.M}$, vanishes.

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 recursive and non-recursive linear models (Spirtes, 1995). Further, all vanishing partial correlations implied by a SEM can be obtained using d-separation (Pearl, 2000). Finally, in models with independent error terms, these vanishing partial correlations represent all of the model's testable implications (Pearl, 2000).

Identification

A model parameter is *identified* if it is uniquely determined from the covariance matrix. 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]. For example, consider the model represented by Figure 4a. Using Wright's equations we obtain the following equalities:

$$\sigma_{XY} = a + bc \tag{18}$$
$$\sigma_{ZX} = b \tag{19}$$
$$\sigma_{ZY} = c + ba \tag{20}$$

As a result, b is uniquely identified with $b = \sigma_{ZX}$. Now, substituting σ_{ZX} for b into the other two equations we obtain:

$$\sigma_{XY} = a + \sigma_{ZX}c \tag{21}$$
$$\sigma_{ZY} = c + \sigma_{ZX}a \tag{22}$$

Both a and c are identified with $a = \sigma_{XY} - \sigma_{ZX}\sigma_{ZY}$ and $c = \sigma_{ZY} - \sigma_{ZX}\sigma_{XY}$. We see that all model parameters have unique solutions in terms of the covariance matrix, and hence the model is identified.

In contrast, the model depicted in Figure 9 is not identified. Using Wright's rules we obtain a single equation: $a + C_{XY} = \sigma_{XY}$. Since there are infinite values for a and C_{XY} that satisfy this equation, neither parameter is identified and the model is not identified.

Many SEM researchers identify structural models by submitting the specification and data to an SEM program, which attempts to minimize a fitting function. If there is difficulty in this computation, the program warns that the model may not be identified. While convenient, there are disadvantages
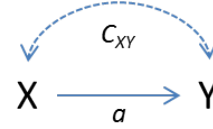

Figure 9. A diagram representing an unidentified model

to using typical SEM software to determine model identifiability. Kenny and Milan (2011) list the following drawbacks:

(i) If poor starting values are chosen, the program could mistakenly conclude the model is not identified when in fact it may be identified.

(ii) The program is not very helpful in indicating which parameters are not identified.

(iii) Most importantly, the program only gives an answer after the researcher has taken the time to collect data.

We add two additional drawbacks to this list:

(iv) If poor starting values are chosen, the program may exit with parameter values at a local minimum of the fitting function rather than the global minimum, giving incorrect values to the parameters.

(v) If even one coefficient is not identifiable, most software[11] are unable to identify any of the path coefficients.

In this section, we give graphical criteria that allows the modeler to determine the identifiability of individual parameters from mere inspection of the path diagram. Further, our criteria also give the values of the identified parameters in terms of the entries of the covariance matrix. 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 well known recursive and null rules (Bollen, 1989) and the regression rule (Kenny and Milan, 2011).

A Simple Criterion for Identifying Individual Coefficients

In Figure 4a, Z is a common cause of both X and Y and is often called a *confounder*. In epidemiology and other areas, it is well known that, in order to estimate α using regression, we must "adjust for" Z by including it in the regression equation. However, common causes are not always available

---

[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]According to Kenny and Milan (2011), AMOS is the only program that attempts to identify parameters when the model is under-identified.
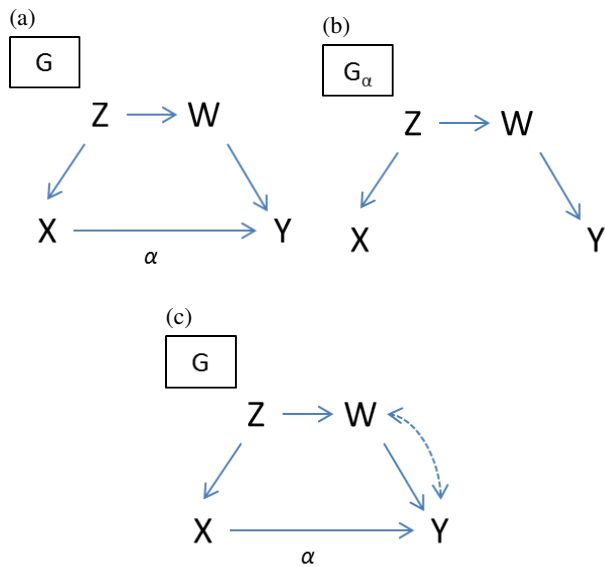
Figure 10. Diagrams illustrating identification by the single-door criterion (a) α is identified by adjusting for Z or W (b) The graph G_α used in the identification of α (c) α is identified by adjusting for Z (or Z and W) but not W alone

for measurement or adjustment. Instead, proxies are often used. For instance, in Figure 10a, W might well serve to replace Z in the adjustment and, thus, *deconfound* the relationship between X and Y. The question arises, how can we, in general, determine whether a set of variables is adequate for adjustment when attempting to identify a given coefficient? In other words, when would the regression coefficient of X in the regression of Y on X and Z be equal to the path coefficient from X to Y? The following criterion, called *single-door*, allows the modeler to answer this question by inspection of the path diagram.

Theorem 2. *(Pearl, 2000) (Single-door Criterion) Let G be any recursive causal graph in which α is the path coefficient associated with link $X \to Y$, 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 10a, we see that Z blocks the spurious path $X \leftarrow Z \to W \to Y$ and X is d-separated from Y by Z in Figure 10b. Therefore, $\alpha = \beta_{YX.Z}$. This is to be expected since Z is a common cause of X and Y. Theorem 2 tells us, however, that W can also be used for adjustment since W also d-separates X from Y in Figure 10b, and we obtain $\alpha = \beta_{YX.W}$. Moreover, we will see in a subsequent section
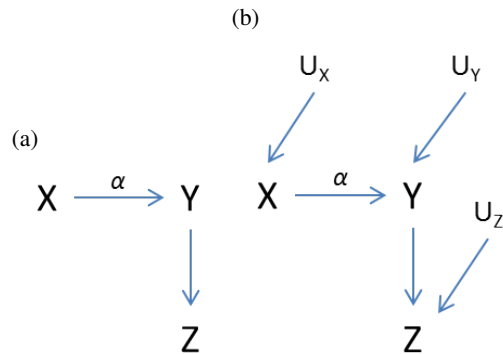

Figure 11. Example showing that adjusting for a descendant of Y induces bias in the estimation of α

that the choice of W is superior to that of Z in terms of estimation power. Consider, however, Figure 10c. 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[12]. In conclusion, α is equal to $\beta_{YX.Z}$ in Figures 10a and 10c. However, α is equal to $\beta_{YX.W}$ in Figure 10a only.

It is well known that estimating α using regression requires that X be uncorrelated with the error term of Y. Notice that the single-door criterion gives the graphical conditions for when conditioning on a set Z renders X and the error term of Y uncorrelated. Whenever X is d-separated from Y when the edge $X \to Y$ is removed then X must also be d-separated from the error term of Y in the original graph since Y acts as a collider. For example, in Figure 11b, X is d-separated from U_Y and X is d-separated from Y when $X \to Y$ is removed.

The intuition for the requirement that Z not be a descendant of Y is depicted in Figures 11a and 11b. We typically do not display the error terms, which can be understood as latent causes. In Figure 11b, 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.

No matter how complex the model, the given single-door criterion gives us a quick and reliable criterion, sufficient for identification of a structural parameter using regression. It allows us to choose a variety of conditioning sets 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.
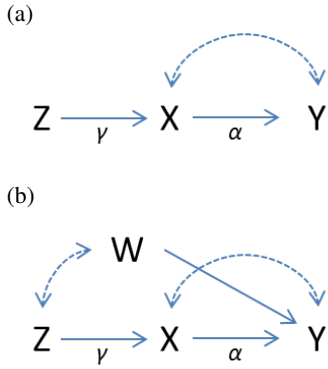
Figure 12. (a) Z qualifies as an instrumental variable (b) Z is a conditional instrument (given W)
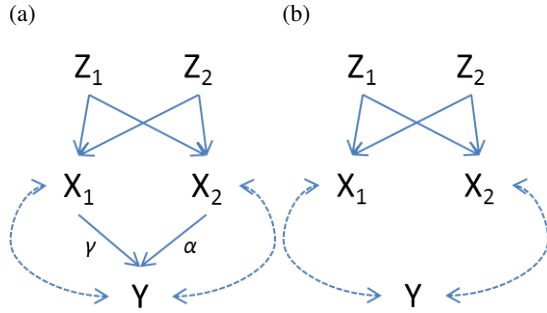

Figure 13. Diagrams illustrating instrumental sets

Instrumental Variables

It is well known that an *instrumental variable*, Z, can be used to identify a path coefficient, α, even when there is unobserved confounding between X and Y, as in Figure 12a. The following is a standard definition of an instrumental variable, adapted from Wikipedia (2014):

Definition 1. *A variable Z qualifies as an instrumental variable for a coefficient from X to Y if*

(i) Z is correlated with X

(ii) Z is uncorrelated with the error term of Y

Explanation typically stops here and the modeler has to determine judgmentally whether condition (ii) is satisfied in a complex model containing multiple equations and error terms. The major obstacles in such judgment are:

(i) To understand what is meant by the "the error term" of Y within a system of many equations.

(ii) To judge whether Z is uncorrelated with that "error term" within a system of correlated errors.

As a result, instrumental variables are often incorrectly determined. In this section, we show how to utilize the path diagram to determine by inspection whether a variable is an instrument. Bollen and Bauer (2004) called such variables "model-implied" instruments and used algebraic methods for their identification. Kyono (2010), on the other hand, used graphical methods similar to those described here[13]. Finally, we also show how to find *conditional instrumental variables* and *instrumental sets* (Brito and Pearl, 2002a), allowing us to identify α in Figure 12b and both γ and α in Figure 13a.

Theorem 3. *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 $X \rightarrow Y$ from G and

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

When Z is an instrument for α then $\alpha = \frac{\beta_{ZY}}{\beta_{ZX}}$.

Z in Figure 12a is an example of an instrumental variable since Z is d-separated from Y but still d-connected to X when we remove the edge associated with α. Using Wright's path tracing rules, it is easy to see that $\frac{\rho_{ZY}}{\rho_{ZX}} = \frac{\gamma\alpha}{\gamma} = \alpha$.

Theorem 3 clarifies the circumstances for which Z qualifies as an instrument. Even if the underlying structure is not clear to the modeler, he or she is able to consider competing explanations for the data generating process and determine immediately which of them qualifies Z as an instrument.

In Figure 12b, Z is not an instrument because it is d-connected to Y even after deleting the edge from X to Y. However, we can condition on W to block the spurious path from Z to Y through W and obtain $\alpha = \frac{\beta_{ZY.W}}{\beta_{ZX.W}}$. Thus, we see that in some cases, variables may become instrumental variables by conditioning on other variables.

Theorem 4. *(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

When Z is a conditional instrument for α given W then $\alpha = \frac{\beta_{ZY.W}}{\beta_{ZX.W}}$.

---

[12] It is for this reason that the direct effect cannot be defined by conditioning on a mediator but must instead invoke intervention (Pearl, 2013, 2014b), as we did earlier.

[13] Kyono (2010) also released software that implements d-separation, the single-door criterion, graphical techniques for identifying instruments, and more. Much of this functionality is also implemented in DAGitty (Textor et al., 2011), which includes a convenient GUI for manipulating graphs.

Additionally, it may be possible to use multiple variables as an *instrumental set* in order to identify parameters when, individually, none of the variables qualify as an instrument. In Figure 13a, neither Z_1 nor Z_2 are conditional 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:

$$\sigma_{Z_1 Y} = \sigma_{Z_1 X_1} \gamma + \sigma_{Z_1 X_2} \alpha$$
$$\sigma_{Z_2 Y} = \sigma_{Z_2 X_1} \gamma + \sigma_{Z_2 X_2} \alpha$$

As a result, we are able to obtain two linearly independent equations with two unknowns and solve for γ and α. We call a set of variables that enables a solution in this manner an *instrumental set*.

Definition 2. $\{Z_1, Z_2, ..., Z_n\}$ *is an* instrumental set *for the coefficients* $\alpha_1, ..., \alpha_n$ *associated with edges* $X_1 \rightarrow Y, ..., X_n \rightarrow Y$ *if the following conditions are satisfied.*

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

(ii) *There exists paths* $p_1, p_2, ..., p_n$ *such that* p_i *is a path from* Z_i *to Y that includes edge* $X_i \rightarrow Y$ *and if paths* p_i *and* p_j *have a common variable V, then either*

    (a) *both* $p_i[Z_i...V]$ *and* $p_j[V...Y]$ *point to V or*

    (b) *both* $p_j[Z_j...V]$ *and* $p_i[V...Y]$ *point to V.*

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

The second condition in Definition 2 can be understood as requiring that two paths p_i and p_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. This condition is illustrated in the example below.

Theorem 5. *Let* $\{Z_1, Z_2, ..., Z_n\}$ *be an instrumental set for the coefficients* $\alpha_1, ..., \alpha_n$ *associated with edges*

$$X_1 \rightarrow Y, ..., X_n \rightarrow Y.$$

Then the linear equations,

$$\sigma_{Z_1 Y} = \sigma_{Z_1 X_1} \alpha_1 + \sigma_{Z_1 X_2} \alpha_2 + ... + \sigma_{Z_1 X_n} \alpha_n$$
$$\sigma_{Z_2 Y} = \sigma_{Z_2 X_1} \alpha_1 + \sigma_{Z_2 X_2} \alpha_2 + ... + \sigma_{Z_2 X_n} \alpha_n$$
$$\vdots$$
$$\sigma_{Z_n Y} = \sigma_{Z_n X_1} \alpha_1 + \sigma_{Z_n X_2} \alpha_2 + ... + \sigma_{Z_n X_n} \alpha_n,$$

are linearly independent for almost all parameterizations of the model.

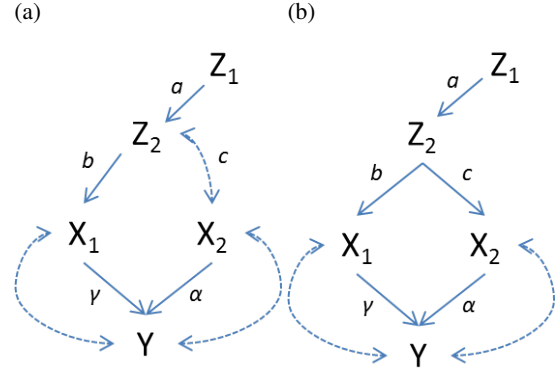*Figure 14.* (a) Z_1 and Z_2 qualify as an instrumental set (b) Z_1 and Z_2 do not qualify as an instrumental set

Like Z_1 and Z_2 in Figure 13a, Z_1 and Z_2 in Figure 14a qualify as an instrumental set. Z_1 and Z_2 are d-separated from Y in the graph $\bar{G}$, where the edges $X_1 \rightarrow Y$ and $X_2 \rightarrow Y$ have been removed. Additionally, we have $p_1 = Z_1 \rightarrow Z_2 \rightarrow X_1 \rightarrow Y$ and $p_2 = Z_2 \leftrightarrow X_2 \rightarrow Y$. Using Wright's rules we obtain

$$\sigma_{Z_1 Y} = ab\gamma = \sigma_{Z_1 X_1} \gamma + 0 \cdot \alpha = \sigma_{Z_1 X_1} \gamma + \sigma_{Z_1 X_2} \alpha \text{ and}$$
$$\sigma_{Z_2 Y} = b\gamma + c\alpha = \sigma_{Z_2 X_1} \gamma + \sigma_{Z_2 X_2} \alpha,$$

in accordance with Theorem 5. Solving the equations identifies α and γ giving:

$$\gamma = \frac{\sigma_{Z_1 Y}}{\sigma_{Z_1 X_1}}$$
$$\alpha = \frac{\sigma_{Z_2 Y}}{\sigma_{Z X_2}} - \frac{\sigma_{Z_2 X_1} \sigma_{Z_1 Y}}{\sigma_{Z_2 X_2} \sigma_{Z_1 X_1}}$$

Notice that p_1 and p_2 satisfy the second condition of Definition 2 because in p_1, the arrow associated with coefficient, a, *entering* the shared node, Z_2, is pointing at Z_2 while in p_2, the arrow associated with parameter, c, *leaving* Z_2 is also pointing at the shared node, Z_2. As a result, if the paths p_1 and p_2 are broken at the common variable, Z_2, and their pieces swapped and rearranged, p_1 will become a blocked path due to the collider at Z_2. Algebraically, this means that $\sigma_{Z_1 Y}$ lacks the influence of the path $Z_2 \leftrightarrow X_2 \rightarrow Y$ and, therefore, does not contain the term $ac\alpha$. $\sigma_{Z_2 Y}$, on the other hand, contains the term $c\alpha$ associated with the path. It is in this way that condition (ii) of Definition 2 allows p_i and p_j to share a node while still ensuring linear independence.

In contrast, consider Figure 14b. 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 $\sigma_{Z_1 Y}$ is not linearly independent of $\sigma_{Z_1 Y}$ with respect to Y's coefficients:

$$\sigma_{Z_1 Y} = b\gamma + c\alpha$$
$$\sigma_{Z_2 Y} = ab\gamma + ac\alpha = a(b\gamma + c\alpha) = a\sigma_{Z_1 Y}$$

In some cases, condition (i) of Definition 2 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 14b consists of three c-components, $\{X_1, X_2, Y\}$, $\{Z_2\}$, and $\{Z_1\}$, while the graph depicted in Figure 16 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 15a consists of the following equations:

$$Z = U_Z$$
$$X = aX + U_X$$
$$W = bW + U_W$$
$$V = U_V$$
$$Y = \alpha X + dV + U_Y$$

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 15a using any of the methods considered thus far.
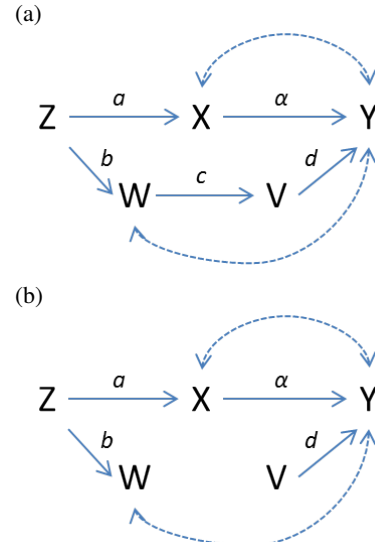

Figure 15. (a) Example illustrating c-component decomposition (b) Sub-graph consisting of c-component, $\{W, X, Y\}$, and its parents, *Z* and *V*.

However, the sub-graph for the c-component, $\{W, X, Y\}$, depicted in Figure 15b, 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 15b shows that $\alpha = \frac{\beta_{ZY}}{\beta_{ZX}}$. However, this is clearly not the case in Figure 15a. 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.

Advanced 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 a recursive SEM.
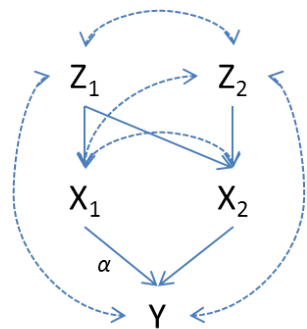
Figure 16. A bow-free graph; the absence of a 'bow' pattern assures identification

This algorithm was generalized by Foygel et al. (2012) to determine identifiability of a greater set of graphs[14]. Additionally their criterion, called the half-trek criterion, applies to both recursive and non-recursive models. The half-trek algorithm was further generalized by Chen et al. (2014) to identify more coefficients in under-identified 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, 2000) and non-parametric algorithms for identifying causal effects (Tian and Pearl, 2003; 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.

A Simple Criterion for Model Identification

In order to determine identifiability of the model using the single-door criterion or instrumental variables, the modeler must check the identifiability of each path 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 a recursive 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 16 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 4b, the variables X and Y create a bow-arc.

Theorem 6. *(Brito and Pearl, 2002b) (Bow-free Rule) Every recursive model whose path diagram lacks bow-arcs is identified.*

The bow-free rule is able to identify models that the single-door criterion is not. In Figure 16, 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 16 does not contain any bow-arcs it is identified according to Theorem 6. Finally, since the single-door criterion is unable to identify any model that contain bow-arcs[15], 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.)

Total Effects

When the model is not identifiable, modelers typically consider research with SEMs "impossible" (Kenny and Milan, 2011) 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 (2011)), it is often possible to identify some of the model coefficients even when the model as a whole is not identifiable. Further, we show in this section that it is often not necessary to identify all coefficients or even coefficients along a causal path in order to identify the causal effect of interest[16]. For example, in Figure 17a, the total effect of X on Y, $\frac{\partial}{\partial x}E[Y|do(X = x)]$, is identified and equal to β_{YX} even though it is unclear how to identify b, d, or e. The back-door criterion, given below, is a sufficient condition for the identification of a total effect.

Theorem 7. *(Pearl, 2000) (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 measurements Z such that*

---

[14]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.

[15]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.

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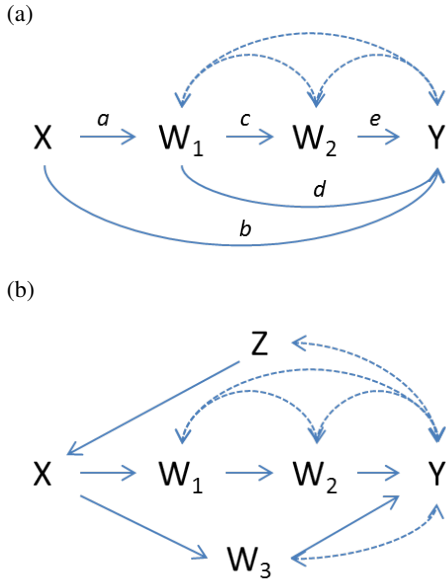
(a)

(b)

Figure 17. Unidentified graphs for which the total effect of *X* on *Y* is identified

(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 17a we see that the total of effect of *X* on *Y*, $\frac{\partial}{\partial x}E[Y|do(X = x)]$, is β_{YX} while the total effect of *X* on *Y* in Figure 17b is $\beta_{YX.Z}$.

Do-calculus (Pearl, 2000) and the aforementioned non-parametric algorithms (Tian and Pearl, 2003; Shpitser and Pearl, 2006; Huang and Valtorta, 2006) can also be used to identify total effects in linear models.

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 accurately reflects the data generating mechanism. 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, 1997). 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. Finally, if the model is very large and complex, it is possible that a global chi-square test
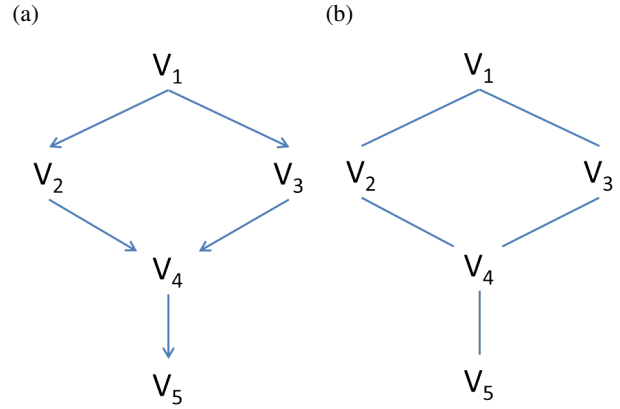
(a) (b)

Figure 18. (a) Example illustrating vanishing partial correlation (b) The skeleton of the model in (a)

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 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. Path diagrams allow modelers to identify vanishing partial correlations by inspection, provide a necessary and sufficient condition for equivalence among recursive models with uncorrelated error terms (often called *Markovian* models), and permit us to predict new types of constraints, beyond the vanishing correlation variety.

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 (Pearl, 2000). For the example depicted in Figure 18a, 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 recursive models with or without correlated error terms.

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 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 18a is Figure 18b.

Theorem 8. *(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 8 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, 2000).

The graphs in Figures 19b, 19c, and 19d are equivalent because they are all compatible with the graph in Figure 19a, which displays the skeleton and v-structures. Note 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 8 is necessary and sufficient for equivalence between Markovian models. It is a necessary condition for equivalence between non-recursive models and models with correlated error terms since d-separation in the graph implies vanishing partial correlation in the covariance matrix. (Models that are either non-recursive or have correlated error terms are called *non-Markovian*. Non-Markovian models that are recursive are called *semi-Markovian*.) In contrast, the more prevalent replacement criterion (Lee and Hershberger, 1990) is not always valid[17]. Pearl (2012) gave the following example depicted in Figure 20. 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
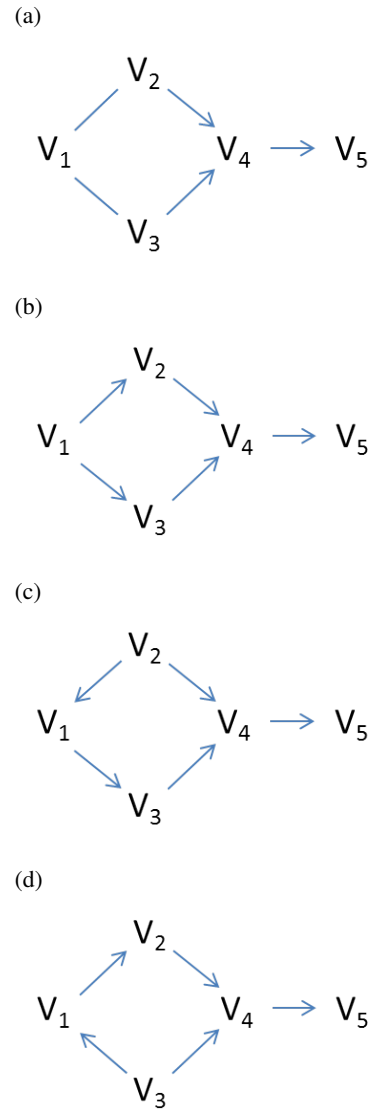
(a)

(b)

(c)

(d)

Figure 19. Models (b), (c), and (d) are equivalent. The arrows in (a) cannot be reversed.
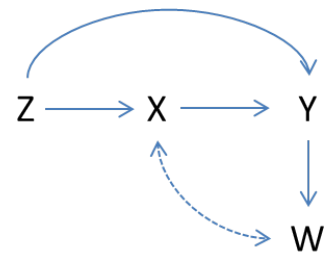
Figure 20. Counterexample to the standard Replacement Rule; The arrow $X \rightarrow Y$ cannot be replaced.

---

[17]The replacement rule violates the transitivity of equivalence (Hershberger, 2006), yet it is still used in most of the SEM literature (Mulaik, 2009; Williams, 2011, pp. 247-260).
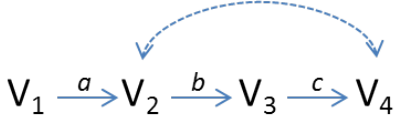
Figure 21. A graph illustrating a Verma constraint

is not imposed by the original model. This can be seen from the fact that, conditioned on Y, the path $Z \to Y \leftarrow X \leftrightarrow W$ is unblocked and becomes blocked if replaced by $Z \to Y \leftrightarrow X \leftrightarrow W$. The same applies to path $Z \to X \leftrightarrow W$, since Y would cease to be a descendant of X.

Testable Implications in Semi-Markovian Models

In the case of non-Markovian models, additional testable implications may be present, which are not revealed by d-separation. First noted by Verma and Pearl (1990), these constraints, often called *Verma constraints* in the non-parametric literature, impose invariance rather than conditional independence restrictions. Algorithms that enumerate certain types of Verma constraints for semi-Markovian, non-parameteric SEMs are given by Tian and Pearl (2002) and Shpitser and Pearl (2008).

In Figure 21, for example, one can show that regardless of the structural equations, 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.

Testable implications can also be obtained by *overidentifying* model parameters[18]. In some cases, these constraints will be conditional independence constraints while in others they are not. The Verma constraint shown in Figure 21 is obtainable using overidentification. Using the single-door criterion (Pearl, 2000) and Wright's path-tracing rules, we identify two expressions for c in terms of the covariance matrix:

$$c = \beta_{43.2} = \frac{\sigma_{43.2}}{1 - \sigma_{32}^2}, \text{ and}$$

$$c = \frac{abc}{ab} = \frac{\sigma_{41}}{\sigma_{31}}$$

Equating the two expressions, we obtain the following constraint:

$$\frac{\sigma_{43.2}}{1 - \sigma_{32}^2} = \frac{\sigma_{41}}{\sigma_{21}\sigma_{32}}.$$

This constraint is equivalent to the one obtained using the non-parametric methods of Tian and Pearl (2002) and Shpitser and Pearl (2008). An algorithm for systematically discovering constraints by overidentifying coefficients using the causal graph is given by Chen et. al. (2014). As yet, 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 recursive models. See (Cooper, 1999), (Pearl, 2000, ch. 2), and (Spirtes et al., 2000, chs. 5 and 6)[19] for examples. For non-recursive models, Hoover and Phiromswad (2013) make use of overidentifying constraints resulting from both multiple instruments and 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, 2010). 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, 2010) for linear SEMs.

Theorem 9. $\beta_{YX.Z} = \beta_{YX.W}$ *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.

---

[18]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). However, expressing a parameter in terms of the covariance matrix in more than one way does not necessarily mean that equating the two expressions actually constrains the covariance matrix. See (Pearl, 2004) for a formal definition of parameter overidentification.

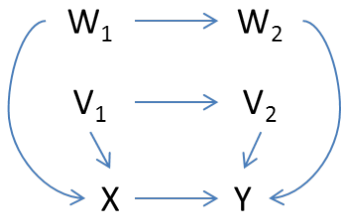[19]Software implementing these algorithms is available from the TETRAD Project (http://www.phil.cmu.edu/projects/tetrad/).

Figure 22. $\{V_1, W_2\}$ and $\{V_2, W_1\}$ are c-equivalent but not $\{W_1\}$ and $\{W_2\}$
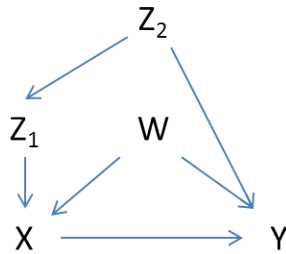

Figure 23. Graph illustrating preference to Z_1 over Z_2; $Var[\beta_{YX.WZ_2}] \le Var[\beta_{YX.WZ_1}]$

Consider the graph depicted in Figure 22. 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.

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 10. *Suppose that sets $\{W, Z_1\}$ and $\{W, Z_2\}$ satisfy the back-door criterion relative to (X, Y) in a linear-normal SEM. If $\{W, Z_1\}$ d-separates X from Z_2 and $\{X, Z_2, W\}$ d-*

separates Y from Z_1, then $Var[\beta_{YX.WZ_2}] \le Var[\beta_{YX.WZ_1}]$. 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 23, 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.

The following model, depicted in Figure 24a, represents an "encouragement design" (Holland, 1988; Pearl, 2014b) 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 4.

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

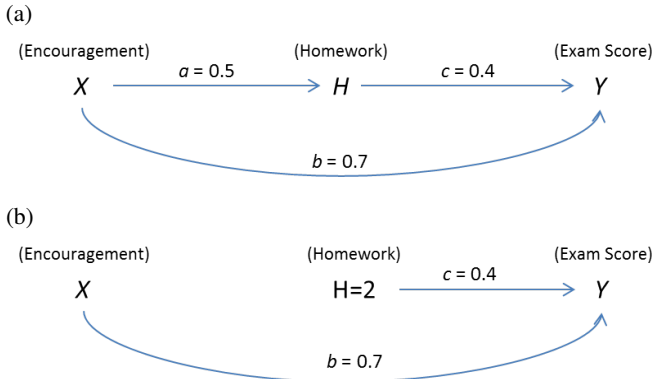We also give the values for the coefficients (which can be

(a)

(b)

Figure 24. 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:

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

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 24b. Finally, we compute the value of Y in our modified model using the updated U values giving:

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

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, 2000):

(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 $E[B_{A=a}|E = e]$ may not be identifiable, even if we have the luxury of running experiments (Pearl, 2009). In linear models, however, any counterfactual of the form, $E[Y_{X=x}|E = e]$, with e an arbitrary evidence, is identified whenever $E[Y|do(X = x)]$ is identified (Pearl, 2009, p. 389). As a result, if the data generating mechanism is linear, any counterfactual quantity is identifiable whenever the model parameters are identifiable[20].

Theorem 11. *(Pearl, 2009) Let T be the slope of the total effect of X on Y, $\frac{\partial}{\partial x}E[Y|do(X = x)]$, then $E[Y_{X=x}|E = e] = E[Y|E = e] + T(x - E[X|E = e])$.*

---

[20]Any expectation of the form $E[Y|do(X = x)]$ can, of course, be identified experimentally by randomizing the value of X and computing the average of Y over the population for which $X = x$.
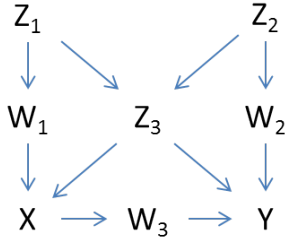
Figure 25. Graph corresponding to Model 5 in text

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 11 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 nontrivial problems that SEM researchers are likely to encounter.

Model 5.

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

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 25 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.
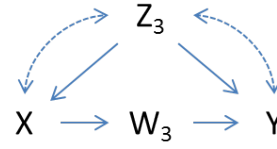
Solutions:

Figure 26. Graph representing Model 5 when Z_1, W_1, Z_2, and W_2 are unobserved

 (i) Figure 25 shows that $\{W_1, Z_3, W_2, W_3\}$ blocks all paths between 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, $\sigma_{XZ_1.W_1,Z_3} = 0$ and $\sigma_{YZ_2.Z_3W_2} = 0$.

 (ii) When X, Y, W_3, and Z_3 are latent variables, Model 5 is equivalent to the graph in Figure 26. 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 $X \rightarrow W_3$, X is d-separated from W_3. Likewise, a can be identified using the single-door criterion. When we remove the edge $W_3 \rightarrow Y$, 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) $\beta_{YX.W_1Z_3} = \beta_{YX.W_1Z_3Z_1}$ 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. This paper demonstrates

how graphs can compute the testable implications of modeling assumptions and how they can 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 far exceed 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 Greenland, 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). We hope that the potential of these tools will be recognized by SEM researchers as well.

Acknowledgments

References

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

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

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

BOLLEN, K. A. and BAUER, D. J. (2004). Automating the selection of model-implied instrumental variables. *Sociological methods & research* **32** 425–452.

BOLLEN, K. A. and PEARL, J. (2013). Eight myths about causality and structural equation models. In *Handbook of Causal Analysis for Social Research*. Springer, 301–328.

BRITO, C. (2004). *Graphical methods for identification in structural equation models*. Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.

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.

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., TIAN, J. and PEARL, J. (2014). Testable implications of linear structural equation models. R-428, <http://ftp.cs.ucla.edu/pub/stat_ser/r428.pdf>, CA. To appear in Proceedings of AAAI-2014.

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.

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. (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.

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.

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

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.

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. (2011). Identification: A non-technical discussion of a technical issue. *Handbook of Structural Equation Modeling (R. Hoyle, ed.). Guilford Press, New York* .

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.

KYONO, T. (2010). Commentator: A front-end user-interface module for graphical and structural equation modeling. Tech. Rep. R-364, <http://ftp.cs.ucla.edu/pub/stat_ser/r364.pdf>, Master Thesis, Department of Computer Science, University of California, Los Angeles, CA.

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.

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.

OSTER, E. (2013). Unobservable selection and coefficient stability: Theory and validation. Tech. rep., National Bureau of Economic Research.

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

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

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. H. Hoyle, ed.). Guilford Press, New York. In press.

PEARL, J. (2013). Interpretation and identification of causal mediation. Tech. Rep. R-389,

<http://ftp.cs.ucla.edu/pub/stat_ser/r389.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Forthcoming *Psychological Methods*, (2014).

PEARL, J. (2014a). The deductive approach to causal inference. Tech. Rep. R-424, <http://ftp.cs.ucla.edu/pub/stat_ser/r424.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Submitted, *Journal of Causal Inference*.

PEARL, J. (2014b). Reply to commentary by Imai, Keele, Tingley, and Yamamoto, concerning causal mediation analysis. Tech. Rep. R-421, <http://ftp.cs.ucla.edu/pub/stat_ser/r421.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Forthcoming *Psychological Methods* (2014) with discussions of "Interpretation and Identification of Causal Mediation," (R-389).

PEARL, J. and PAZ, A. (2010). Confounding equivalence in causal equivalence. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. AUAI, Corvallis, OR, 433–441.

SHIPLEY, B. (1997). An inferential test for structural equation models based on directed acyclic graphs and its non-parametric equivalents. Tech. rep., Department of Biology, University of Sherbrooke, Canada. Also in *Structural Equation Modelling*, 7:206–218, 2000.

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.

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.

TEXTOR, J., HARDT, J. and KNÜPPEL, S. (2011). Dagitty: a graphical tool for analyzing causal diagrams. *Epidemiology* **22** 745.

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. (2002). On the testable implications of causal models with hidden variables. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc.

TIAN, J. and PEARL, J. (2003). On the identification of causal effects. Tech. Rep. R-290-L, Department of Computer Science, University of California, Los Angeles, CA. Http://www.cs.iastate.edu/~jtian/r290-L.pdf.

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

VERMA, T. (1993). Graphical aspects of causal models. Tech. Rep. R-191, UCLA, Computer Science Department.

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.

WIKIPEDIA (2014). Instrumental variable. [Online; accessed 26-March-2014].
URL `http://en.wikipedia.org/wiki/Instrumental_variable`

WILLIAMS, L. (2011). Equivalent models: Concepts, problems, alternatives. *Handbook of Structural Equation Modeling (R. Hoyle, ed.). Guilford Press, New York* .

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

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