Path Diagrams

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Path Diagrams

Introduction

- Path Diagrams play an important role in communication and visualization of structural equation models
- You can sometimes see things easily in a path diagram that are more difficult to see from a set of written equations.
- As useful as they are in some instances, path diagrams can also be counterproductive:
    1. Inconsistent and/or conflicting diagramming conventions can induce confusion and error.
    2. As the number of variables in the diagram increases, the diagram may become visually confusing and/or impossible to read.
- In what follows, I shall discuss the most common path diagramming conventions used in the classic covariance structure modeling framework.
- A thorough understanding of these conventions will prove useful in deciphering more complex types of models.

Path Diagram Basics

- Path diagrams are like flowcharts.
- They show variables interconnected with lines that are used to indicate causal flow.
- In the most general type of diagram, each path involves two variables (in either boxes or ovals) connected by either arrows (lines, usually straight, with an arrowhead on one end) or wires (lines, usually curved, with no arrowhead), or "slings" (with two arrowheads).
- Arrows are used to indicate directed relationships, or linear relationships between two variables. An arrow from X to Y indicates a linear relationship where Y is the dependent variable and X the independent variable.
- Wires or Slings are used to represent undirected relationships, which represent variances (if the line curves back from a variable to itself) or covariances (if the line curves from one variable to another).

Path Diagram Basics

Simple Bivariate Regression

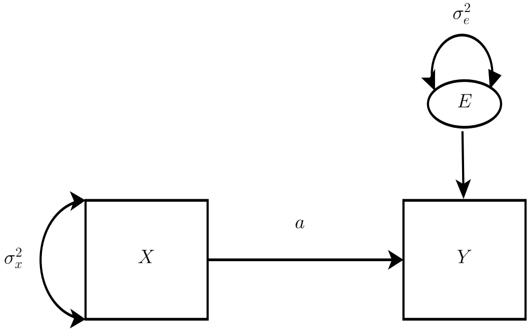
- One can think of a path diagram as a device for showing which variables cause changes in other variables.
- However, path diagrams need not be thought of strictly in this way.
- They may also be given a narrower, more specific interpretation.
- Consider the classic linear regression equation

$$Y = aX + e \tag{1}$$

- Here is the path diagram representation.

Path Diagram Basics

Simple Bivariate Regression



Path Diagram Basics

Simple Bivariate Regression

- Such diagrams establish a simple isomorphism.
- All variables in the equation system are placed in the diagram, either in boxes or ovals.
- Manifest variables are in boxes, latent variables (including error terms) in ovals.
- Each equation is represented on the diagram as follows:
  1. All independent variables (the variables on the right side of an equation) have arrows pointing to the dependent variable.
  2. The weighting coefficients are placed in clear proximity to the arrows.

Path Diagram Basics

Variable Types

- Besides being either manifest or latent, a variable is either exogenous or endogenous.
- A variable is endogenous if and only if it has at least one arrow pointing to it, i.e., if it is "on the receiving end" of a directed relationship.
- A variable is exogenous if an only if it has no arrows pointing to it and at least one arrow pointing away from it.
- Note — a variable can be endogenous and still have arrows pointing away from it.

Path Diagram Basics

Variable Types

- Note, then, that a variable is one of 4 types:
    1. Manifest-Exogenous
    2. Manifest-Endogenous
    3. Latent-Exogenous
    4. Latent-Endogenous

Path Diagram Basics

Variable Types

- If random variables are related by linear equations, then variables which are endogenous have variances and covariances which are determinate functions of the variables on which they regress. For example, if X and Y are orthogonal and

$$W = aX + bY \tag{2}$$

then it must be the case that

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 \tag{3}$$

Path Diagram Basics

Variable Types

- One way of guaranteeing that a diagram can account for variances and covariances among all its variables is to require:
  1. Representation of all variances and covariances among exogenous variables,
  2. No variances or covariances to be directly represented in the diagram for endogenous variables, and
  3. All variables in the diagram be involved in at least one relationship.

Path Diagram Basics

Variable Types

- Unfortunately, there is a significant practical problem with many path diagrams — lack of space.
- In many cases, there are so many exogenous variables that there is simply not enough room to represent, adequately, the variances and covariances among them.
- Diagrams which try often end up looking like piles of spaghetti.
- One way of compensating for this problem is to include rules for default variances and covariances which allow a considerable number of them to be represented implicitly in the diagram.
- One consistent standard for basic structural equation models is represented in the rules shown in the next section

Path Diagram Rules I

1. Manifest variables are always represented in boxes (squares or rectangles) while latent variables are always in ovals or circles.
2. Each directed relationship is represented explicitly by an arrow between two variables.
3. Undirected relationships need not be represented explicitly. (See rule 9 below regarding implicit representation of undirected relationships.)
4. Undirected relationships, when represented explicitly, are shown by a sling (two-headed arrow) or wire from a variable to itself, or from one variable to another.
5. Endogenous variables may never have slings/wires connected to them.
6. Free parameter numbers or labels for a sling/wire or arrow are always represented with integers or labels placed on, near, or slightly above the middle of the wire or arrowline. A free parameter is a number whose value is estimated by the program. Two free parameters having the same parameter number or label are required to have the same value.

Path Diagram Rules II

7. Fixed values for a wire or arrow are always represented with a floating point number containing a decimal point. The number is generally placed on, near, or slightly above the middle of the sling/wire or arrow line. A fixed value is assigned by the user. (There are default values that are applicable in various situations.)
8. Different statistical populations are represented by a line of demarcation and the words Group 1 (for the first population or group), Group 2, etc., in each diagram section.
9. All exogenous variables must have their variances and covariances represented either explicitly or implicitly by either free parameters or fixed values. If variances and covariances are not represented explicitly, then the following rules hold:
    - Among latent exogenous variables, variances not explicitly represented in the diagram are assumed to be fixed values of 1.0, and covariances not explicitly represented are assumed to be fixed values of 0.

Path Diagram Rules III

- Among manifest exogenous variables, variances and covariances not explicitly represented are assumed to be free parameters each having a different parameter number. These parameter numbers are not equal to any number appearing explicitly in the diagram.
- Constants (intercept terms) are generally represented as triangular exogenous variables.

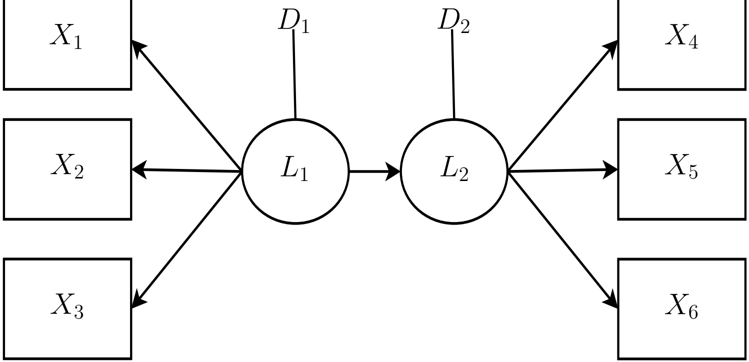
Exceptions and Ambiguities

- By adopting a consistent standard for path diagrams, we can facilitate clear communication of path models, regardless of what system is used to analyze them.
- The typical beginning student of SEM will attempt to reproduce results from published papers employing a wide variety of standards for their path diagrams.
- In some cases this approach will create no problems. However, experience indicates that it is often useful to translate published diagrams into diagram that obeys rules 1–9 in the previous section, before specifying the model for estimation.
- Frequently the translation process will draw attention to errors or ambiguities in the published diagram.

Exceptions and Ambiguities

- The figure on the next slide shows a portion of a path diagram which is quite typical of what is found in the literature.
- This is not a complete diagram and it does not conform to diagramming rules in the preceding section.

Exceptions and Ambiguities



Exceptions and Ambiguities

- Some of the diagram is clear and routine, but what do we make of the symbols D_1 and D_2?
- Variable L_1 is a latent exogenous variable. It has arrows pointing away from it and no arrows pointing to it.
- Since, by rule 9 for diagrams (see above), all exogenous variables must have their variances and covariances explained, the most reasonable assumption is that D_1 stands for the variance of latent variable L_1.

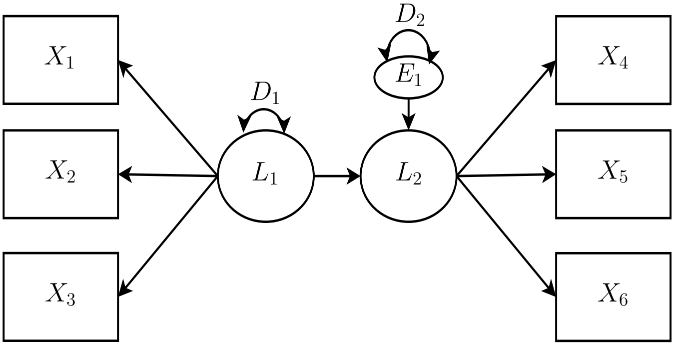
Exceptions and Ambiguities

- Hence, the diagram is modified to make $D1$ a parameter attached to a wire from $L1$ to itself.
- But what is the status of $D2$? In the diagram it looks just like $D1$, but closer inspection reveals it must mean something different.
- $D2$ is connected to $L2$, and $L2$ is an endogenous latent variable.
- Consequently, the most reasonable interpretation is that $D2$ represents an error variance for latent variable $L2$. It is represented with an error latent variable $E2$ with variance $D2$.

Exceptions and Ambiguities

- The revised path diagram, more accurately reflecting the author's model, is shown in the figure on the next slide.
- Notice, however, that the diagram is still not fully explicit. Each of the manifest variables is endogenous, and, as such, needs an error (or residual) variance.
- Many path diagrams, for the sake of compactness, will not include these paths.

Exceptions and Ambiguities



Exceptions and Ambiguities

- In some cases you will have to be creative, tenacious, and lucky to figure out what the author of a path diagram intended.
- Even the most accomplished and generally careful authors will leave out paths, forget to mention that some values were fixed rather than free parameters, or simply misrepresent the model actually tested.
- Sometimes the only way to figure out what the author actually did is to try several models, until you find coefficients which agree with the published values.
- These difficulties are compounded by the occasional typographical errors that appear in published covariance and correlation matrices.

Exceptions and Ambiguities

The RAM Diagramming System

- Some path diagrams do not represent the error variance attached to endogenous latent variables at all — they leave this to the reader to figure out for him/herself.
- Whenever an endogenous latent variable has no error term, you should suspect that an error latent variable has been left out, especially if your degrees of freedom don't agree with those of the published paper.
- The later version of the RAM system developed by Jack McArdle adds an additional twist to this.
- A residual variable is an exogenous variable that has a directed path to one (and only one) endogenous variable.
- In the RAM system, residual variables are not represented explicitly in the diagram.
- Rather, their variances are shown as two-headed arrows (or "slings") attached to the variable they point to.
- Of course, this means that slings can mean one thing in one part of the diagram, and another thing in a different part of the same diagram.