Taking Apart the Pieces of SEM:
Path Diagrams, Path Coefficients, and Model Construction

Outline of the Basics
1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

Basic Terminology

Representing Unexplained Correlation

- This is multiple regression!
- <-> Can represent many things
  - Uncertain causal relationship
  - Common driver

Unexplained Correlation
Representing Unexplained Correlation

- Really, a correlation between residual variance
- Convention: show correlation between endogenous errors but not exogenous – still there, though!

Path Diagrams and Causality

1. Sewell Wright’s intention was to describe (1) causal relationships and (2) strength of associations.

2. Explicit consideration of causation languished for 70 years. Judea Pearl and others have revived it in the science of artificial intelligence.

3. Pearl argues that regular mathematics is unable to express the needed expressions to represent causation. “=” versus “→”


Practical Criteria for Supporting Causal Assumptions

1. A manipulation of $x$ would result in a subsequent change in the values of $y$

2. OR the values of $x$ serve as indicators for the values of $x$ that existed when effects on $y$ were being generated.

3. Models are properly specified to extract causal information

Can my model be fit?

*Identification*

<table>
<thead>
<tr>
<th>Equation</th>
<th>Identified</th>
<th>Underidentifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3 = a + b$</td>
<td>$3 = a + b + c$</td>
<td>$a$, $b$, and $c$ have no unique solution</td>
</tr>
<tr>
<td>$4 = 2a + b$</td>
<td>$4 = 2a + b + 3c$</td>
<td></td>
</tr>
</tbody>
</table>

*a* and *b* have unique solutions

<table>
<thead>
<tr>
<th>Equation</th>
<th>Overidentifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3 = a + b$</td>
<td>$3 = a + b + c$</td>
</tr>
<tr>
<td>$4 = 2a + b$</td>
<td>$4 = 2a + b + 3c$</td>
</tr>
<tr>
<td>$7 = 3b + a$</td>
<td>$7 = 3b + a + 3c$</td>
</tr>
</tbody>
</table>
Structure of Models and Identification

Unsaturated or Overidentified

[Diagram showing relationships between variables]

Saturated or Just Identified

[Diagram showing relationships between variables]

Oversaturated or Underidentified

[Diagram showing relationships between variables]

Identification in SEM

# of Parameters v. # of Variables

We need enough information to estimate unique parameters

\[
\text{# params} \leq \text{# variables} \times (\text{# variables}+1)/2
\]

T-Rule: \( t \leq p(p+1)/2 \)

Feedbacks

Recursive

[Diagram showing recursive relationships]

Nonrecursive

[Diagram showing nonrecursive relationships]

Recursive (each item in a series is directly determined by the preceding item).

Breaking Feedbacks using Time

Nonrecursive

[Diagram showing nonrecursive relationships]

Recursive

[Diagram showing recursive relationships]

• Feedback assumes equilibrium
• Often concerned with dynamic processes
• Solution: Build in the lag
What is this?

Questions?

Outline of the Basics

1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data

Variance of a Normal Distribution

\[ \text{VAR}_x = \frac{\sum (x - \overline{x})^2}{n-1} \]
\[ \text{SD}_x = \sqrt{\text{VAR}_x} \]
**Covariance and Correlation**

\[
COV_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n-1}
\]

\[
r_{xy} = \frac{COV_{xy}}{SD_x \times SD_y}
\]

**Correlations and Regression**

Correlations contain less information, but are directly comparable

**Example:** these regressions having same unstandardized slope and intercept

Regression: \( y = a + bx \)

Standardized Coefficients: \( r = b \times \frac{sd(x)}{sd(y)} \)

**Covariances and Correlations**

We often use covariances to fit models, but standardized covariances – i.e. correlations – for interpretation.

\[
r_{xy} = \frac{COV_{xy}}{SD_x \times SD_y}
\]

\( r_{xy} = \text{correlation} / \text{std covariance} \)

\( COV = \text{covariances (ave cross product)} \)

\( SD = \text{std. dev.} = VAR^{1/2} \)

**Raw Covariance Matrix**

<table>
<thead>
<tr>
<th></th>
<th>x₁</th>
<th>x₂</th>
<th>y₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁</td>
<td>3.2</td>
<td>0.65</td>
<td>1.98</td>
</tr>
<tr>
<td>x₂</td>
<td>0.65</td>
<td>0.8</td>
<td>1.19</td>
</tr>
<tr>
<td>y₁</td>
<td>1.98</td>
<td>1.19</td>
<td>4.8</td>
</tr>
</tbody>
</table>

**Standardized Covariance Matrix**

<table>
<thead>
<tr>
<th></th>
<th>x₁</th>
<th>x₂</th>
<th>y₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>x₂</td>
<td>0.40</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>y₁</td>
<td>0.50</td>
<td>0.60</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**First Rule of Path Coefficients:** the path coefficients for unanalyzed relationships (curved arrows) between exogenous variables are simply the correlations (standardized form) or covariances (unstandardized form).
Second Rule of Path Coefficients: when variables are connected by a single causal path, the path coefficient is simply the standardized or unstandardized regression coefficient (note that a standardized regression coefficient is a simple correlation.)

\[
\begin{align*}
\gamma_{11} &= 0.50 \\
\beta_{21} &= 0.60
\end{align*}
\]

γ (gamma) used to represent effect of exogenous on endogenous.
β (beta) used to represent effect of endogenous on endogenous.

Third Rule of Path Coefficients: strength of a compound path is the product of the coefficients along the path.

\[
\begin{align*}
x_1 &\rightarrow y_1 &\rightarrow y_2
\end{align*}
\]

Thus, in this example the effect of \(x_1\) on \(y_2\) = 0.5 x 0.6 = 0.30

Since the strength of the indirect path from \(x_1\) to \(y_2\) equals the correlation between \(x_1\) and \(y_2\), we say \(x_1\) and \(y_2\) are conditionally independent.

The inequality implies that the true model is

\[
\begin{align*}
x_1 &\rightarrow y_1 &\rightarrow y_2
\end{align*}
\]

Fourth Rule of Path Coefficients: when variables are connected by more than one causal pathway, the path coefficients are "partial" regression coefficients.

Which pairs of variables are connected by two causal paths?
answer: \(x_1\) and \(y_2\) (obvious one), but also \(y_1\) and \(y_2\), which are connected by the joint influence of \(x_1\) on both of them.
And for another case:

A case of shared causal influence: the unanalyzed relation between \( x_2 \) and \( y_1 \) represents the effects of an unspecified joint causal process. Therefore, \( x_2 \) and \( y_1 \) are connected by two causal paths. \( x_2 \) and \( y_1 \) likewise.

How to Interpret Partial Path Coefficients: The Concept of Statistical Control

The effect of \( y_1 \) on \( y_2 \) is controlled for the joint effects of \( x_1 \).

With all other variables in model held to their means, how much does a response variable change when a predictor is varied?

Fifth Rule of Path Coefficients: paths from error variables represent prediction error (influences from other forces).

### Equation for path from error variable

\[
\text{Implied correlation between } y_1 \text{ and } y_2 = 0.50 \times 0.40 = 0.20.
\]

Sixth Rule of Path Coefficients: unanalyzed residual correlations between endogenous variables are partial correlations or covariances.
the partial correlation between $y_2$ and $y_2$ is typically represented as a **correlated error term**

This implies that some other factor is influencing $y_1$ and $y_2$

Note that total correlation between $y_1$ and $y_2 = 0.50 \times 0.40 + 0.86 \times 0.50 \times 0.92 = 0.60$ (the observed corr)

**Seventh Rule of Path Coefficients**: total effect one variable has on another equals the sum of its direct and indirect effects.

**Eighth Rule of Path Coefficients**: sum of all pathways between two variables (directed and undirected) equals the correlation/covariance.

**Outline of the Basics**

1. Terminology and housekeeping
   - Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data
What are the goals of the analysis?

Purpose of modeling effort:
- discovery?
- testing hypotheses?
- making predictions?

Focus of modeling effort:
- driver focused?
- response focused?
- mediation focused?
- theory testing focused?

Span of inference:
- doing inferential estimation?
- learning about processes?

The Ebb & Flow of Model Building

- Start with big ideas and basic theory
- Focus ideas on a targeted area
- Expand conceptual model to encompass the details of the problem
  - Be thorough, otherwise you may miss important elements of suppression or confounding variables
- Prune unnecessary details
- Confront your model with data and expand and contract it as needed…
The Structural Equation Meta-Model (SEMM)

Concept A → Causal Process 1 → Concept C

Causal Process 2

Concept B

My Research Program as a Meta-Model

Food Web Structure → Ecosystem Function

Environmental Change

Targeting Your Question

Food Web Structure → Ecosystem Function

Environmental Change

Focused on driver & Mediator

Targeting Your Question

Environment Abundance

Climate Change
Take a Step Back...

Remove Drivers Not of Interest

Identify Relevant Processes

- Food Web Structure
  - Habitat, Food, Light
  - Foundation Species
  - Abundance
  - Waves
  - Climate Change
Outline of the Basics

1. Terminology and housekeeping
   – Introduction to Causal Path Diagrams
2. The basics of path coefficients
3. The Structural Equation Meta-Model
4. Confronting your meta-model with data
Confront Model with Natural History

1. Kelp moderates disturbance
   - More Kelp = Smaller Disturbance?
   - BUT no effect on kelp that isn’t present...

2. Kelp regrows quickly
   - Dense beds after storms if nutrients present

Measuring Realized Disturbance via Satellite Measurements

Incorporate Interacting Variables

Incorporate Natural History of Disturbance
Satellite Derived Kelp Cover (# Pixels/250m^2)

Summer Kelp Abundance (#/sq m)

Food Web Network Metrics

Maximum Winter Wave Force (orbital velocity)

Previous Year Summer Kelp Abundance (#/sq m)

Consider Alternate Models

1. Full Model
2. No waves -> Food Web effect
3. No effect of previous year’s kelp -> Food Web
4. No effect of either waves or previous kelp -> Food Web

Confront your models with data!
(then have lunch)