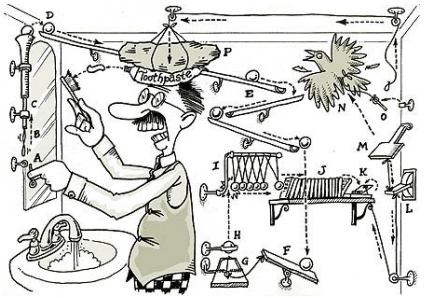
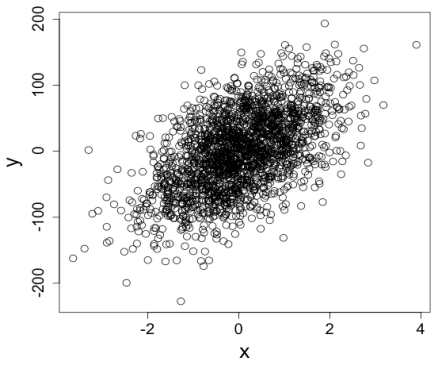
Anatomy of Path Diagrams



Overview

1. Covariance and Correlation
2. Pieces of a Path Diagram
3. Model Structure and Identification

Covariance and correlation



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

$$r_{xy} = \frac{COV_{xy}}{SD_x SD_y}$$

Covariance and correlation

We often use covariances to calculate slopes, but standardized covariances – i.e. correlations – for interpretation.

Raw Covariance Matrix

	x_1	x_2	y_1
x_1	0.81		
x_2	0.87	1.63	
y_1	0.88	1.80	4.98

variance covariance

Standardized Covariance Matrix

	x_1	x_2	y_1
x_1	1.0		
x_2	0.76	1.0	
y_1	0.44	0.63	1.0

correlation

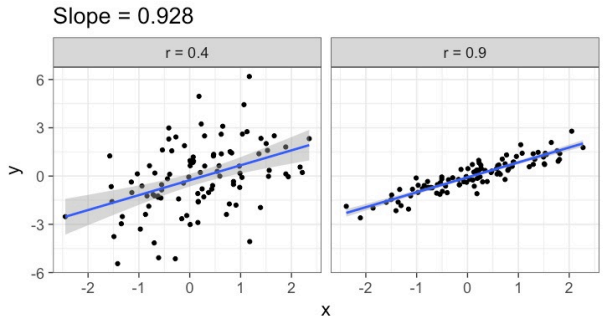
Standardization

- *Unstandardized coefficient* = absolute strength of the pathway
  - " An 1 unit change in *X* results in some unit change in *Y* "

$$\beta_{xy\ std} = b_{xy} * sd_x/sd_y$$

- *Standardized coefficient* = relative strength of the pathway
  - " A 1 standard deviation change in *X* results in some standard deviation change in *Y* "
  - Path Coefficient

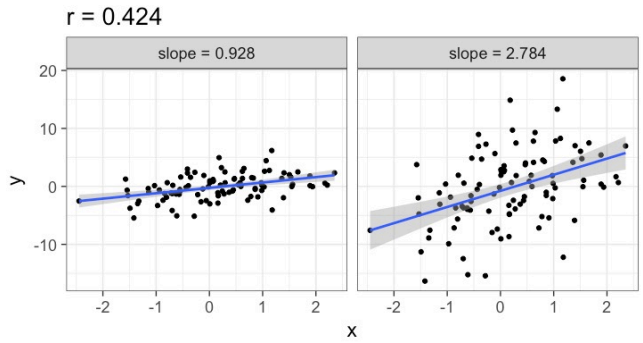
Same Slope, Different Correlation



Regression: y = a + bx

Standardized Coefficients: r = b * sd(x)/sd(y)

Different Slope, Same Correlation



Regression: y = a + bx

Standardized Coefficients: r = b * sd(x)/sd(y)

Standardization

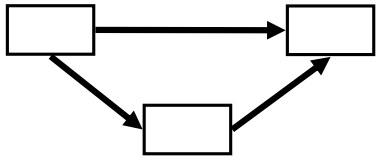
Unstandardized	Standardized
Good for prediction: coefficients are in raw units	Good for ranking: coefficients are in equivalent units
Has direct real world meaning	Less clear real world meaning
Can be compared across pathways or models that have identical units	Can be compared across all pathways in all models

Overview

1. Covariance and Correlation
2. Pieces of a Path Diagram
3. Model Structure and Identification

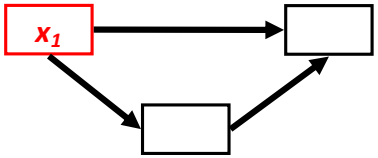
Terms & Definitions.

- Structural equation model = observed, latent, composite
- Direct acyclic graph (DAG) = observed
- Path diagram = observed, …



Terms & Definitions.

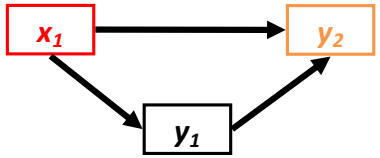
Exogenous variable = independent variable, predictor



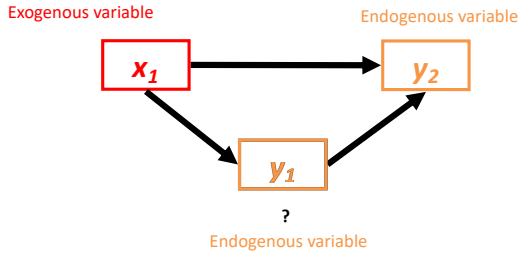
Terms & Definitions.

Exogenous variable

Endogenous variable = dependent variable, response

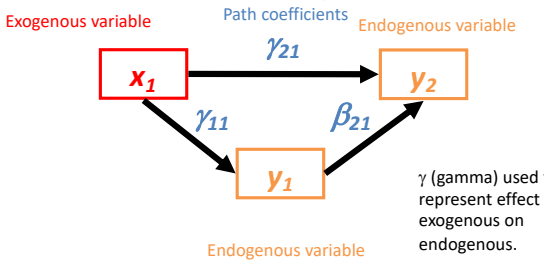


Terms & Definitions.



Terms & Definitions.

Assumes linear relationships!*

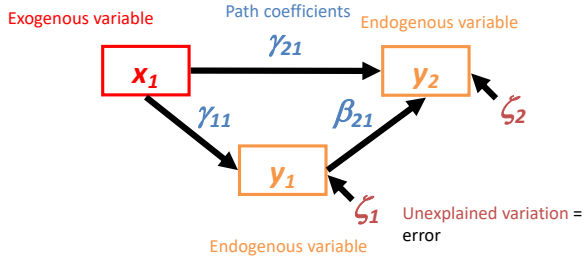


γ (gamma) used to represent effect of exogenous on endogenous.

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

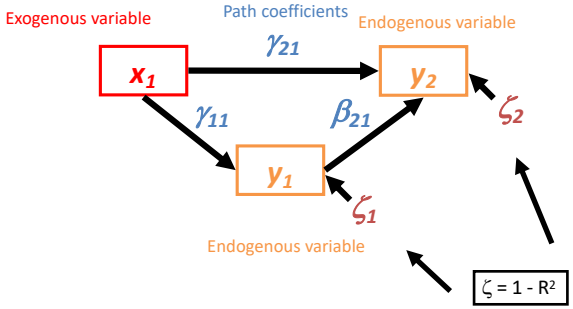
Coefficients show, with all other variables in model held to their means, how much a response changes when a predictor is varied

Terms & Definitions.

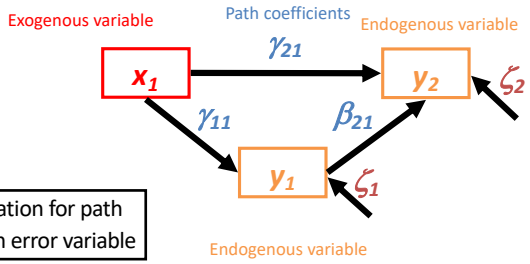


Unexplained variation = error

Terms & Definitions: Error coefficients



Terms & Definitions: Error coefficients

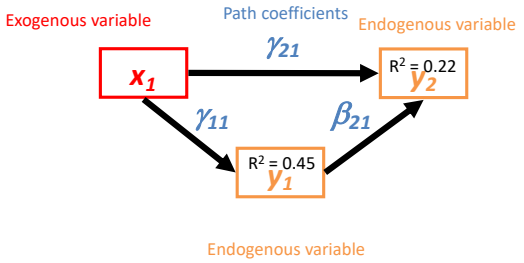


equation for path from error variable

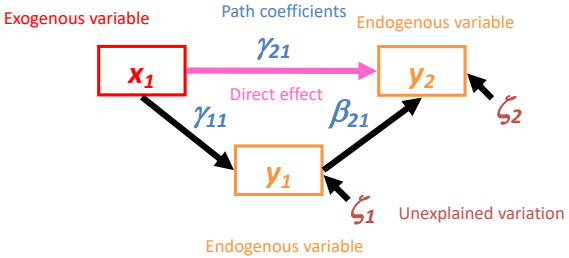
$$\zeta = \sqrt{1 - R^2_{y_i}}$$

This makes path coefficients comparable to those between variables

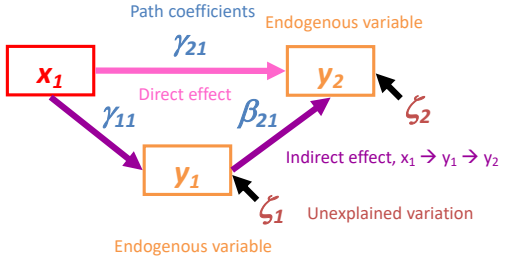
Terms & Definitions: Error coefficients



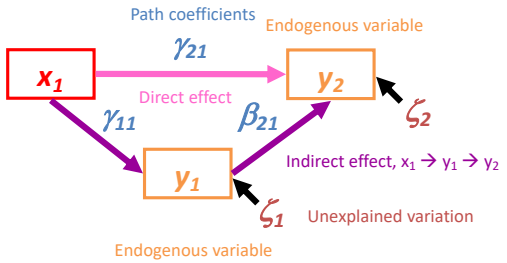
Terms & Definitions.



Terms & Definitions.

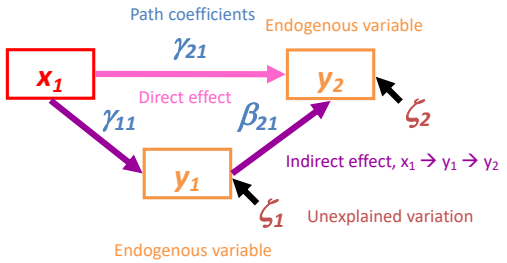


Terms & Definitions.



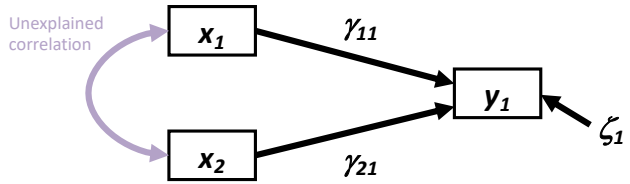
Indirect Effect = $\gamma_{11} * \beta_{21}$

Terms & Definitions.

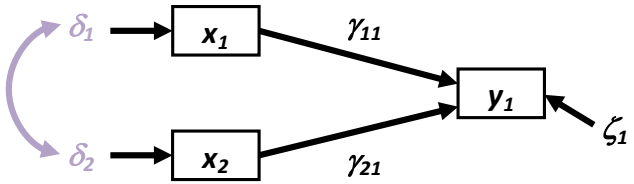


Total Effect = $\gamma_{21} + \gamma_{11} * \beta_{21}$

Terms & Definitions.

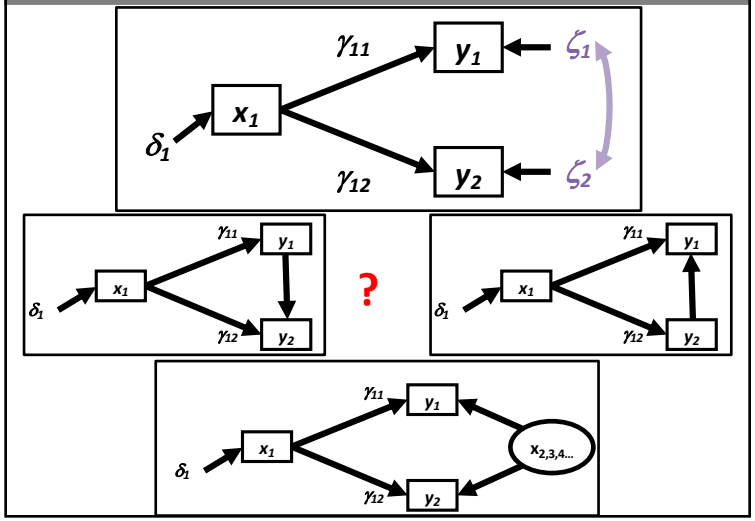


Terms & Definitions.



- Uncertain causal relationship ($x_1 \rightarrow x_2$, or $x_2 \rightarrow x_1$, common driver)
- We do not care if variables are exogenous (but check for collinearity)
- *Convention:* show correlation between endogenous errors but not exogenous – still there, though!

Equivalent Models



Overview

1. Covariance and Correlation
2. Pieces of a Path Diagram
3. Model Structure and Identification

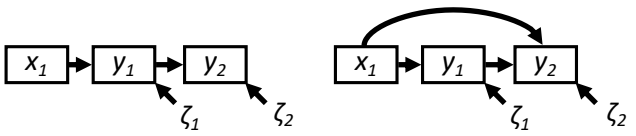
Identification. Can I fit my model?

3 = a + b 4 = 2a + b	*a* and *b* have unique solutions	**Just identified**
3 = a + b + c 4 = 2a + b + 3c	*a*, *b*, and *c* have no unique solution	**Underidentified**
3 = a + b 4 = 2a + b 7 = 3b + a	*a* and *b* have unique solutions, more knowns than unknowns	**Overidentified**

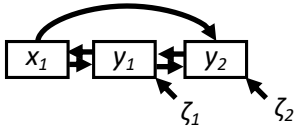
Identification. Can I fit my model?

Overidentified (Unsaturated)

Just Identified (Saturated)



Underidentified (Oversaturated)



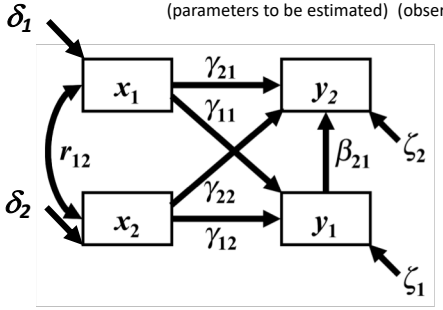
Identification. The *t*-rule

Test whether the model can be estimated based on available data:

t-rule = $t \leq n(n+1)/2$

t = # of unknowns (parameters to be estimated)

n = # of knowns (observed variables)

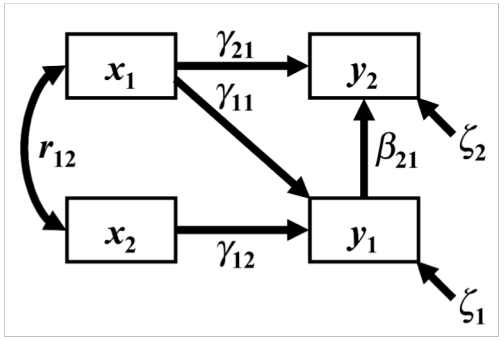


Knowns = x_1, x_2, y_1, y_2
$n = 4$
$4(4+1)/2 = 10$

Unknowns = γ_{21}, γ_{11}, γ_{22}, γ_{12}, β_{21}, r_{12} + variances on 4 observed variables (δ_1, δ_2, γ_1, γ_1)
$t = 10$

Identification. The *t*-rule

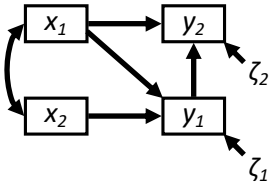
t-rule = $t \leq n(n+1)/2$



$T = 9 \leq 10$, *unsaturated*

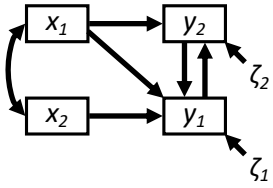
Identification. Feedbacks

Recursive



Recursive = each item in a series is directly determined by the preceding item

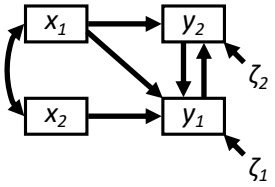
Non-recursive



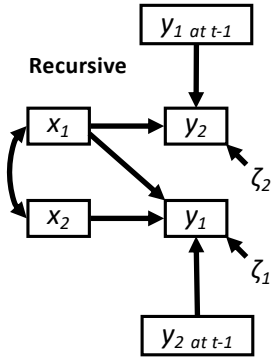
Non-recursive = there is bidirectionality (feedbacks) implicit in the model

Identification. Breaking Feedbacks with Time

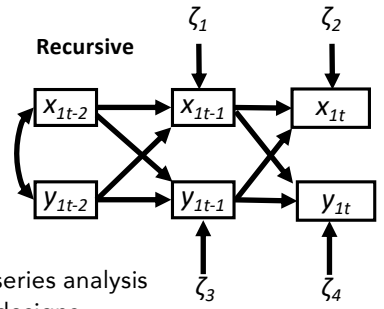
Non-recursive



Recursive



Identification. Cross-Lagged Panel Models.



- Time series analysis
- BACI designs
- Etc…

Exercise: Start Thinking About Your System

1. Sketch a model of 4-5 variables of your system
   - Think fast!
   - This does not have to be COMPLETE
2. Label exogenous and endogenous variables
3. Is your model identified? Fix if not!
4. Is it recursive? Can you break recursive relationships? If so, redraw.
5. Write out paths of indirect effects
6. Are any of your variances linked to other parts of the system?

You've learned the pieces – take a break!
Then let's learn how we put them together!