2. The Basics

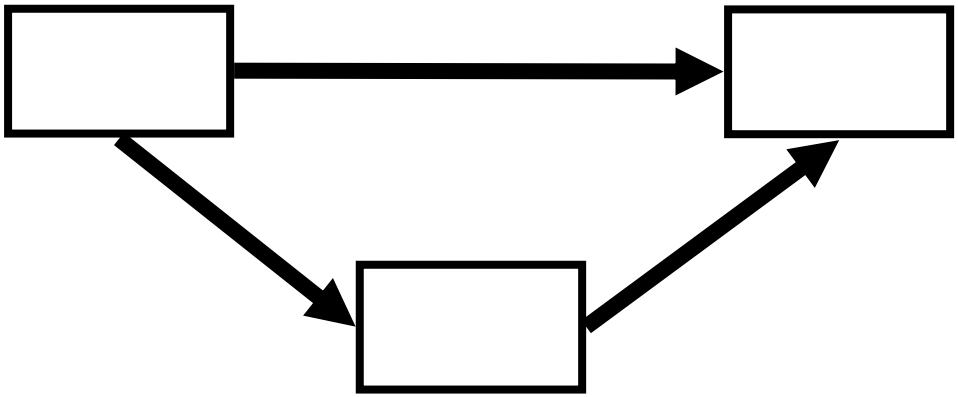
Overview

1. Terms and Definitions
2. Model Identification
3. Path Coefficients

2.1 Terms and Definitions

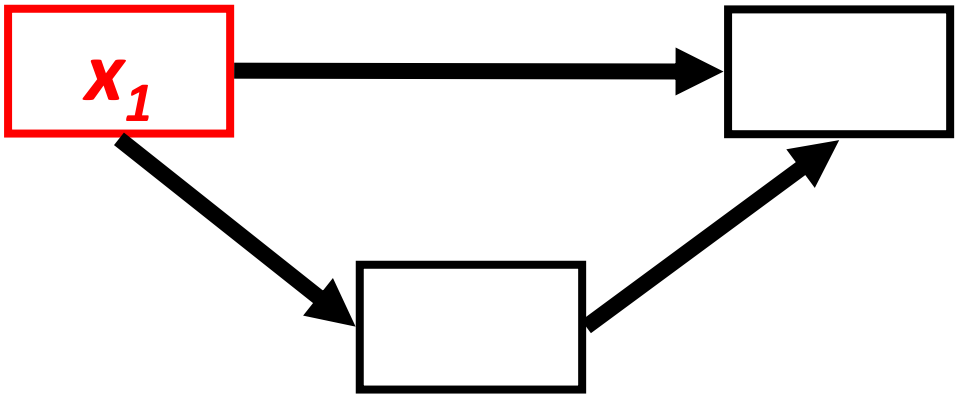
2.1 Terms & Definitions.

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



2.1 Terms & Definitions.

Exogenous variable = independent variable, predictor

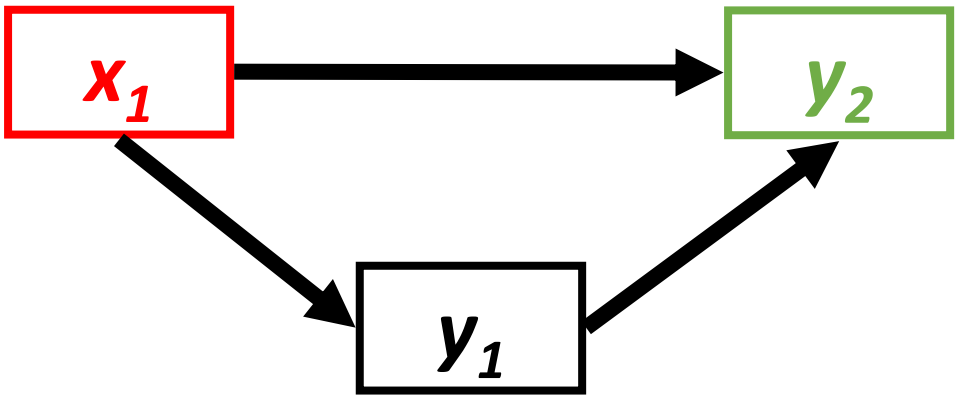


2.1 Terms & Definitions.

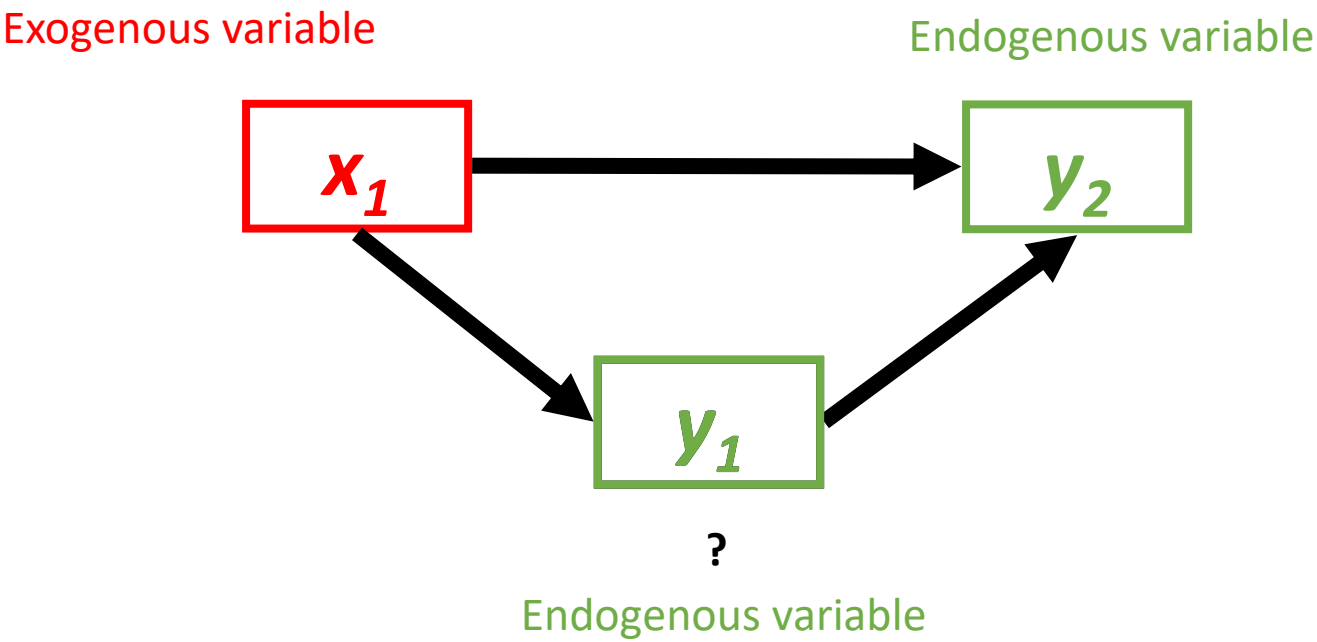
Exogenous variable

Endogenous variable =

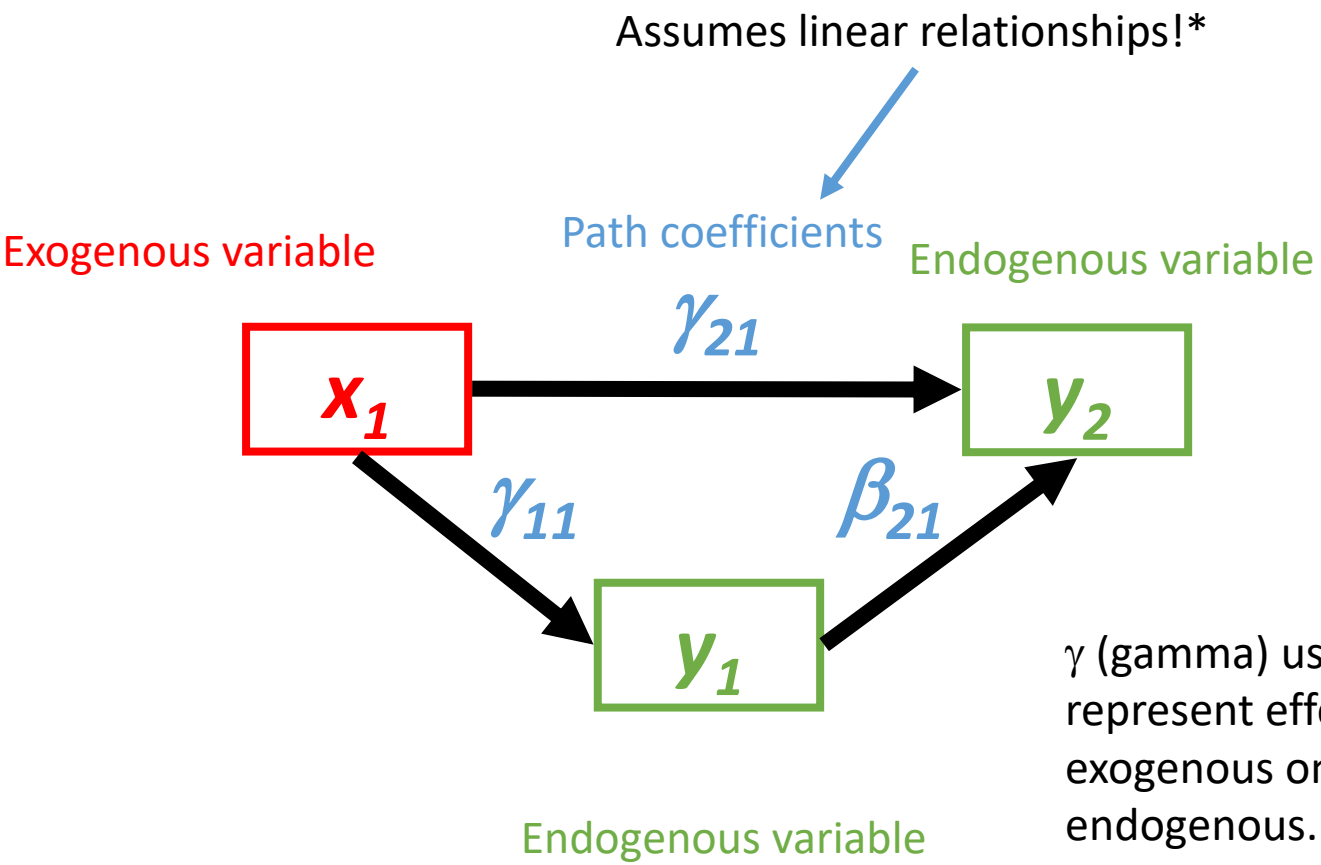
dependent variable,
response



2.1 Terms & Definitions.



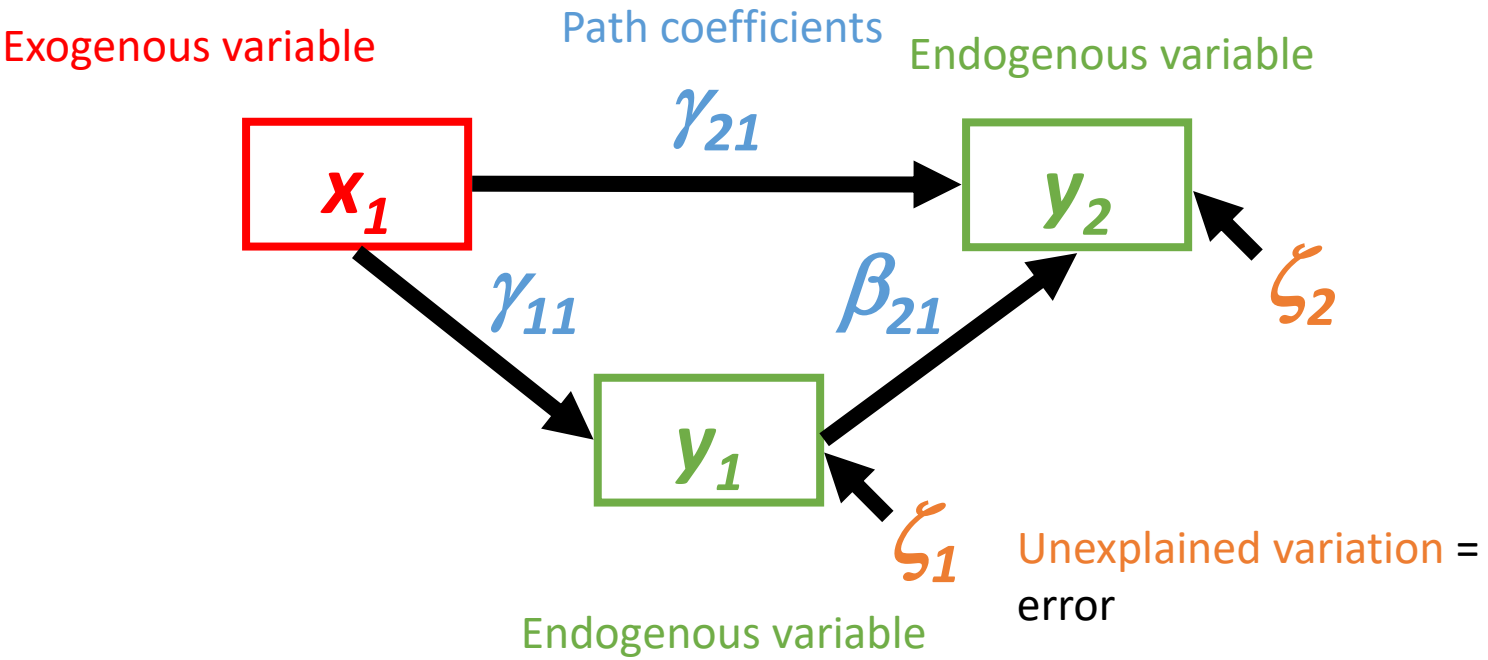
2.1 Terms & Definitions.



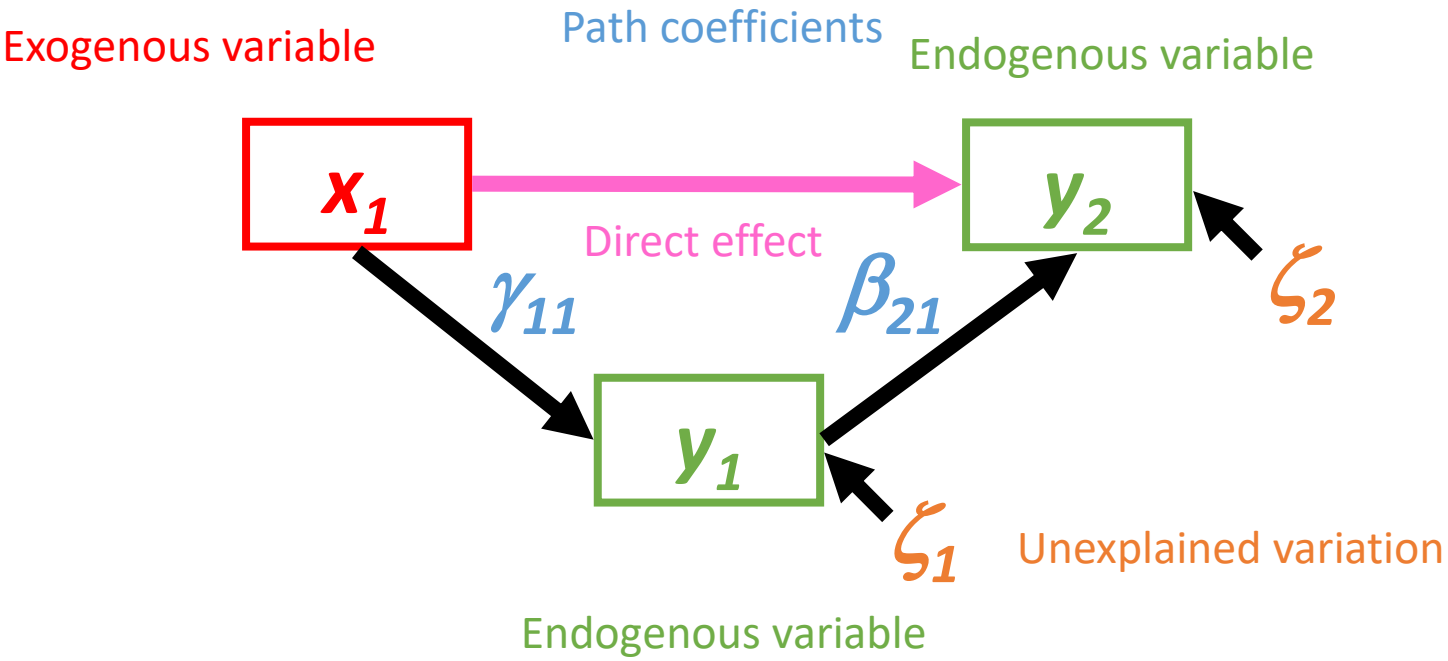
γ (gamma) used to represent effect of exogenous on endogenous.

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

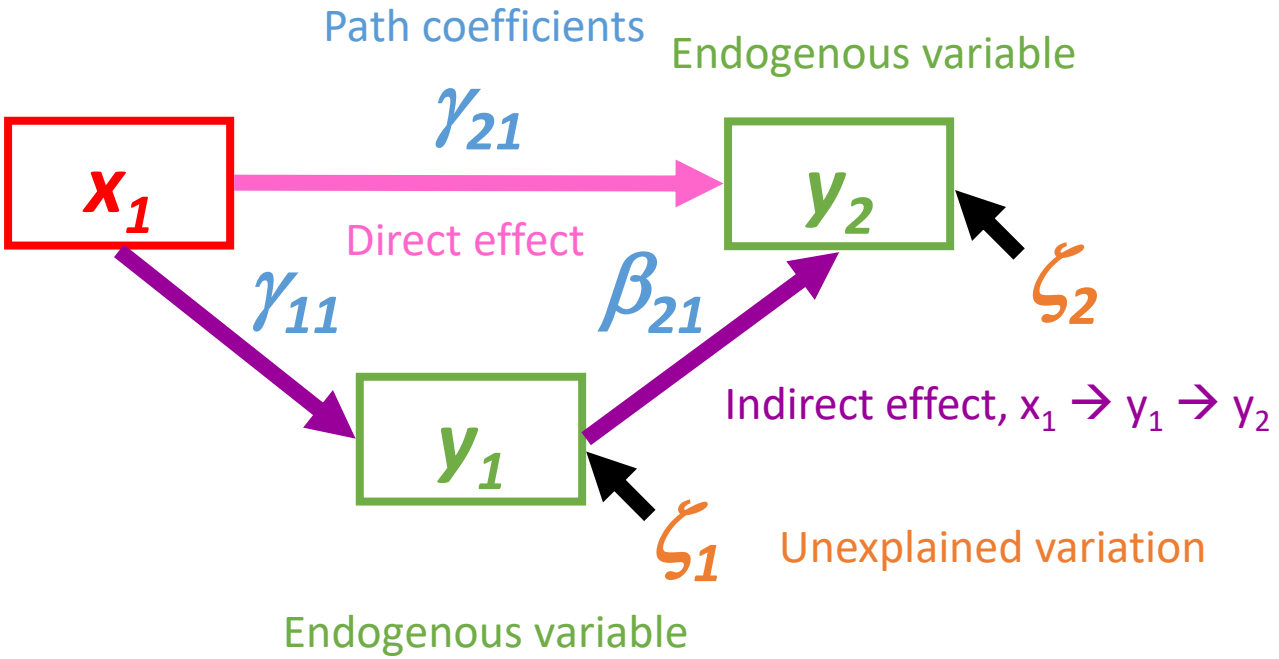
2.1 Terms & Definitions.



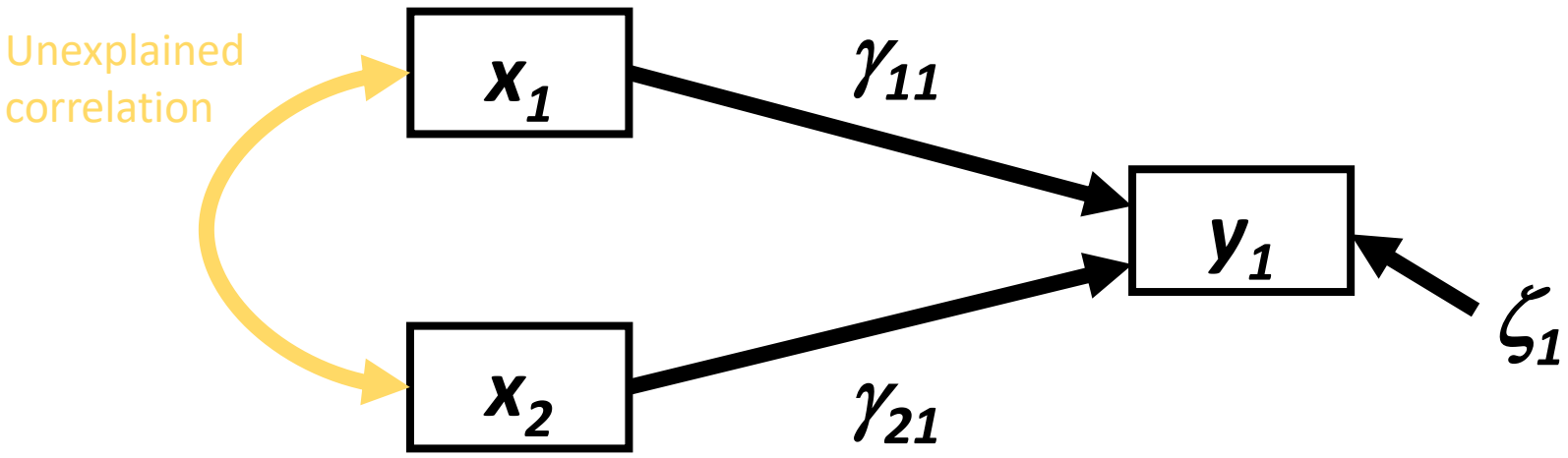
2.1 Terms & Definitions.



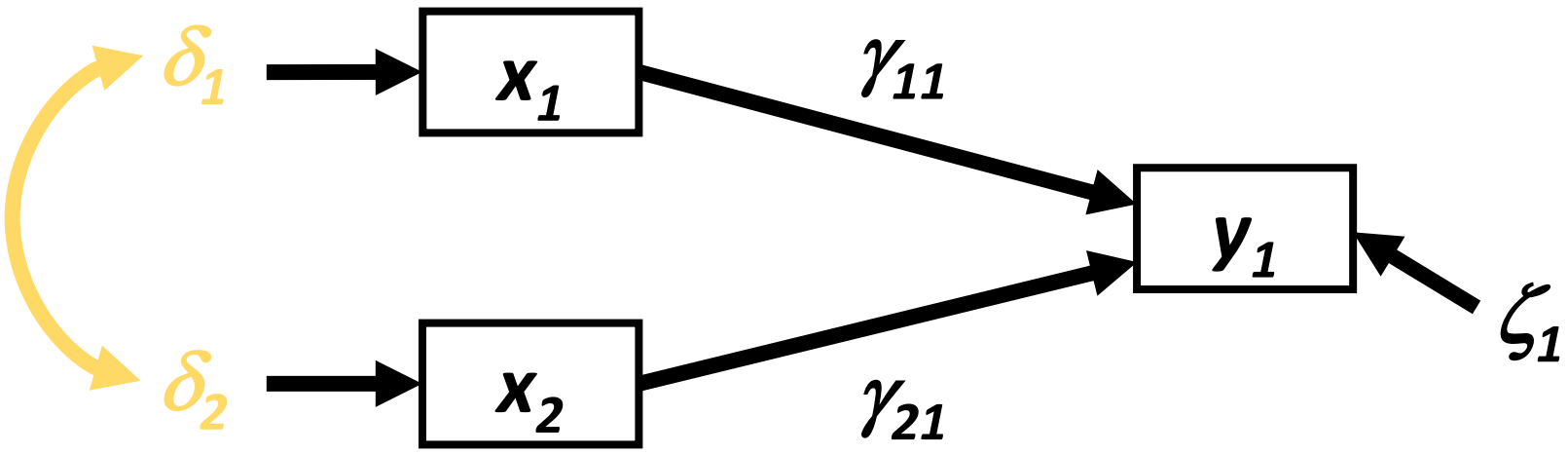
2.1 Terms & Definitions.



2.1 Terms & Definitions.



2.1 Terms & Definitions.



- Uncertain causal relationship ($x_1 \rightarrow x_2$, or $x_2 \rightarrow x_1$?)
- Common driver (correlated error)
- *Convention:* show correlation between endogenous errors but not exogenous – still there, though!

2.2 Model Identification

2.2 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

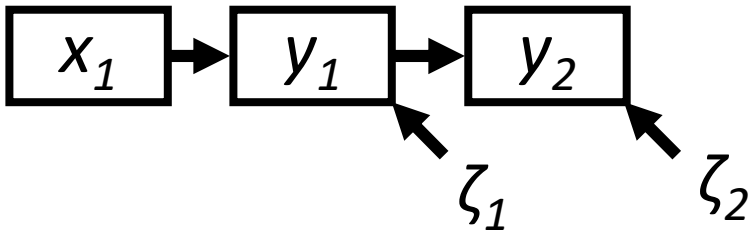
$3 = a + b$
$4 = 2a + b$
$7 = 3b + a$

a, *b*, and *c* have unique solutions, more knowns than unknowns

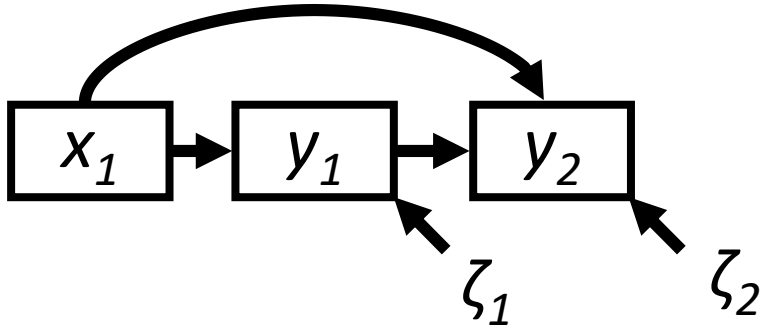
Overidentified

2.2 Identification. Can I fit my model?

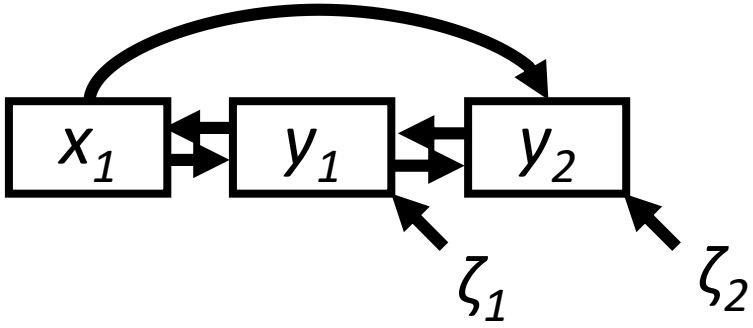
Overidentified
(Unsaturated)



Just Identified
(Saturated)



Underidentified (Oversaturated)



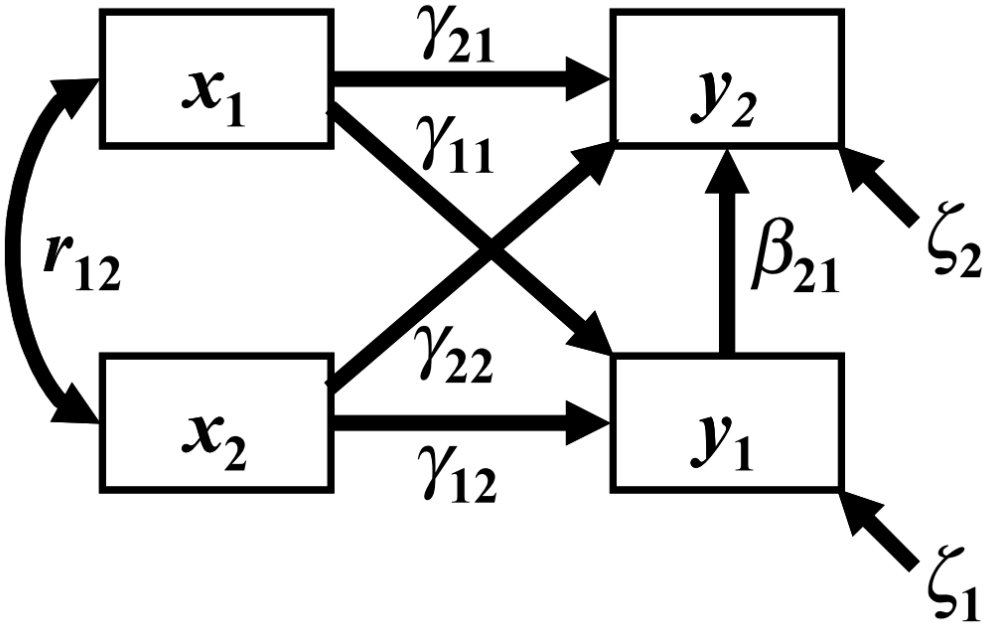
2.2 Identification. The *t*-rule

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

$$\underline{t\text{-rule}} = t \leq n(n+1)/2$$

t = # of unknowns (parameters to be estimated)

n = # of knowns (observed variables)



Knowns = x_1, x_2, y_1, y_2

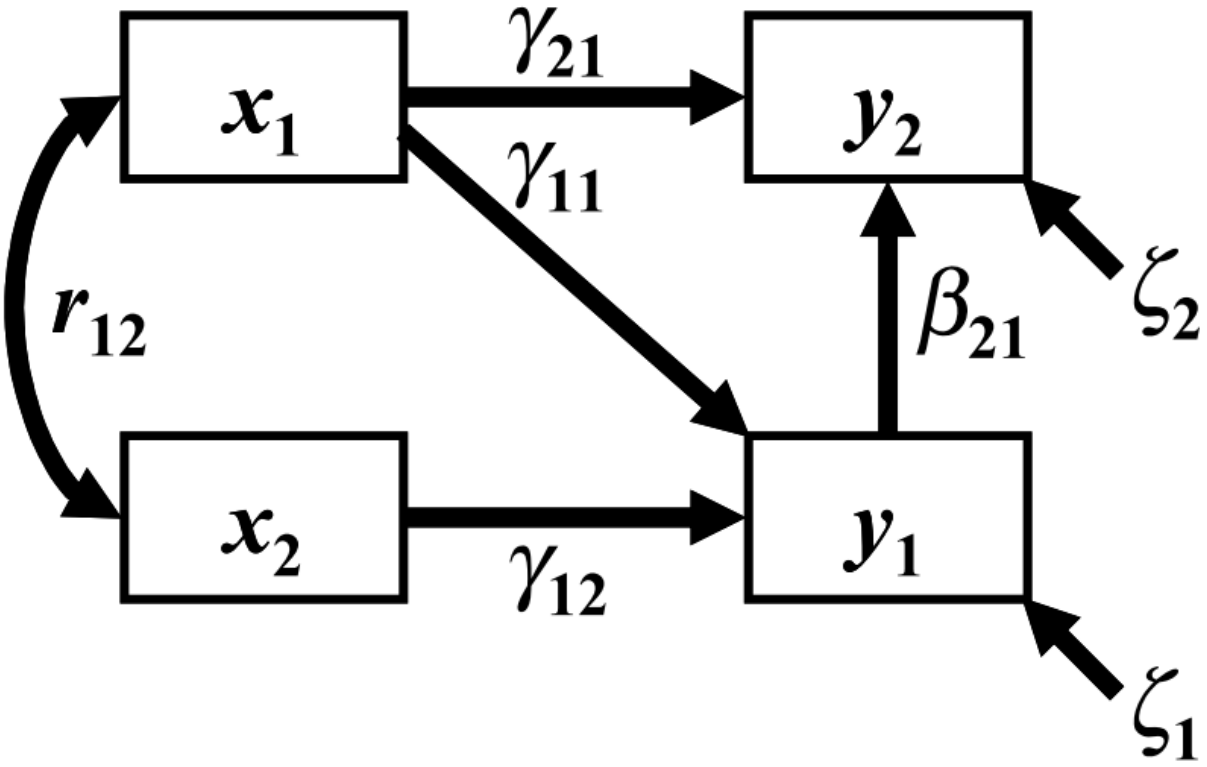
n = 4

4(4 + 1)/2 = 10

Unknowns = γ_{21}, γ_{11}, γ_{22}, γ_{12}, β_{21}, r_{12} + variances on 4 observed variables

t = 10

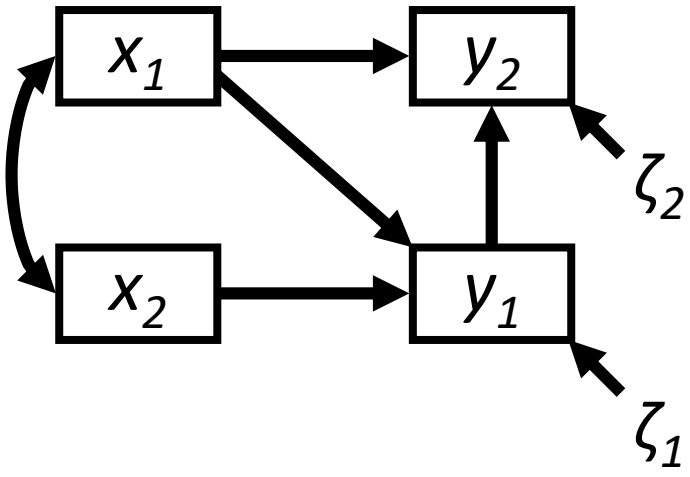
2.2 Identification. The *t*-rule



T = 9 ≤ *n* 10, *overidentified*

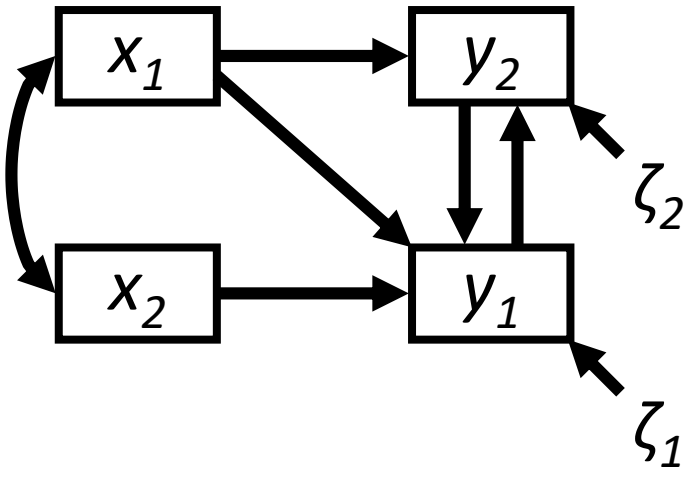
2.2 Identification. Feedbacks

Recursive



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

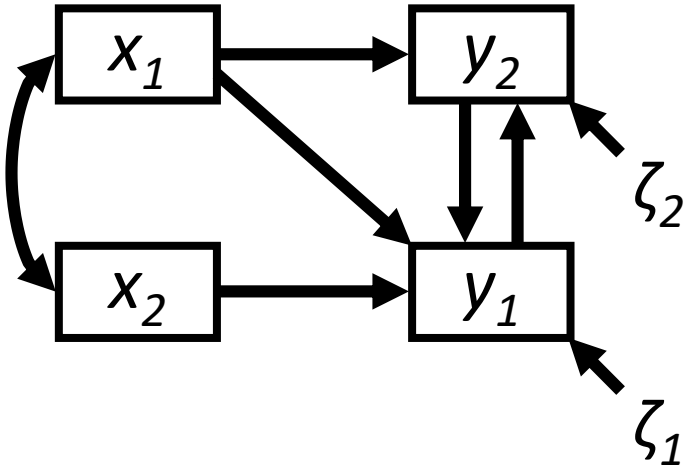
Non-recursive



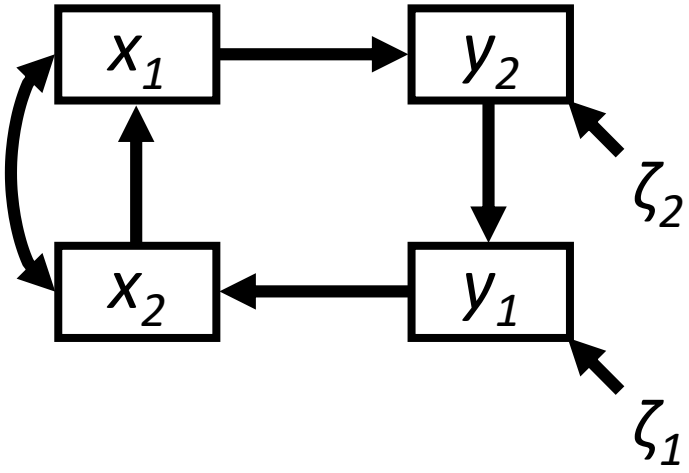
Non-recursive = there is bidirectionality (feedbacks) implicit in the model

2.2 Identification. Loops

Non-recursive

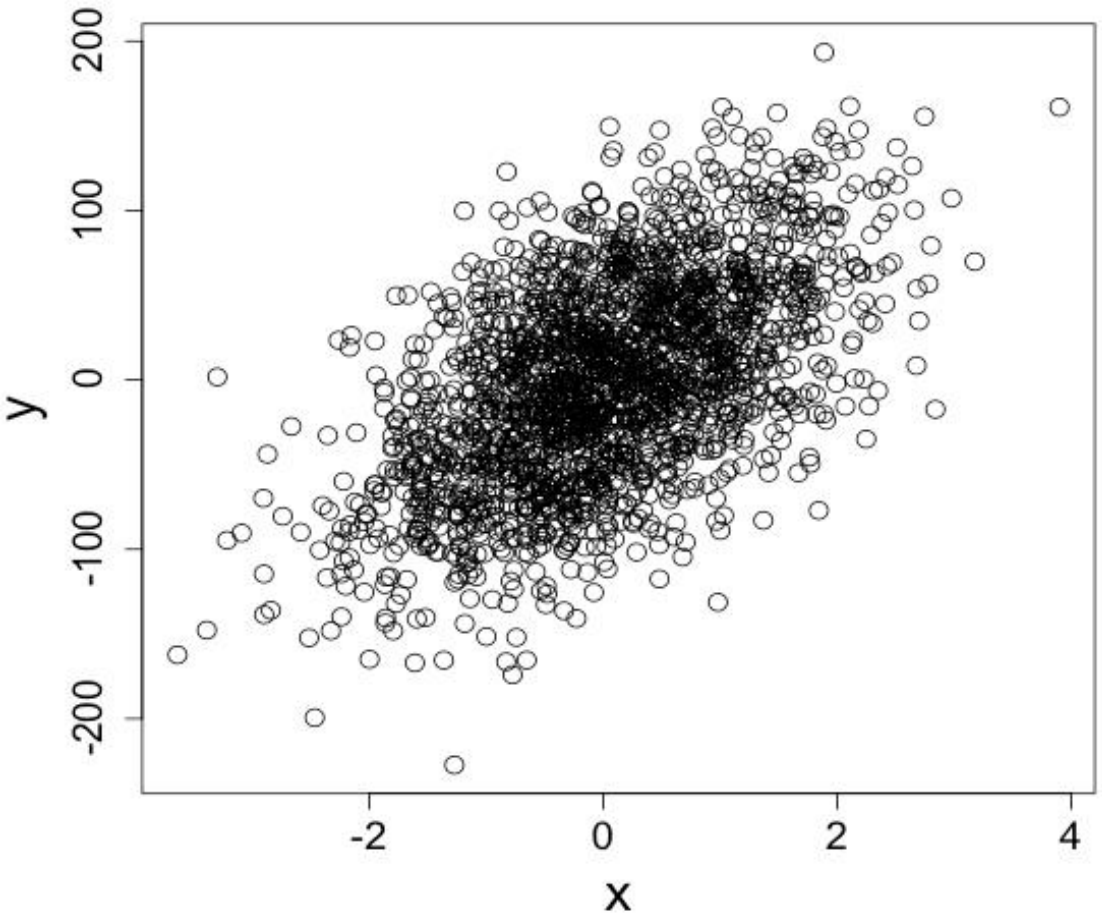


Non-recursive



2.3 Path Coefficients

2.3 Coefficients. Covariance and correlation



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

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

2.3 Coefficients. Covariance and correlation

We often use covariances to fit models, 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

2.3 Coefficients. Standardization

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

$$\beta_{std} = B_{xy} * \frac{sd(x)}{sd(y)}$$

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

2.3 Coefficients. 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

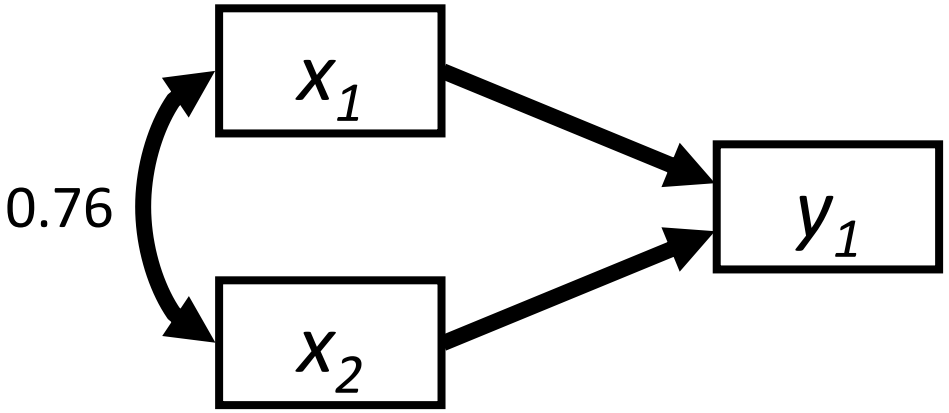
2.3 Coefficients.

The 8 Rules of Path Coefficients

2_The_Basics.R

2.3 Coefficients. Rule #1 of path coefficients

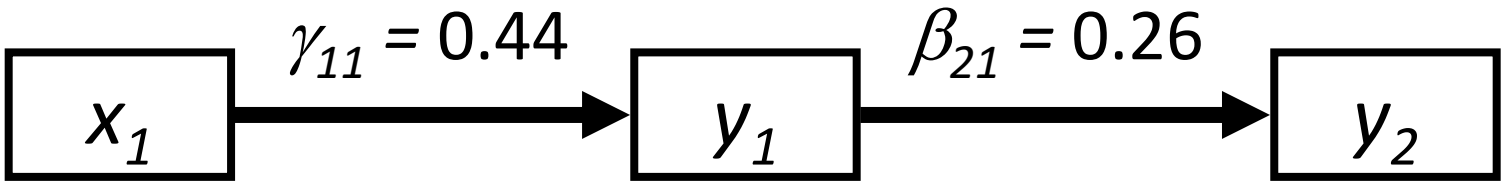
First Rule of Path Coefficients: path coefficients for unanalyzed relationships (double-headed arrows) between exogenous variables are simply the correlations (standardized form) or covariances (unstandardized form)



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

2.3 Coefficients. Rule #2 of path coefficients

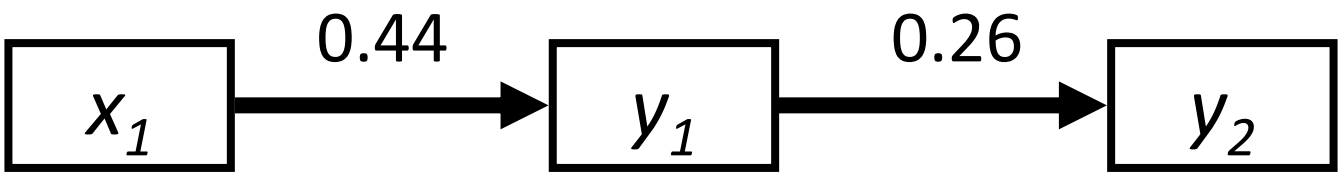
Second Rule of Path Coefficients: when variables are connected by a *single* causal path, the (standardized) path coefficient is simply the correlation coefficient



	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0

2.3 Coefficients. Rule #3 of path coefficients

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

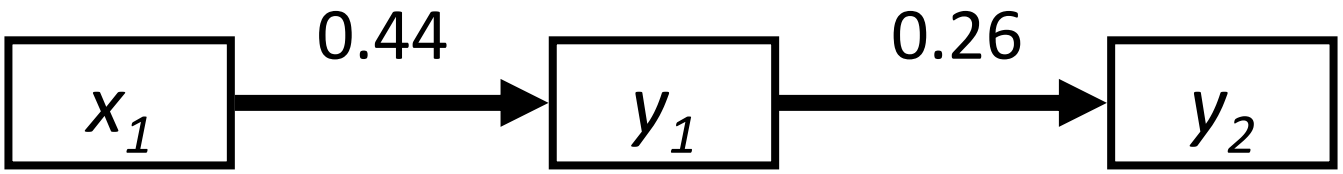


The effect of x_1 on y_2 = 0.44 * 0.26 = 0.11

If 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*.

2.3 Coefficients. Rule #3 of path coefficients

What does it mean when two separated variables are *not* conditionally independent?

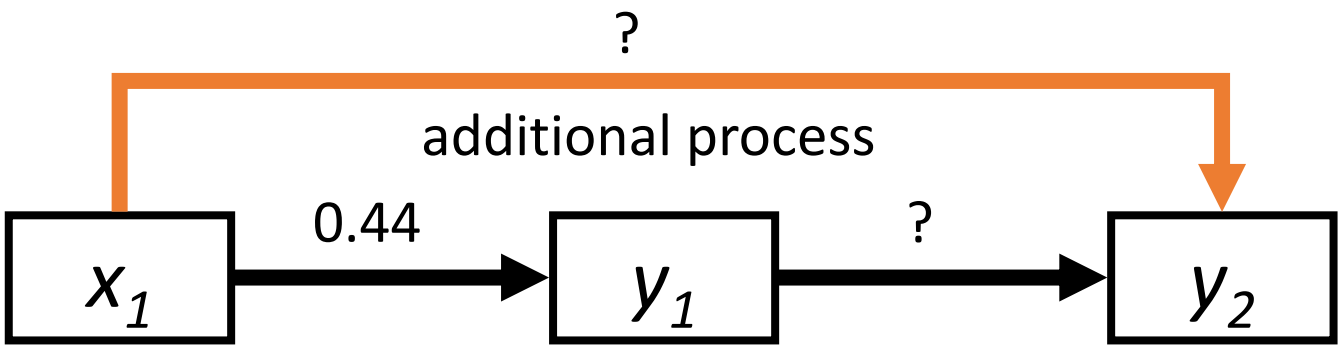


0.44 * 0.26 = 0.11, which is <u>not equal</u> to 0.31

	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0

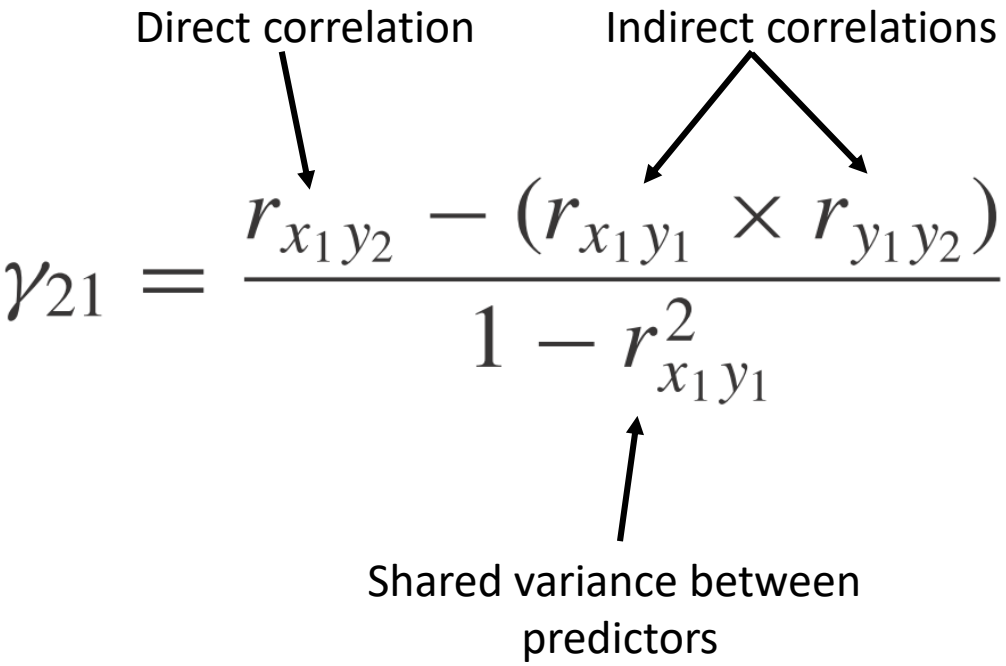
2.3 Coefficients. Rule #4 of path coefficients

The inequality implies that the true model is:



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

2.3 Coefficients. What is a partial coefficient?

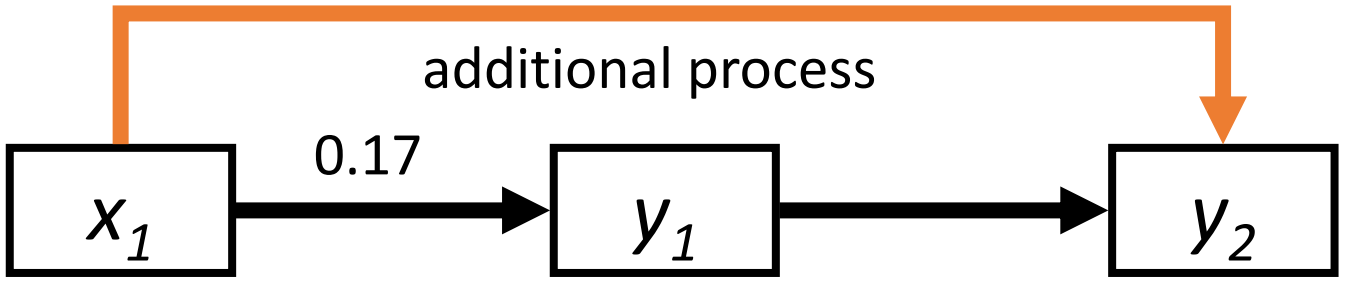


2.3 Coefficients. What is a partial coefficient?

$$\gamma_{21} = \frac{r_{x_1 y_2} - (r_{x_1 y_1} \times r_{y_1 y_2})}{1 - r^2_{x_1 y_1}}$$

	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0

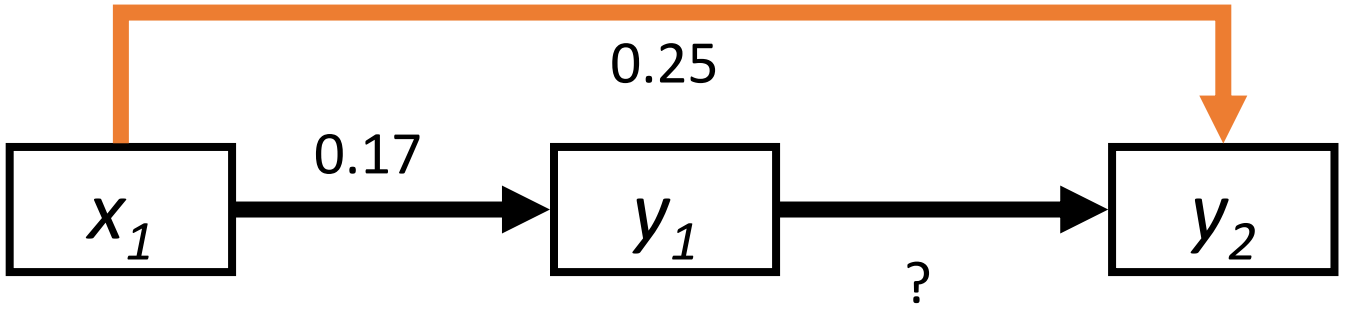
$$\beta_{12} = \frac{(0.31 - (0.26 * 0.44))}{1 - 0.44^2} = 0.25$$



2.3 Coefficients. What is a partial coefficient?

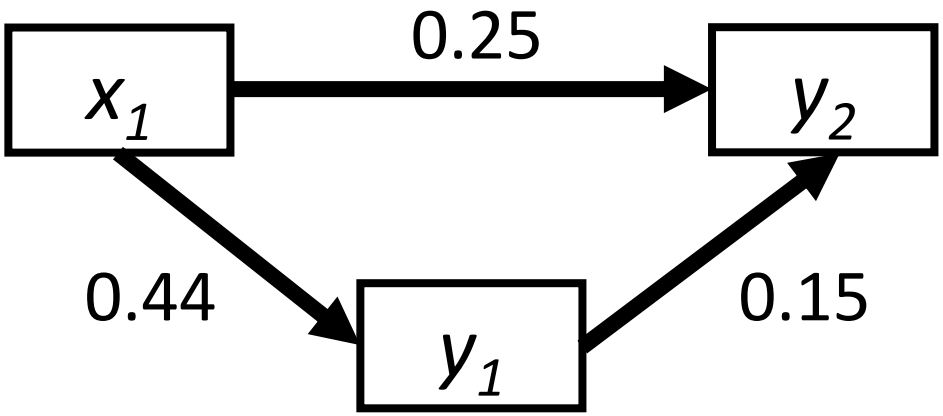
$$\gamma_{21} = \frac{r_{x_1 y_2} - (r_{x_1 y_1} \times r_{y_1 y_2})}{1 - r^2_{x_1 y_1}}$$

	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0



$$\gamma_{21} = \frac{(0.26 - (0.44 * 0.31))}{1 - 0.44^2} = 0.15$$

2.3 Coefficients. Statistical control

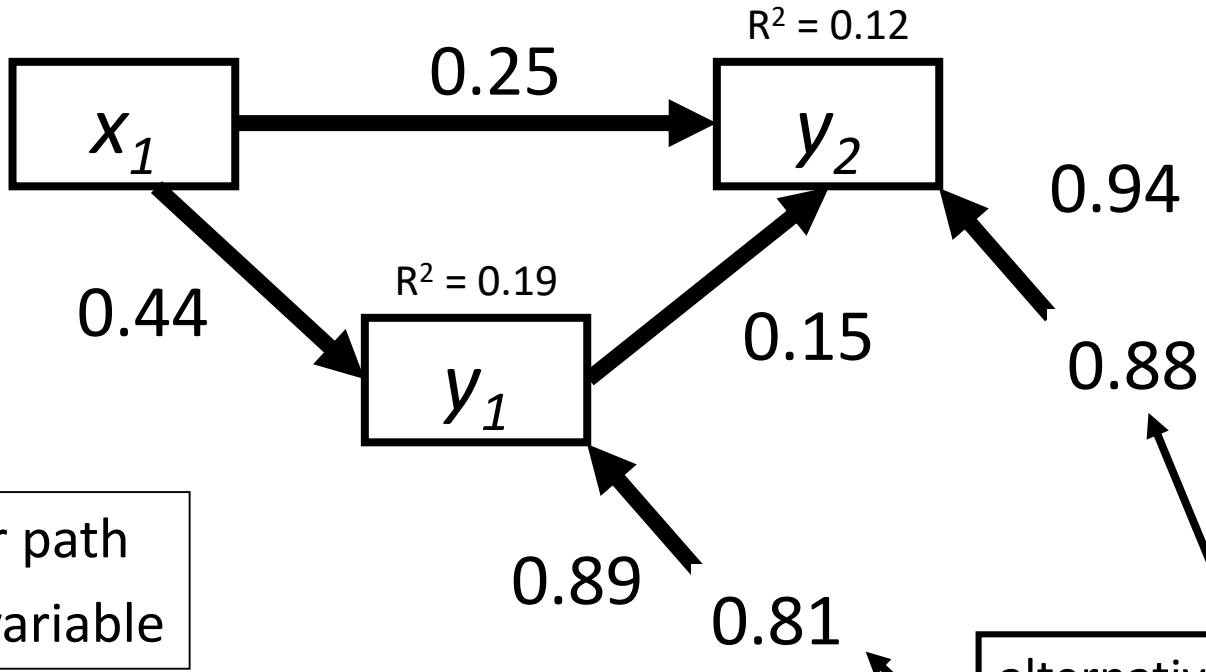


The effect of y_1 on y_2 is controlled for the joint effects of x_1.

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

2.3 Coefficients. Rule #5 of path coefficients

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

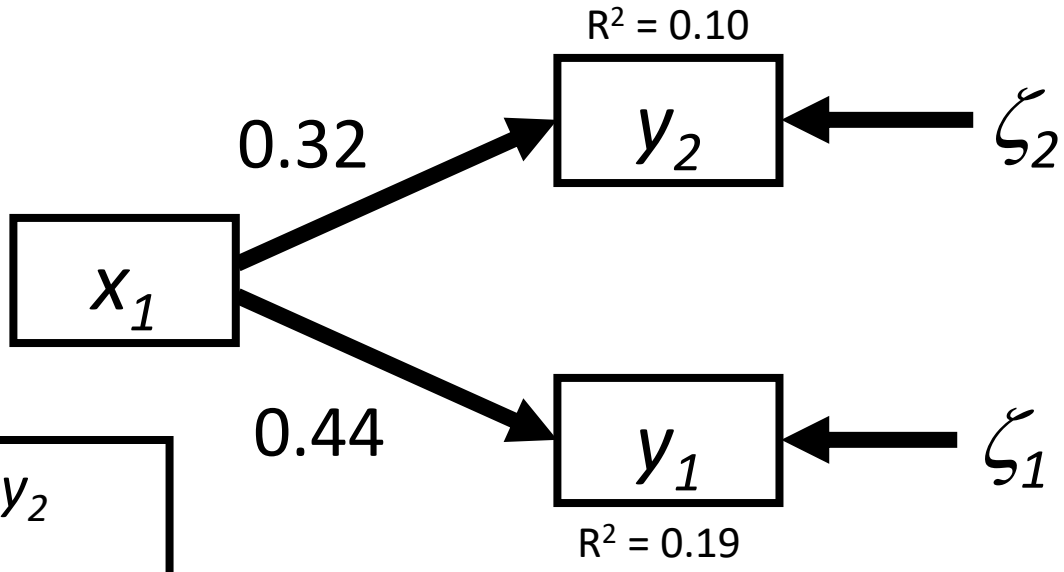


equation for path from error variable

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

alternative is to show values for zetas, which = 1 - R^2

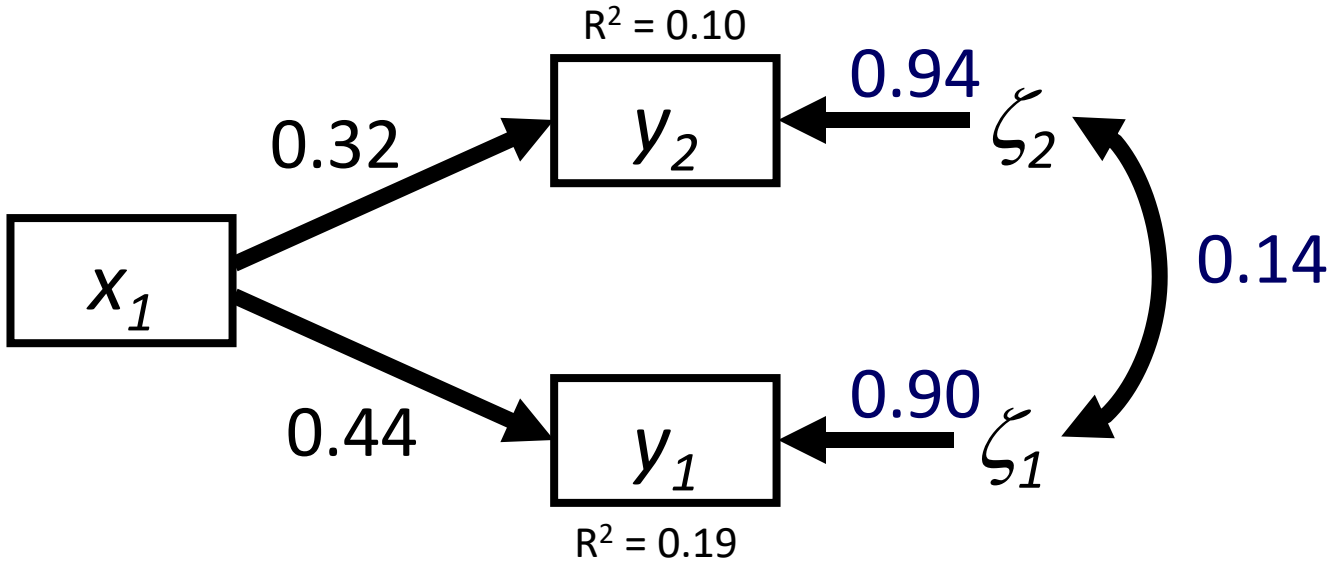
2.3 Coefficients. Rule #6 of path coefficients



	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0

Sixth Rule of Path Coefficients: unanalyzed residual correlations between endogenous variables are partial correlations or covariances.

2.3 Coefficients. Rule #6 of path coefficients



The partial correlation between y_1 and y_2 is typically represented as a correlated error term:

This implies that some other factor is influencing y_1 and y_2

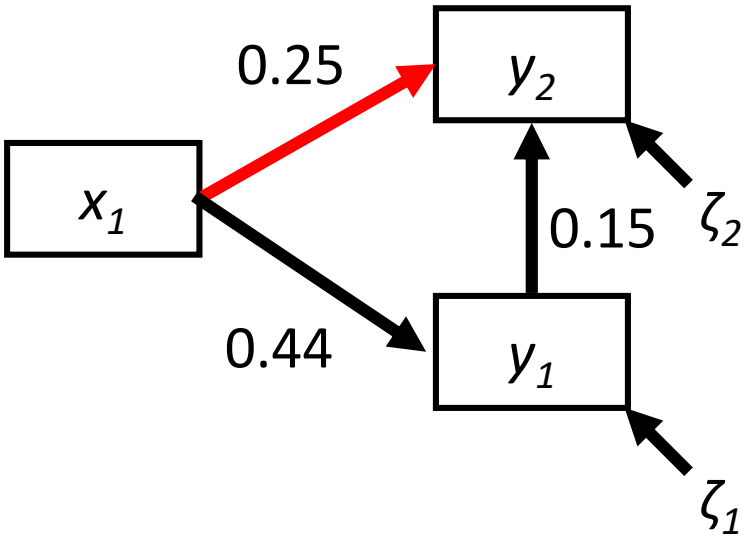
Note that total correlation between y_1 and y_2 =

0.32 * 0.44 + 0.94 * 0.14 * 0.90 = 0.26 (the observed corr)

2.3 Coefficients. Rule #7 of path coefficients

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

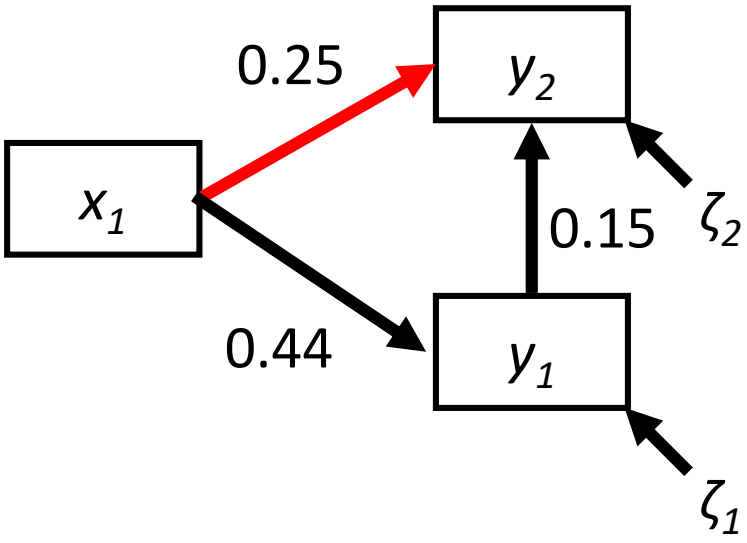
Total Effects:



0.15 + 0.44 * 0.25 = 0.31

2.3 Coefficients. Rule #8 of path coefficients

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



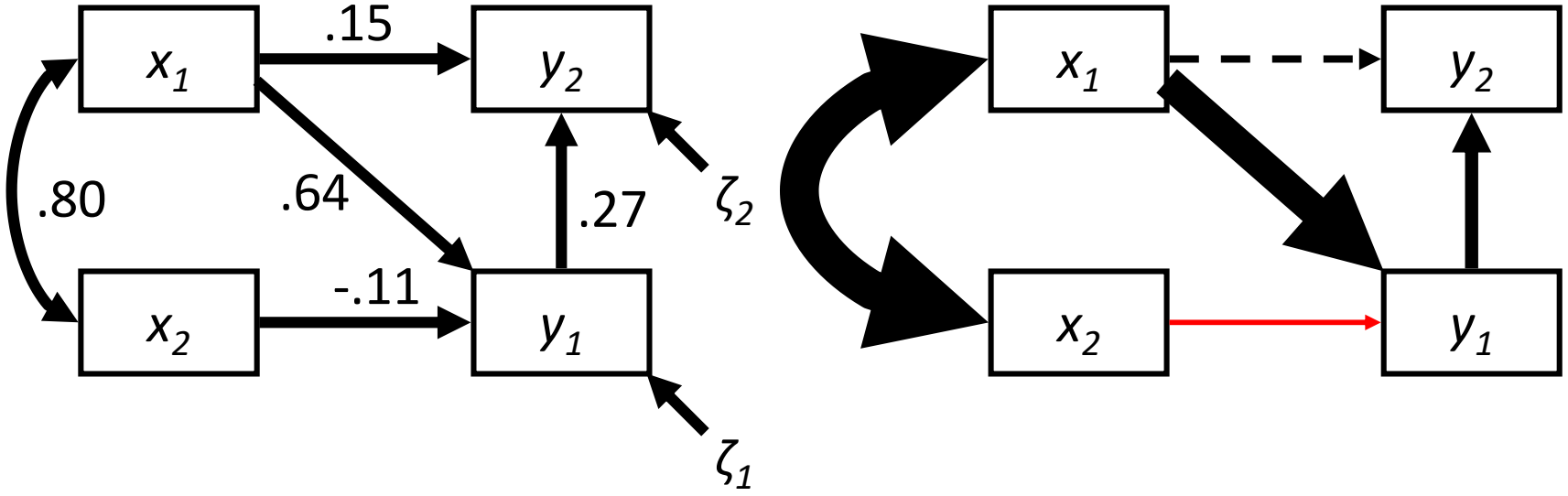
Total Effects:

	x_1	y_1	y_2
x_1	1.0		
y_1	0.44	1.0	
y_2	0.31	0.26	1.0

0.15 + 0.44 * 0.25 = 0.31

2.3 Coefficients. Presentation conventions

- Arrow width is scaled by the size of the effect
- Arrow color = direction of effect
- Dashed lines = non-significant paths
- Color denotes positive/negative
- Coefficients reported on diagram or in table



2.3 Coefficients. Summary

- The path coefficient represents the *independent contribution* of the predictor on the response = statistical controls
- The *indirect effects* can be quantified by multiplying compound paths
- Partial correlations among exogenous variables is the *correlated error* representing a joint external influence

2.3 Coefficients. Presentation

- A few options for presentation:
  - *semPlots* package in R
  - Powerpoint
  - Adobe Illustrator/Inkscape


BREAK TIME
CATNIP IF YOU GOT 'EM
quickmeme.com