Introduction to Path Analysis

Session 2, Lecture 2
11/01/06

Outline

- Path Diagrams
- Direct, Indirect, Total Effects
- Recursive and Non-recursive models
- Identification
- Estimation

Path diagrams: pictoral representations of associations

Key characteristics:

1. As developed by Wright, refer to models that are linear in the parameters (but they can be nonlinear in the variables) $Y = \beta_0 + \beta_1 x^2$ (OK)

   $Y = \beta_0 + \beta_1^2 x$ (not OK)

1. Exogenous variables: their causes lie outside the model
2. Endogenous variables: determined by variables within the model
3. May or may not include latent variables

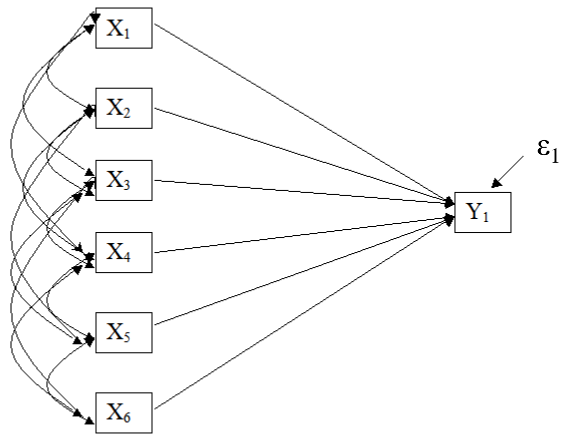
(right now, only manifest (observed) variables)

Standard equation format for a regression equation:

$y_1 = \alpha + \gamma_{11}x_1 + \gamma_{12}x_2 + \gamma_{13}x_3 + \gamma_{14}x_4 + \gamma_{15}x_5 + \gamma_{16}x_6 + \zeta_1$

Path model depicting a regression equation:

$$y_1 = \alpha + \gamma_{11}x_1 + \gamma_{12}x_2 + \gamma_{13}x_3 + \gamma_{14}x_4 + \gamma_{15}x_5 + \gamma_{16}x_6 + \varepsilon_1$$



Common notation:

1. Noncausal associations between exogenous variables indicated by two-headed arrows

2. Causal associations represented by unidirectional arrows extended from each determining variable to each variable dependent on it

3. Residual variables are represented by unidirectional arrows leading from the residual variable to the dependent variable

Notation:
X – Exogenous observed variable
Y – Endogenous observed variable

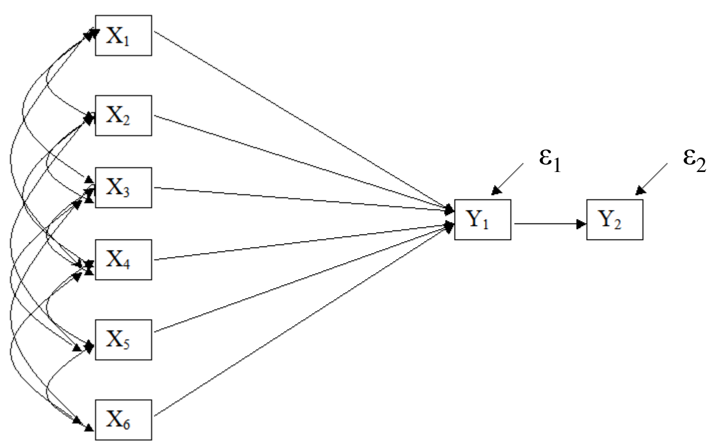
Gamma (γ): Coefficient of association between an exogenous and endogenous variable

Beta (β): Coefficient of association between two endogenous variables

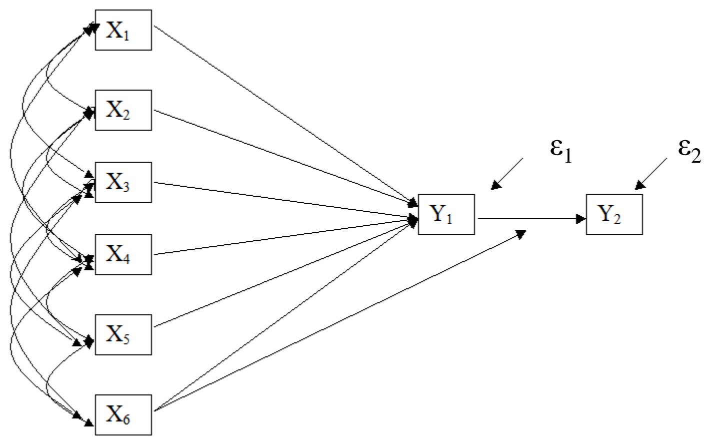
Epsilon (ε) Error term for observed endogenous variable

Subscript protocol: first number refers to the 'destination' variable, while second number refers to the 'origination' variable.

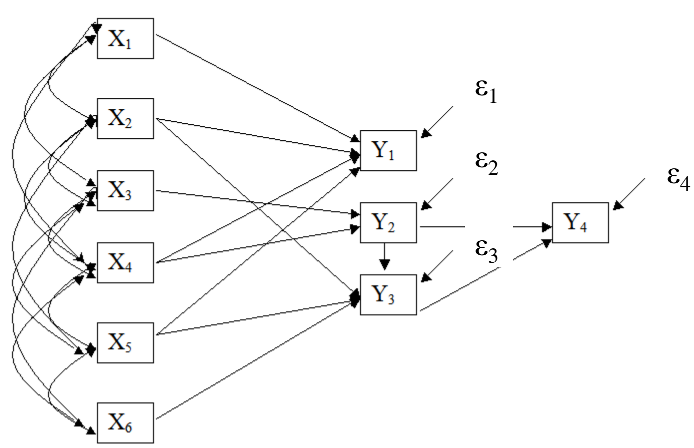
A Path model depicting mediation:



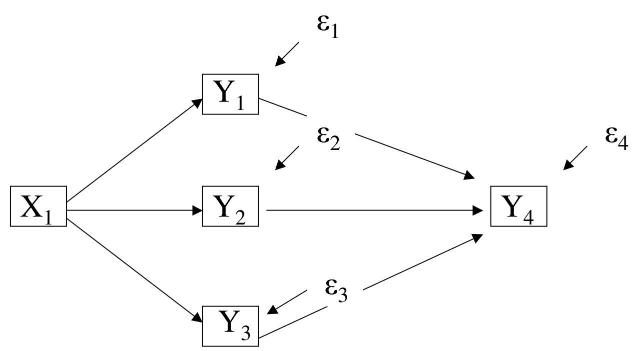
A Path model depicting moderation (AKA interaction, effect modification) :



Another path model depicting mediation:



Yet another path model depicting mediation:



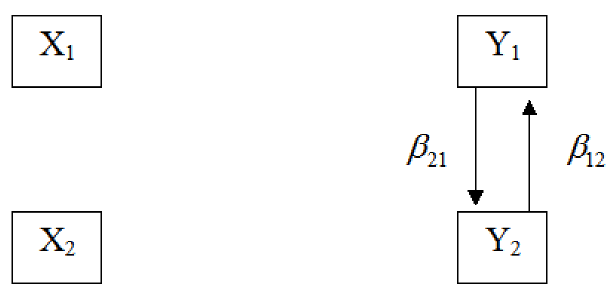
Can you depict a path diagram for this system of equations?

$y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$

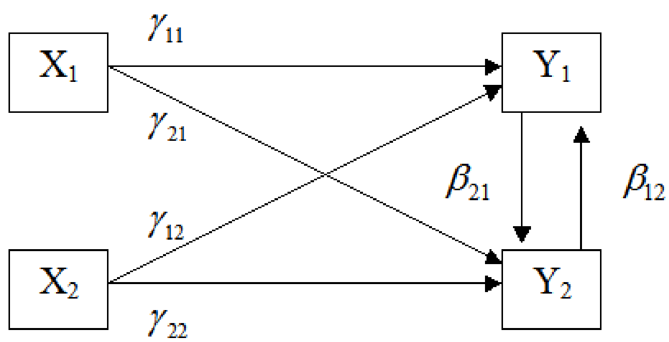
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \beta_{21}y_1 + \varepsilon_2$

$$y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$$
$$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \beta_{21}y_1 + \varepsilon_2$$

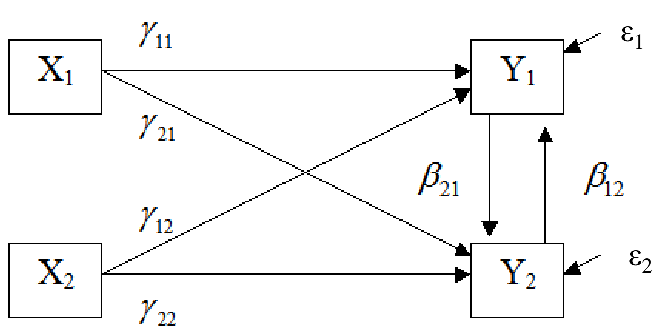
$$y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$$
$$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \beta_{21}y_1 + \varepsilon_2$$



$$y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$$
$$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \beta_{21}y_1 + \varepsilon_2$$



$$y_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$$
$$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \beta_{21}y_1 + \varepsilon_2$$



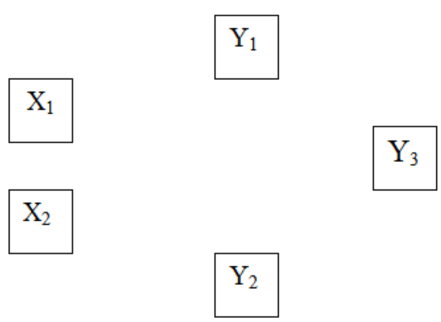
How about this one?

$y_1 = \gamma_{12}x_2 + \beta_{12}y_2 + \zeta_1$

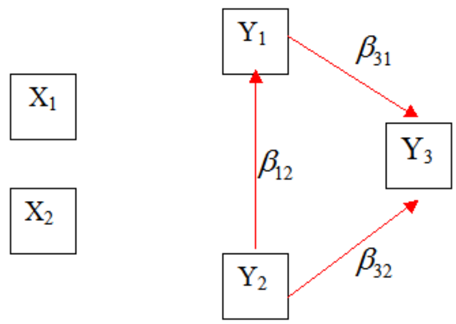
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \zeta_2$

$y_3 = \gamma_{31}x_1 + \gamma_{32}x_2 + \beta_{31}y_1 + \beta_{32}y_2 + \zeta_3$

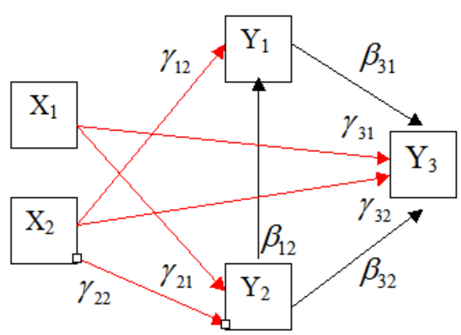
$y_1 = \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \varepsilon_2$
$y_3 = \gamma_{31}x_1 + \gamma_{32}x_2 + \beta_{31}y_1 + \beta_{32}y_2 + \varepsilon_3$



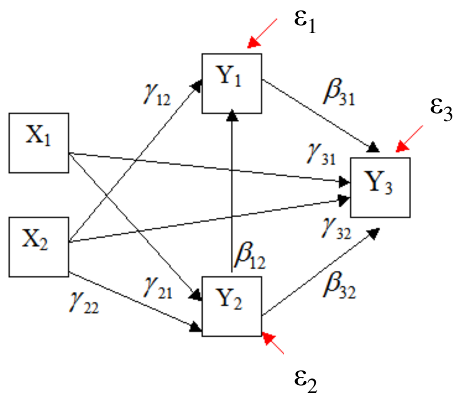
$y_1 = \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \varepsilon_2$
$y_3 = \gamma_{31}x_1 + \gamma_{32}x_2 + \beta_{31}y_1 + \beta_{32}y_2 + \varepsilon_3$



$y_1 = \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \varepsilon_2$
$y_3 = \gamma_{31}x_1 + \gamma_{32}x_2 + \beta_{31}y_1 + \beta_{32}y_2 + \varepsilon_3$



$y_1 = \gamma_{12}x_2 + \beta_{12}y_2 + \varepsilon_1$
$y_2 = \gamma_{21}x_1 + \gamma_{22}x_2 + \varepsilon_2$
$y_3 = \gamma_{31}x_1 + \gamma_{32}x_2 + \beta_{31}y_1 + \beta_{32}y_2 + \varepsilon_3$



Key assumptions of path analysis:

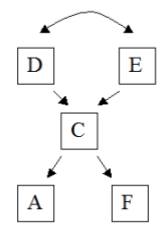
1. $E(\varepsilon_i)=0$: mean value of disturbance term is 0

2. $cov(\varepsilon_i, \varepsilon_j)=0$: no autocorrelation between the disturbance terms

3. $var(\varepsilon_i|X_i)= \sigma^2$

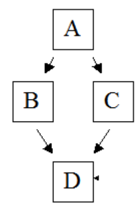
4. $cov(\varepsilon_i,X_i)=0$

Indirect paths: three rules for their computation

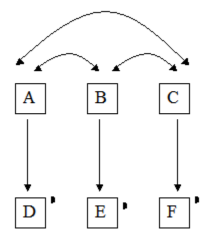


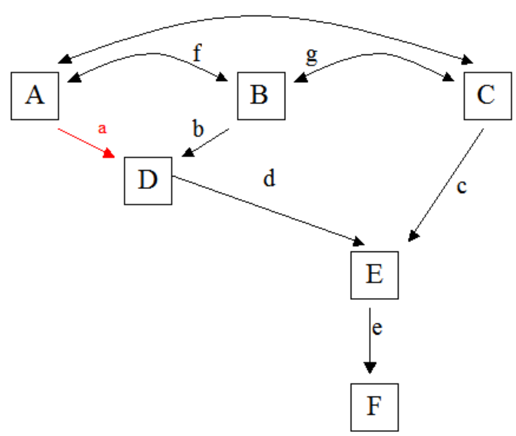
1) no loops – can’t go through the same variable twice – ACF is OK, ACDECF not OK

2) no going forward then backward. backward than forward is OK (events can be connected by common causes, but not common outcomes)
BAC is OK, BDC not OK

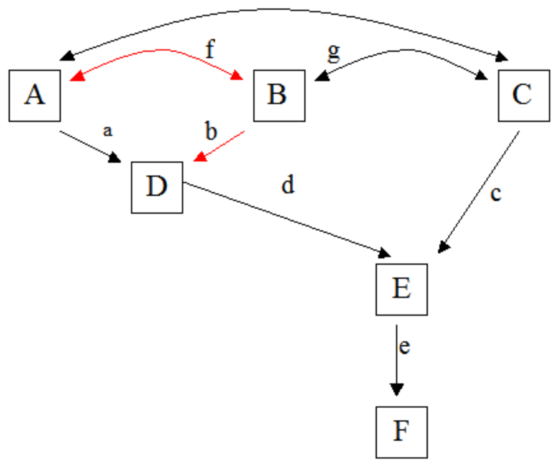


3) a maximum of one curved arrow per path
DACF OK, DABCF not OK
DABE OK, DACBE not OK

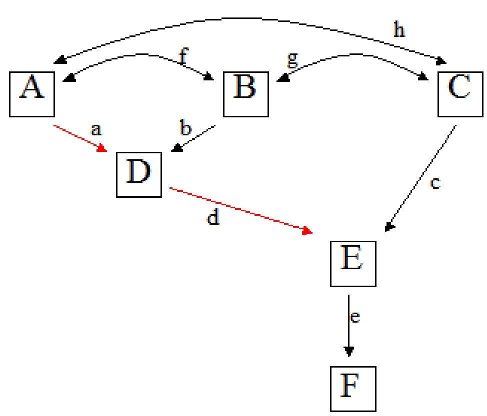




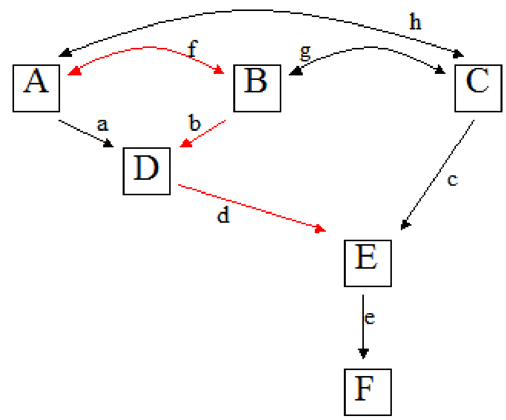
What is total effect of A on D? a



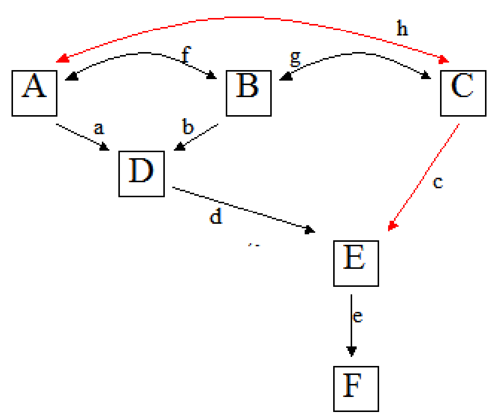
What is total effect of A on D? a + fb



What is total effect of A on E? ad



What is total effect of A on E?   ad + fbd



What is total effect of A on E?   ab + fbd + hc

Optional Home Exercise (Answers posted online Friday)

1. A on F
2. D on E
3. E on F
4. B on F
5. C on F
6. D on F

A path model is a structural equation model if it depicts a causal relationship (as defined by Bollen)

1. Not all path models are structural equation models
2. Not all structural equation models are represented as path models
3. Whether or not a model is a structural equation model depends on the researcher's theory

4. Structural equation models require prior information on causality, or a strong theory, particularly when using cross-sectional data

5. Theory is paramount for a structural equation model

6. Good structural equation models represent causal paths that are "undebatable"

7. Path diagrams and structural equation models highlight model assumptions

Direct, Indirect, and Total Effects:

Direct effects: association of one variable with another (a single arrow between variables)

Indirect effects: indirect paths composed entirely of direct effects (i.e., no double-headed arrows) (because SEMs are CAUSAL models)

Total effect: direct effect plus indirect effect(s)

Direct, Indirect, and Total Effects, Example:

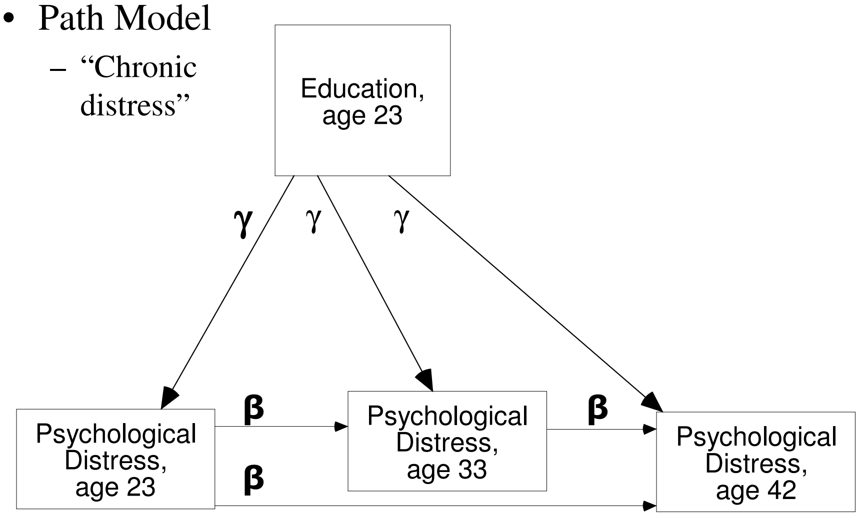
- Does association of education with adult depression stem from chronic or short-term depression?
  - Chronic depression: same people account for depression disparities as a cohort ages
  - Short-term depression: depression disparities represent staggered onset of short-term depression

Direct, Indirect, and Total Effects, Example:

- Most longitudinal studies show that education affects depression, but not vice-versa
- Use data from the National Child Development Survey, which assessed a birth cohort of about 10,000 individuals for depression at age 23, 33, and 42.

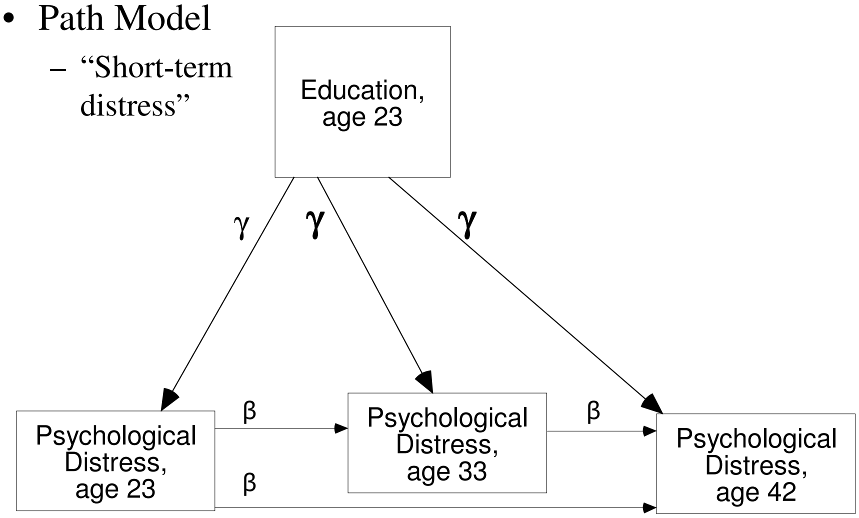
Direct, Indirect and Total Effects, example

- Path Model
  - "Chronic distress"



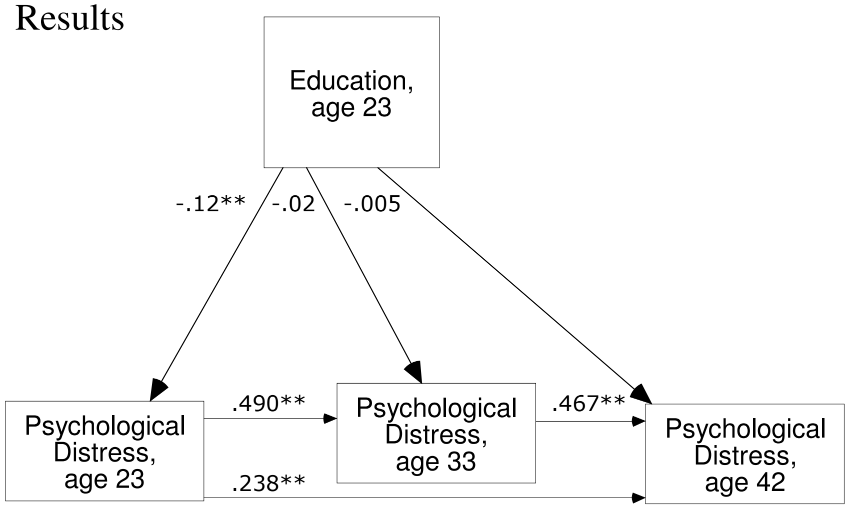
Direct, Indirect and Total Effects, example

- Path Model
  - "Short-term distress"



Direct, Indirect and Total Effects, example

Results



Estimation of Path Diagrams:

$$\Sigma = \Sigma(\theta)$$

Where Σ (sigma) is the population covariance matrix of observed variables and Σ(θ) is a covariance matrix written as a function of the model parameters (θ)

Rules for Variances

- $Var(A + B)=Var(A)+Var(B)+2Cov(A,B)$
- $Var(A - B)= Var(A)+Var(B)-2Cov(A,B)$

Let c be any constant (like, 5)

- – $Var(c+A)=0 + Var(A)$ (constants don't vary)
- – $Var(cA) = c^2Var(A)$

Rules for Coariances

- $Cov(X,Y)=E(XY)-E(X)E(Y)$

- $Cov(c,X)=0$
- $Cov(cX,Y)=cCov(X,Y)$
- $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$

Estimation of Path Diagrams

1) Solve system of equations

$$
\begin{tikzcd}
\boxed{X_1} \arrow[r, "\gamma_{11}"] & \boxed{Y_1} \arrow[r, "\beta_{21}"] & \boxed{Y_2} \\
& \varepsilon_1 \arrow[u] & \varepsilon_2 \arrow[u]
\end{tikzcd}
$$

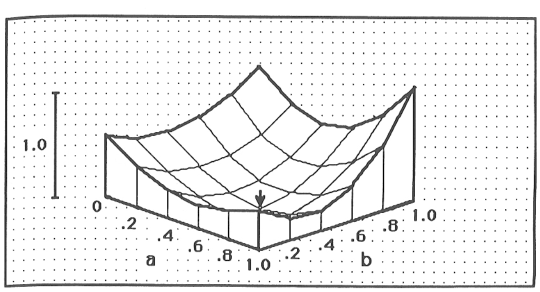
$$Y_1 = \gamma_{11} X_1 + \varepsilon_1$$
$$Y_2 = \beta_{21} Y_1 + \varepsilon_2$$

$$
\begin{bmatrix}
var(Y_1) & & \\
cov(Y_2,Y_1) & var(Y_2) & \\
cov(X_1,Y_1) & cov(X_1,Y_2) & var(X_1)
\end{bmatrix}
=
\begin{bmatrix}
\gamma_{11}^2 var(X_1) + var(\varepsilon_1) & & \\
\beta_{21} var(Y_1) & \beta_{21}^2 var(Y_1) + var(\varepsilon_2) & \\
\gamma_{11} var(X_1) & \beta_{21}\gamma_{11} var(X_1) & var(X_1)
\end{bmatrix}
$$

Estimation of Path Diagrams:

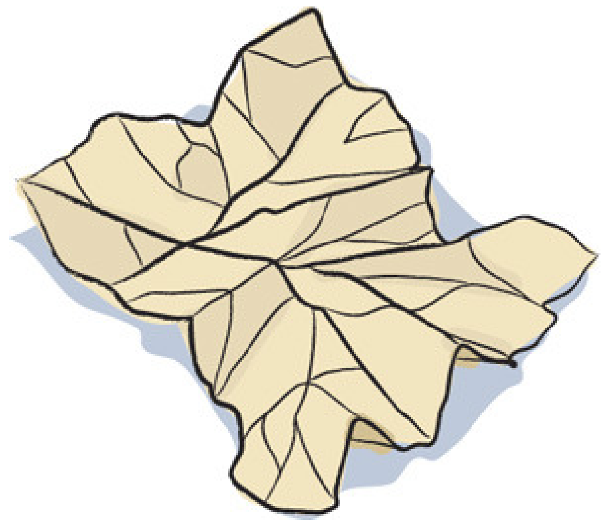
2) Write a computer program to estimate every single possible combination of parameters possible, and see which fits best

3) Use an iterative procedure:



(graph from Loehlin, 1998)

What your parameter space probably *really* looks like…



Identifiability

- Identification: A model is identified if is theoretically and empirically (estimability) possible to estimate one and only one set of parameters.

- T-rule
- Null-B rule
- Recursive Rule

T-Rule

- necessary, but not sufficient rule
- the number of unknown parameters to be solved for cannot exceed the number of observed variances and covariances.
- number of parameters to be estimated:
  - variances of exogenous variables
  - variances of errors for endogenous variables
  - direct effects and double-headed arrows
- number of variances and covariances computed as: n(n+1)/2, where n is the number of observed exogenous and endogenous variables.

T-Rule (why?)

- Why n(n+1)/2?
- The top matrix has 4*4=16 elements, but 6 are redundant, so that leaves you with 10.

$$\begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \sigma_{x_1y_1} & \sigma_{x_1y_2} \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & \sigma_{x_2y_1} & \sigma_{x_2y_2} \\ \sigma_{x_1y_1} & \sigma_{x_2y_1} & \sigma_{y_1}^2 & \sigma_{y_1y_2} \\ \sigma_{x_1y_2} & \sigma_{x_2y_2} & \sigma_{y_1y_2} & \sigma_{y_2}^2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma_{x_1}^2 & & & \\ \sigma_{x_1x_2} & \sigma_{x_2}^2 & & \\ \sigma_{x_1y_1} & \sigma_{x_2y_1} & \sigma_{y_1}^2 & \\ \sigma_{x_1y_2} & \sigma_{x_2y_2} & \sigma_{y_1y_2} & \sigma_{y_2}^2 \end{bmatrix}$$

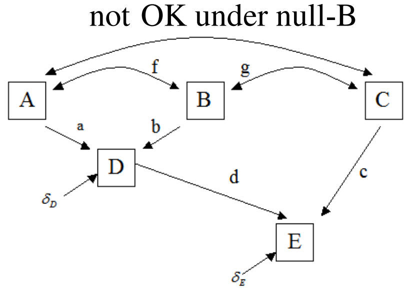
Two forms of T-rule

- Bollen: $n(n+1)/2 \geq$ number of exogenous variances + number of error variances + number of direct effects + number of double-headed arrows.
- Maruyama: $n(n-1)/2 \geq$ number of direct effects + number of double-headed arrows.
- Both yield the same result.

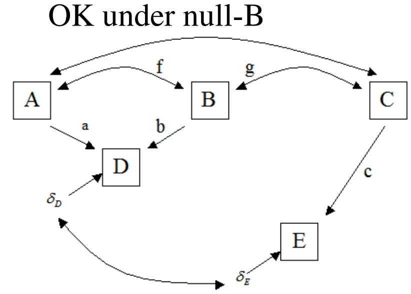
Null-B

- a sufficient, but not necessary
- when no endogenous variable affects any other endogenous variable then the model is identified.
  - e.g. No Betas.
- error of endogenous variables can be correlated and the model will still be identified, as long as the endogenous variables themselves are not correlated.

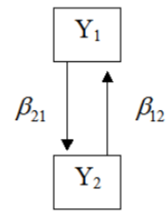
not OK under null-B



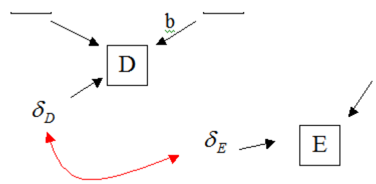
OK under null-B



Recursive



- sufficient, but not necessary
- To be recursive, must have NEITHER
  - reciprocal causation
  - NOR correlated disturbances (AKA errors, i.e., deltas)



Is this model identified?

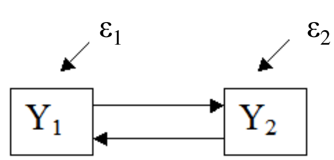
~~Null B Rule~~

~~Recursive Rule~~

~~T Rule:~~

vars of exog. vars: 0
vars of errors for endog: 2
direct effects: 2
double-headed arrows: 0 +
Total to be Estimated: **4**

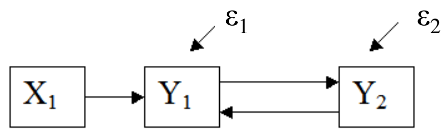
Observed Vars and Covars:
n(n+1)/2 = (2*3)/2 = **3**



~~Null B Rule~~

~~Recursive Rule~~

T Rule:



vars of exog. vars: 1
vars of errors for endog: 2
direct effects: 3
double-headed arrows: 0 +
Total to be Estimated: **6**

Observed Vars and Covars:
$n(n+1)/2 = (3*4)/2 =$ **6**