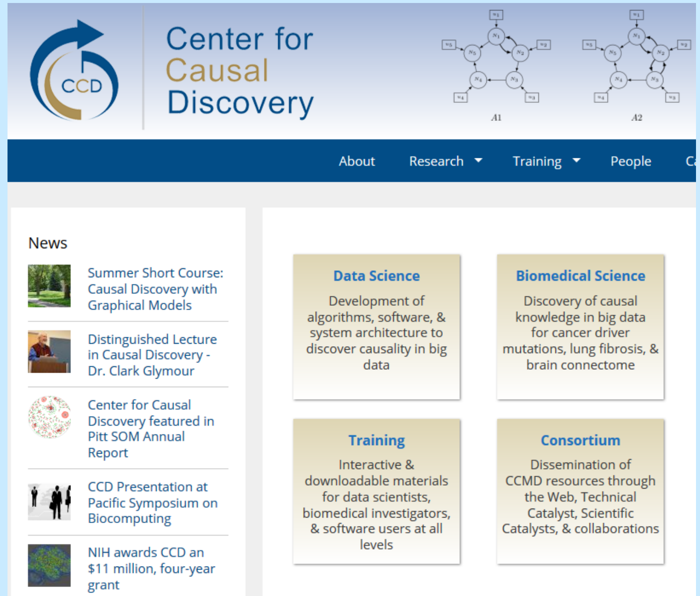
Center for Causal Discovery



Day 3: Search Continued

June 15, 2015

Carnegie Mellon University

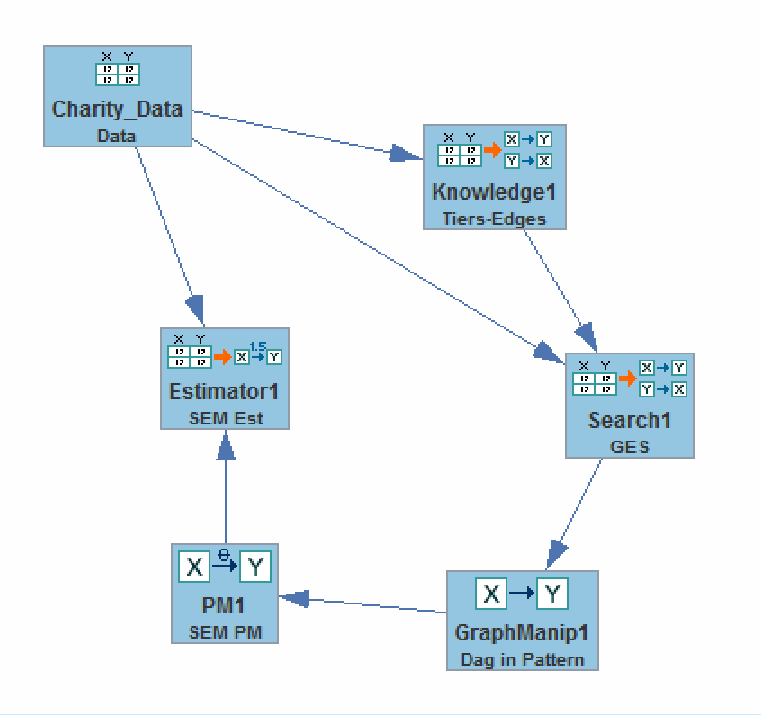
Outline

Models ← Data

1) Bridge Principles: Markov Axiom and D-separation
2) Model Equivalence
3) Model Search
    A. For Patterns
    B. For PAGs
4) Multiple Regression vs. Model Search
5) Measurement Issues and Latent Variables

Search Results?

1) Charitable Giving
2) Lead and IQ
3) Timberlake and Williams



Constraint-based Search for Patterns

1) Adjacency phase

2) Orientation phase

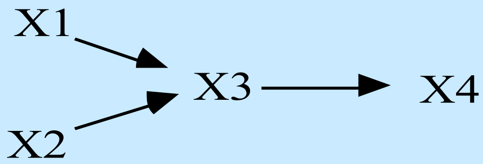
Constraint-based Search for Patterns: Adjacency phase

X and Y are not adjacent if they are independent conditional on any subset that doesn't X and Y

1) Adjacency

- Begin with a fully connected undirected graph
- Remove adjacency X-Y if X _||_ Y | any set S

Causal Graph



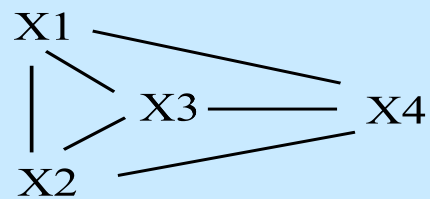
Independencies

$X1 \perp\!\!\!\perp X2$

$X1 \perp\!\!\!\perp X4 \mid \{X3\}$

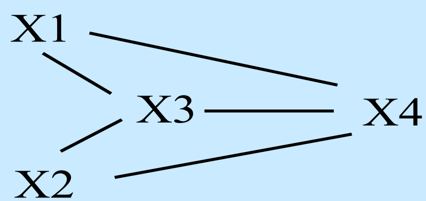
$X2 \perp\!\!\!\perp X4 \mid \{X3\}$

Begin with:



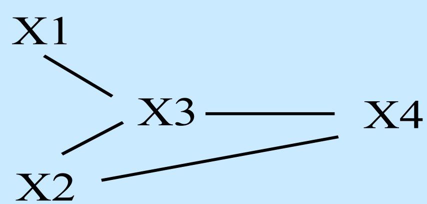
From

$X1 \perp\!\!\!\perp X2$



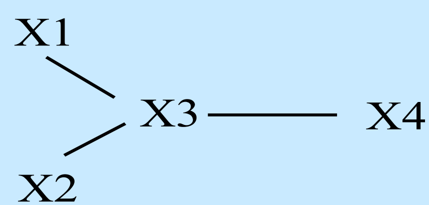
From

$X1 \perp\!\!\!\perp X4 \mid \{X3\}$



From

$X2 \perp\!\!\!\perp X4 \mid \{X3\}$



Constraint-based Search for Patterns: Orientation phase

2) Orientation

- Collider test:
  Find triples X – Y – Z, orient according to whether the set that separated X-Z contains Y

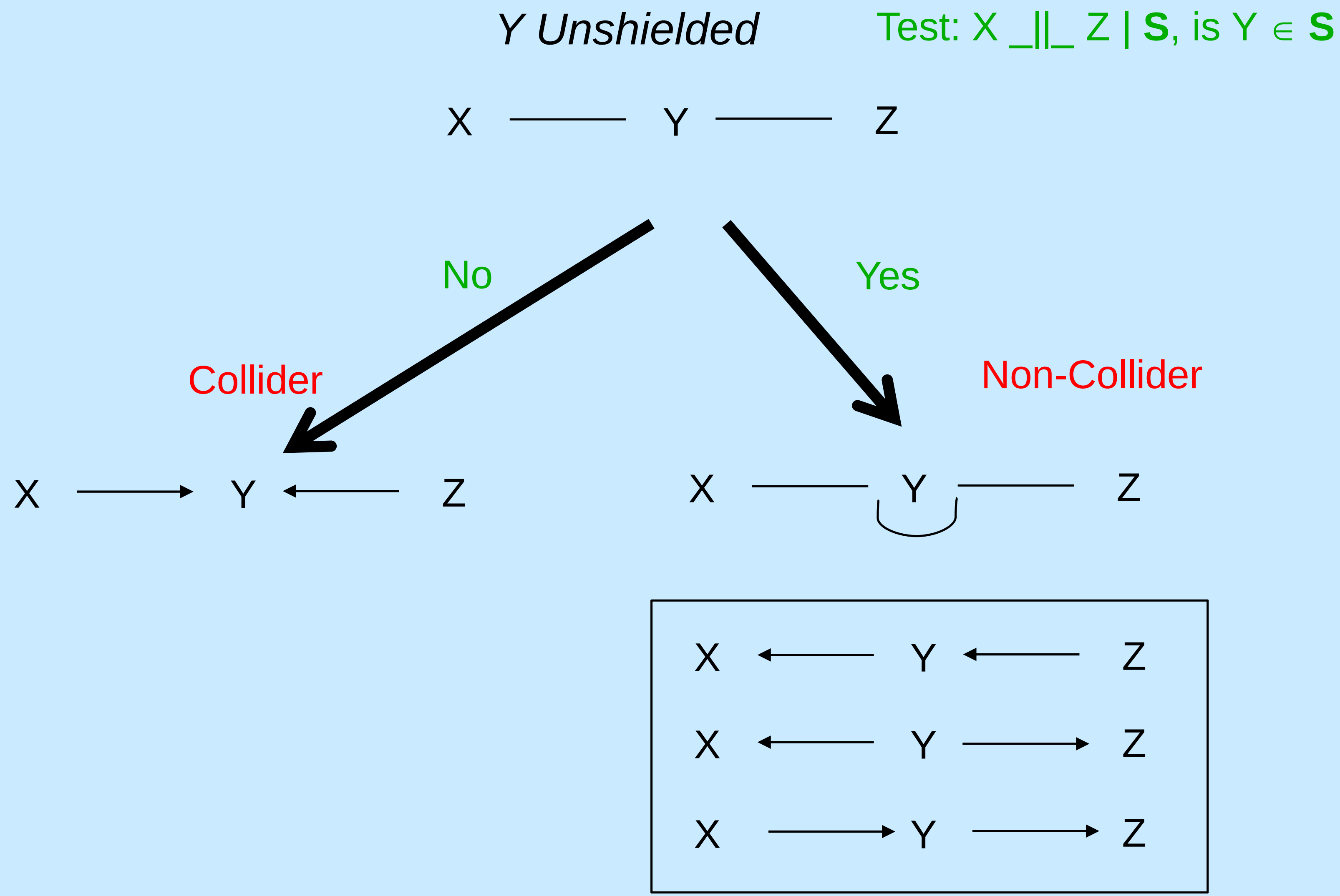
- Away from collider test:
  Find triples X → Y – Z, orient Y – Z connection via collider test

- Repeat until no further orientations

- Apply Meek Rules

Search: Orientation

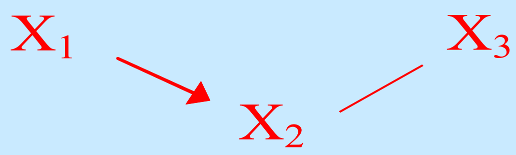
Patterns



Search: Orientation

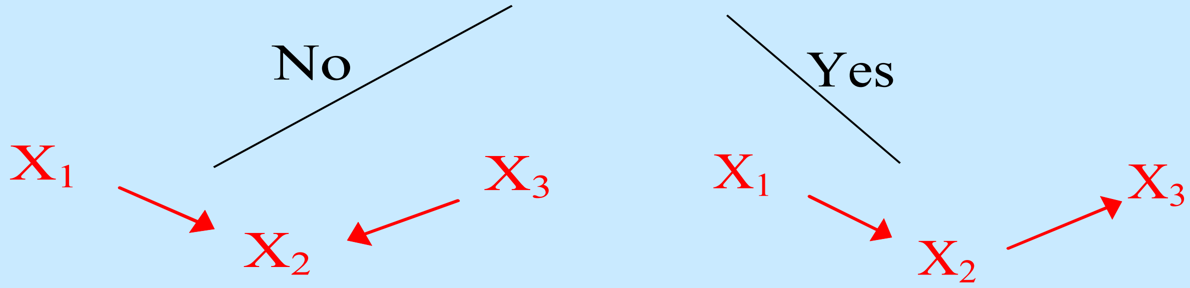
Away from Collider

Test Conditions



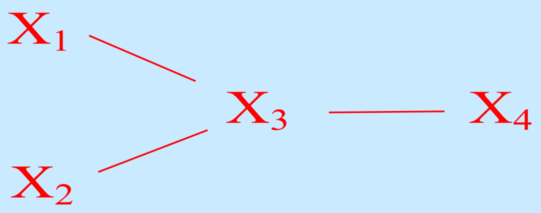
1) X_1 - X_2 adjacent, and *into* X_2.
2) X_2 - X_3 adjacent
3) X_1 - X_3 not adjacent

Test $X_1 \perp\!\!\!\perp X_3 \mid \mathbf{S}, X_2 \in \mathbf{S}$



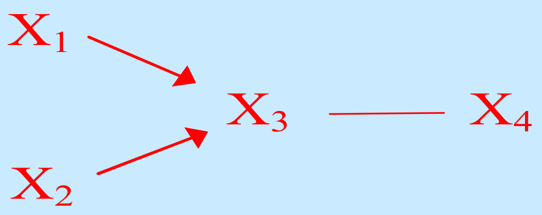
Search: Orientation

After Adjacency Phase



Collider Test: X1 – X3 – X2

X1 _||_ X2

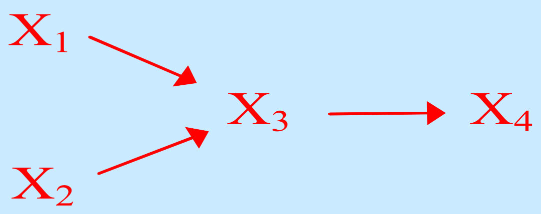


Away from Collider Test:

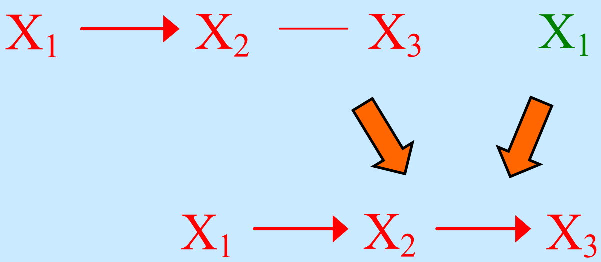
X1 → X3 – X4   *X2 → X3 – X4*

X1 _||_ X4 | X3

X2 _||_ X4 | X3



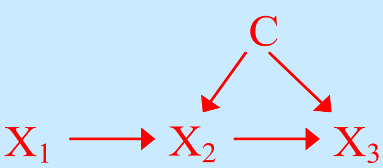
Away from Collider Power!



$X_1 \_||\_ \ X_3 \mid \mathbf{S}, X_2 \in \mathbf{S}$

$X_2 – X_3$ oriented as $X_2 \rightarrow X_3$

Why does this test also show that X_2 and X_3 are *not confounded*?



$X_1 \_||\_ \ X_3 \mid \mathbf{S}, X_2 \in \mathbf{S}$

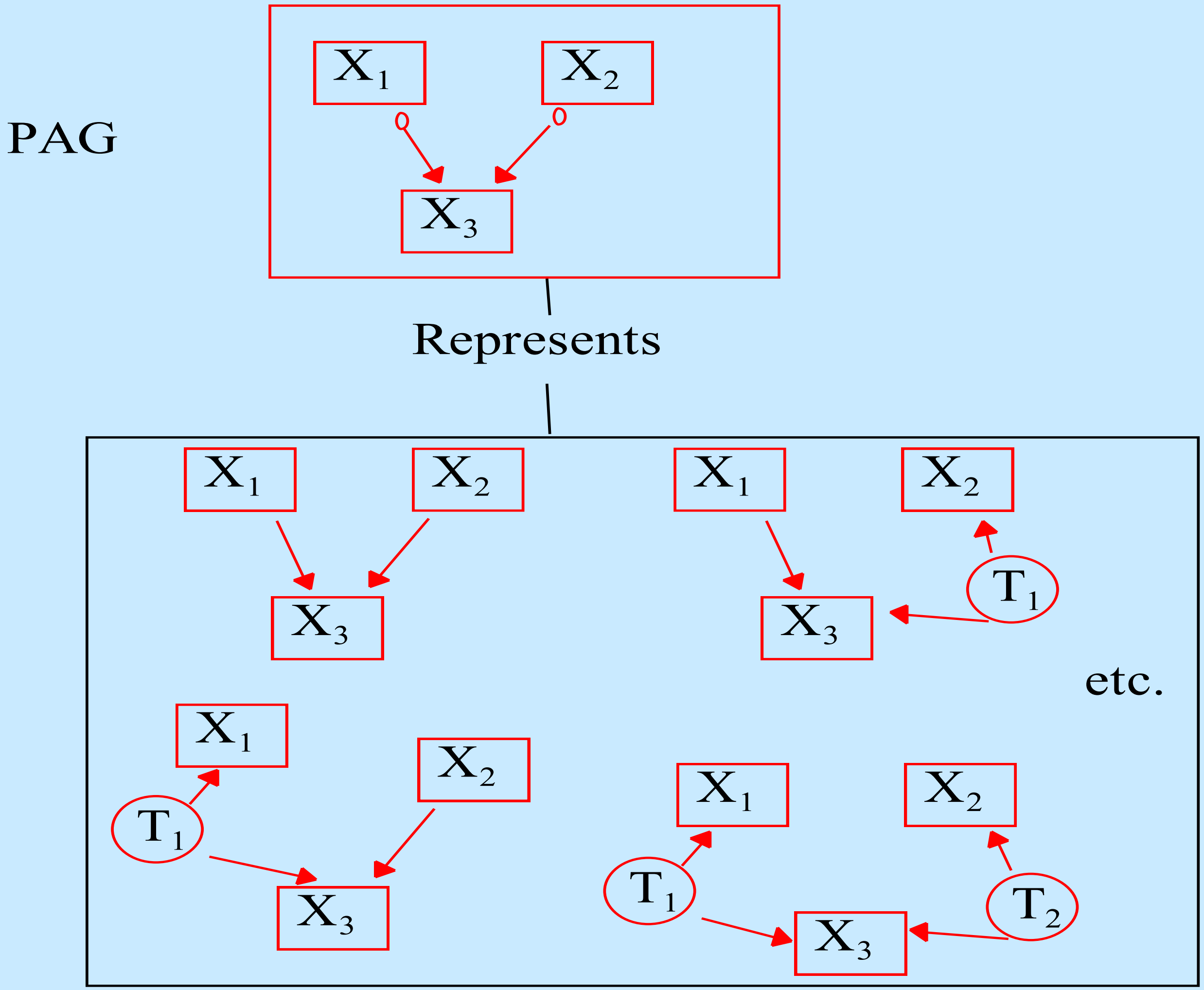
$X_1 \_||\_ \ X_3 \mid \mathbf{S}, X_2 \in \mathbf{S}, \mathbf{C} \notin \mathbf{S}$

Independence Equivalence Classes: Patterns & PAGs

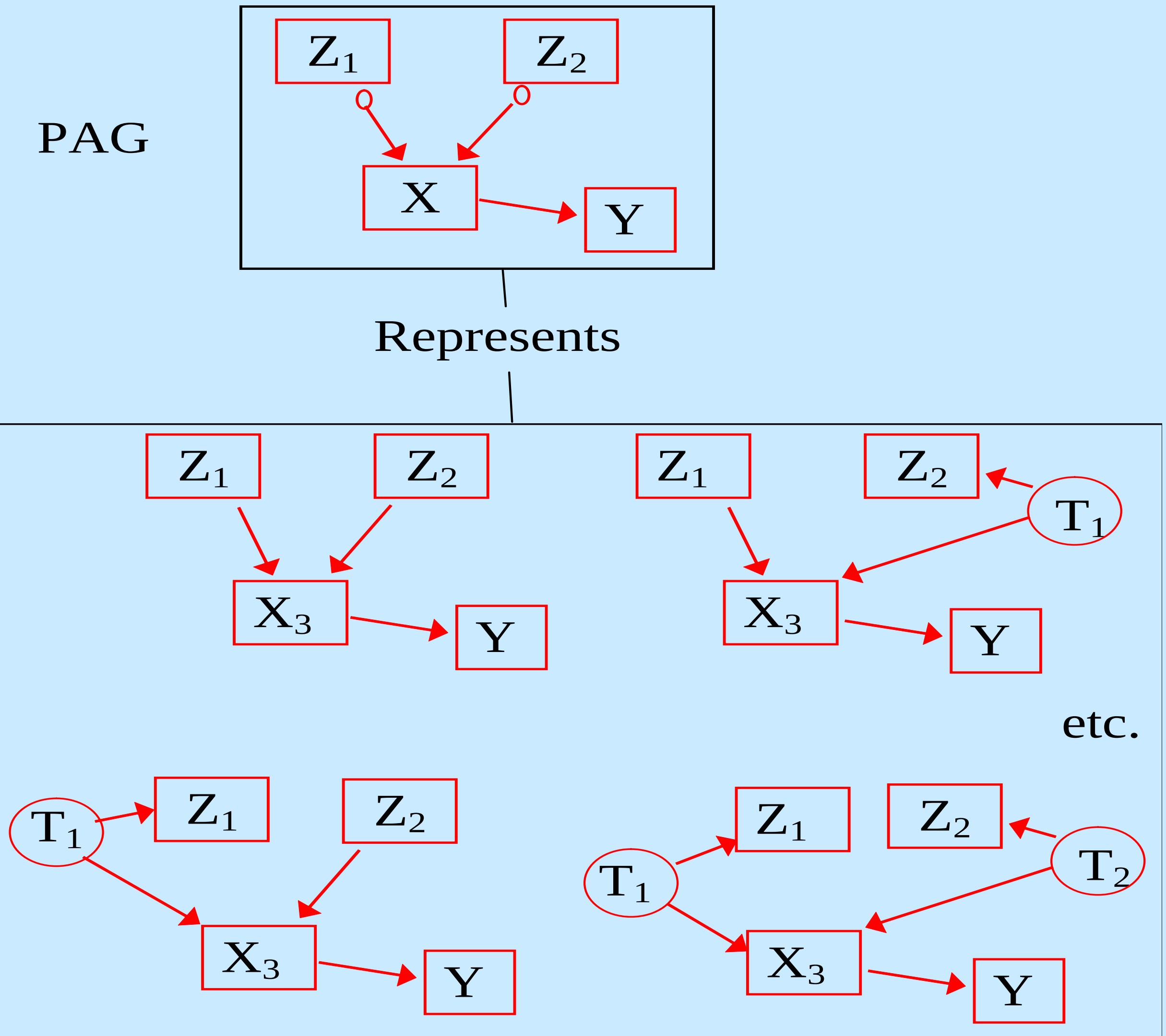
- Patterns (Verma and Pearl, 1990): graphical representation of d-separation equivalence among models with no latent common causes

- PAGs: (Richardson 1994) graphical representation of a d-separation equivalence class that includes models with *latent common causes* and *sample selection bias* that are d-separation equivalent over a set of measured variables **X**

PAGs: Partial Ancestral Graphs

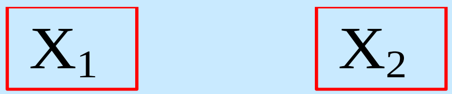


PAGs: Partial Ancestral Graphs

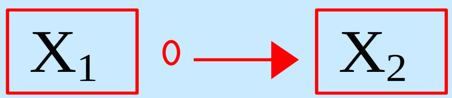


PAGs: Partial Ancestral Graphs

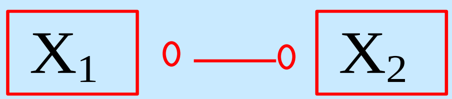
What PAG edges mean.



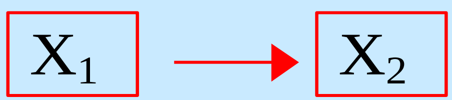
X_1 and X_2 are not adjacent



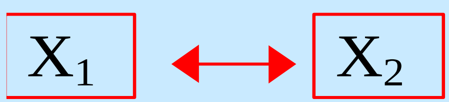
X_2 is not an ancestor of X_1



No set d-separates X_2 and X_1



X_1 is a cause of X_2

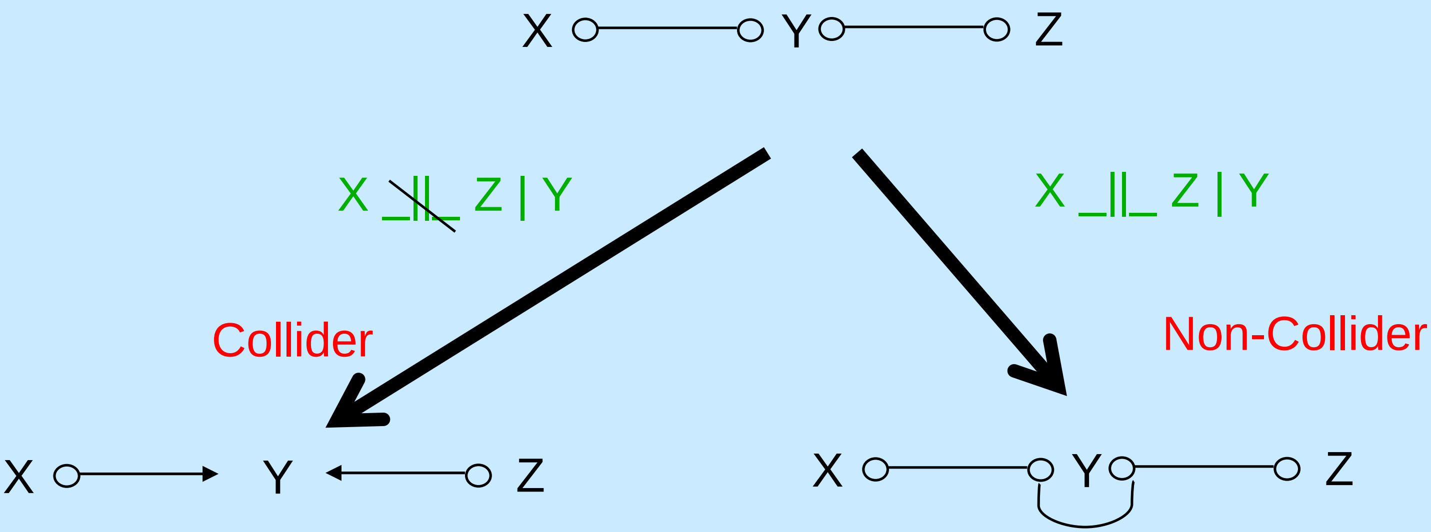


There is a latent common cause of X_1 and X_2

PAG Search: Orientation

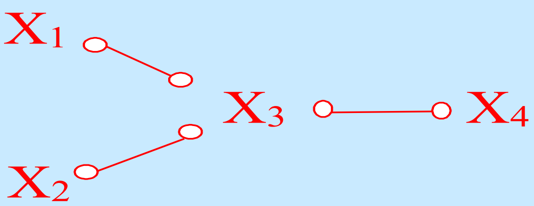
PAGs

Y Unshielded



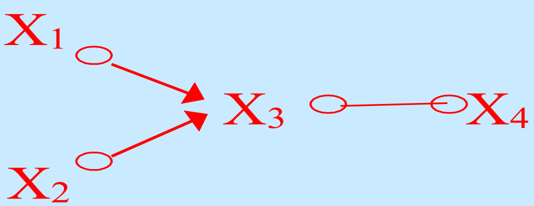
PAG Search: Orientation

After Adjacency Phase



Collider Test: X1 – X3 – X2

X1 _||_ X2

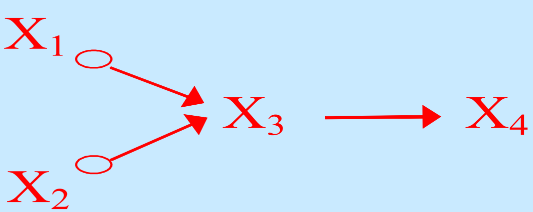


Away from Collider Test:

X1 → X3 – X4   X2 → X3 – X4

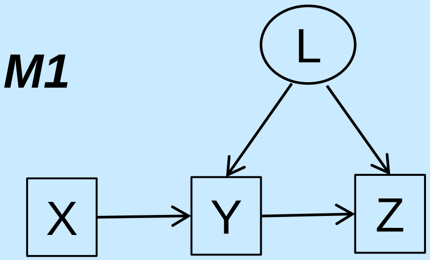
X1 _||_ X4 | X3

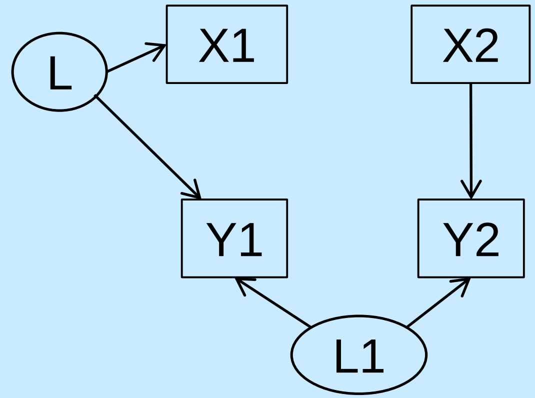
X2 _||_ X4 | X3



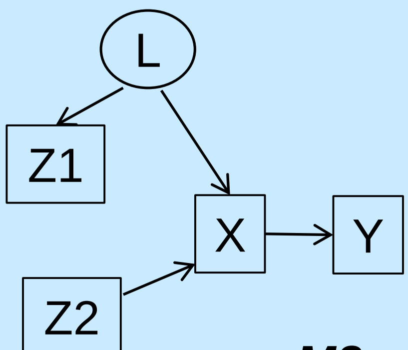
Interesting Cases

M1

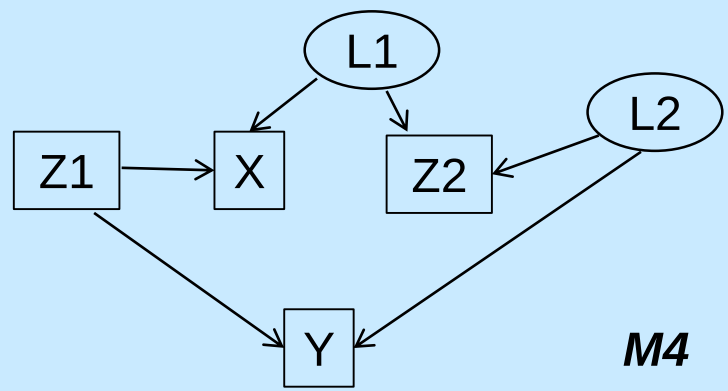




M2



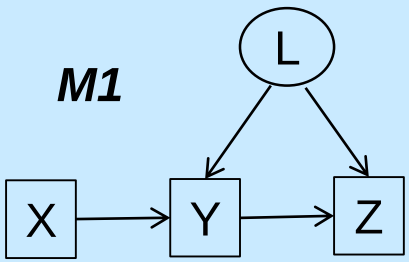
M3

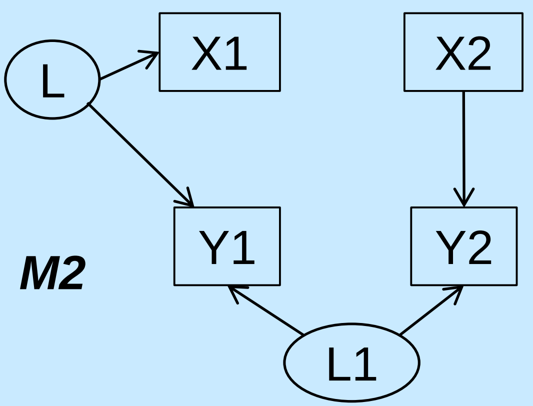


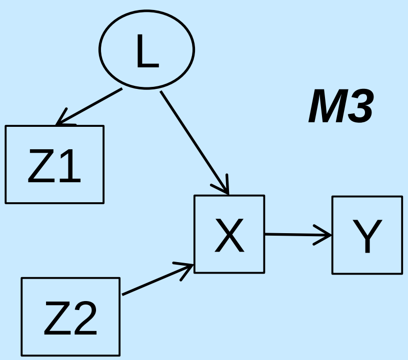
M4

Tetrad Demo and Hands-on

1) Create new session
2) Select “Search from Simulated Data” from Template menu
3) Build graphs for M1, M2, M3 “interesting cases”, parameterize, instantiate, and generate sample data N=1,000.
4) Execute PC search, α = .05
5) Execute FCI search, α = .05







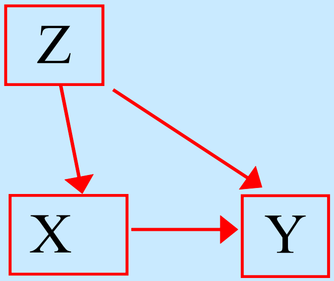
Regression
&
Causal Inference

Regression & Causal Inference

Typical (non-experimental) strategy:

1. Establish a prima facie case (X associated with Y)

But, omitted variable bias



2. So, identifiy and measure potential confounders **Z:**
   a) prior to X,
   b) associated with X,
   c) associated with Y

3. Statistically adjust for **Z** (multiple regression)

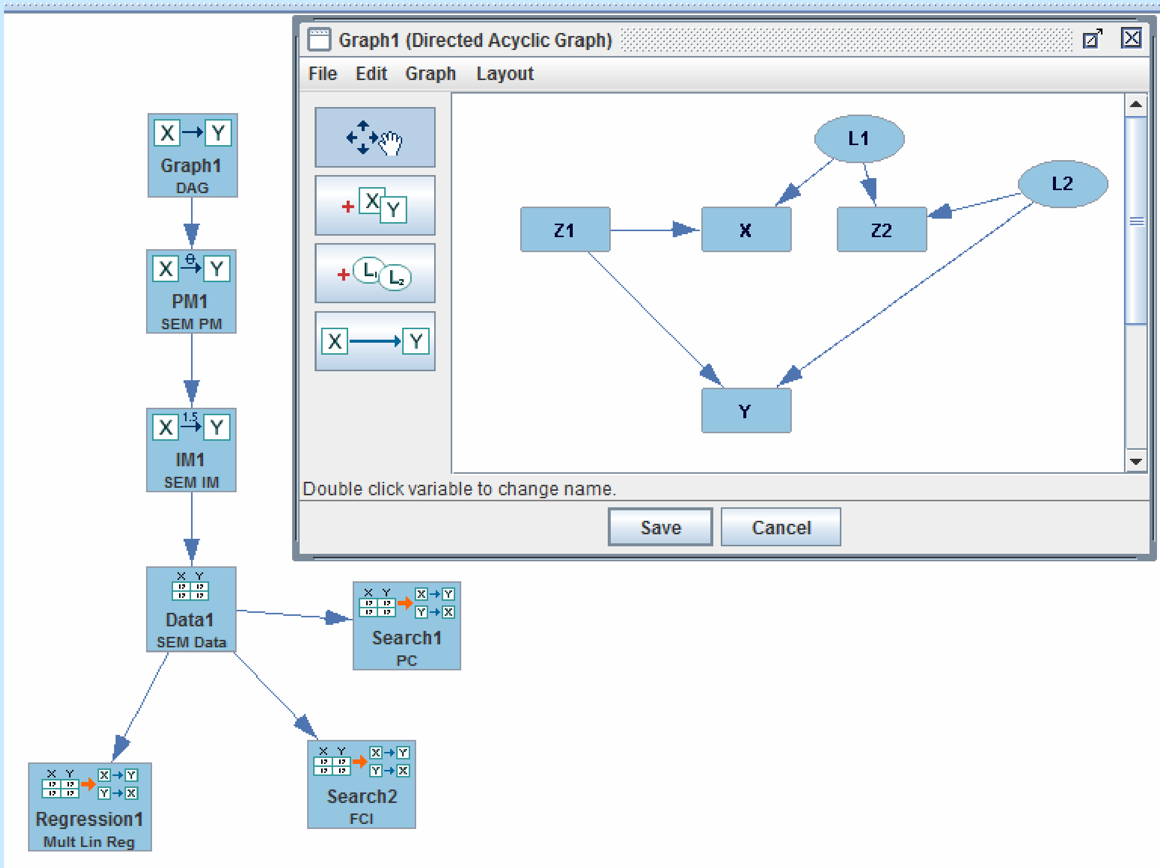
Regression & Causal Inference

Multiple regression or any similar strategy is provably unreliable for causal inference regarding X → Y, with covariates **Z**, unless:

- X truly prior to Y
- X, **Z**, and Y are causally sufficient (no confounding)

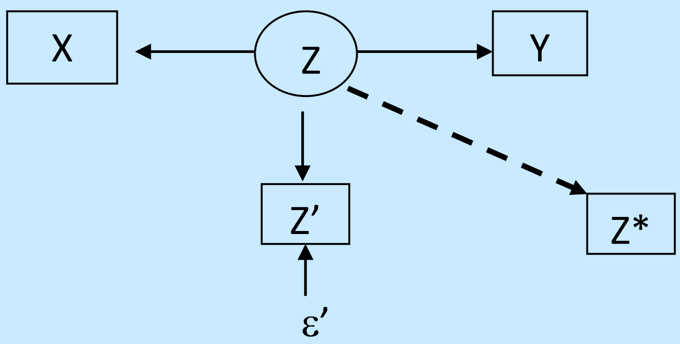
Tetrad Demo and Hands-on

1) Create new session
2) Select “Search from Simulated Data” from Template menu
3) Build a graph for M4 “interesting cases”, parameterize as SEM, instantiate, and generate sample data N=1,000.
4) Execute PC search, $\alpha = .05$
5) Execute FCI search, $\alpha = .05$



Measurement

Measurement Error and Coarsening Endanger conditional Independence!



X _||_ Y | Z

Coarsening: $-\infty < Z < 0 \rightarrow Z^* = 0$

$0 \le Z < i \rightarrow Z^* = 1$

$i \le Z < j \rightarrow Z^* = 2$

..

$k \le Z < \infty \rightarrow Z^* = k$

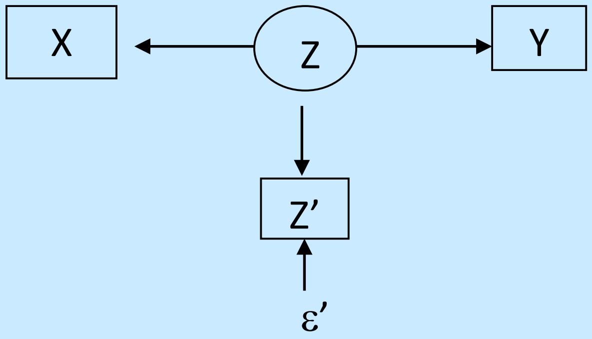
Measurement Error: $Z' = Z + \varepsilon$

X ~~_||_~~ Y | Z' (unless Var(ε') = 0)

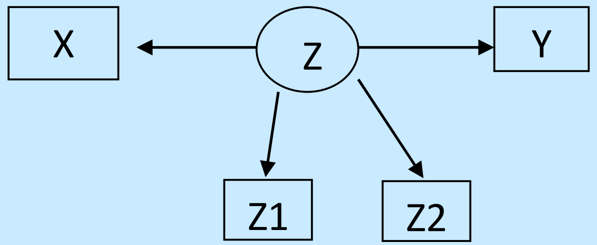
X ~~_||_~~ Y | Z* (almost always)

Strategies

1. Parameterize measurement error:
- Sensitivity Analysis
- Bayesian Analysis
- Bounds

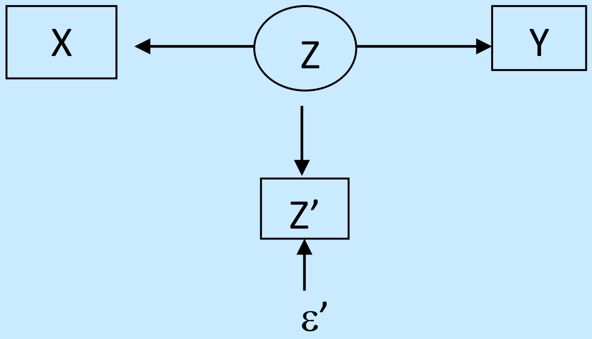


2. Multiple Indicators:

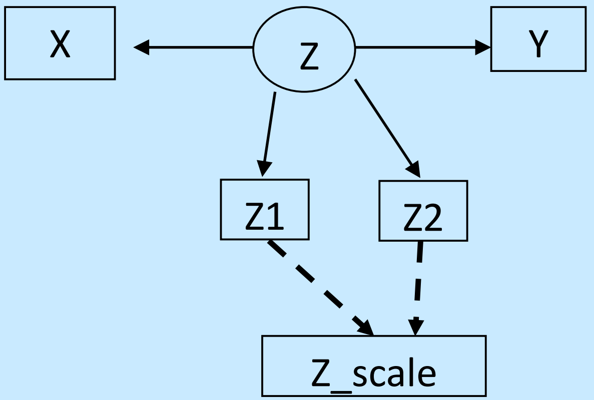


Strategies

1. Parameterize measurement error:
- Sensitivity Analysis
- Bayesian Analysis
- Bounds

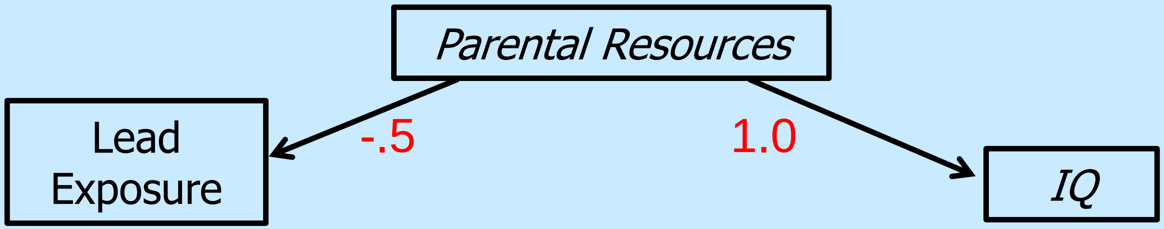


2. Multiple Indicators:
- Scales



X _||_ Y | Z_scale

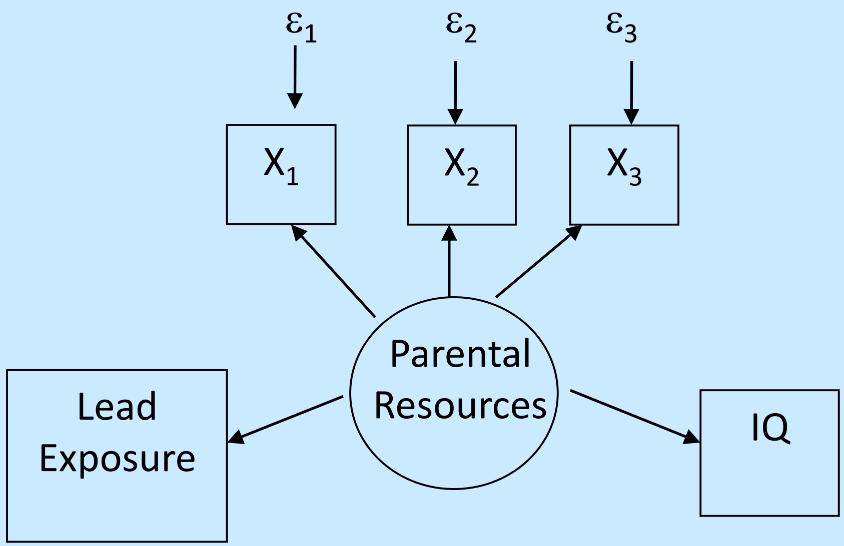
Psuedorandom sample: N = 2,000



Regression of IQ on Lead, PR

Independent Variable	Coefficient Estimate	p-value
PR	0.98	0.000
Lead	-0.088	0.378

Multiple Measures of the Confounder

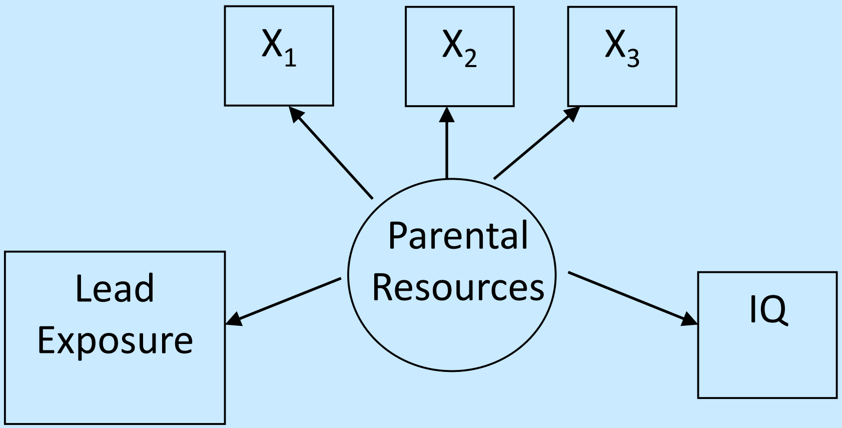


$$X_1 := \gamma_1 * \text{Parental Resources} + \varepsilon_1$$

$$X_2 := \gamma_2 * \text{Parental Resources} + \varepsilon_2$$

$$X_3 := \gamma_3 * \text{Parental Resources} + \varepsilon_3$$

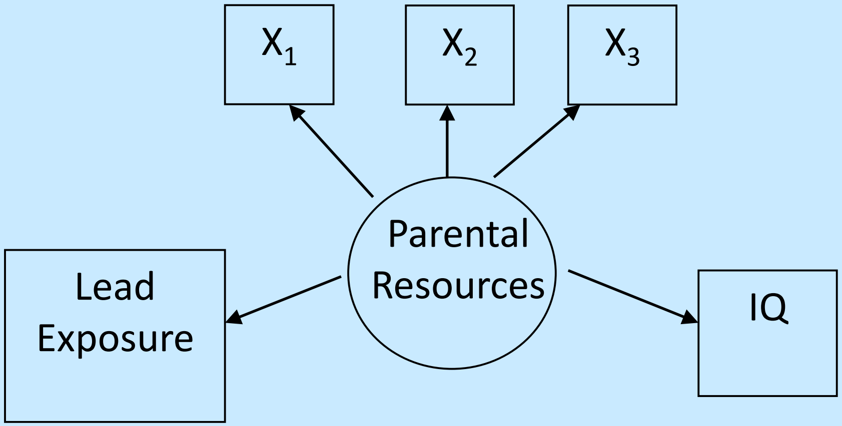
Scales don't preserve conditional independence



$$PR\_Scale = (X_1 + X_2 + X_3) / 3$$

Independent Variable	Coefficient Estimate	p-value
PR_scale	0.290	0.000
Lead	-0.423	0.000

Indicators Don't Preserve Conditional Independence

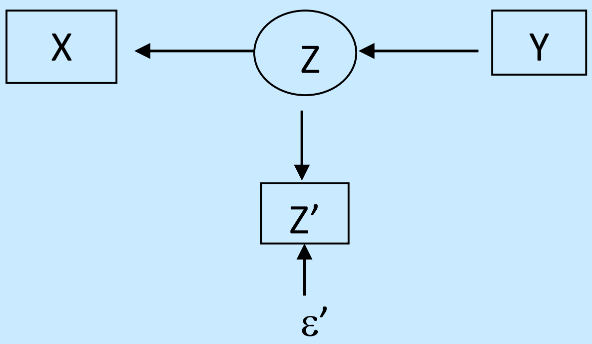


Regress IQ on: Lead, X_1, X_2, X_3

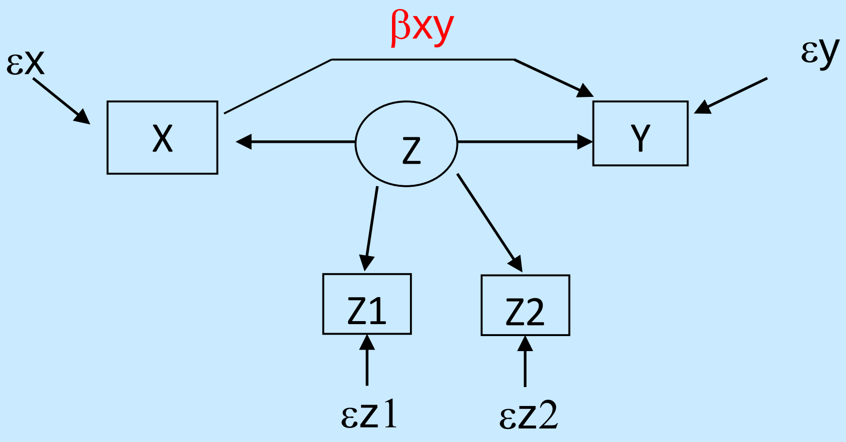
Independent Variable	Coefficient Estimate	p-value
X_1	0.22	0.002
X_2	0.45	0.000
X_3	0.18	0.013
Lead	-0.414	0.000

Strategies

1. Parameterize measurement error:
- Sensitivity Analysis
- Bayesian Analysis
- Bounds

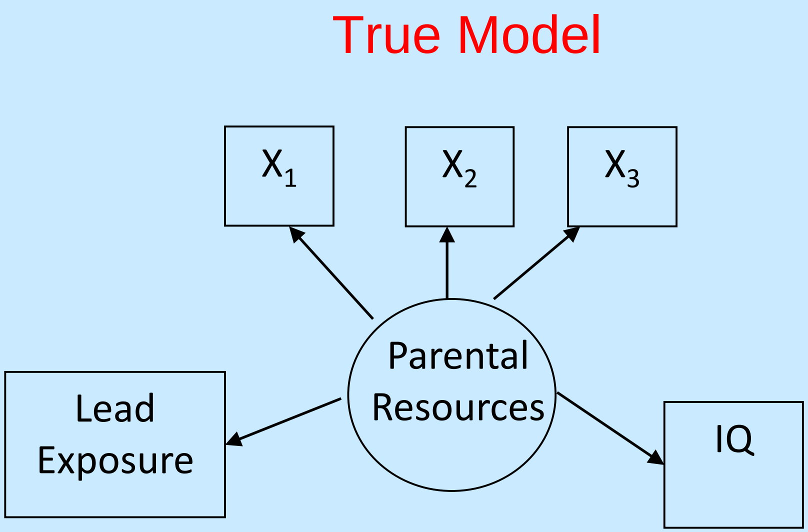


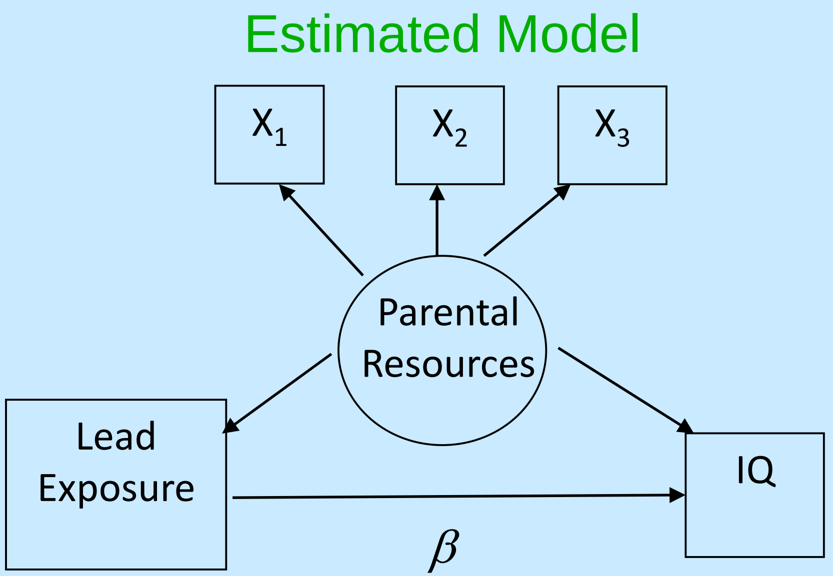
2. Multiple Indicators:
- Scales
- SEM



$E(\hat{\beta}_{yx}) = 0 \quad \Leftrightarrow$ X _||_ Y | Z

Structural Equation Models Work

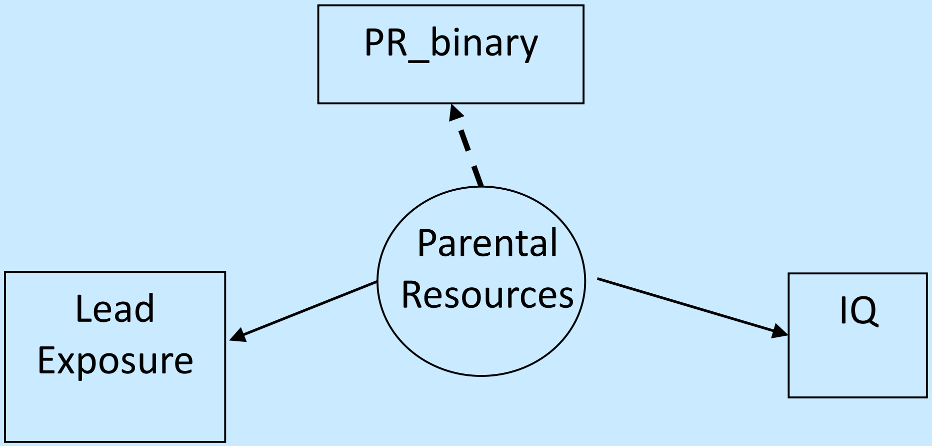




In the Structural Equation Model

- $E(\hat{\beta}) = 0$
- $\hat{\beta} = .07$ (p-value = .499)
- Lead and IQ "screened off" by PR

Coarsening is Bad

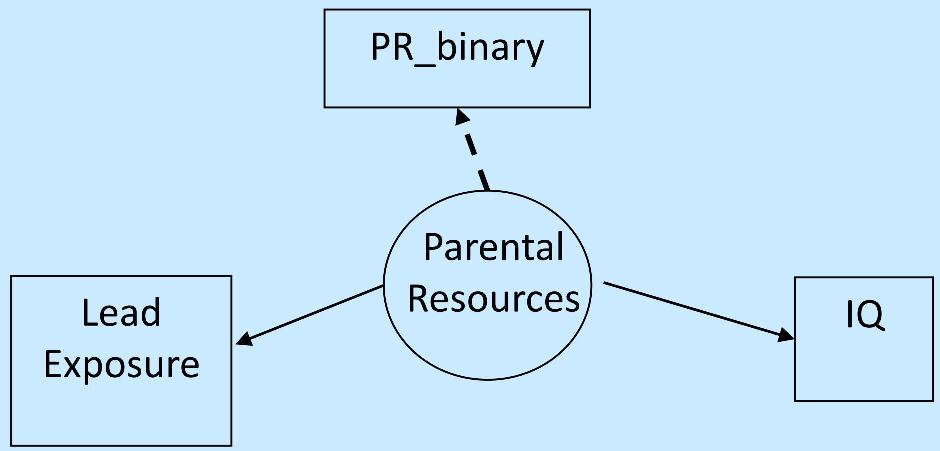


Parental Resources < m(PR) → PR_binary = 0

Parental Resources ≥ m(PR) → PR_binary = 1

Independent Variable	Coefficient Estimate	p-value	Screened-off at .05?
PR_binary	3.53	0.000	No
Lead	-0.56	0.000	No

Coarsening is Bad



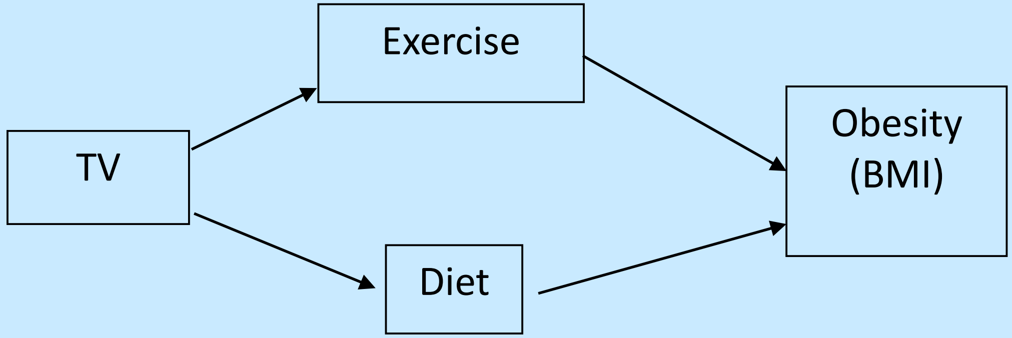
Parental Resources < m(PR) → PR_binary = 0

Parental Resources ≥ m(PR) → PR_binary = 1

Independent Variable	Coefficient Estimate	p-value	Screened-off at .05?
PR_binary	3.53	0.000	No
Lead	-0.56	0.000	No

TV → Obesity

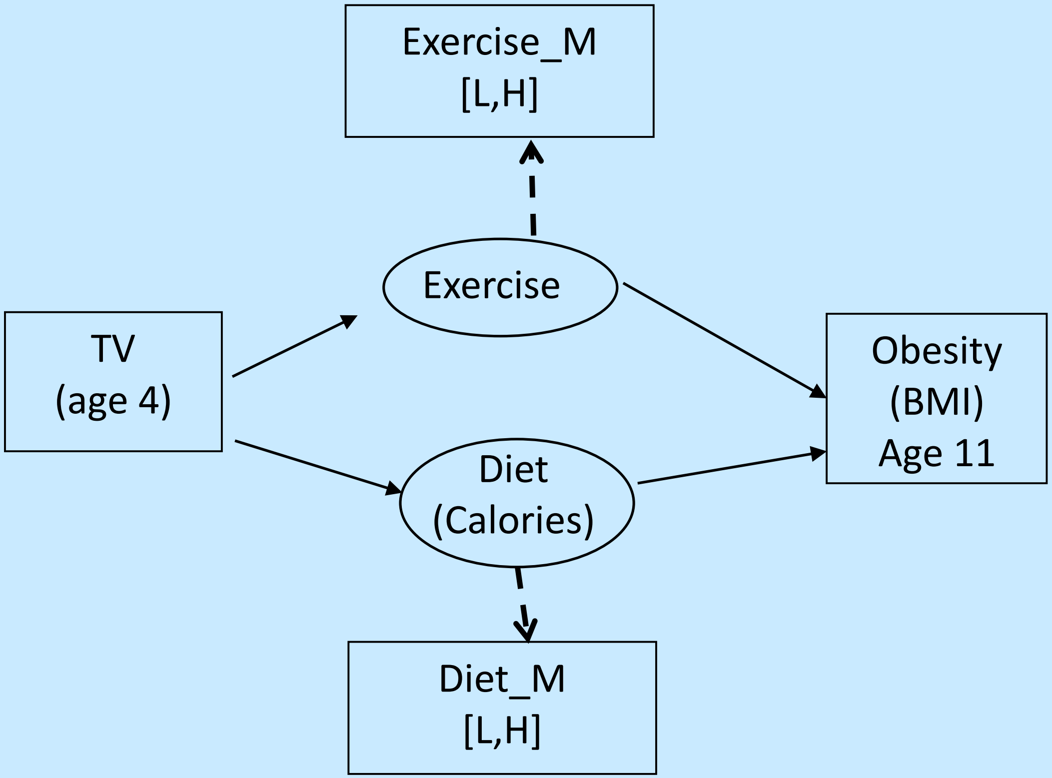
Proctor, et al. (2003). Television viewing and change in body fat from preschool to early adolescence: The Framingham Children's Study *International Journal of Obesity*, 27, 827-833.



Goals:

- Estimate the influence of TV on BMI
- Tease apart the mechanisms (diet, exercise)

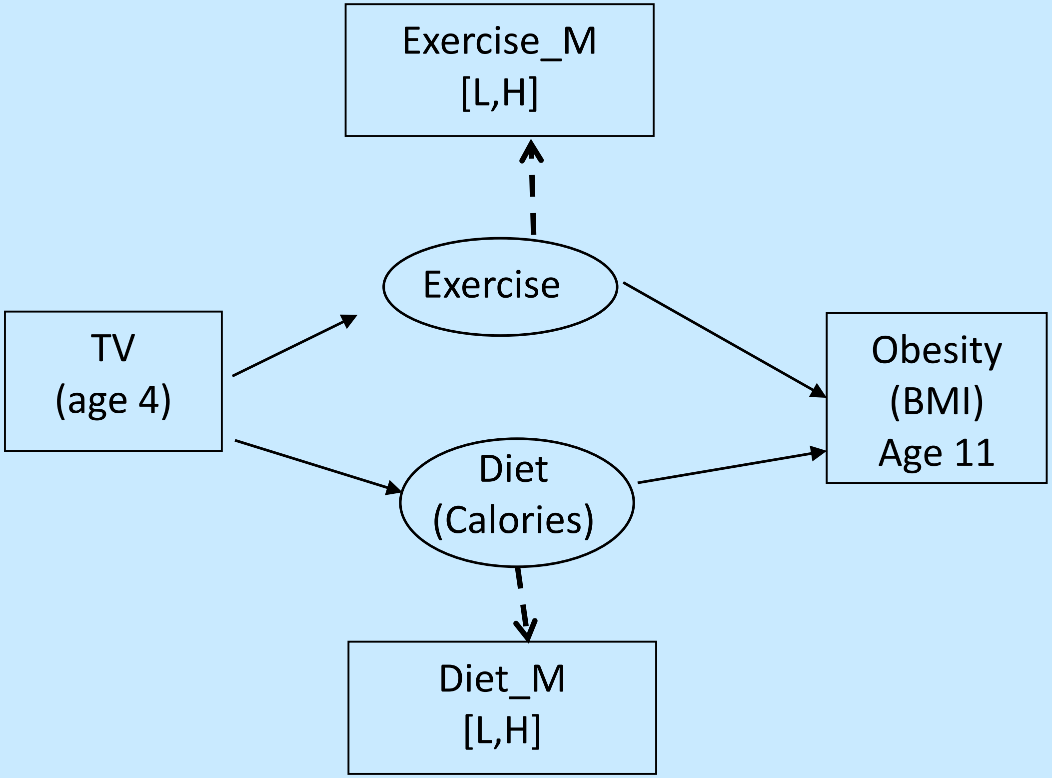
Measures of Exercise, Diet



Exercise_M: L ← Calories expended in exercise in bottom two tertiles
Exercise_M: H ← Calories expended in exercise in top tertile

Diet_M: L ← Calories consumed in bottom two tertiles
Diet_M: H ← Calories consumed in top tertile

Measures of Exercise, Diet

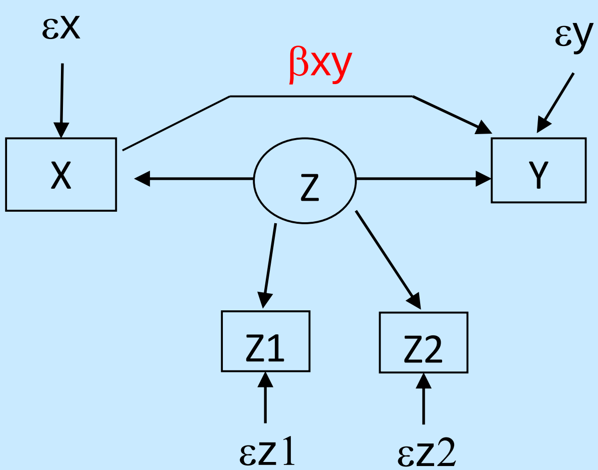


Findings:

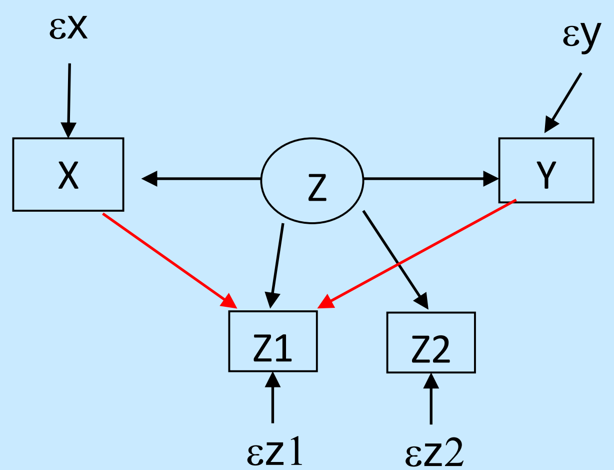
- TV and Obesity NOT screened off by Exercise_M & Diet_M
- Bias in mechanism estimation unknown

Problems with Latent Variable SEMs

Specified Model



True Model

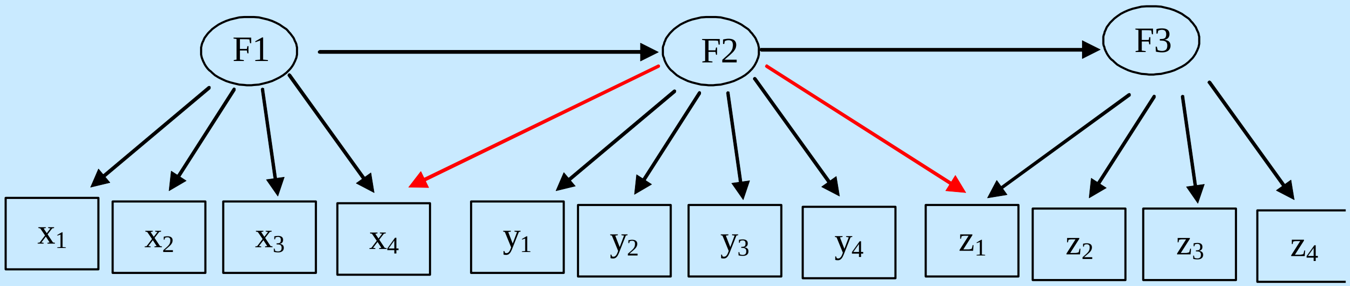


$E(\hat{\beta}_{yx}) = 0$ $\not\Leftrightarrow$ X _||_ Y | Z

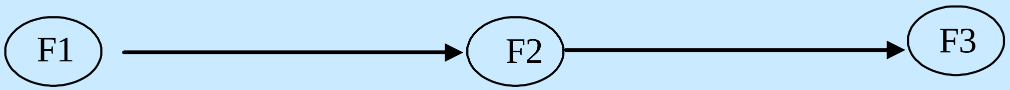
$E(\hat{\beta}_{yx}) \neq \beta xy$

Latent Variable Models

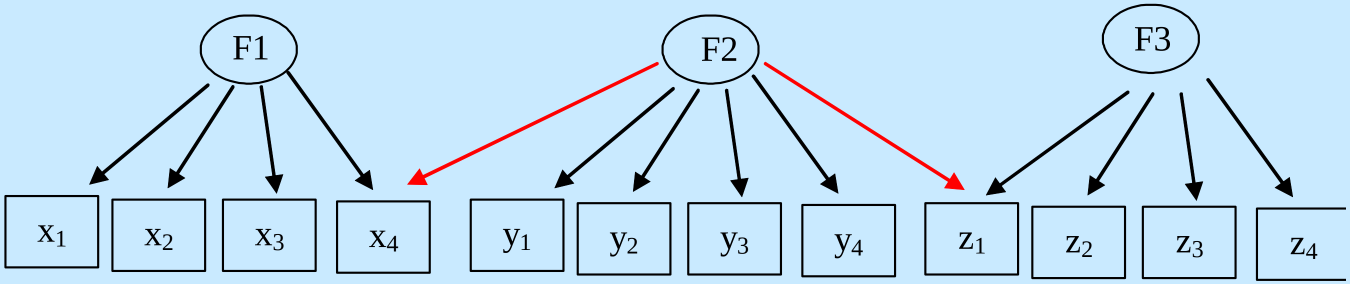
Full Model



Structural Model

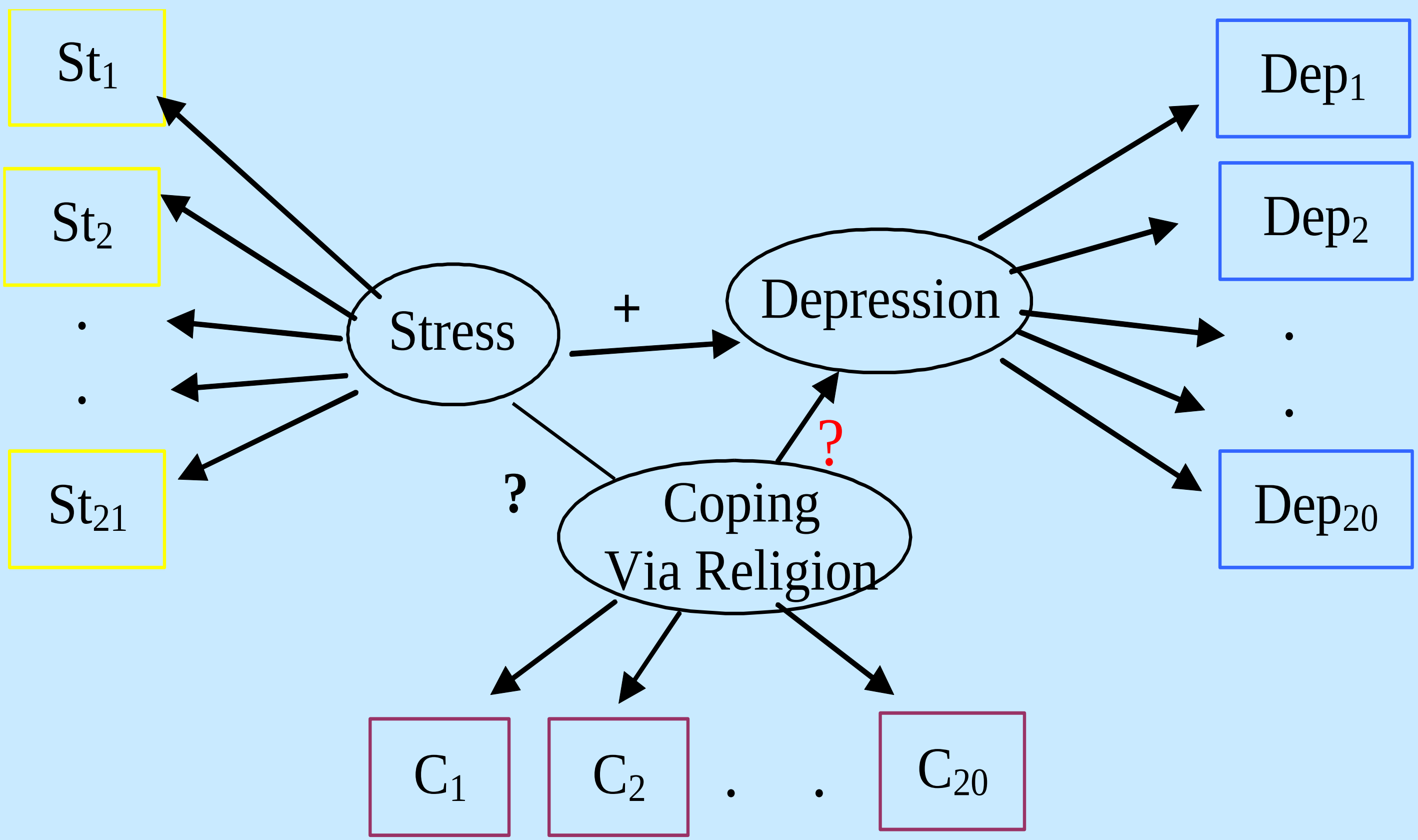


Measurement Model



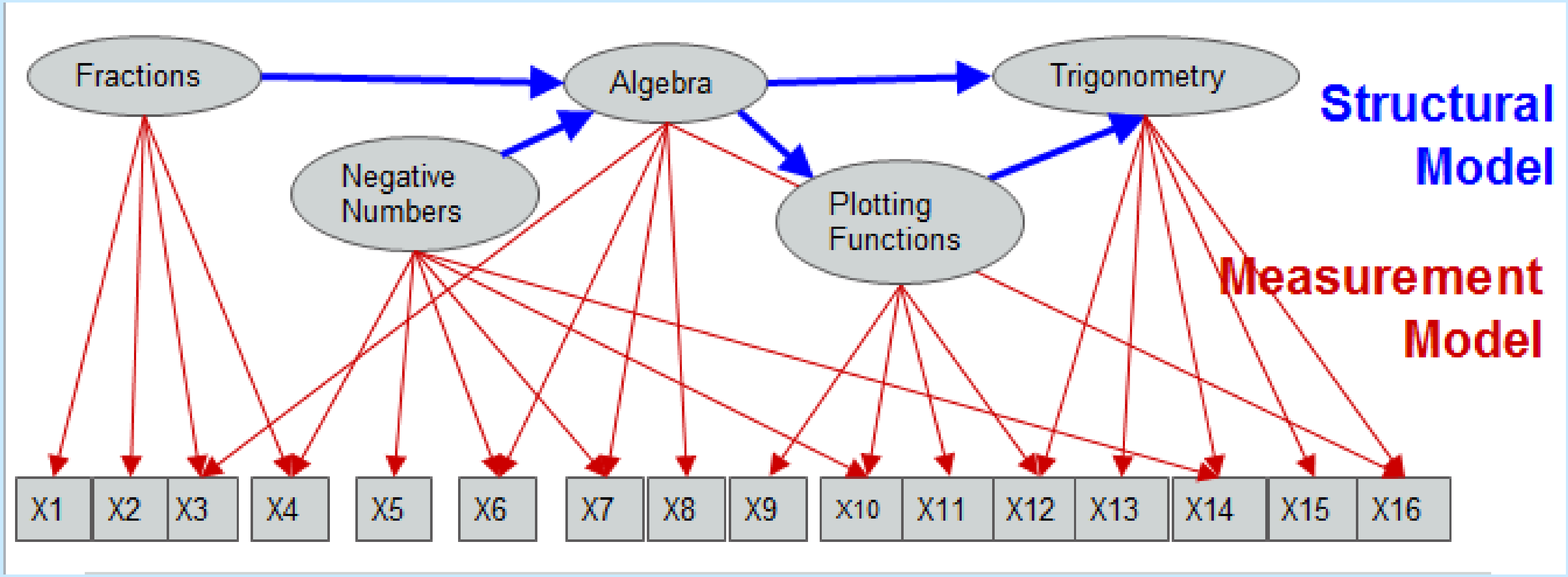
Psychometric Models

Social/Personality Psychology



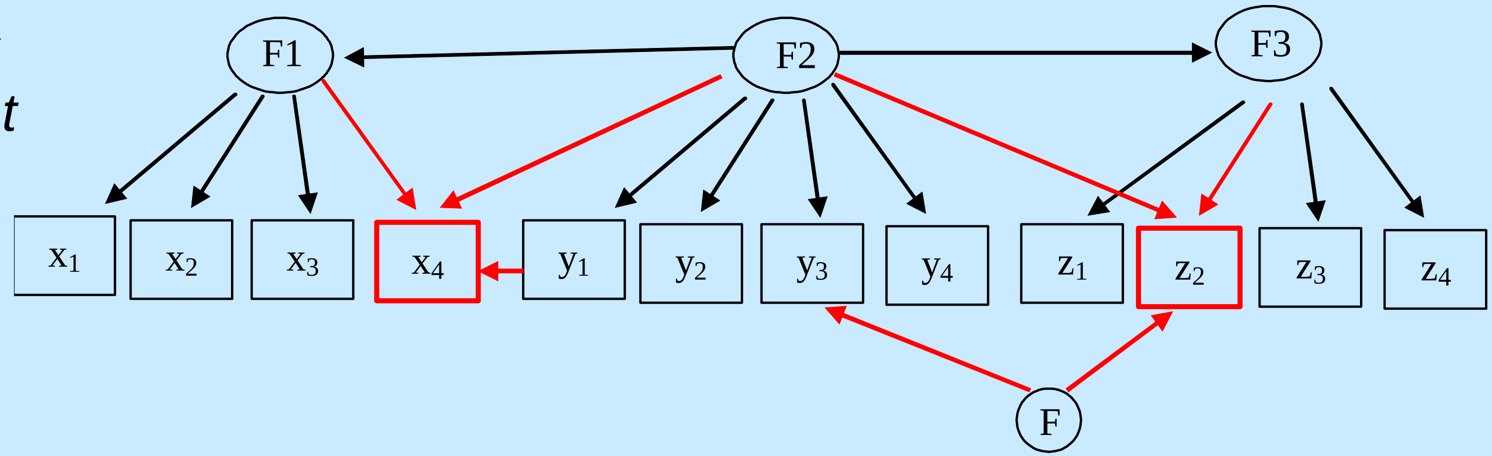
Psychometric Models

Educational Research

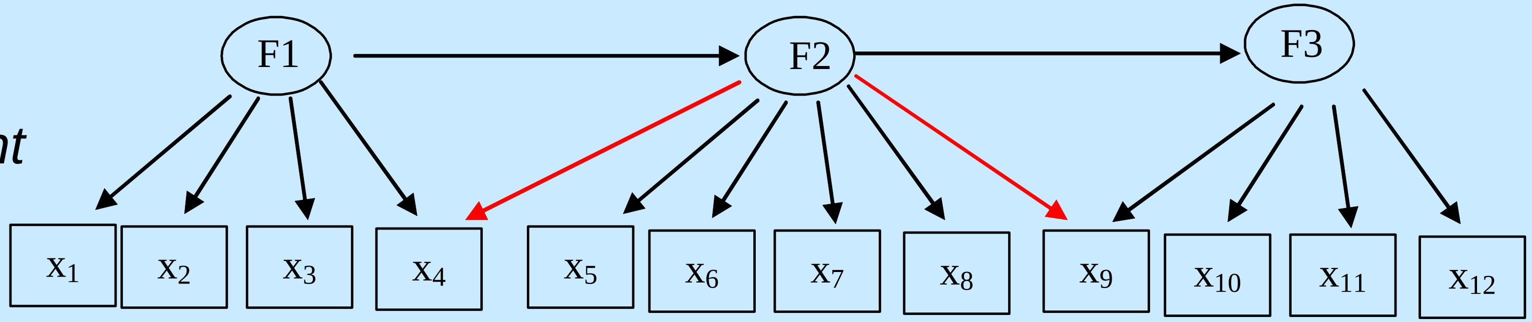


Local Independence / Pure Measurement Models

Not Locally Independent



Locally Independent



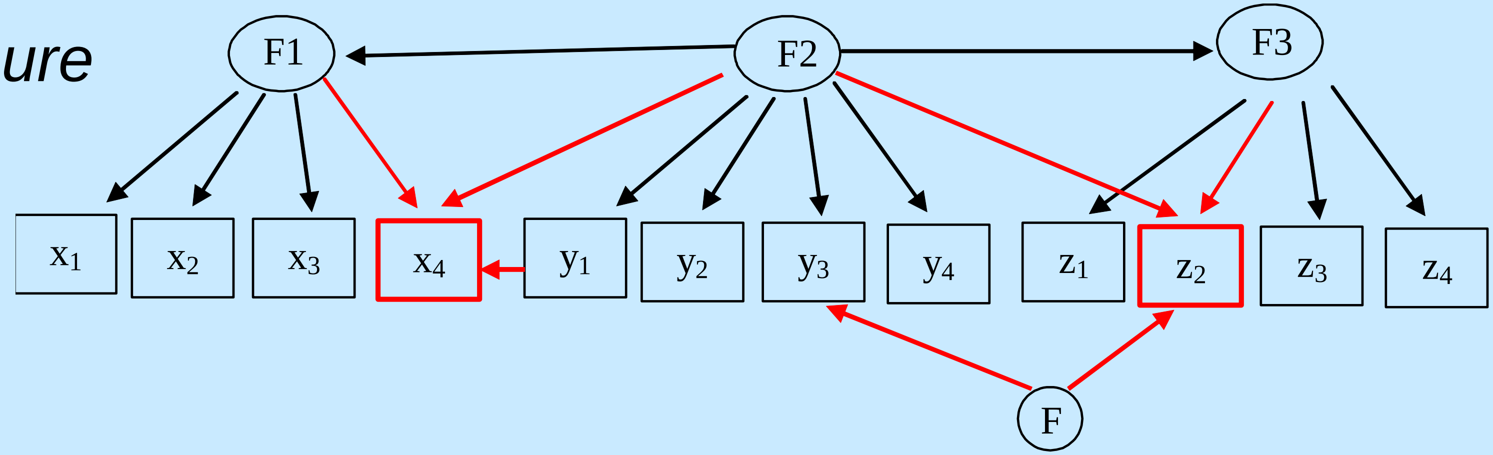
Local Independence:

For every pair of measured items x_i, x_j:

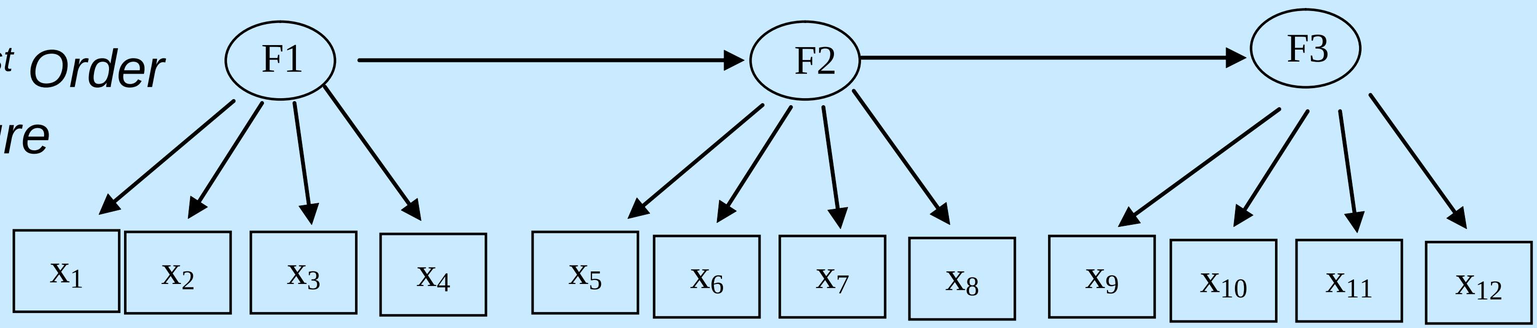
$x_i \perp\!\!\!\perp x_j \mid$ *modeled latent parents of* x_i

Local Independence / Pure Measurement Models

Impure

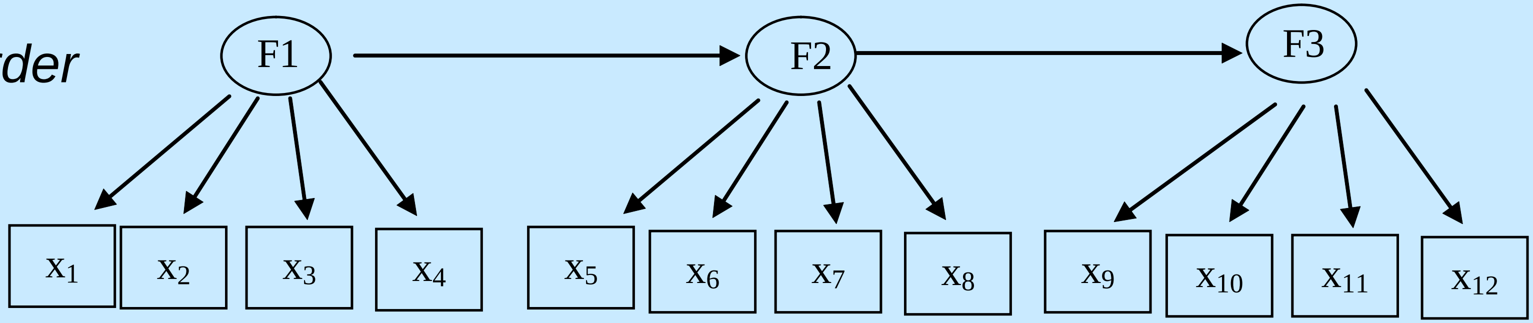


1st Order Pure

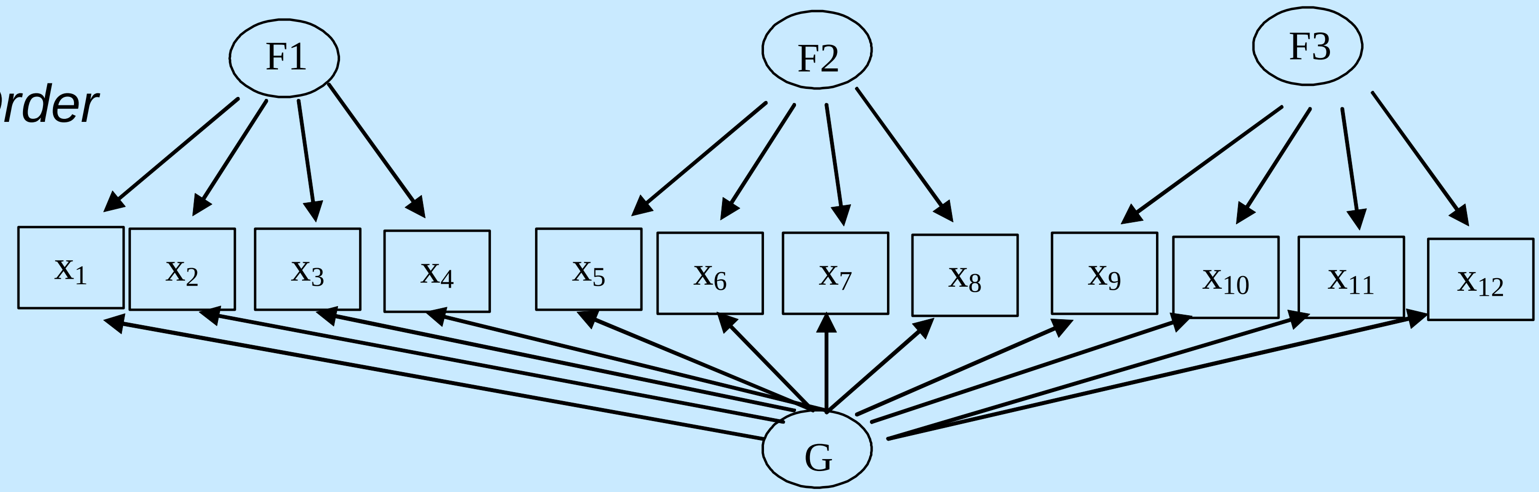


Local Independence / Pure Measurement Models

1st Order Pure

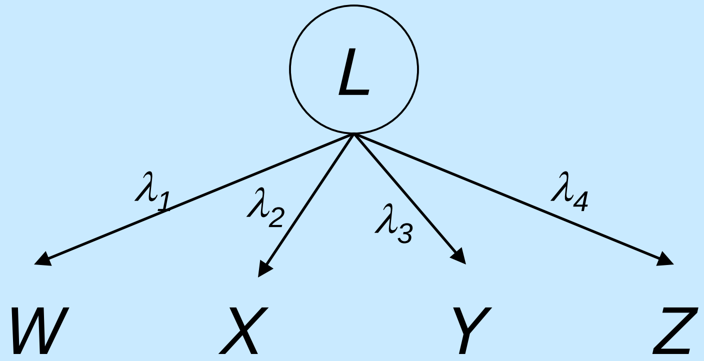


2nd Order Pure



Rank 1 Constraints: Tetrad Equations

- Fact: given



$$W = \lambda_1 L + \varepsilon_1$$
$$X = \lambda_2 L + \varepsilon_2$$
$$Y = \lambda_3 L + \varepsilon_3$$
$$Z = \lambda_4 L + \varepsilon_4$$

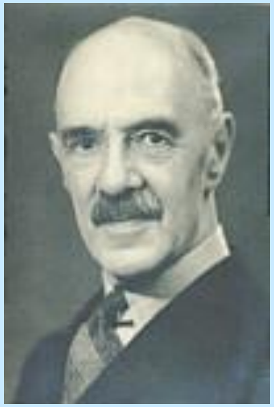
- it follows that

$$Cov_{WX}Cov_{YZ} = (\lambda_1\lambda_2\sigma^2_L)(\lambda_3\lambda_4\sigma^2_L) =$$
$$= (\lambda_1\lambda_3\sigma^2_L)(\lambda_2\lambda_4\sigma^2_L) = Cov_{WY}Cov_{XZ}$$

tetrad constraints

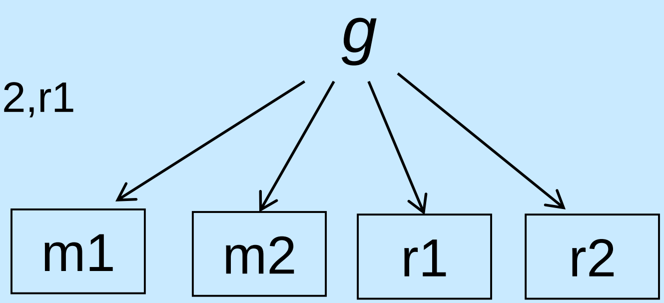
$$\sigma_{WX}\sigma_{YZ} = \sigma_{WY}\sigma_{XZ} = \sigma_{WZ}\sigma_{XY}$$

Charles Spearman (1904)



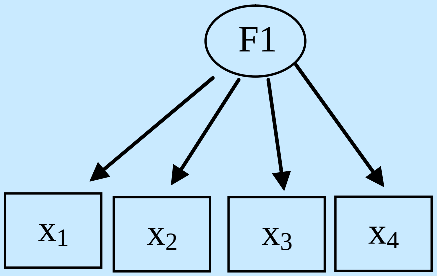
Statistical Constraints → Measurement Model Structure

$$\rho_{m1,m2} * \rho_{r1,r2} = \rho_{m1,r1} * \rho_{m2,r2} = \rho_{m1,r2} * \rho_{m2,r1}$$

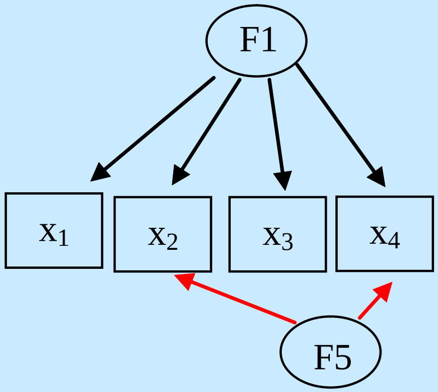


Impurities/Deviations from Local Independence defeat tetrad constraints selectively

Truth



Truth



$\rho_{x1,x2} * \rho_{x3,x4} = \rho_{x1,x3} * \rho_{x2,x4}$

$\rho_{x1,x2} * \rho_{x3,x4} = \rho_{x1,x4} * \rho_{x2,x3}$

$\rho_{x1,x3} * \rho_{x2,x4} = \rho_{x1,x4} * \rho_{x2,x3}$

$\rho_{x1,x2} * \rho_{x3,x4} \neq \rho_{x1,x3} * \boxed{\rho_{x2,x4}}$

$\rho_{x1,x2} * \rho_{x3,x4} = \rho_{x1,x4} * \rho_{x2,x3}$

$\rho_{x1,x3} * \boxed{\rho_{x2,x4}} \neq \rho_{x1,x4} * \rho_{x2,x3}$

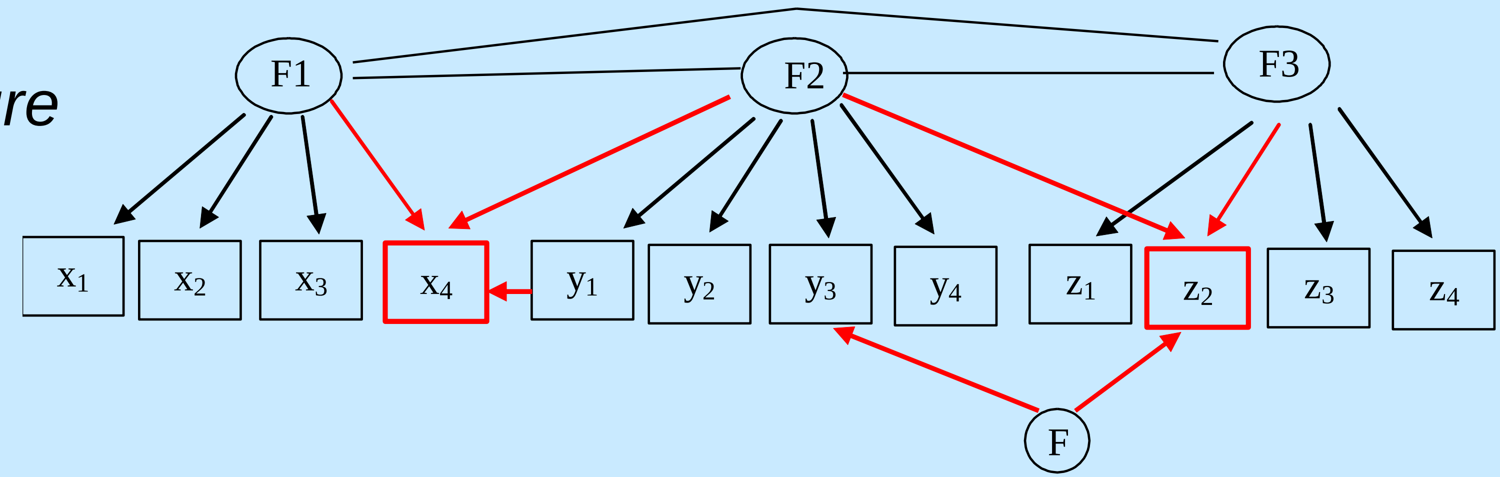
Strategies

1. Cluster and Purify MM first
    1. Use rank constraints to find item subsets that form n^{th} order pure clusters
    2. Using Pure MM : Search for Structural Model by testing independence relations among latents *via SEM estimation*

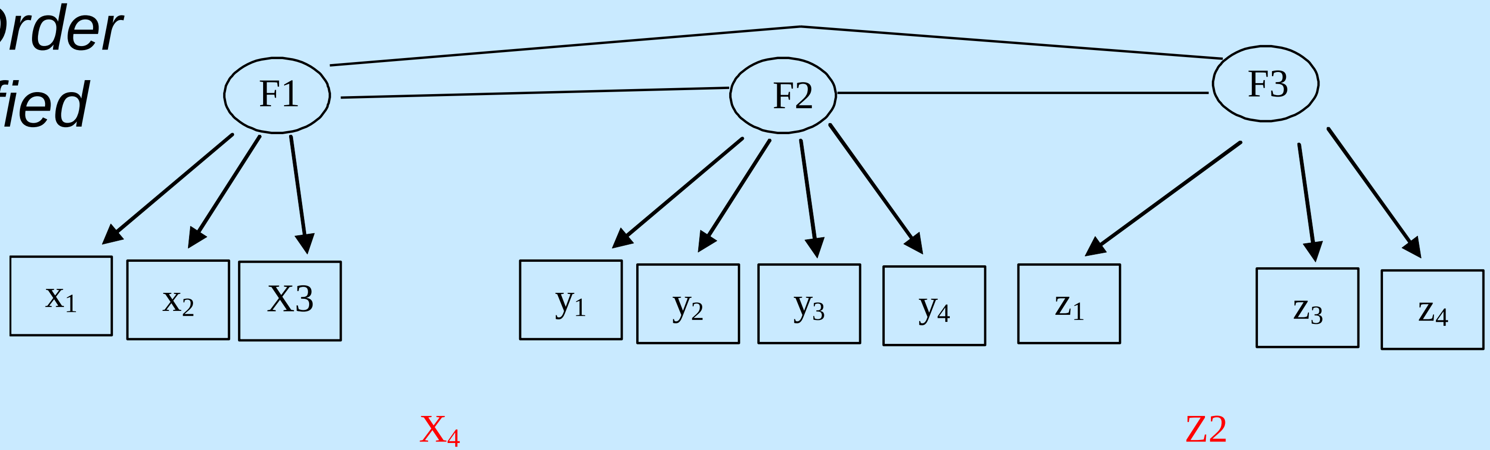
2. Specify Impure Measurement Model
    1. Specify Measurement Model for all items
    2. Using Specified MM: Search for Structural Model by testing independence relations among latents *via SEM estimation*

Purify

Impure



1st Order Purified



Search for Measurement Models

BPC, FOFC: Find One Factor Clusters

Input: Covariance Matrix of measured items:

Output: Subset of items and clusters that are 1st Order Pure

FTFC: Find Two Factor Clusters

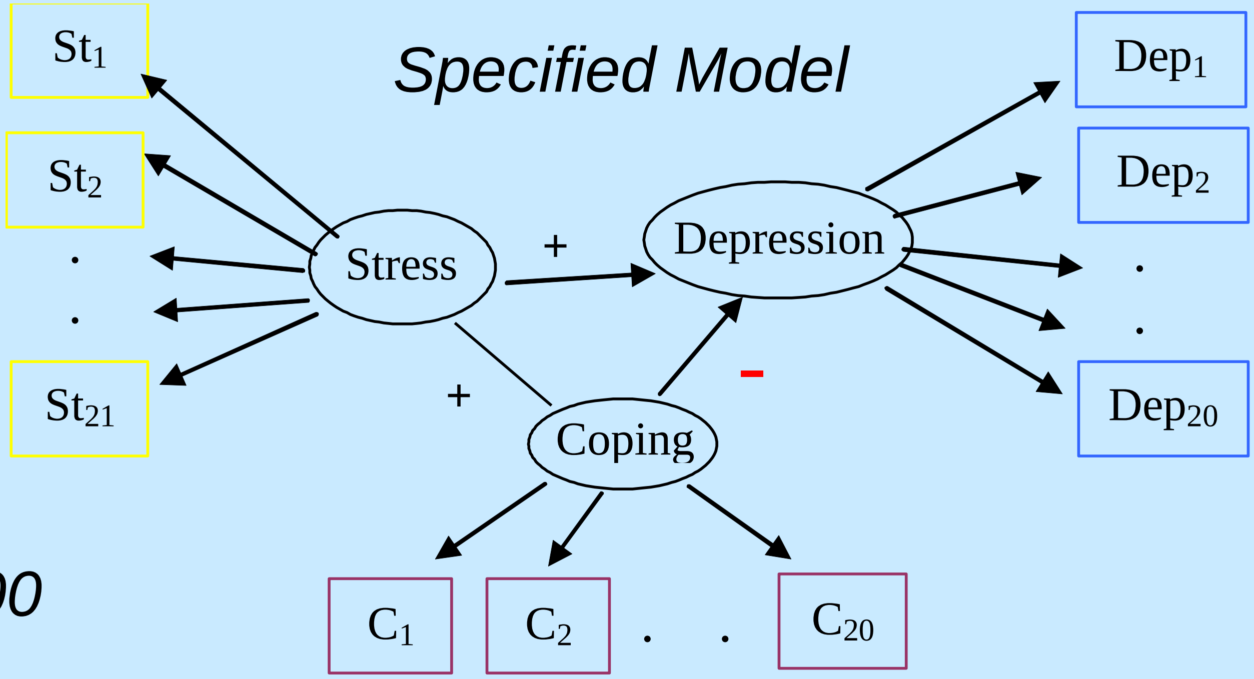
Input: Covariance Matrix of measured items:

Output: Subset of items and clusters that are 2^{nd} Order Pure

BPC Case Study: Stress, Depression, and Religion

Masters Students (N = 127) 61 - item survey (Likert Scale)

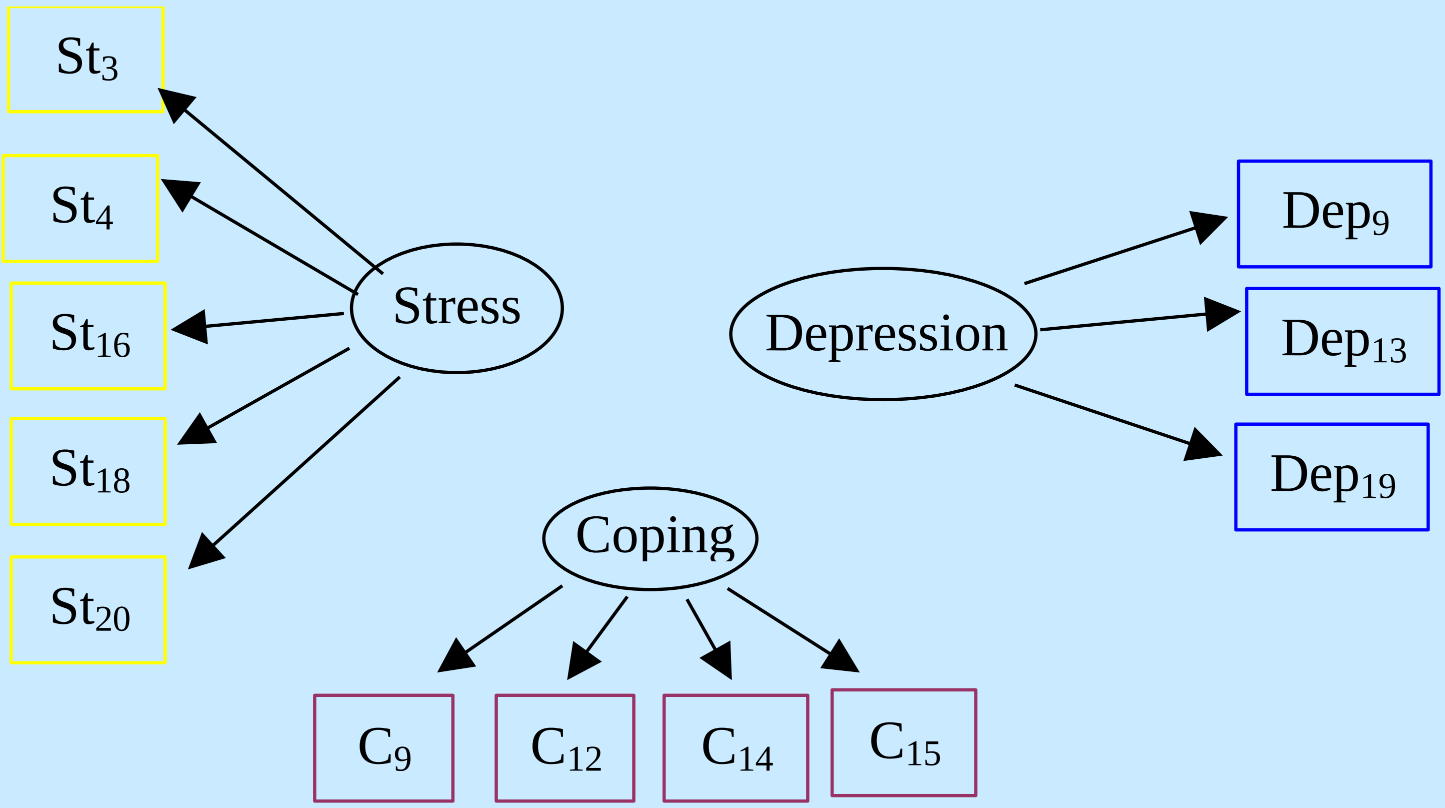
- *Stress: St_1 - St_{21}*
- *Depression: D_1 - D_{20}*
- *Religious Coping: C_1 - C_{20}*



$P(\chi 2) = 0.00$

Case Study: Stress, Depression, and Religion

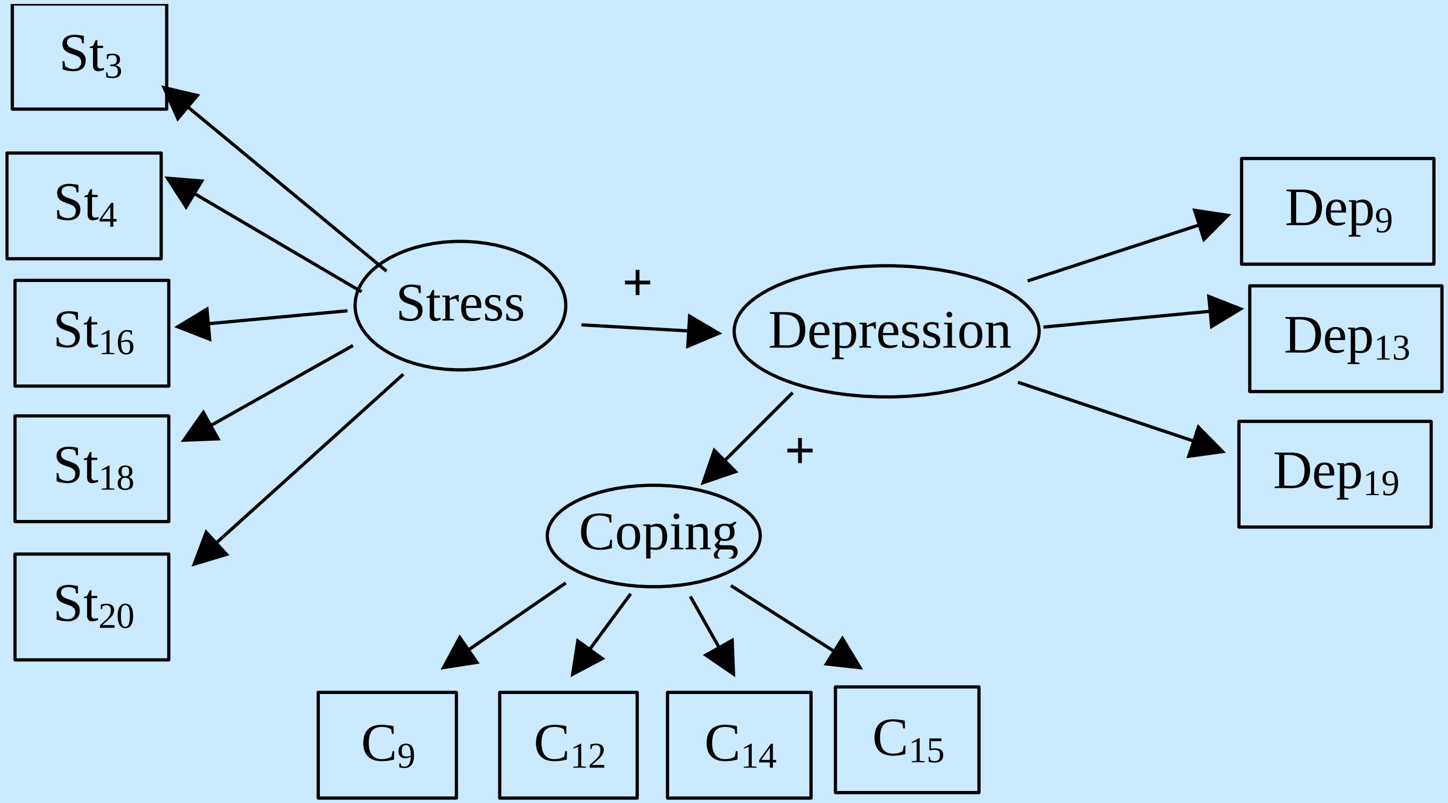
Build Pure Clusters



Case Study: Stress, Depression, and Religion

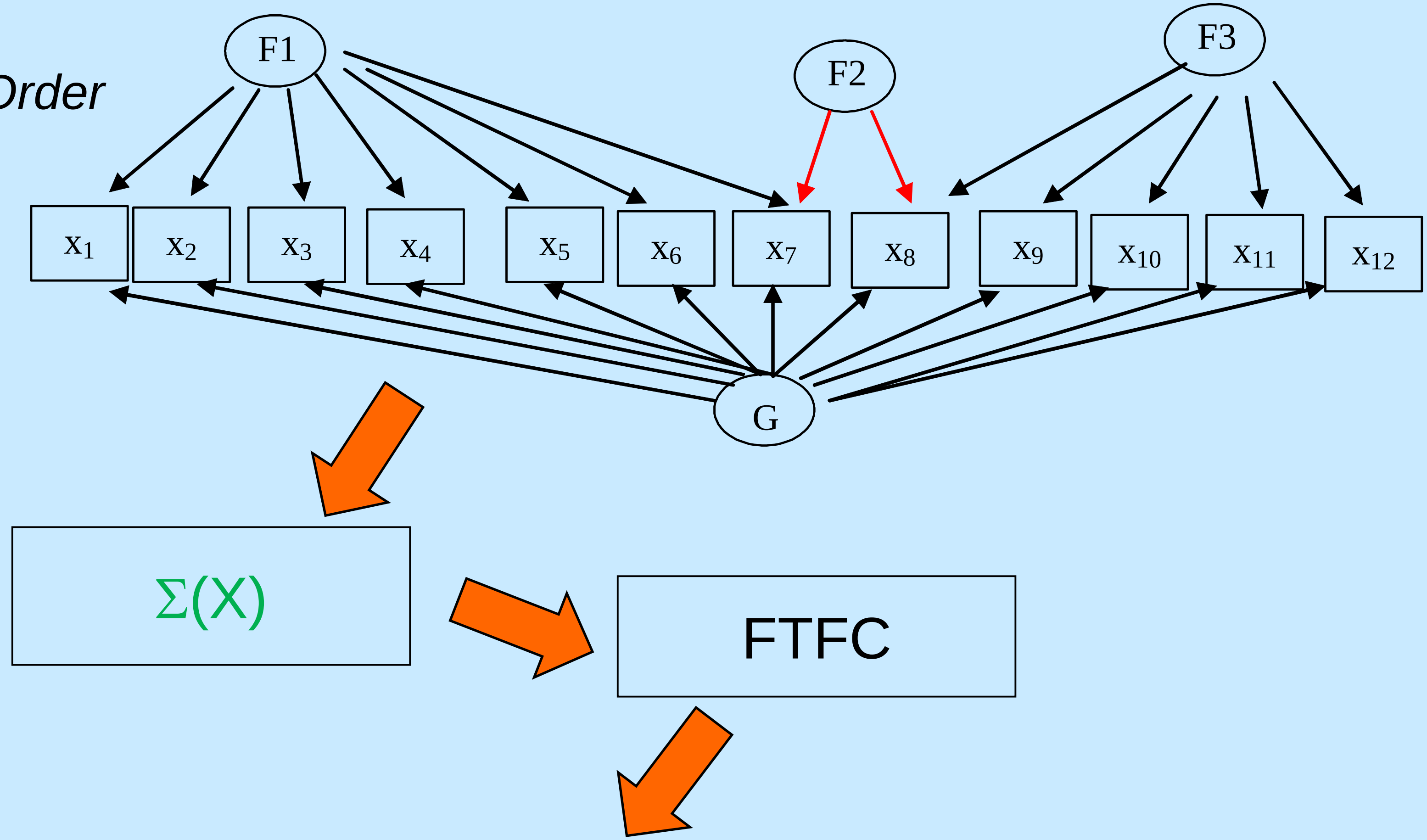
Assume : Stress causally prior to Depression

Find : Stress _||_ Coping | Depression



P(χ2) = 0.28

2^{nd} Order Pure



2nd-Order Pure Clusters:

- $\{X_1, X_2, X_3, X_4, X_5, X_6\}$
- $\{X_8, X_9, X_{10}, X_{11}, X_{12}\}$

Summary of Search

Causal Search
from Passive Observation

- PC, FGS → *Patterns* (Markov equivalence class - no latent confounding)
- FCI → *PAGs* (Markov equivalence - including confounders and selection bias)
- CCD → Linear cyclic models (no confounding)
- Lingam → unique DAG (no confounding – linear non-Gaussian – faithfulness not needed)
- BPC, FOFC, FTFC → (Equivalence class of linear latent variable models)
- LVLingam → set of DAGs (confounders allowed)
- CyclicLingam → set of DGs (cyclic models, no confounding)
- Non-linear additive noise models → unique DAG
- Most of these algorithms are pointwise consistent – uniform consistent algorithms require stronger assumptions

Causal Search
from Manipulations/Interventions

What sorts of manipulation/interventions have been studied?

- Do(X=x) : replace P(X | parents(X)) with P(X=x) = 1.0
- Randomize(X): (replace P(X | parents(X)) with $P_M(X)$, e.g., uniform)
- Soft interventions (replace P(X | parents(X)) with $P_M(X \mid parents(X), I)$, $P_M(I)$)
- Simultaneous interventions (reduces the number of experiments required to be guaranteed to find the truth with an independence oracle from N-1 to 2 log(N)
- Sequential interventions
- Sequential, conditional interventions
- Time sensitive interventions