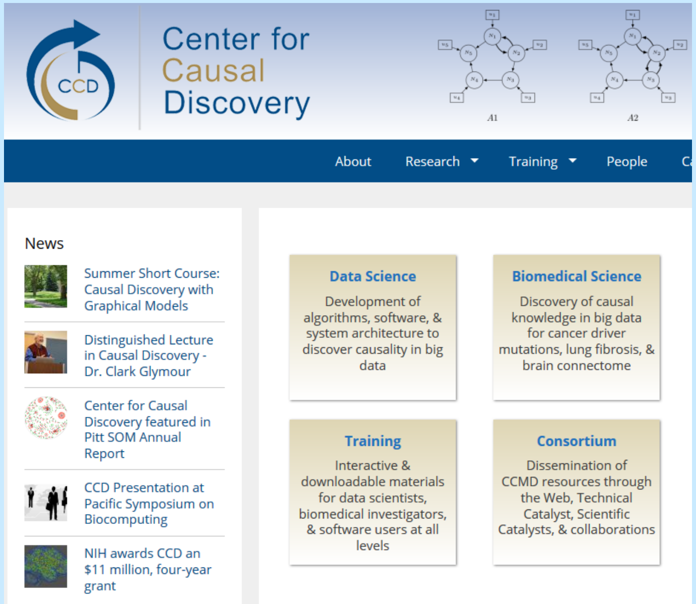
Center for Causal Discovery:

Summer Short Course/Datathon - 2016



June 13-18, 2015

Carnegie Mellon University

Outline

Models → Data

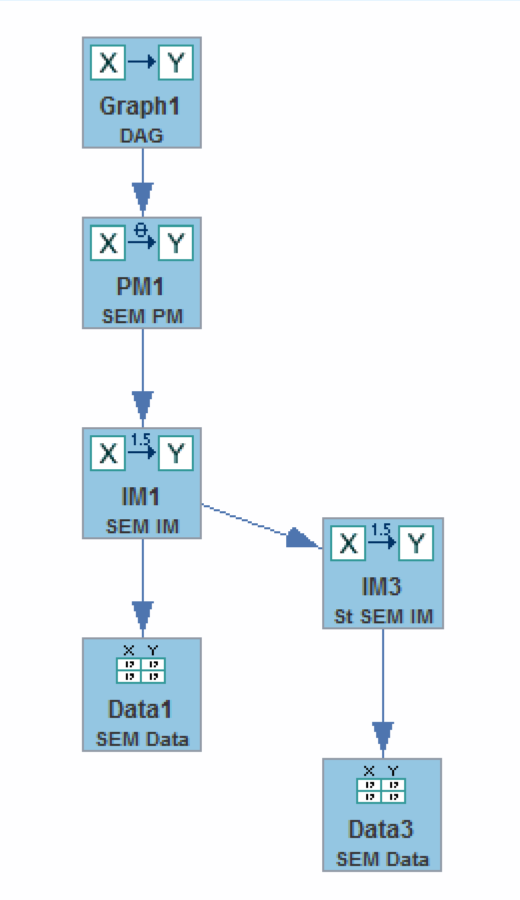
1) Representing/Modeling Causal Systems
2) Estimation and Model fit
3) Hands on with Real Data

Models ← Data

1) Markov Axiom and D-separation
2) Model Equivalence
3) Model Search

Standardized SEMs

1) Attach a SEM PM to your 3-4 variable graph
2) Attach a SEM IM to the SEM PM
3) Change the coefficient values.
4) Attach a Standardized SEM IM to the SEM PM, or the SEM IM
5) Compare the Implied Matrices

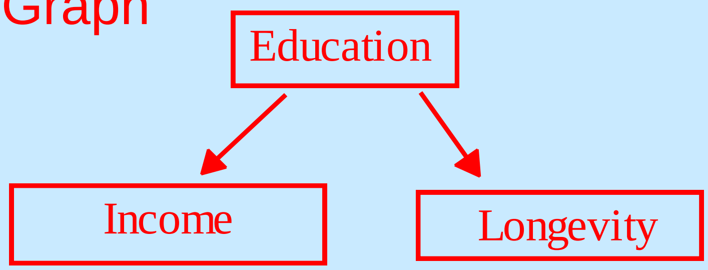


Tetrad Demo & Hands-On

Generalized SEM

1) The Generalized SEM is a generalization of the linear SEM model.
2) Allows for arbitrary connection functions
3) Allows for arbitrary distributions
4) Simulation from cyclic models supported.

Causal Graph



SEM Equations:

$\text{Education} := \varepsilon_{\text{Education}}$

$\text{Income} := \beta_1 \, \text{Education} + \varepsilon_{\text{income}}$

$\text{Longevity} := \beta_2 \, \text{Education} + \varepsilon_{\text{Longevity}}$

$P(\varepsilon_{ed}, \varepsilon_{\text{Income}}, \varepsilon_{\text{Income}}) \sim N(0, \Sigma^2)$

<u>Generalized</u> SEM Equations:

$\text{Education} := \varepsilon_{\text{Education}}$

$\text{Income} := \beta_1 \, \text{Education}^2 + \varepsilon_{\text{income}}$

$\text{Longevity} := \beta_2 \, \text{In}(\text{Education}) + \varepsilon_{\text{Longevity}}$

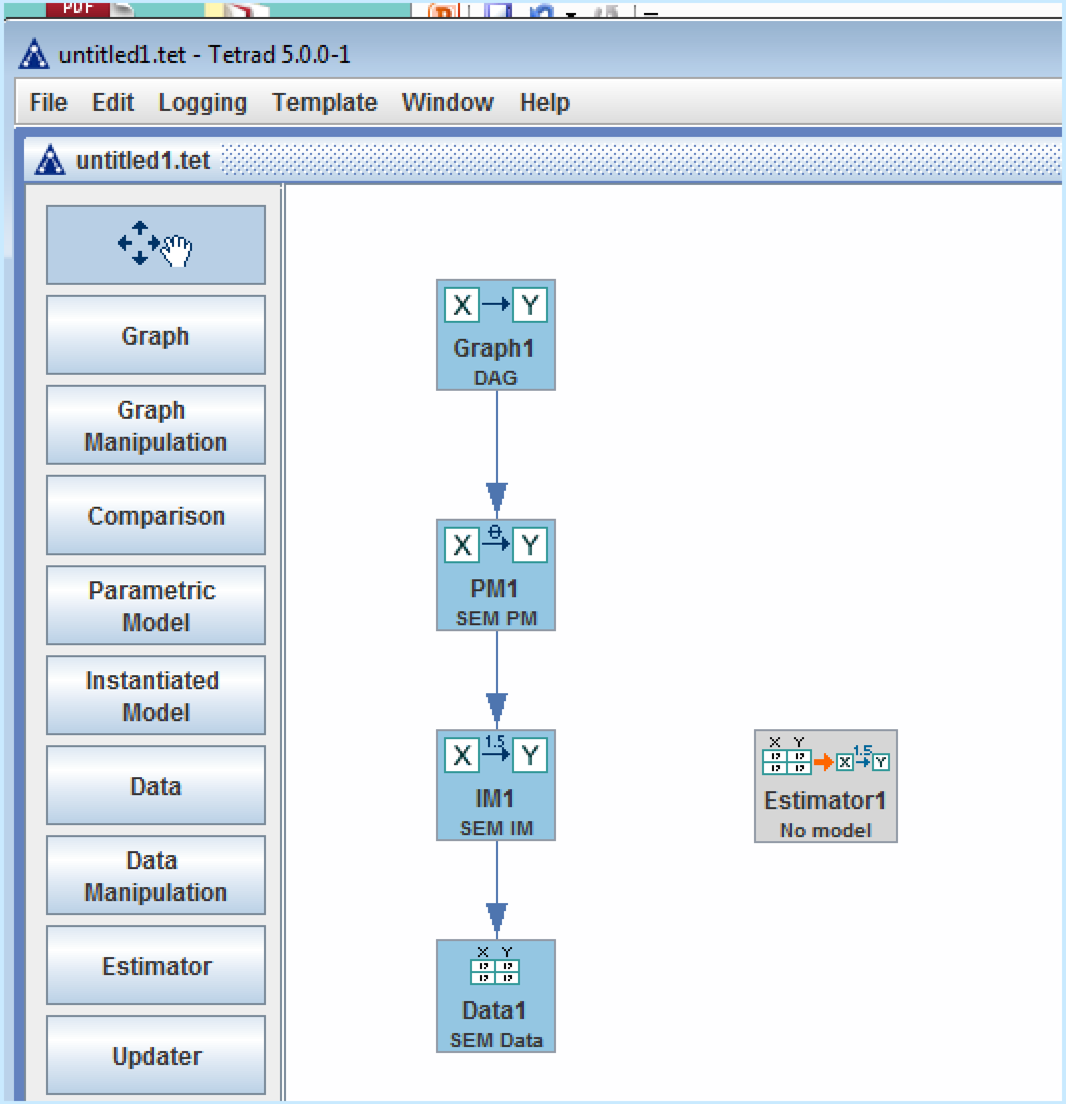
$P(\varepsilon_{ed}, \varepsilon_{\text{Income}}, \varepsilon_{\text{Income}}) \sim U(0,1)$

Hands On

1) Create a DAG.
2) Parameterize it as a Generalized SEM.
3) In PM – select from Tools menu “show error terms”

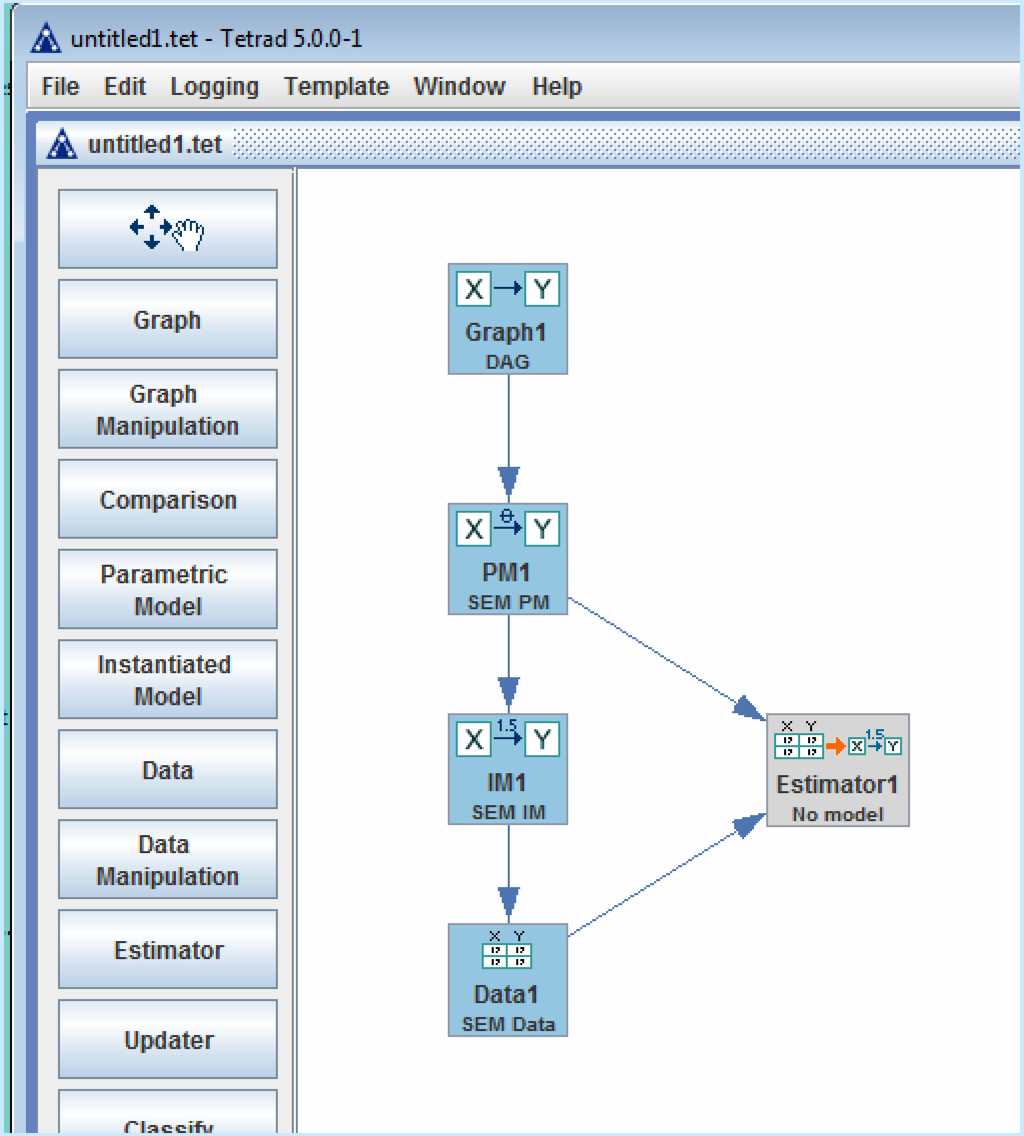
   Click on error term, change its distribution to Uniform
4) Make at least one function non-linear
5) Make at least one function interactive
6) Save the session as “generalizedSEM”.

Estimation



untitled1.tet - Tetrad 5.0.0-1
File Edit Logging Template Window Help
untitled1.tet
Graph
Graph Manipulation
Comparison
Parametric Model
Instantiated Model
Data
Data Manipulation
Estimator
Updater
Graph1
DAG
PM1
SEM PM
IM1
SEM IM
Data1
SEM Data
Estimator1
No model

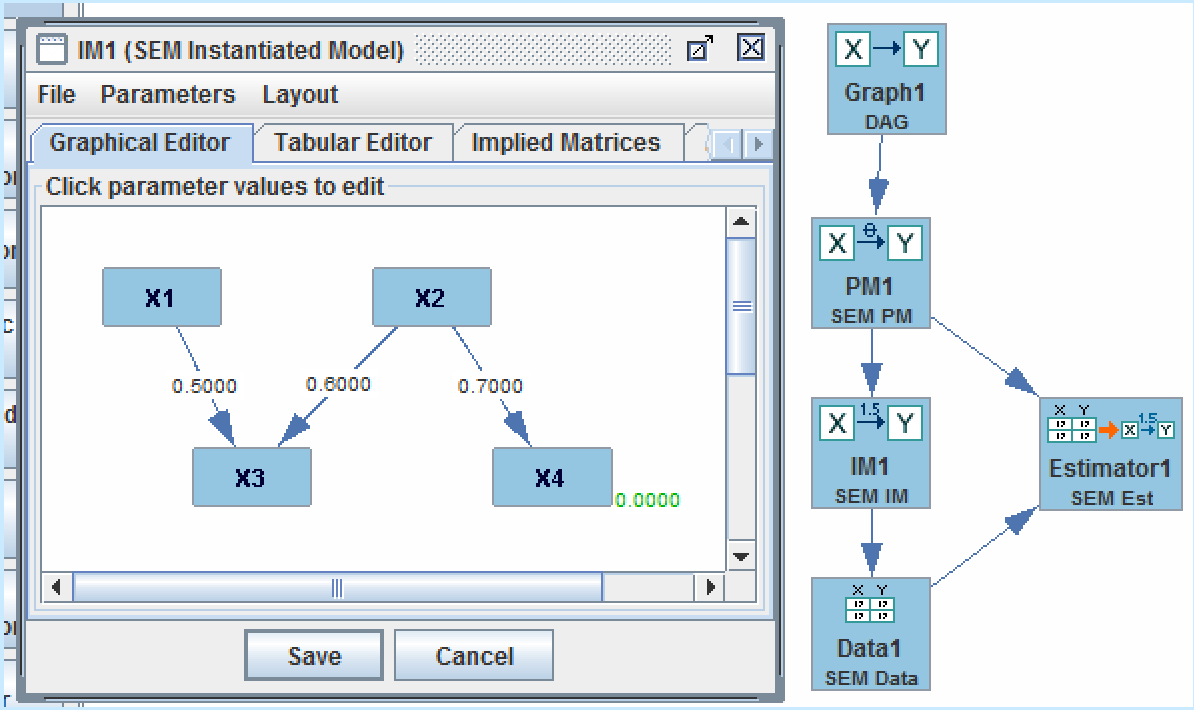
Estimation



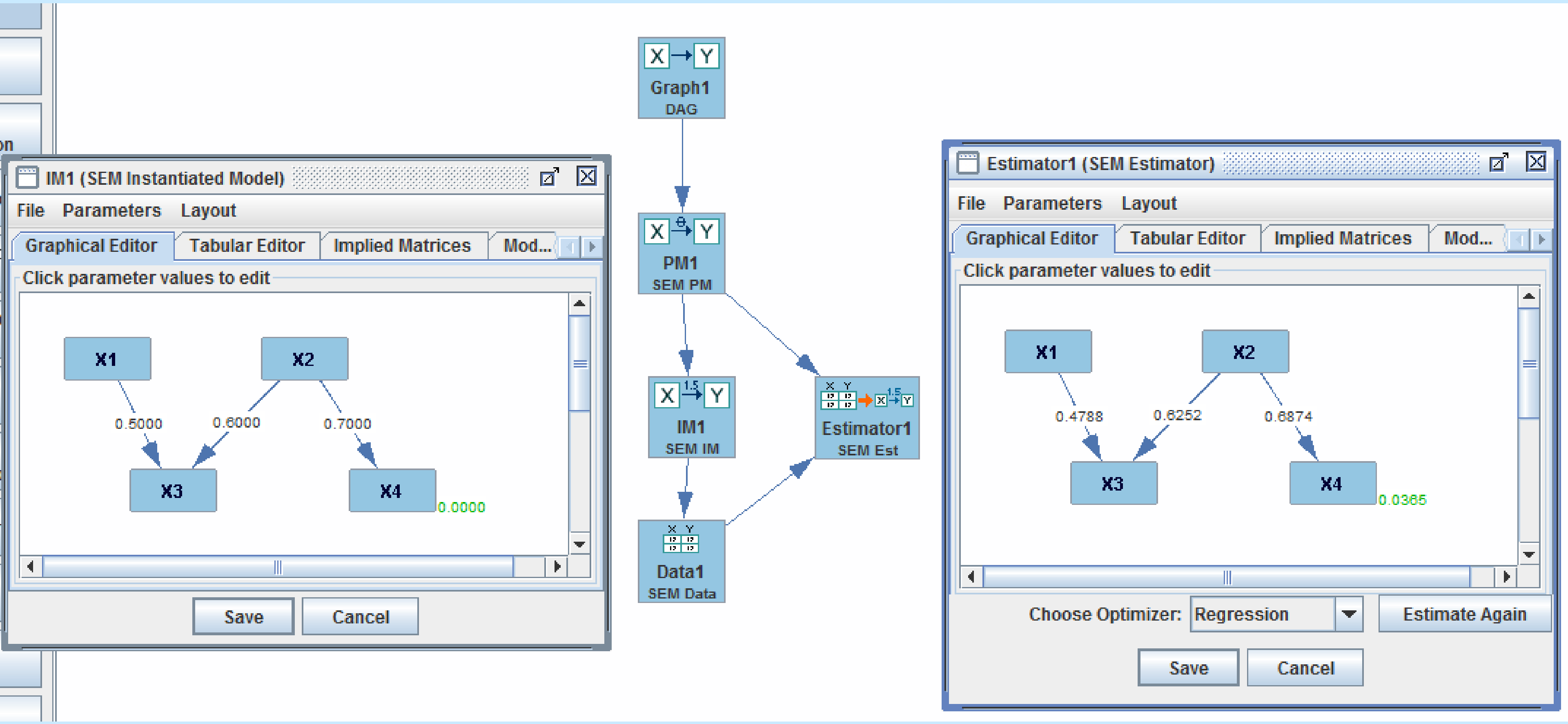
untitled1.tet - Tetrad 5.0.0-1
File Edit Logging Template Window Help
untitled1.tet
Graph
Graph Manipulation
Comparison
Parametric Model
Instantiated Model
Data
Data Manipulation
Estimator
Updater
Graph1
DAG
PM1
SEM PM
IM1
SEM IM
Data1
SEM Data
Estimator1
No model

Tetrad Demo and Hands-on

1) Select Template: “Estimate from Simulated Data”
2) Build the standardized SEM IM shown below
3) Generate simulated data N=1000
4) Estimate model.
5) Save session as “Estimate1”



Estimation



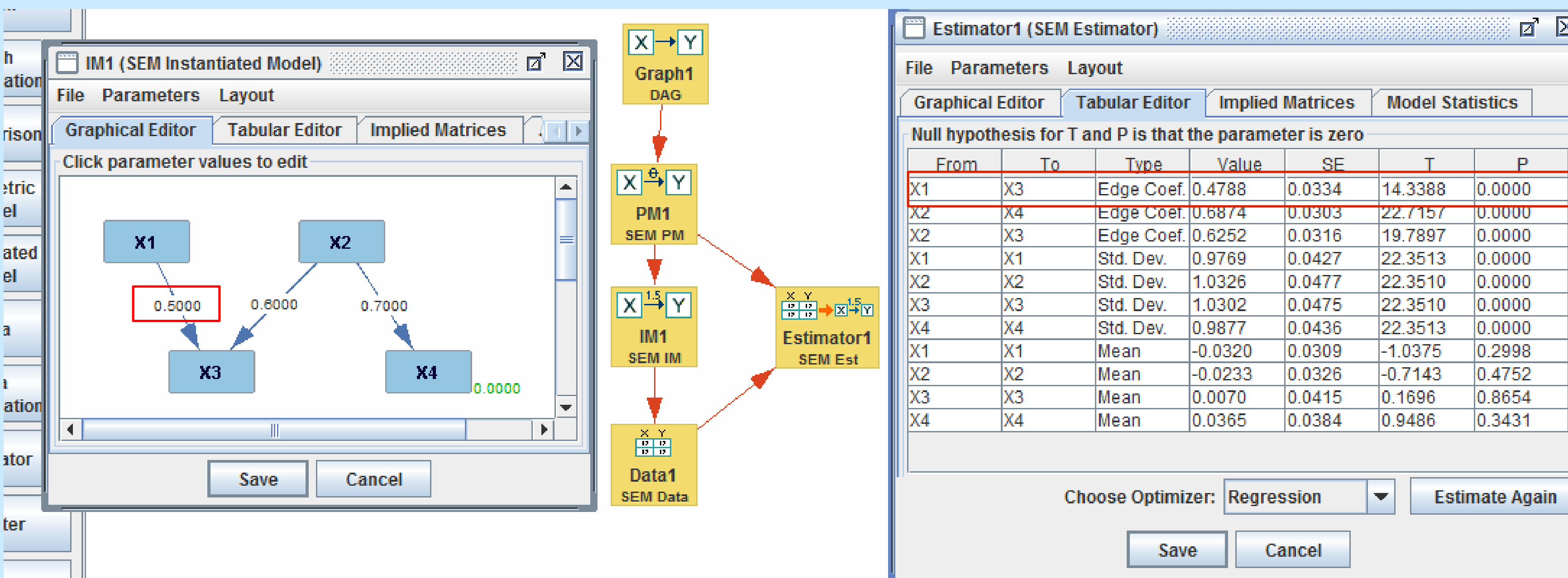
IM1 (SEM Instantiated Model)
File Parameters Layout
Graphical Editor Tabular Editor Implied Matrices Mod...
Click parameter values to edit
X1
X2
0.5000
0.6000
0.7000
X3
X4
0.0000
Save Cancel
Graph1
DAG
PM1
SEM PM
IM1
SEM IM
Data1
SEM Data
Estimator1
SEM Est
Estimator1 (SEM Estimator)
File Parameters Layout
Graphical Editor Tabular Editor Implied Matrices Mod...
Click parameter values to edit
X1
X2
0.4788
0.6252
0.6874
X3
X4
0.0365
Choose Optimizer: Regression
Estimate Again
Save Cancel

Coefficient inference vs. Model Fit

Coefficient Inference: Null: coefficient = 0, e.g., $\beta_{X1\rightarrow X3} = 0$

p-value = p(Estimated value $\hat{\beta}_{X1\rightarrow X3} \geq .4788 \mid \beta_{X1\rightarrow X3} = 0$ & *rest of model correct*)

Reject null (coefficient is "significant") when p-value < α, α usually = .05



Coefficient inference vs. Model Fit

Coefficient Inference: Null: coefficient = 0, e.g., $\beta_{X1 \rightarrow X3} = 0$

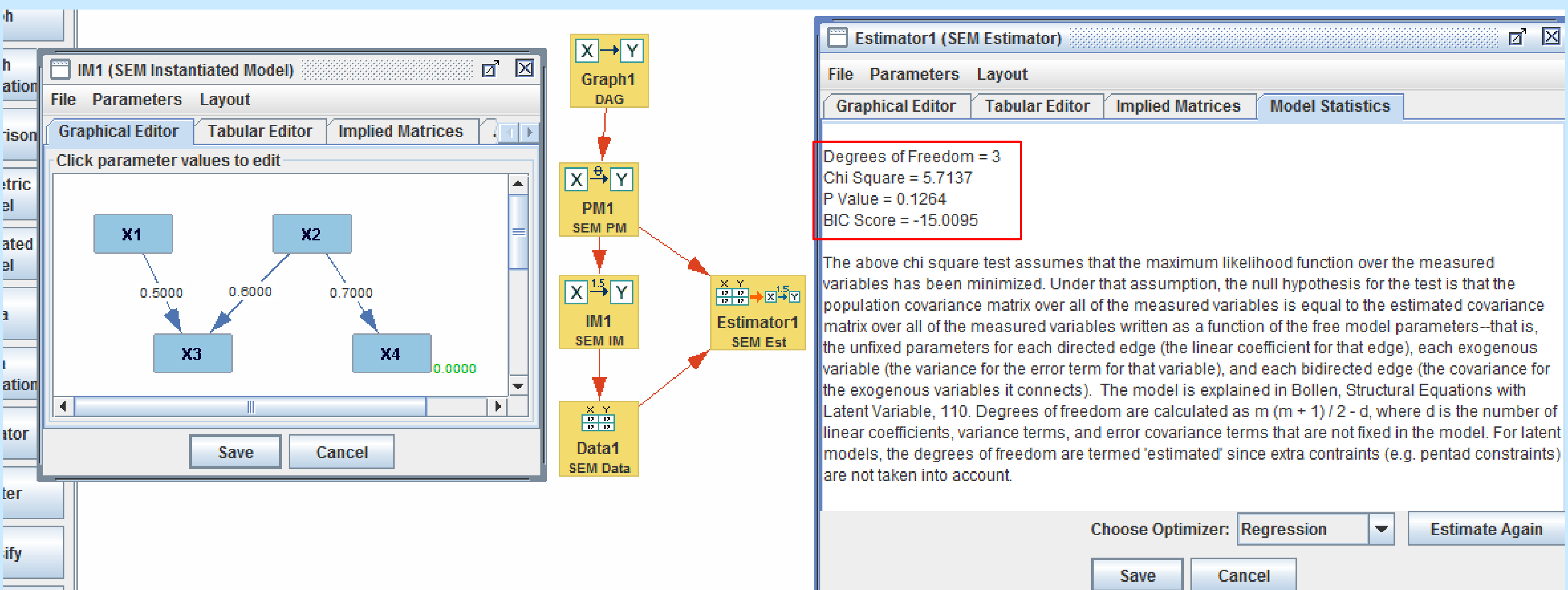
p-value = p(Estimated value $\widehat{\beta}_{X1 \rightarrow X3} \geq .4788 \mid \beta_{X1 \rightarrow X3} = 0$ & *rest of model correct*)

Reject null (coefficient is “significant”) when p-value < < α, α usually = .05,

---

Model fit: Null: Model is *correctly specified* (constraints true in population)

p-value = p(f(Deviation(Σ_{ml},S)) $\geq$ 5.7137 | Model correctly specified)

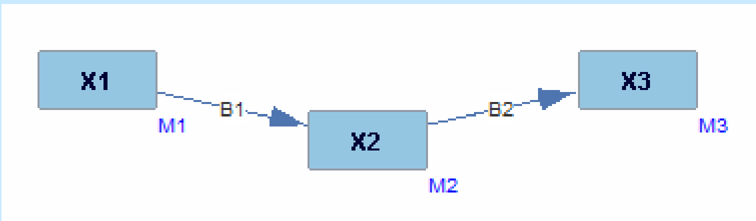


Coefficient inference vs. Model Fit

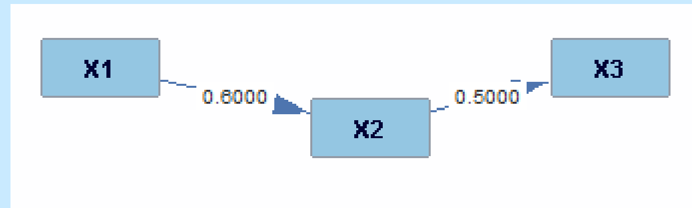
	coefficient $\hat{\beta}_{X1 \rightarrow X3}$ Null: $\beta_{X1 \rightarrow X3} = 0$	Model fit χ^2_{df} Null: Model is correctly specified
p-value < .05	Can reject 0 Significant edge	Can reject correct specification, Model not correctly specified
p-value > .05	Can't reject 0, insignificant edge	Can't reject correct specification, model *may be* correctly specified

Model Fit

Specified Model



True Model



Implied Covariance Matrix

	X1	**X2**	**X3**
X1	1		
X2	β1	1	
X3	β1*β2	β2	1

Population Covariance Matrix

	X1	**X2**	**X3**
X1	1		
X2	.6	1	
X3	.3	.5	1

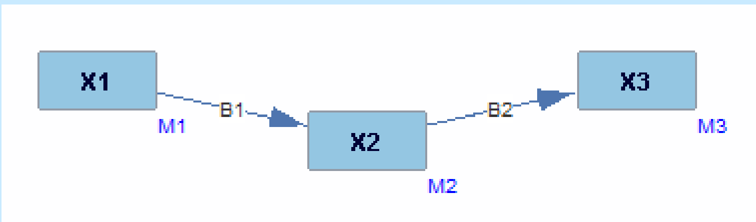
$$\widehat{\beta 1} = r_{X1,X2} = \sim .6$$

$$\widehat{\beta 2} = r_{X2,X3} = \sim .5$$

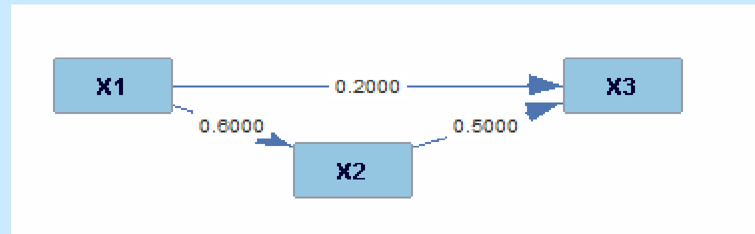
$$\widehat{\beta 1}\ \widehat{\beta 2} = \sim .3 = \widehat{\rho}_{X1,X3}$$

Model Fit

Specified Model



True Model



Implied Covariance Matrix

	X1	**X2**	**X3**
X1	1		
X2	$\beta1$	1	
X3	$\beta1*\beta2$	$\beta2$	1

Population Covariance Matrix

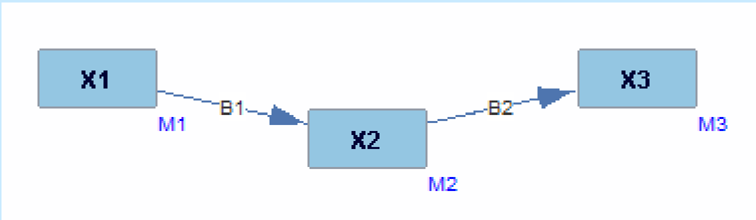
	X1	**X2**	**X3**
X1	1		
X2	.6	1	
X3	.5	.5	1

Unless $r_{x1,x3} = r_{x1,x2}\, r_{x2,x3}$

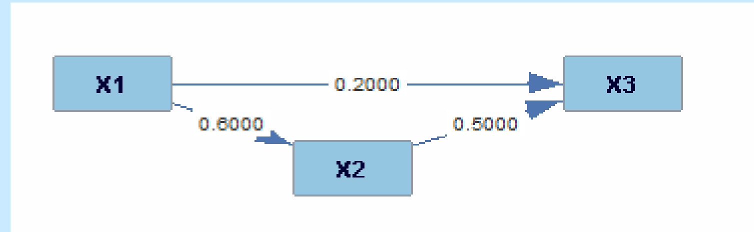
Estimated Covariance Matrix $\neq$ Sample Covariance Matrix

Model Fit

Specified Model



True Model



Implied Covariance Matrix

	X1	**X2**	**X3**
X1	1		
X2	$\beta 1$	1	
X3	$\beta 1 * \beta 2$	$\beta 2$	1

Population Covariance Matrix

	X1	**X2**	**X3**
X1	1		
X2	.6	1	
X3	.32	.5	1

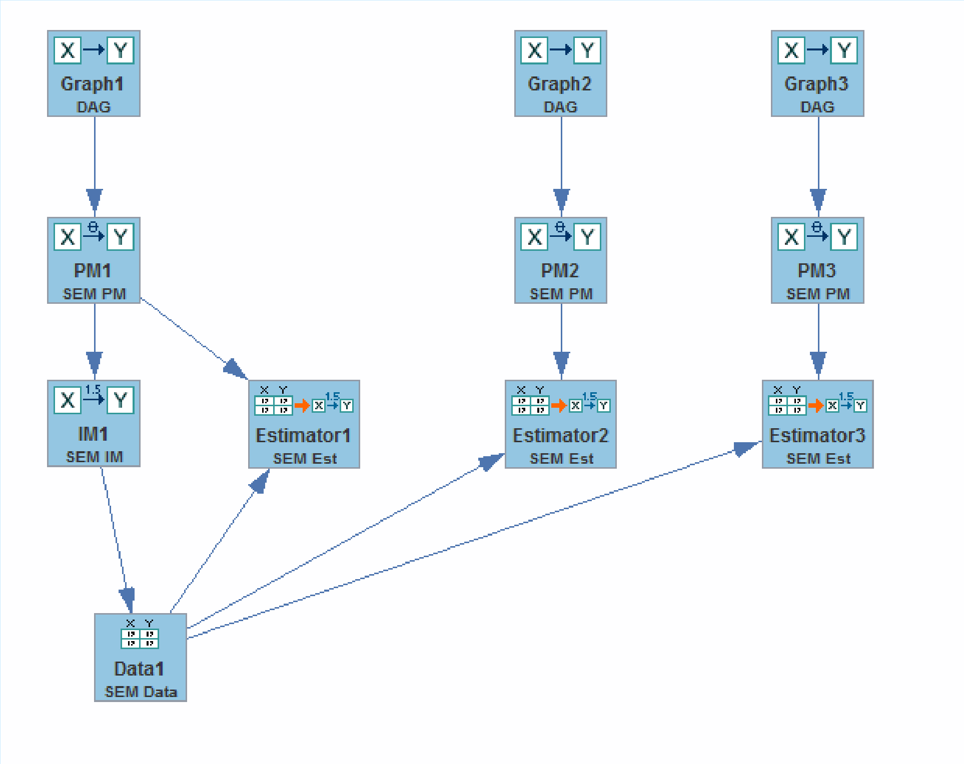
Model fit: Null: Model is *correctly specified* (constraints true in population)

$$\rho_{X1,X3} = \rho_{X1,X2}\, \rho_{X2,X3}$$

p-value = p(f(Deviation(Σ_{ml},S)) ≥ $\chi 2$ | Model correctly specified)

Tetrad Demo and Hands-on

1) Create two DAGs with the same variables – each with one edge flipped, and attach a SEM PM to each new graph (copy and paste by selecting nodes, Ctl-C to copy, and then Ctl-V to paste)
2) Estimate each new model on the data produced by original graph
3) Check p-values of:
   a) Edge coefficients
   b) Model fit
4) Save session as: "estimation2"



Charitable Giving

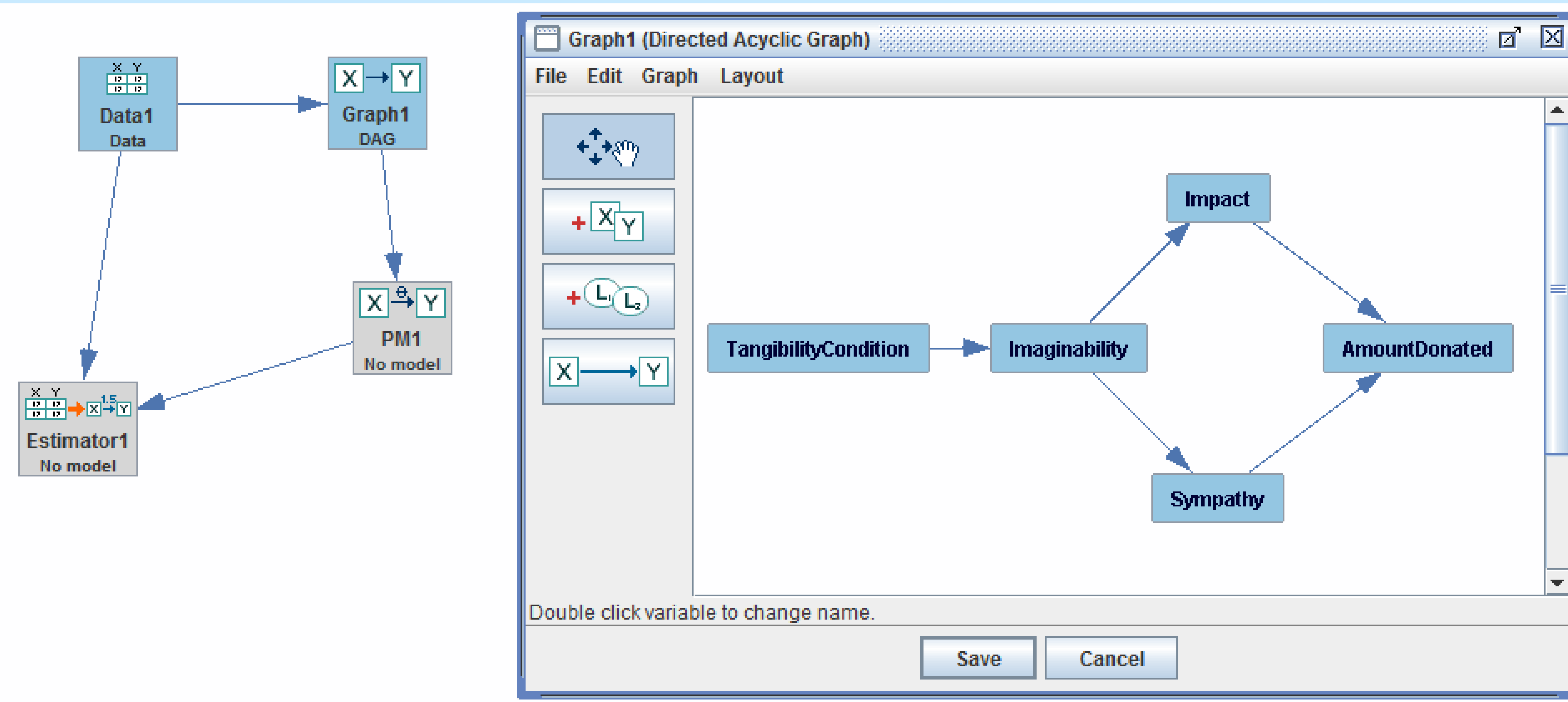
What influences giving? Sympathy? Impact?

"The Donor is in the Details", Organizational Behavior and Human Decision Processes, Issue 1, 15-23, C. Cryder, with G. Loewenstein, R. Scheines.

N = 94

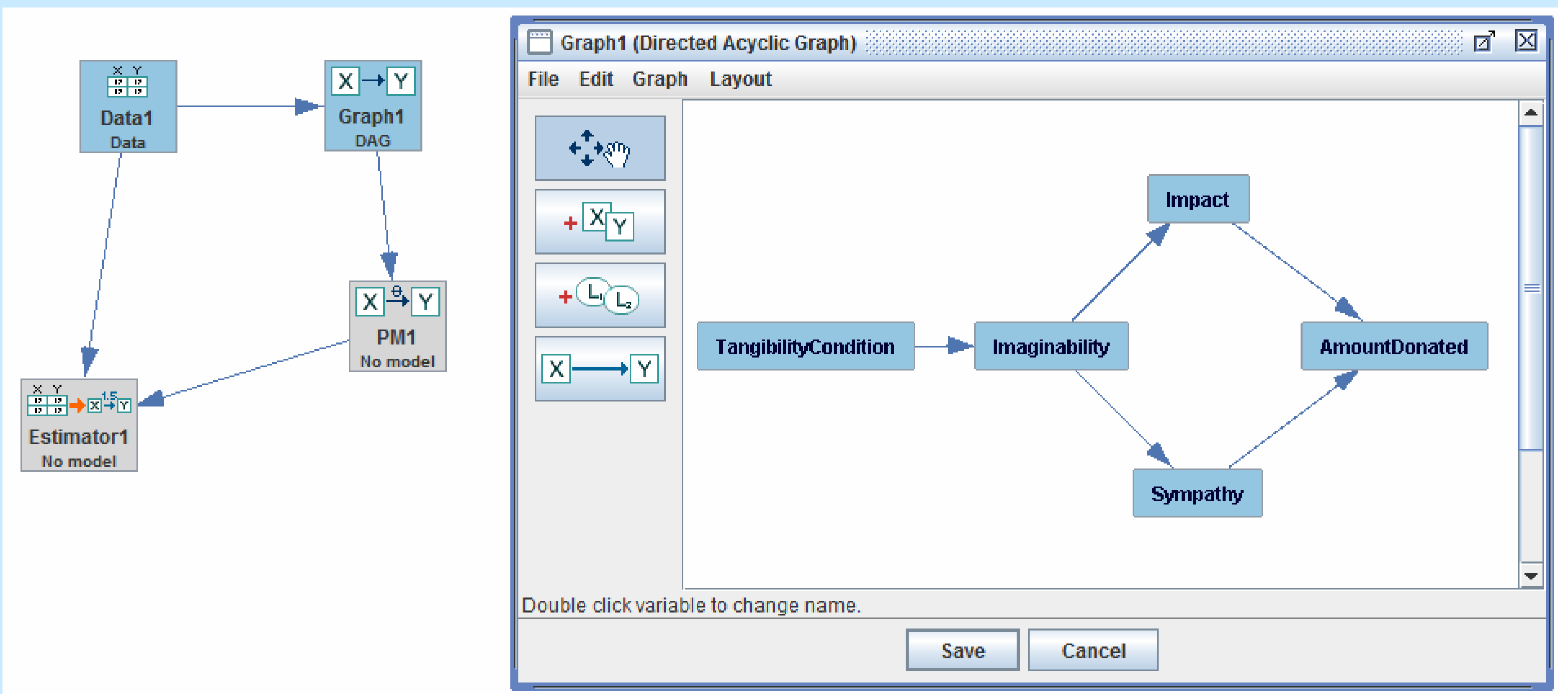
TangibilityCondition	[1,0]	Randomly assigned experimental condition
Imaginability	[1..7]	How concrete scenario I
Sympathy	[1..7]	How much sympathy for target
Impact	[1..7]	How much impact will my donation have
AmountDonated	[0..5]	How much actually donated

Theoretical Hypothesis



Tetrad Demo and Hands-on

1) Load charity.txt (tabular – not covariance data)
2) Build graph of theoretical hypothesis
3) Build SEM PM from graph
4) Estimate PM, check results



Foreign Investment

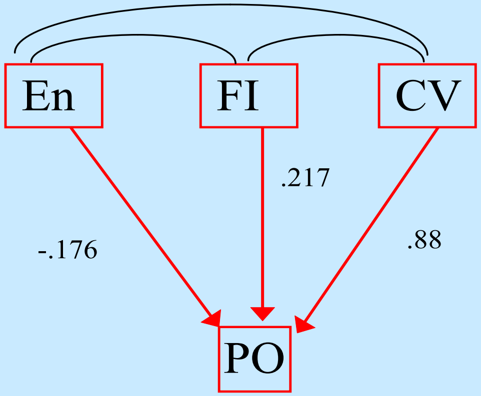
Does Foreign Investment in 3rd World Countries inhibit Democracy?

Timberlake, M. and Williams, K. (1984). Dependence, political exclusion, and government repression: Some cross-national evidence. American Sociological Review 49, 141-146.

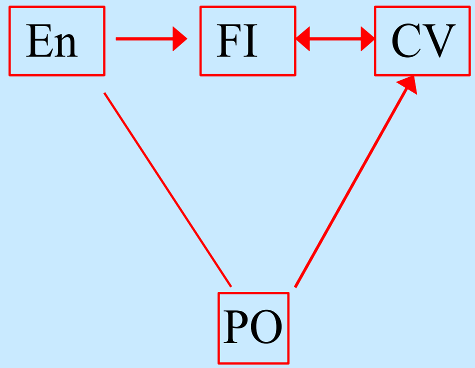
N = 72

PO	degree of political exclusivity
CV	lack of civil liberties
EN	energy consumption per capita (economic development)
FI	level of foreign investment

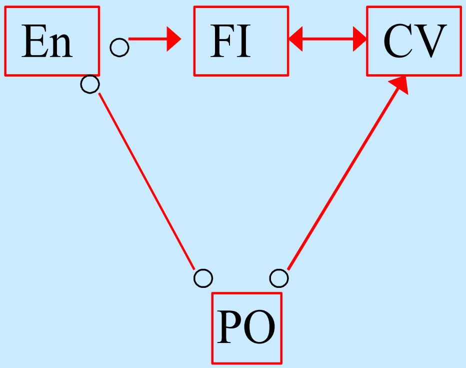
Case Study: Foreign Investment *Alternative Models*



Regression

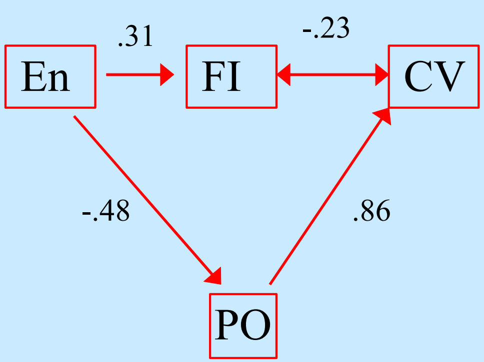


Tetrad - PC



Tetrad - FCI

There is no model with testable constraints (df > 0) that is not rejected by the data, in which FI has a positive effect on PO.



Fit: df=2, $\chi 2$=0.12,
p-value = .94

Tetrad Demo and Hands-on

1) Load tw.txt (this IS covariance data)
2) Do a regression
3) Build an alternative hypothesis, Graph - SEM PM, SEM IM
4) Estimate PM, check results

Hands On
Lead and IQ

Lead:	Lead concentration in baby teeth
CIQ:	child’s IQ score at 7
PIQ:	Parent’s average IQ
MED:	mother’s education (years)
NLB:	number of live births prior to child
MAB:	mother’s age at birth of child
FAB:	father’s age at birth of child

Hands On
Lead and IQ

1) Load leadiq1.tet
2) Specify different hypotheses, test the model fit on each
3) See if you can find a model (without using search), that is not rejected by the data