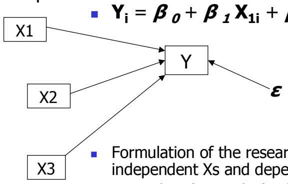
Structural Equation Modeling

Mgmt 290
Lecture 2 – Path Analysis
Oct. 3, 2005

Regression Modeling

- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

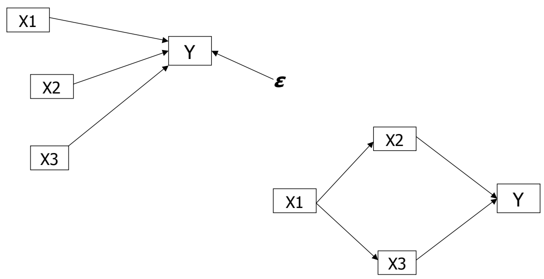


- Formulation of the research question in terms of independent Xs and dependent variable Y
- To explain (or predict) Y by Xs (one equation)
- Explain the variance of Y
- Assumptions: measured without errors, linear, …

In the real world

- No clear differentiation between independent variables and dependent variables
- How the Xs relate to each other may be important to our research
- Not all Xs affect Y directly
- Variables may be measured with error

Need to Move from Regression to SEM



(1) Selecting SEM Models to Fit Your Data

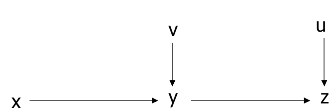
- # of models = $4^{n!/2!(n-2)!}$

# Vars	# Models
2	4
3	64
4	4096
5	1048576
6	1073741824

(2) Estimating the Coefficients of SEM Model

y = a1 + b1 * x + v

z = a2 + b2 * y + u



FOUR Elements of a Method

- 1) Equation (model representation and model types)
- 2) Estimation (identification issue and estimation options)
- 3) Errors (between data and assumptions) and Evaluation of SEM Models (Fit Indexes)
- 4) Explanation of the Results
- *In these four areas, methods diff from each other.*

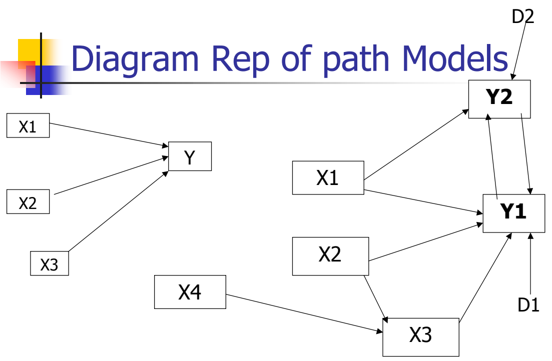
4 Elements of Regression

- 1) Equation - $\mathbf{Y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} \mathbf{X_{1i}} + \boldsymbol{\beta_2} \mathbf{X_{2i}} + \boldsymbol{\beta_3} \mathbf{X_{3i}} + \boldsymbol{\varepsilon_i}$
- 2) Estimation – OLS
- 3) Errors (Assumptions and Diagnostics) and Evaluation (R^2)
- 4) Using B_i, t statistics for results Explanation

Part 1: Equations and More
-- The Representation of Path Models

3 ways: diagram, equation, table (matrix)

Diagram Rep of path Models



A picture is worth a thousand words

Equation Rep of Path Models

- **SEM**
- $Y2_i = \beta_0 + \beta_1 Y_{1i} + \beta_2 X_{1i} + \varepsilon_i$
- $Y1_i = \text{Ø}_0 + \text{Ø}_1 X_{1i} + \text{Ø}_2 X_{2i} + \text{Ø}_3 X_{3i} + \varepsilon 1_i$
- $X3_i = C_0 + C_2 X_{2i} + C_3 X4_i + \varepsilon 2_i$

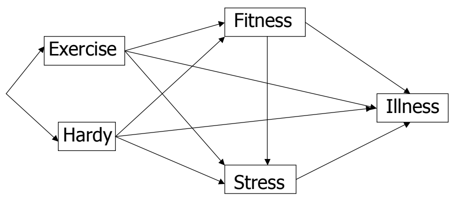
- ***Regression***
- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

Table Rep of Path Models

	X1	X2	X3	X4	Y1
Y1	1	1	1	0	1
Y2	1	0	0	0	1
X3	0	1	1	1	0

- Called as "system matrix"

Example - Diagram



- Kline 2005, p67

Example - equations

- Fitness = β_0 + β_1 Exercise + β_2 Hardiness
- Stress = $Ø_0$ + $Ø_1$ Exercise + $Ø_2$ Hardiness + $Ø_3$ Fitness
- Illness = C_0 + C_1 Exercise + C_2 Hardiness + C_3 Fitness + C_4 Stress

Example - table

	Exercise	Hardiness	Fitness	Stress
Fitness	1	1	1	0
Stress	1	1	1	1
Illness	1	1	1	1

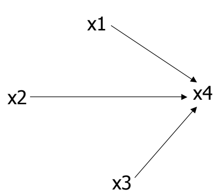
Building Blocks of Path Model

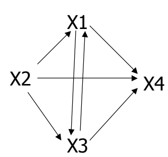
- 1) Variables – exogenous, endogenous
- 2) Relationships – direct, indirect, reciprocal,
- 3) Residual – disturbance and (measurement errors)

- Similar to regression

Model Complexity 1

- X1 → X2 → X3 → X4
- 4 X 3 X 2 X 1 = 24

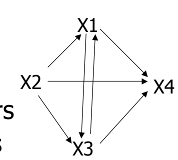




Model Complexity 2

	x1	x2	x3	x4
x1	1	.23	.17	.45
x2	.23	1	.56	.09
x3	.17	.56	1	.33
x4	.45	.09	.33	1

-
- Number of observations
- = v(v+1) / 2



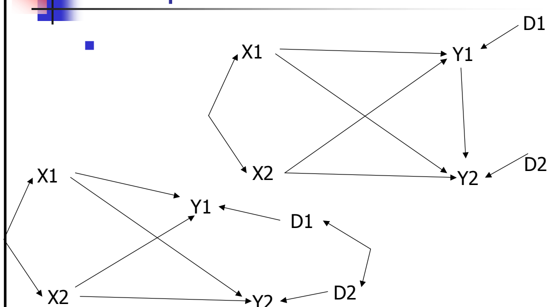
- Number of model parameters
- <= number of observations
- # parameters = p + e + d
- (p - # paths, e - #obs for exogenous, d - #obs for disturbance)

Two Types of Path Models

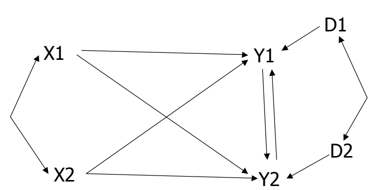
- Recursive
- 1) disturbances uncorrelated
- 2) no feedback loops

- Non-recursive
- 1) have feedback loops
- 2)or have correlated disturbances

Examples of Recursive Models



Examples of Non-Recursive Models



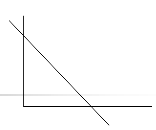
Part 2: Estimation Methods for Path Models

Identification Issue and
Estimation Options

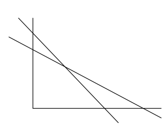
The Identification Issue

- whether or not theoretically identifiable
- (whether or not the model parameters can be represented by observations – covariances)
- Or whether a set of equations can be solved.
- Diff $\Sigma (y_i - b_0 - b_1 * x_i)^2$ against b_0 and b_1, then solve two equations to get b0 and b1 for simple regression
- similar for SEM (see Duncan's book)

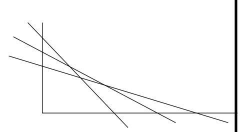
Examples



- Under-identified
- x+y = 6
- Just-identified
- x+y=6
- 2x+y=10



- Over-identified
- x+y=6
- 2x+y=10
- 3x+y=12

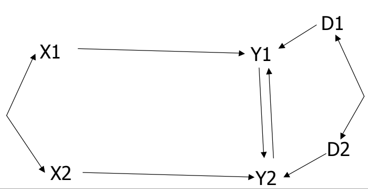


Conditions for Identifiable

Pattern of disturbance correlation	Feedback loops?	Conditions for identification	Necessary or sufficient
All possible pairwise	Yes or no	Parameters <= observations Order condition Rank condition	Nec Nec suf
All possible pairwise within recursively related blocks	Yes or no	Parameters <= observations Order condition Rank condition in each block	Nec Nec suf
Any other pattern	Yes or no	Parameters <= observations Model-fitting program yields a solution that passes checks or uniqueness	Nec Nec

The Order Condition

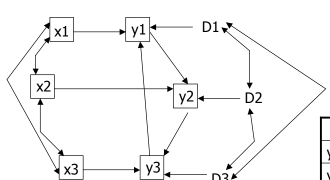
- Number of excluded variables >=
- (number of endogenous variables – 1)
- For each endogenous variable



(Kline p306-307)

The Rank Condition

- Use the system matrix
- rank >= #endogenous vars - 1



	x1	x2	x3	y1	y2	y3
y1	1	0	0	1	0	1
y2	0	1	0	1	1	0
y3	0	0	1	0	1	1

Estimation Options

- OLS
- ML

- OLS and ML produces the identical results for just-identified recursive models, and similar results for over-identified models

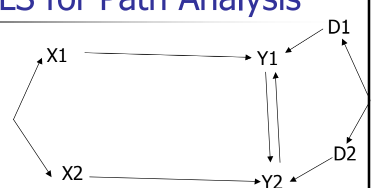
About ML

- Maximum Likelihood Estimation

- Simultaneous
- Iterative
- (starting values)

- Most common used (default method in LISREL and many other programs)

Using OLS for Path Analysis



- Ok for recursive models
- Not for non-recursive (violate the assumption that residuals uncorrelated with ind vars)

Part 3: Errors and Evaluation

The Fit Indexes

- Chi squares – most common used
- d.f. = # obs - # parameters
- RMR (mean squared residual)
- RMSEA (root mean squared error of approximation)
- GFI & AGFT (0 ~ 1)
- (Similar to R^2 and adjusted R^2)

More about Fit indexes

- Dozens available
- LISREL prints 15
- ChiSquares/df < 3
- GFI > .90
- Inspect correlation residuals

The Main Assumptions

- Similar to multiple regression
- Except for
- 1) Correlation between endogenous vars and disturbances
- 2) Multivariate normality

More on assumptions

- Linearity & residuals normally dist. …
- -> affect sig. Tests
- Normality does not affect ML estimates much

Diagnostics

- Multi-collinearity
- Normality
- Linearity
- Homoscedasticity
- Missing Values
- Outliners

Part 4: Explanation of the Findings

Path coefficients

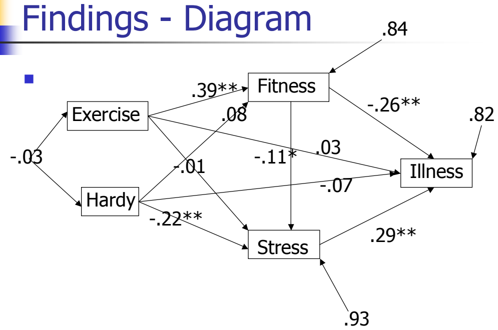
Interpret the Findings

- Path coefficients
- (standardized, un-standardized)
- (direct, indirect, total effects)

- Significance of path coefficients

<10	Small
.30	medium
>.50	large

Findings - Diagram



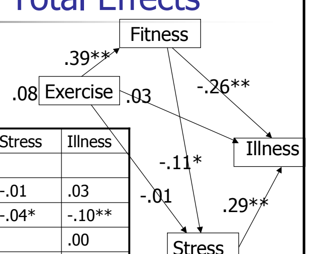
Findings - equations

- **Fitness = β_0 + *.11* Exercise + *.39* Hardiness**
- **Stress = $Ø_0$ - .01 Exercise -.39 Hardiness -.04 Fitness**
- **Illness = C_0 +.32 Exercise –12.14 Hardiness –8.84 Fitness + 27.12 Stress**

Findings - table

	Exercise	Hardiness	Fitness	Stress
Fitness	**1.11**(.01)**	**.39 (.23)**	**0**	**0**
Stress	**-.01(.01)**	**-.39(.09)**	**-.04(.02)**	**0**
Illness	**.32(.48)**	**-12.14(7.98)**	**-8.84(1.75)**	**27.12(4.55)**

Calculation of Indirect and Total Effects



	Fitness	Stress	Illness
Exercise			
Direct Effect	.39**	-.01	.03
Indirect via Fitness		-.04*	-.10**
Indirect via Stress			.00
Indirect via Fitness & Stress			-.01
Total Effect	.39	-.05	-.08

About the assignment # 1

- The study of structural equation models can be divided into two parts: the easy part and the hard part – "talking about it" and "doing it". (Duncan 1975, p149)
- The easy part is mathematical. The hard part is constructing causal models that are consistent with sound theory. (Wolfle 1985, p385)

Assignment # 1

- Introduction
- Argument
- your theory, causal models represented in words
- Variables/Dataset
- (including literature, NO more than 2 pages)