

Structural Equation Modeling Using Stata

Paul D. Allison, Ph.D.

Upcoming Seminar:
August 16-17, 2018, Stockholm

- 1 **Introduction to Structural Equation Modeling Using Stata**
- 2 **Structural Equation Models**
- 3 **What is SEM good for?**
- 4 **SEM**
- 5 **Preview: A Latent Variable SEM**
- 6 **Latent Variable Model (cont.)**
- 7 **Cautions**
- 8 **Outline**
- 9 **Software for SEMs**
- 10 **Favorite Textbook**
- 11
- 12 **Linear Regression in SEM**
- 13 **GSS2014 Example**
- 14 **Linear Regression with Stata**
- 15 **FIML for Missing Data**
- 16 **Further Reading**
- 17 **Assumptions**
- 18 **FIML in Stata**
- 19 **Path Diagram (from Mplus)**
- 20 **Path Analysis of Observed Variables**
- 21 **Some Rules and Definitions**
- 22 **Three Predictor Variables**
- 23 **Two-Equation System**
- 24 **Why combine the two equations?**
- 25 **Calculation of Indirect Effect**
- 26 **A More Complex Model**
- 27 **Decomposition of Direct & Indirect Effects**
- 28 **Standardized Coefficients**

- 29 Numerical Examples
- 30 More Complex Example
- 31 Decomposition of Effects
- 32 Illness Data
- 33 Summary Data
- 34 Covariance Matrix
- 35 Covariance Matrix for Illness Data
- 36 Illness Regression in Stata
- 37 Stata Results - Unstandardized
- 38 Counting Moments & Parameters
- 39 Mplus Results - Standardized
- 40 Illness Model with Indirect Effects
- 41 Model Diagram
- 42 Path Diagrams in Stata
- 43 Results
- 44 More Goodness of Fit Measures
- 45 Identification Status of the Model
- 46 Improving the Model
- 47 GOF for Improved Model
- 48 Estimates for Improved Model
- 49 Indirect Effects
- 50 Indirect Effects in Stata
- 51 Specific Indirect Effects in Stata
- 52 Partial Correlations
- 53 Partial Correlations (cont.)
- 54 Maruyama (1998) Data
- 55 Partial Correlations in Stata
- 56 Partial Correlation Results
- 57 Causal Ordering

- 58 **How to Decide**
- 59 **Nonrecursive Systems**
- 60 **Identification Problem in Nonrecursive Models**
- 61 **Identification Problem (cont.)**
- 62 **A Just-Identified Model**
- 63 **Reduced Form Equations**
- 64 **Solutions for Structural Parameters**
- 65 **Sufficient Condition for Identification**
- 66 **Varieties of Identification**
- 67 **Problems with Instrumental Variables**
- 68 **Example of a Nonrecursive Model**
- 69 **Nonrecursive Example (cont.)**
- 70 **Stata Code for Nonrecursive Model**
- 71 **Nonrecursive Results**
- 72 **GOF for Nonrecursive**
- 73 **Latent Variable Models**
- 74 **Roadmap for Latent Variables**
- 75 **Classical Test Theory**
- 76 **Random Measurement Error**
- 77 **Reliability**
- 78 **Parallel Measures**
- 79 **Tau-Equivalent Measures**
- 80 **Tau-Equivalence: Example**
- 81 **Tau-Equivalence in Stata**
- 82 **Congeneric Tests**
- 83 **Three Congeneric Tests**
- 84 **Three Congeneric Measures (cont.)**
- 85 **Identification in General**
- 86 **Standardized Version**

- 87 **Digression: Tracing Rule for Correlations**
- 88 **Tracing Rule (cont.)**
- 89 **Tracing Rule (cont.)**
- 90 **Standardized Version of 3 Congenerics (cont.)**
- 91 **Three Congenerics: Example**
- 92 **Three Congenerics: Stata**
- 93 **Three Congenerics: Results**
- 94 **Four Congeneric Measures**
- 95 **Overidentification with 4 Congeneric Measures**
- 96 **Four Congeneric Measures with Stata**
- 97 **Estimates for Four Congenerics**
- 98 **GOF for Four Congenerics**
- 99 **Alternative Model for 4 Congeneric Measures**
- 100 **Stata for Alternative Model**
- 101 **Results for Alternative Model**
- 102 **Heywood Case**
- 103 **Factor Models**
- 104 **Factor Models (cont.)**
- 105 **Identification (Standardized)**
- 106 **Identification (cont.)**
- 107 **Two Approaches to Identification Problem**
- 108 **Identification (Unstandardized)**
- 109 **Determining Identification**
- 110 **Normalizing Constraints**
- 111 **Normalizing Constraints**
- 112 **ML Estimation of CFA Models**
- 113 **Multivariate Normality**
- 114 **ML Details**
- 115 **Chi-Square Test**

- 116 **Example: Self-Concept Measurement**
- 117 **Self Concept Path Diagram**
- 118 **Self Concept. Results**
- 119 **Self Concept Results (cont.)**
- 120 **Self Concept Results (cont.)**
- 121 **Global Goodness of Fit Measures**
- 122 **Other Global Measures**
- 123 **Other Global Measures (cont.)**
- 124 **Specific Goodness of Fit Measures**
- 125 **Standardized Residuals for Self-Concept Model**
- 126 **Modification Indices**
- 127 **Mod Indices for Self-Concept**
- 128 **Mod Indices for Self-Concept (cont.)**
- 129 **Freeing Up Parameters**
- 130 **Results from Freeing 1 Parameter**
- 131 **Selected Results (cont.)**
- 132 **Correlated Errors**
- 133 **Two Correlated Errors**
- 134 **A Five-Indicator Model**
- 135 **A Two-Factor Model**
- 136 **Example: Self-Concept Data**
- 137 **Selected Results**
- 138 **The General Structural Equation Model**
- 139 **GSS2014 Example: Stata Code**
- 140 **GSS2014 Example: GOF Results**
- 141 **GSS2014: Standardized Results**
- 142 **GSS2014: Standardized Results**
- 143 **Farm Manager Example (Rock et al. 1977)**
- 144 **Farm Managers Path Diagram**

- 145 **Farm Managers: Stata Code**
- 146 **Farm Managers. Selected Results**
- 147 **Selected Results (cont.)**
- 148 **A Tau-Equivalent Model**
- 149 **Parallel Model**
- 150 **Identification in SEM Models**
- 151 **An Identified SEM**
- 152 **What to Do If Endogenous Variables Aren't Normal**
- 153 **Example: NLSY Data**
- 154 **ML Results for NLSY Data**
- 155 **Both Variables Highly Skewed**
- 156 **Satorra-Bentler Robust SE's**
- 157 **Weighted Least Squares**
- 158 **Weighted Least Squares**
- 159 **WLS Results**
- 160 **Multiple Group Analysis**
- 161 **Subjective Class Example**
- 162 **Reading in the Data in Stata**
- 163 **Subjective Class Models**
- 164 **Stata Code for 2-Group Models**
- 165 **Stata Code (cont.)**
- 166 **Tests for Comparing the Groups**
- 167 **Model 2 Results**
- 168 **Model 2 Results (cont.)**
- 169 **Wald & Score Tests Comparing Groups**
- 170 **Output from estat ginvariant**
- 171 **More output from estat ginvariant**
- 172 **Interactions and Non-Linearities**
- 173 **Ordinal and Binary Data**

- 174 **Special Correlations**
- 175 **Special Correlations**
- 176 **Specialized Models**
- 177 **CFA Model with Categorical Indicators**
- 178 **Selected Results**
- 179 **Other Capabilities of gsem**
- 180 **Cautions About SEMs**
- 181 **Examples I Don't Like**
- 182 **Examples I Like**
- 183 **SEMs and Causality**
- 184 **Exemplary Article**
- 185 **Some Recommendations:
Lest you forget**

Introduction to Structural Equation Modeling Using Stata

Paul D. Allison, Instructor

February 2017

www.StatisticalHorizons.com

Copyright © 2017 Paul Allison

1

Structural Equation Models

The classic SEM includes many common linear models used in the behavioral sciences:

- Multiple regression
- ANOVA
- Path analysis
- Multivariate ANOVA and regression
- Factor analysis
- Canonical correlation
- Non-recursive simultaneous equations
- Seemingly unrelated regressions
- Dynamic panel data models

2

What is SEM good for?

- Modeling complex causal mechanisms.
- Studying mediation (direct and indirect effects).
- Correcting for measurement error in predictor variables.
- Avoiding multicollinearity for predictor variables that are measuring the same thing.
- Analysis with instrumental variables.
- Modeling reciprocal relationships (2-way causation).
- Handling missing data (by maximum likelihood).
- Scale construction and development.
- Analyzing longitudinal data.
- Providing a very general modeling framework to handle all sorts of different problems in a unified way.

3

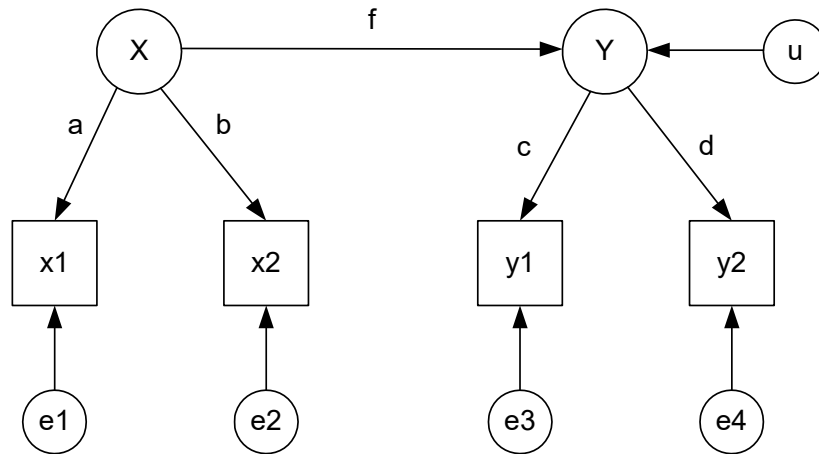
SEM

Convergence of psychometrics and econometrics

- Simultaneous equation models, possibly with reciprocal (nonrecursive) relationships
- Latent (unobserved) variables with multiple indicators.
- Latent variables are the most distinguishing feature of SEM. For example:

4

Preview: A Latent Variable SEM



X and Y are unobserved variables, x1, x2, y1, and y2 are observed indicators, e1-e4 and u are random errors. a, b, c, d, and f are correlation coefficients.

5

Latent Variable Model (cont.)

- If we know the six correlations among the observed variables, simple hand calculations can produce estimates of a through f . We can also test the fit of the model.
- Why is it desirable to estimate models like this?
 - Most variables are measured with at least some error.
 - In a regression model, measurement error in independent variables can produce severe bias in coefficient estimates.
 - We can correct this bias if we have multiple indicators for variables with measurement error.
 - Multiple indicators can also yield more powerful hypothesis tests.

6

Cautions

- Although SEM's can be very useful, the methodology is often used badly and indiscriminately.
 - Often applied to data where it's inappropriate.
 - Can sometimes obscure rather than illuminate.
 - Easy to get sucked into overly complex modeling.

7

Outline

1. Introduction to SEM
2. Linear regression with missing data
3. Path analysis of observed variables
4. Direct and indirect effects
5. Identification problem in nonrecursive models
6. Reliability: parallel and tau-equivalent measures
7. Multiple indicators of latent variables
8. Confirmatory factor analysis
9. Goodness of fit measures
10. Structural relations among latent variables
11. Alternative estimation methods.
12. Multiple group analysis
13. Models for ordinal and nominal data

8

Software for SEMs

LISREL – Karl Jöreskog and Dag Sörbom

EQS – Peter Bentler

PROC CALIS (SAS) – W. Hartmann, Yiu-Fai Yung

Amos – James Arbuckle

Mplus – Bengt Muthén

sem, gsem (Stata)

Packages for R:

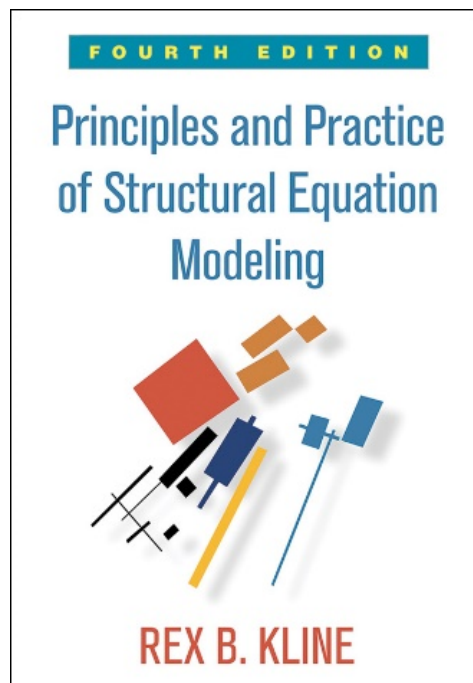
OpenMX – Michael Neale

sem – John Fox

lavaan – Yves Rosseel

9

Favorite Textbook



10