STATISTICAL HORIZONS

Introduction to Structural Equation Modeling

Paul D. Allison, Ph.D.

Upcoming Seminar:
January 31-February 1, 2020, Ft. Myers, Florida

59 **Estimates for Improved Model**

60 **Indirect Effects**

61 **Indirect Estimates**

62 **Indirect Effects in SAS**

63 **Specific Indirect Effects with PARMS**

64 **Indirect Effects in Stata**

65 **Specific Indirect Effects in Stata**

66 **Indirect Effects in lavaan**

67 **Indirect Estimates from lavaan**

68 **Partial Correlations**

69 **Partial Correlations (cont.)**

70 **Maruyama (1998) Data**

71 **Partial Correlations in Mplus**

72 **Results with Partial Correlation**

73 **Results (cont.)**

74 **Partial Correlations in SAS**

75 **Partial Correlations in Stata**

76 **Partial Correlations in lavaan**

77 **Causal Ordering**

78 **How to Decide**

79 **Nonrecursive Systems**

80 **Identification Problem in Nonrecursive Models**

81 **Identification Problem (cont.)**

82 **A Just-Identified Model**

83 **Reduced Form Equations**

84 **Solutions for Structural Parameters**

85 **Sufficient Condition for Identification**

86 **Varieties of Identification**

87 **Problems with Instrumental Variables**

146 **Multivariate Normality**

147 **ML Details**

148 **Chi-Square Test**

149 **Self-Concept Example: Mplus Code**

150 **Self Concept Path Diagram**

151 **SAS Code for Self Concept**

152 **Stata Code for Self Concept**

153 **lavaan Code for Self Concept**

154 **Self Concept: Mplus Results**

155 **Self Concept Results**

156 **Global Goodness of Fit Measures**

157 **Other Global Measures**

158 **Other Global Measures (cont.)**

159 **Specific Goodness of Fit Measures**

160 **Standardized Residuals for Self-Concept Model**

161 **Residuals in SAS**

162 **Residuals in Stata and lavaan**

163 **Modification Indices**

164 **Mod Indices for Self-Concept**

165 **Freeing Up Parameters**

166 **Results from Freeing 1 Parameter**

167 **Selected Results (cont.)**

168 **Correlated Errors**

169 **Two Correlated Errors**

170 **A Five-Indicator Model**

171 **A Two-Factor Model**

172 **Example: Self-Concept Data**

173 **Selected Results**

174 **The General Structural Equation Model**

175 **GSS2014 Example: Mplus Code**

176 **GSS2014 Example: Mplus Results**

177 **GSS2014: Mplus Standardized Results**

178 **GSS2014: Code for Other Packages**

179 **Farm Manager Example (Rock et al. 1977)**

180 **Farm Managers Path Diagram**

181 **Data and Mplus Code**

182 **SAS Code for Farm Managers**

183 **lavaan Code for Farm Managers**

184 **Stata Code for Farm Managers**

185 **Farm Managers: Selected Mplus Results**

186 **Selected Results (cont.)**

187 **Identification in SEM Models**

188 **An Identified SEM**

189 **Alternative Estimation Methods**

190 **GLS Example**

191 **GLS Results**

192 **What to Do If Endogenous Variables Aren't Normal**

193 **Example: NLSY Data**

194 **ML Results for NLSY Data**

195 **Both Variables Highly Skewed**

196 **Other ESTIMATOR Options**

197 **Satorra-Bentler Robust SE's**

198 **Weighted Least Squares**

199 **WLS in Mplus for NLSY Data**

200 **WLS Output**

201 **Multiple Group Analysis**

202 **Subjective Class Example**

203 **Subjective Class Data**

204 ☐ **Subjective Class Models**

205 ☐ **Mplus Code for Model 1**

206 ☐ **Mplus Code (cont.)**

207 ☐ **Model 4 Code**

208 ☐ **Tests for Comparing the Groups**

209 ☐ **SAS Code to Read Data**

210 ☐ **SAS Code – Model 1**

211 ☐ **SAS Code – Model 2**

212 ☐ **SAS Code – Model 3**

213 ☐ **SAS Code – Model 4**

214 ☐ **Reading in the Data in Stata**

215 ☐ **Stata Code for 2-Group Models**

216 ☐ **Stata Code (cont.)**

217 ☐ **Interactions and Non-Linearities**

218 ☐ **Interactions with Latent Variables**

219 ☐ **Interactions with Latent Variables**

220 ☐ **Ordinal and Binary Data**

221 ☐ **Special Correlations**

222 ☐ **Special Correlations**

223 ☐ **Specialized Models**

224 ☐ **Mplus for MIMIC Model with Binary Data**

225 ☐ **Diagram for Probit MIMIC Model**

226 ☐ **Stata for MIMIC Model with Binary Data**

227 ☐ **Probit Results with WLSMV Method**

228 ☐ **Probit Results (cont.)**

229 ☐ **Other Features of Mplus**

230 ☐ **Cautions About SEMs**

231 ☐ **Examples I Don't Like**

232 ☐ **Examples I Like**

233 ☐ **SEMs and Causality**

234 ☐ **Exemplary Article**

Introduction to Structural Equation Modeling

Paul D. Allison, Instructor

www.StatisticalHorizons.com

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

What is SEM good for?

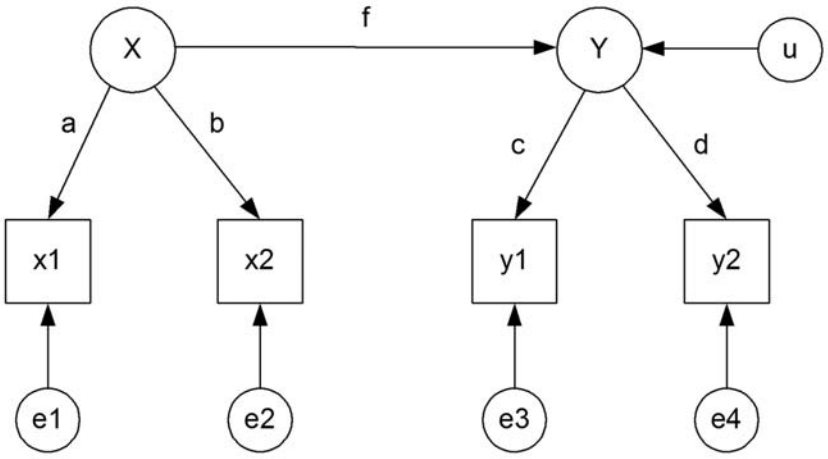
- Modeling complex causal mechanisms.
- Studying mediation (direct and indirect effects).
- Correcting for measurement error in predictor variables.
- Avoiding multicollinearity for predictor variables that are measuring the same thing.
- Analysis with instrumental variables.
- Modeling reciprocal relationships (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.

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:

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.

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.

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.

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

Software for SEMs

LISREL – Karl Jöreskog and Dag Sörbom

EQS – Peter Bentler

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

Amos – James Arbuckle

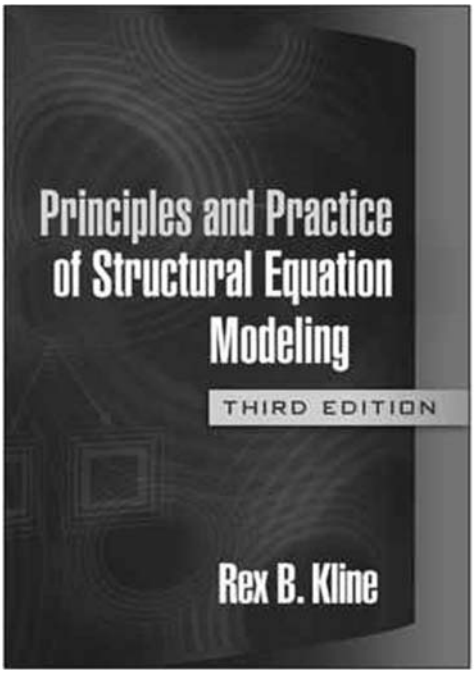
Mplus – Bengt Muthén

sem, gsem (Stata)

Packages for R:

- **OpenMX** – Michael Neale
- **sem** – John Fox
- **lavaan** – Yves Rosseel

Favorite Textbook



Linear Regression in SEM

The standard linear regression model is just a special case of SEM:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

We make the usual assumptions about ε:

- uncorrelated with the x's.
- mean of 0
- homoskedastic (variance is constant)
- normally distributed.

By default, all SEM programs do maximum likelihood (ML) estimation. Under these assumptions, ML is equivalent to ordinary least squares (OLS).

Why do it in SEM? Because SEM can handle missing data by maximum likelihood—one of the best methods available.

GSS2014 Example

Data from the 2014 General Social Survey (GSS). There were a total of 2538 respondents. Here are the variables that we will use, along with their ranges and the number of cases with data missing:

AGE	Age of respondent (18-89), 9 cases missing
ATTEND	Frequency of attendance at religious services (0-8), 13 cases missing
CHILDS	Number of children (0-8), 8 cases missing
EDUC	Highest year of school completed (0-20), 1 case missing
FEMALE	1=female, 0=male
HEALTH	Condition of health (1 excellent – 4 poor), 828 cases missing; 824 of these were not asked the question
INCOME	Total family income (in thousands of dollars), 224 cases missing
MARRIED	1=married, 0=unmarried, 4 cases missing
PAEDUC	Father's highest year school completed, father (0 – 20), 653 cases missing
PARTYID	Political party identification (1 strong democrat – 6 strong republican); 88 cases missing
POLVIEWS	Think of self as liberal or conservative (1 liberal – 7 conservative) 89 cases missing
PROCHOICE	Scale of support for abortion rights (1 – 6), 1033 cases missing; 824 of these were not asked the question **(dependent variable)**
WHITE	1=white race, 0= non-white