Introduction to Structural Equation Modeling

Paul D. Allison, Ph.D.

Upcoming Seminar:
January 31-February 1, 2020, Ft. Myers, Florida
Introduction to Structural Equation Modeling

Structural Equation Models

What is SEM good for?

SEM

Preview: A Latent Variable SEM

Latent Variable Model (cont.)

Cautions

Outline

Software for SEMs

Favorite Textbook

Linear Regression in SEM

GSS2014 Example

Regression with Mplus

Mplus Output

Linear Regression with Stata

Linear Regression with SAS

Linear Regression with lavaan

FIML for Missing Data

Further Reading

Assumptions

FIML in SAS

FIML in Stata

FIML in lavaan

FIML in Mplus

Mplus “Problem”

Path Diagram from Mplus

Path Analysis of Observed Variables

Some Rules and Definitions

Three Predictor Variables
SEMs and Causality

Exemplary Article
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:
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
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

Principles and Practice of Structural Equation Modeling
Third Edition
Rex B. Kline
Linear Regression in SEM

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

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \]

We make the usual assumptions about \( \epsilon \):
- 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:

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