Structural Equation Modeling: what is it and what can we use it for?

Professor Patrick Sturgis
What is SEM?
What is SEM?

• SEM is not one statistical ‘technique’
What is SEM?

- SEM is not one statistical ‘technique’
- It integrates a number of different multivariate techniques into one model fitting framework
What is SEM?

- SEM is not one statistical ‘technique’
- It integrates a number of different multivariate techniques into one model fitting framework
- It is an integration of:
  - Measurement theory
  - Factor (latent variable) analysis
  - Path analysis
  - Regression
  - Simultaneous equations
Useful for Research Questions that..
Useful for Research Questions that..

• Involve complex, multi-faceted constructs that are measured with error
Useful for Research Questions that..

• Involve complex, multi-faceted constructs that are measured with error

• That specify ‘systems’ of relationships rather than a dependent variable and a set of predictors
Useful for Research Questions that..

• Involve complex, multi-faceted constructs that are measured with error

• That specify ‘systems’ of relationships rather than a dependent variable and a set of predictors

• Focus on indirect (mediated) as well as direct effects of variables on other variables
Also Known as
Also Known as

- Covariance Structure Analysis
Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures
Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures
- Analysis of Linear Structural Relationships (LISREL)
Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures
- Analysis of Linear Structural Relationships (LISREL)
- Causal Modeling
Software for SEM

- There are a lot of software packages that can fit SEMs
Software for SEM

• There are a lot of software packages that can fit SEMs

• The original and best known is Lisrel, developed by Joreskog and Sorbom
Software for SEM

- There are a lot of software packages that can fit SEMs
- The original and best known is Lisrel, developed by Joreskog and Sorbom
- Mplus, EQS, Amos, Calis, Mx, SEPATH, Tetrad, R, stata
Software for SEM

• There are a lot of software packages that can fit SEMs
• The original and best known is Lisrel, developed by Joreskog and Sorbom
• Mplus, EQS, Amos, Calis, Mx, SEPATH, Tetrad, R, stata
• Some have downloadable student versions
SEM can be thought of as Path Analysis using Latent Variables
What are Latent Variables?
What are Latent Variables?

• Most social scientific concepts are not directly observable, e.g. intelligence, social capital
What are Latent Variables?

• Most social scientific concepts are not directly observable, e.g. intelligence, social capital.

• This makes them hypothetical or ‘latent’ constructs.
What are Latent Variables?

- Most social scientific concepts are not directly observable, e.g. intelligence, social capital
- This makes them hypothetical or ‘latent’ constructs
- We can measure latent variables using observable indicators
What are Latent Variables?

• Most social scientific concepts are not directly observable, e.g. intelligence, social capital

• This makes them hypothetical or ‘latent’ constructs

• We can measure latent variables using observable indicators

• We can think of the variance of a questionnaire item as being caused by:
  – The latent construct we want to measure
  – Other factors (error/unique variance)
True score and measurement error

\[ x = t + e \]

- **Measured**
- **True Score**
- **Error**
  - Mean of Errors ≠ 0
  - Mean of Errors = 0
- True value on construct
\[ X = t + e \]
\[ X = t + e \]
\( X = t + e \)
\[ X = t + e \]
\[ X = t + e \]
$X = t + e$

Observed item

True score

error
$X = t + e$

- **Observed item**
- **True score**
- **Error**
$X = t + e$

- **Observed item**
- **True score**
- **Error**
\[ X = t + e \]

Problem – with one indicator, the equation is **unidentified**
\[ X = t + e \]

Problem – with one indicator, the equation is **unidentified**

We can’t separate true score and error
Multiple Indicator Latent Variables
Multiple Indicator Latent Variables

• To identify t & e components we need multiple indicators of the latent variable
Multiple Indicator Latent Variables

• To identify t & e components we need multiple indicators of the latent variable

• With multiple indicators we can use a latent variable model to partition variance
Multiple Indicator Latent Variables

- To identify t & e components we need multiple indicators of the latent variable.
- With multiple indicators we can use a latent variable model to partition variance.
- E.g. principal components analysis transforms correlated variables into uncorrelated components.
Multiple Indicator Latent Variables

• To identify t & e components we need multiple indicators of the latent variable
• With multiple indicators we can use a latent variable model to partition variance
• e.g. principal components analysis transforms correlated variables into uncorrelated components
• We can then use a reduced set of components to summarise the observed associations
A Common Factor Model

\[ \eta \]

\[ \lambda_1 \hspace{2cm} \lambda_2 \hspace{2cm} \lambda_3 \hspace{2cm} \lambda_4 \]

\[ e_1 \rightarrow x_1 \] \[ e_2 \rightarrow x_2 \] \[ e_3 \rightarrow x_3 \] \[ e_4 \rightarrow x_4 \]

\[ \lambda = \text{Factor loadings} = \text{correlation between factor & indicator} \]
Benefits of Latent Variables
Benefits of Latent Variables

• Most social concepts are complex and multi-faceted
Benefits of Latent Variables

• Most social concepts are complex and multi-faceted
• Using single measures will not adequately cover the full conceptual map
Benefits of Latent Variables

• Most social concepts are complex and multi-faceted
• Using single measures will not adequately cover the full conceptual map
• Removes/reduces random error in measured construct
Benefits of Latent Variables

- Most social concepts are complex and multi-faceted
- Using single measures will not adequately cover the full conceptual map
- Removes/reduces random error in measured construct
- Random error in dependent variables -> estimates unbiased but less precise
Benefits of Latent Variables

• Most social concepts are complex and multifaceted
• Using single measures will not adequately cover the full conceptual map
• Removes/reduces random error in measured construct
• Random error in dependent variables -> estimates unbiased but less precise
• Random error in independent variables -> attenuates regression coefficients toward zero
Remember

SEM can be thought of as
Path Analysis
using
Latent Variables
Remember

SEM can be thought of as Path Analysis using Latent Variables

We now know about latent variables, what about path analysis?
Path Analysis
Path Analysis

• The diagrammatic representation of a theoretical model using standardised notation
Path Analysis

• The diagrammatic representation of a theoretical model using standardised notation

• Regression equations specified between measured variables
Path Analysis

• The diagrammatic representation of a theoretical model using standardised notation

• Regression equations specified between measured variables

• ‘Effects’ of predictor variables on criterion/dependent variables can be:
  – Direct
  – Indirect
  – Total
Path Diagram notation
Path Diagram notation

- Measured latent variable
- Observed / manifest variable
Path Diagram notation

- Measured latent variable
- Observed / manifest variable
- Error variance / disturbance term
Path Diagram notation

- Measured latent variable
- Observed / manifest variable
- Error variance / disturbance term
- Covariance / non-directional path
Path Diagram notation

- Measured latent variable
- Observed / manifest variable
- Error variance / disturbance term
- Covariance / non-directional path
- Regression / directional path
PDI: Single Cause
Two correlated causes
Indirect Effect
Indirect Effect
Indirect Effect

\[ \beta_1 = \text{direct effect of } X_1 \text{ on } Y \]
Indirect Effect

\[ \beta_1 = \text{direct effect of } X_1 \text{ on } Y \]
\[ \beta_2 = \text{direct effect of } X_1 \text{ on } X_2 \]
Indirect Effect

\[ \beta_1 = \text{direct effect of } X_1 \text{ on } Y \]

\[ \beta_2 = \text{direct effect of } X_1 \text{ on } X_2 \]

\[ \beta_3 = \text{direct effect of } X_2 \text{ on } Y \]
Indirect Effect

\[ \beta_1 = \text{direct effect of } X_1 \text{ on } Y \]

\[ \beta_2 = \text{direct effect of } X_1 \text{ on } X_2 \]

\[ \beta_3 = \text{direct effect of } X_2 \text{ on } Y \]

\[ \beta_2 \times \beta_3 = \text{indirect effect of } X_1 \text{ on } Y \]
Indirect Effect

\[ \beta_1 = \text{direct effect of } X_1 \text{ on } Y \]

\[ \beta_2 = \text{direct effect of } X_1 \text{ on } X_2 \]

\[ \beta_3 = \text{direct effect of } X_2 \text{ on } Y \]

\[ \beta_2 \times \beta_3 = \text{indirect effect of } X_1 \text{ on } Y \]

\[ \beta_1 + (\beta_2 \times \beta_3) = \text{total effect of } X_1 \text{ on } Y \]
So a path diagram with latent variables...
So a path diagram with latent variables...
So a path diagram with latent variables...

...is a SEM
For more information contact
ncrm.ac.uk