
Structural Equation Modeling (SEM)

—An Overview of Theories & Applications

Outline

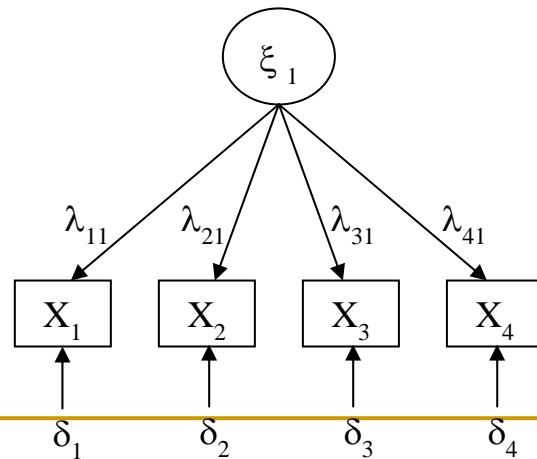
- Origin of SEM
 - Problems Relevant to SEM
 - Mathematical Statistics Issues
 - Interpretations for the Outputs from SEM
 - My Understanding of Statistics
-

1. Origin of SEM

- 1960s ~ 70s
 - Two critical problems in social studies
 - Measurement ==> Latent Variables
 - Causal Relationship ==> Structural Model, Path Analysis, etc
-

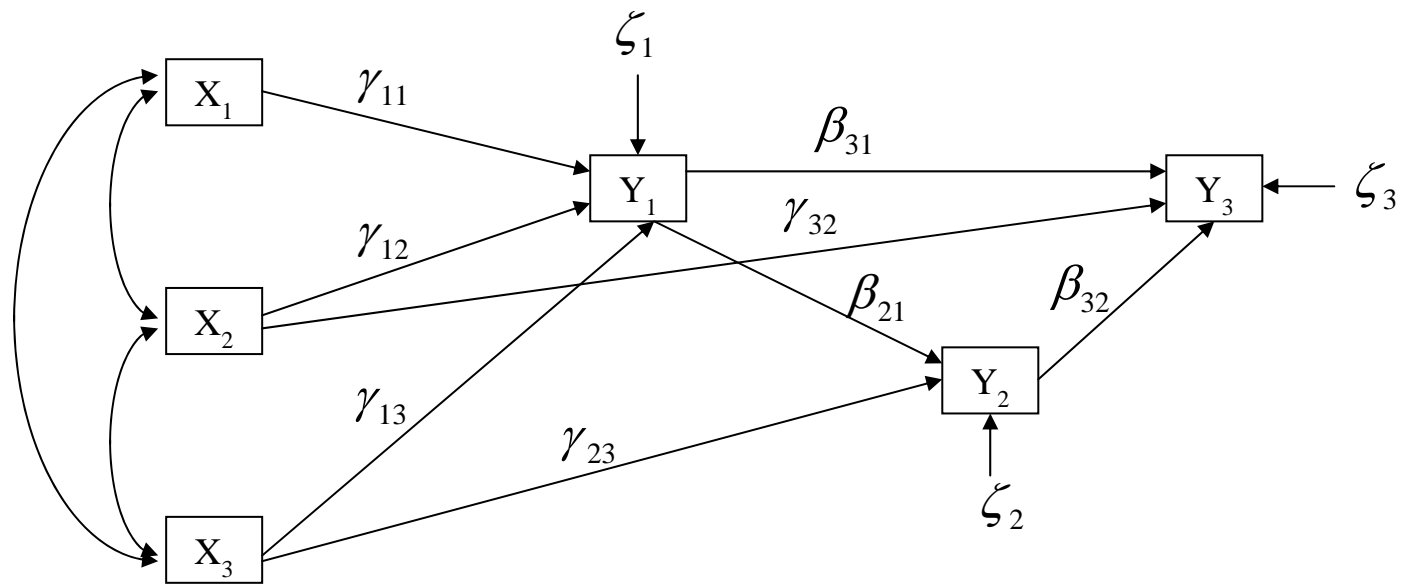
1.1 Latent Variables

- Variables that can not be measured DIRECTLY; e.g. psychological motivations, moral level, intelligence, etc
 - Underlying factors measured by observed indicators
 - Errors of measurement also under consideration

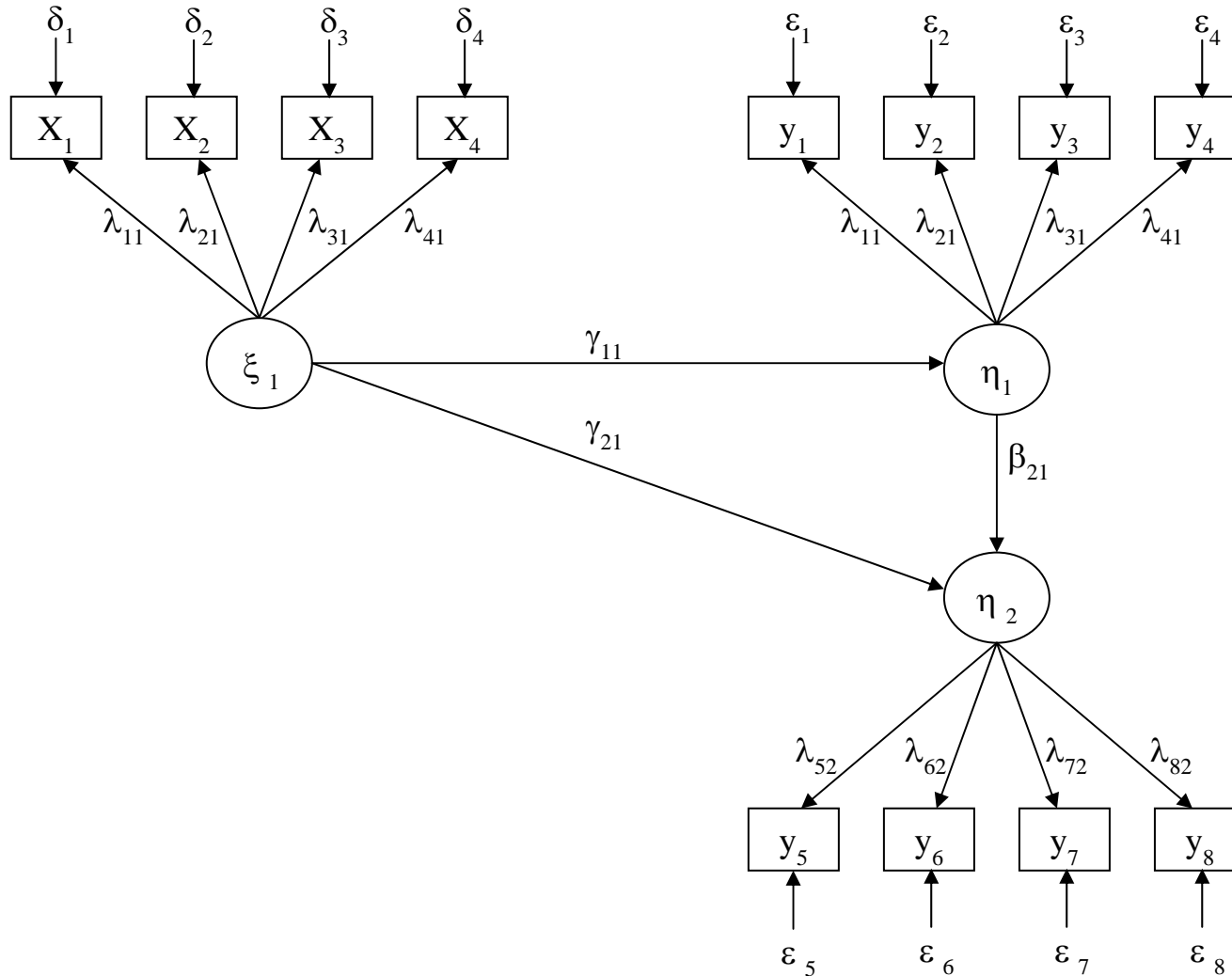


1.2 Causal Model & Path Analysis

- Regression models are too simple for social studies, especially when there are complex causal relationships!



1.3 A Typical Model



1.3 A Typical Model (cont.)

- It's also appropriate if you regard SEM as the combination of **FACTOR ANALYSIS & PATH ANALYSIS**



2. What does SEM do?

- Keep in mind that **RELATIONSHIP** is a keyword in SEM (my personal opinion)
 - SEM solved two great obstacles in social studies, but it is NOT everything!
 - SEM mainly focuses on causal relationships (as well as some measurement issues, just like factor analysis)
-

3. Mathematical Statistics Issues

- The form of SEM (consisted of two models)
- Derivation of parameter estimation (MLE as an example)

3.1 Two sub-models in SEM

- Measurement Model

$$x = \Lambda_x \xi + \delta \text{ ----- (1)}$$

$$y = \Lambda_y \eta + \varepsilon \text{ ----- (2)}$$

- Structural Model

$$\eta = B\eta + \Gamma \xi + \zeta \text{ ----- (3)}$$

- Some assumptions

3.2 Criterion for Parameter Estimation

- Actually, SEM is just a process of **hypothesis testing**.
 - Null hypothesis: $\Sigma = \Sigma(\theta)$
 - Alternative hypothesis: $\Sigma \neq \Sigma(\theta)$
- The aim of estimation is to minimize the difference between Σ and $\Sigma(\theta)$



3.3 Maximum Likelihood Estimation

- If the observed variables are normally distributed, we can use MLE to solve SEM.



3.3 MLE (cont.)

$$\begin{aligned}\Sigma &= \text{Var}(\mathbf{z}) = \text{Var}[(\mathbf{y}', \mathbf{x}')] \\ &= \begin{pmatrix} \text{Cov}(\mathbf{y}', \mathbf{y}') & \text{Cov}(\mathbf{y}', \mathbf{x}') \\ \text{Cov}(\mathbf{x}', \mathbf{y}') & \text{Cov}(\mathbf{x}', \mathbf{x}') \end{pmatrix} \\ &= \begin{pmatrix} \Lambda_y \mathbf{A}(\Gamma \Phi \Gamma' + \Psi) \mathbf{A}' \Lambda_y' + \Phi_\varepsilon & \Lambda_y \mathbf{A} \Gamma \Phi \Lambda_x' \\ \Lambda_x \Phi \Gamma' \mathbf{A}' \Lambda_y' & \Lambda_x \Phi \Lambda_x' + \Phi_\delta \end{pmatrix}\end{aligned}$$

3.3 MLE (cont.)

- Discrepancy Functions (Browne, 1984)

$$F = \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - (p + q) + \ln |\boldsymbol{\Sigma}| - \ln |\mathbf{S}|$$



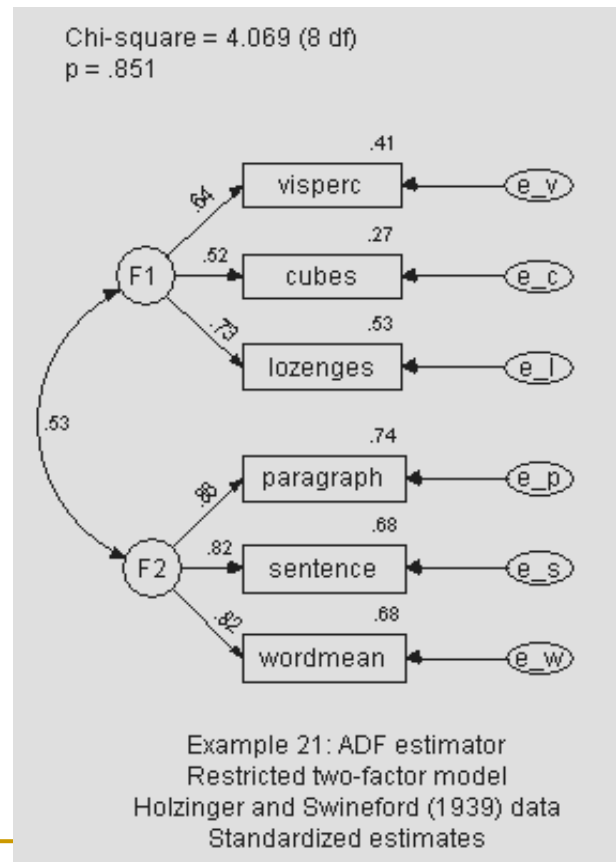
4. Interpretations for the Outputs

- What do we care most?
- Other indicators in the outputs



4.1 Outputs -- Coefficients

- Coefficients in SEM (unknown parameters estimated by various methods)



4.2 Other Indicators

- χ^2 -statistics (smaller value \implies better model)
- GFI, AGFI, RMR, etc

5. Something about Statistics

- Location of statistics in my mind -- A Useful Tool
 - What does this “useful tool” do for us? -- Tell us some truths, but it’s always hard! ...
 - Statistics alone can NOT play all the game!
 - Constraints of statistics (from a mathematical view)
-

5. Something about Statistics (cont.)

- Warnings about statistical software
 - Contents of statistics
 - Descriptive statistics (mean, variance, mode, ...)
 - Correlation and Regression
 - Compare means (t-test and ANOVA, etc)
 - Cluster analysis
 - Factor analysis
 - Time series
 - ...
-

Thanks!

Yihui XIE, School of Statistics

For further information, please visit <http://www.yihui.name>
