Lecture 18: Structural Equation Modeling

- ## 18. 1 What is Structural Equation Modeling (SEM)?

- A collection of statistical techniques to test relations among variables

- Primarily confirmatory, in the hypothetico-deductive tradition

- It combines ANOVA, path regression, and factor analysis models

- It typically uses path diagrams as graphic representation of models

- SEM usually deals with unobserved –or latent– constructs

- SEM is performed on a covariance matrix, but can also use a correlation matrix or raw data

- ## 18. 2 Research Questions

- Most questions examined using traditional techniques plus other ideas involving more complex interrelations among variables

- The nature of latent variables and the relations among them

- The distinction between a measure and the attribute being measured

- Theoretical ideas directly from data

- **18. 3 Latent Variables**

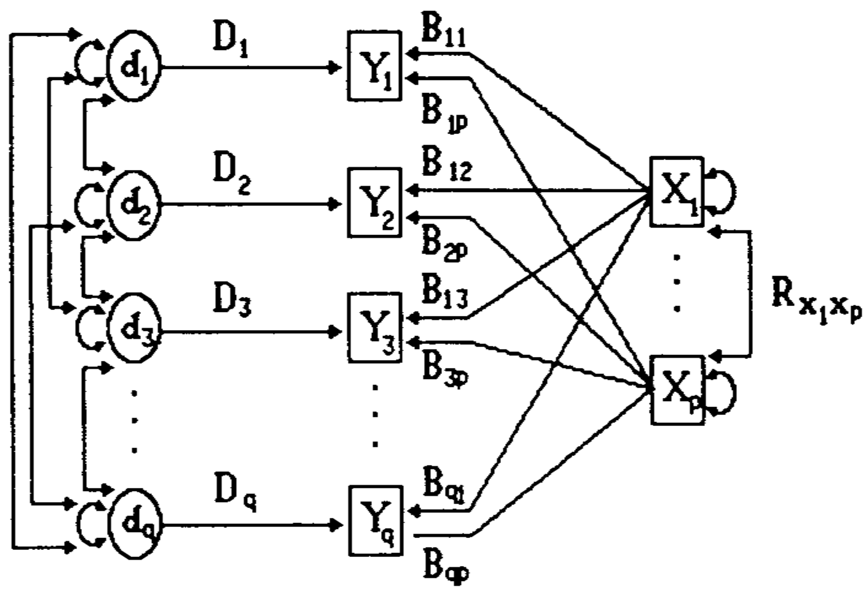
- Latent variables are specification features used in SEM to represent hypothetical constructs

- LVs are included in path models to examine practical problems created by incompleteness of a model

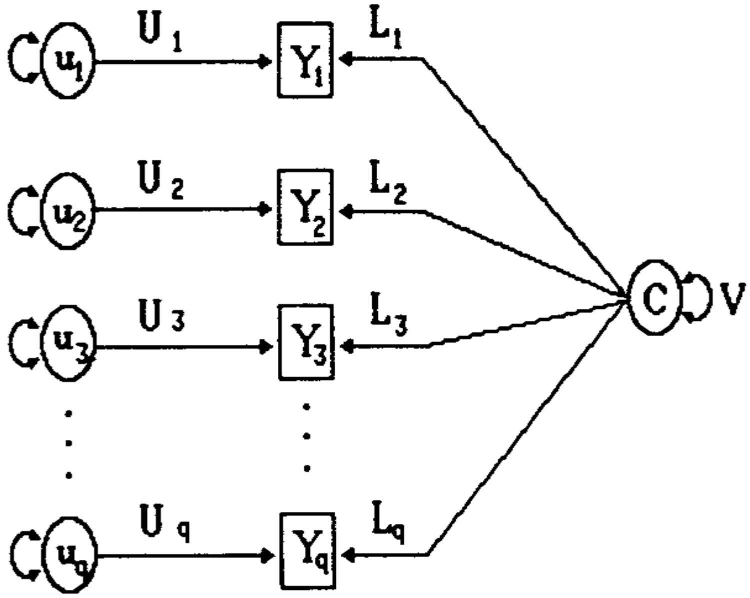
- LVs are used in many ways in a model to express the organization and impacts of observed variables

- LVs can represent testable (i.e., rejectable) hypotheses about empirical data

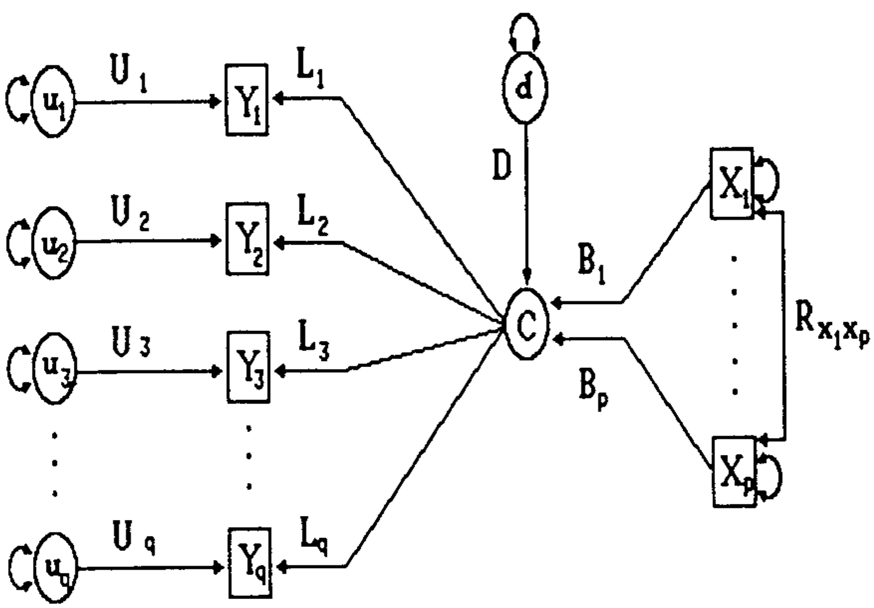
3.1 Multiple Regression



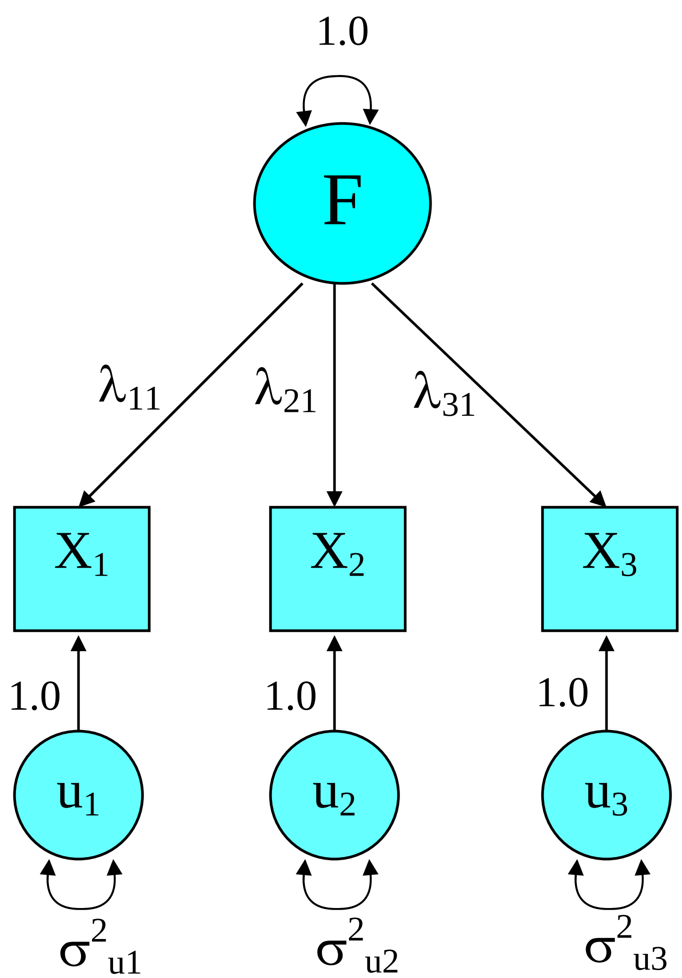
3.2 Factor Analysis



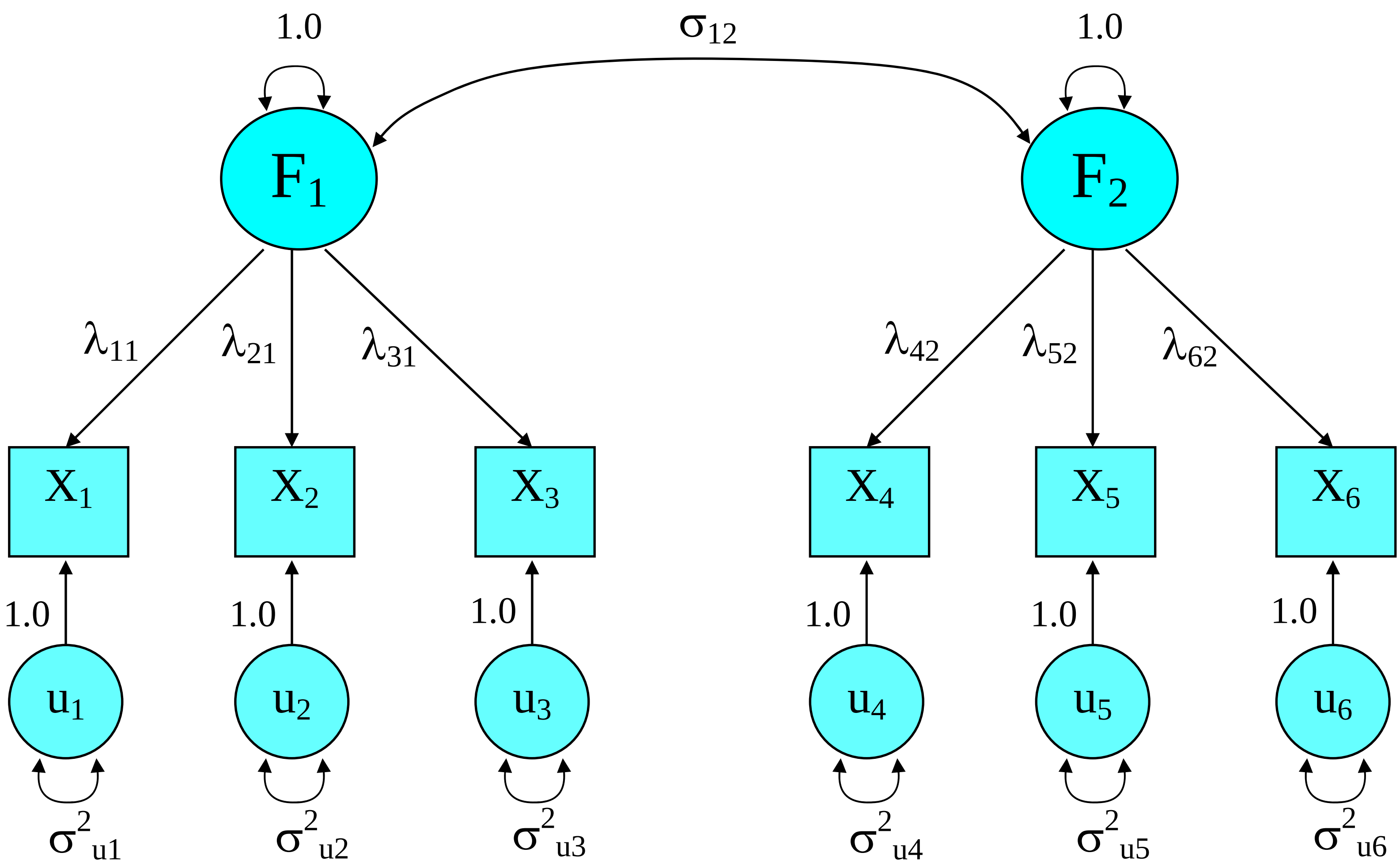
3.4 Latent Path Model



Factor Analysis – One Factor Model



Factor Analysis – Two Factor Model



- Two Factor Model (Factor Analysis)

$$\boldsymbol{\Lambda} = \begin{array}{c|cc} & F_1 & F_2 \\ \hline X_1 & \lambda_{11} & 0 \\ X_2 & \lambda_{21} & 0 \\ X_3 & \lambda_{31} & 0 \\ X_4 & 0 & \lambda_{42} \\ X_5 & 0 & \lambda_{52} \\ X_6 & 0 & \lambda_{62} \end{array} \qquad \boldsymbol{\Psi} = \begin{array}{c|cccccc} & u_1 & u_2 & u_3 & u_4 & u_5 & u_6 \\ \hline u_1 & \sigma^2_{u1} & 0 & 0 & 0 & 0 & 0 \\ u_2 & 0 & \sigma^2_{u2} & 0 & 0 & 0 & 0 \\ u_3 & 0 & 0 & \sigma^2_{u3} & 0 & 0 & 0 \\ u_4 & 0 & 0 & 0 & \sigma^2_{u4} & 0 & 0 \\ u_5 & 0 & 0 & 0 & 0 & \sigma^2_{u5} & 0 \\ u_6 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{u6} \end{array}$$

$$\boldsymbol{\Phi} = \begin{array}{c|cc} & F_1 & F_2 \\ \hline F_1 & 1 & \sigma_{12} \\ F_2 & \sigma_{12} & 1 \end{array}$$

Covariance Structure

$$\boldsymbol{E} = \boldsymbol{\Lambda\Phi\Lambda'} + \boldsymbol{\Psi}$$

• 18. 4 Conceptual Basis of SEM

- Given a set of scores ***Y*** for ***N*** subjects measured on ***M*** variables

- A model for ***Y*** is formulated as a linear equation

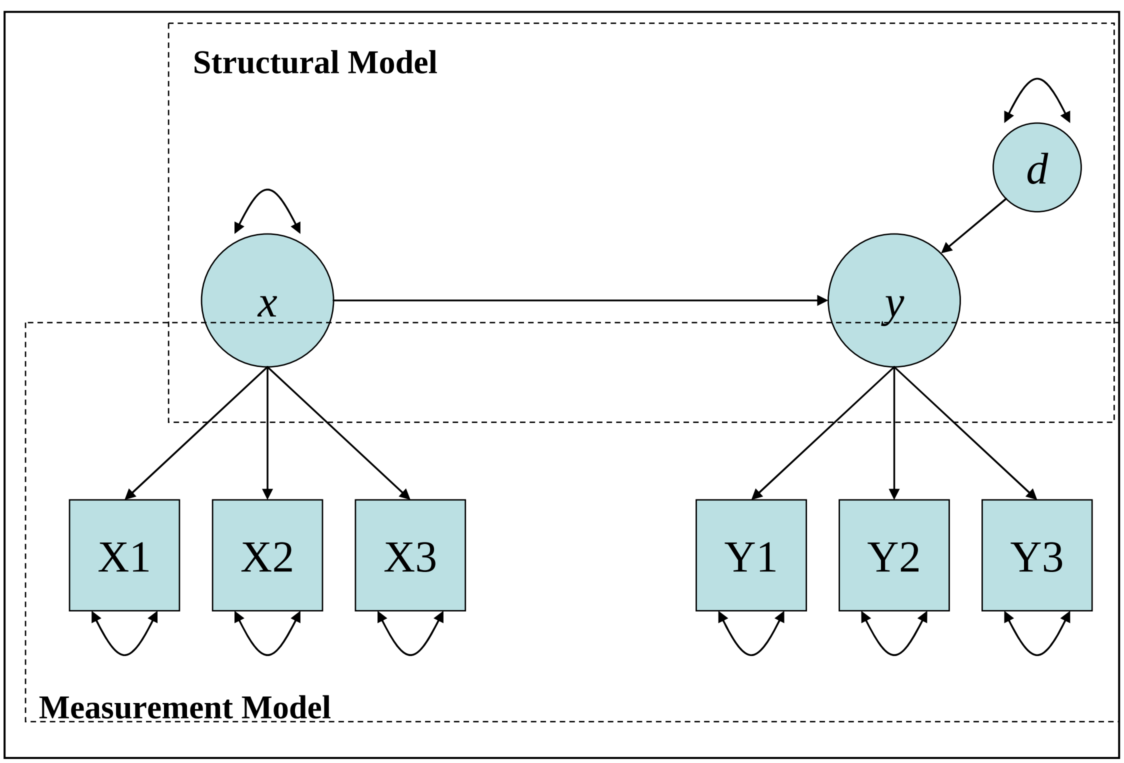
$$\boldsymbol{Y} = \mathbf{b}\boldsymbol{X} + \mathbf{e}$$

- This model leads to a set of expectations ***E***, which are compared with the observed data ***S***

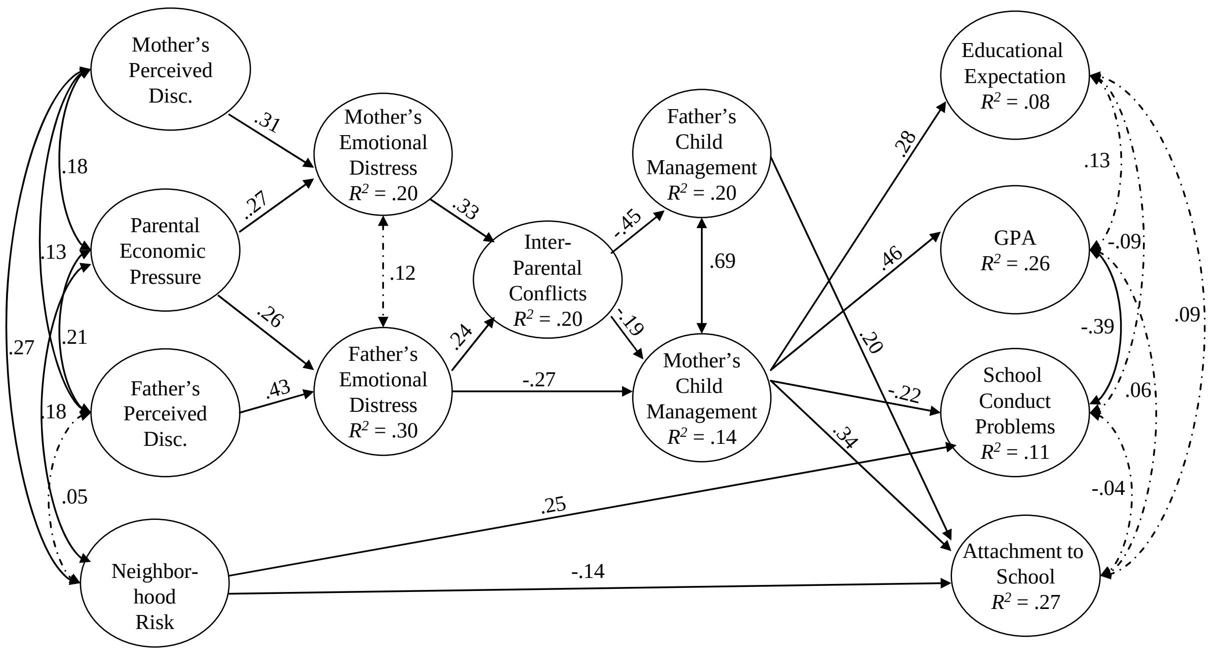
- The discrepancy between ***E*** and ***S*** is evaluated to assess whether the model is a reasonable representation of the data

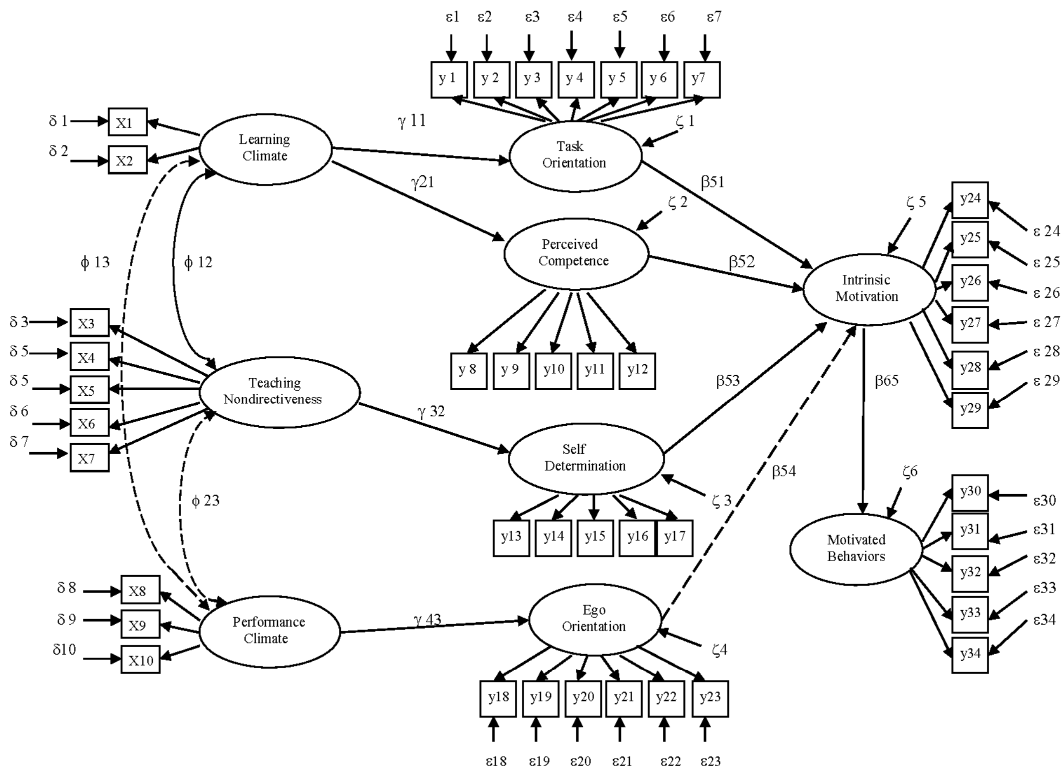
• 18. 5 Model Components

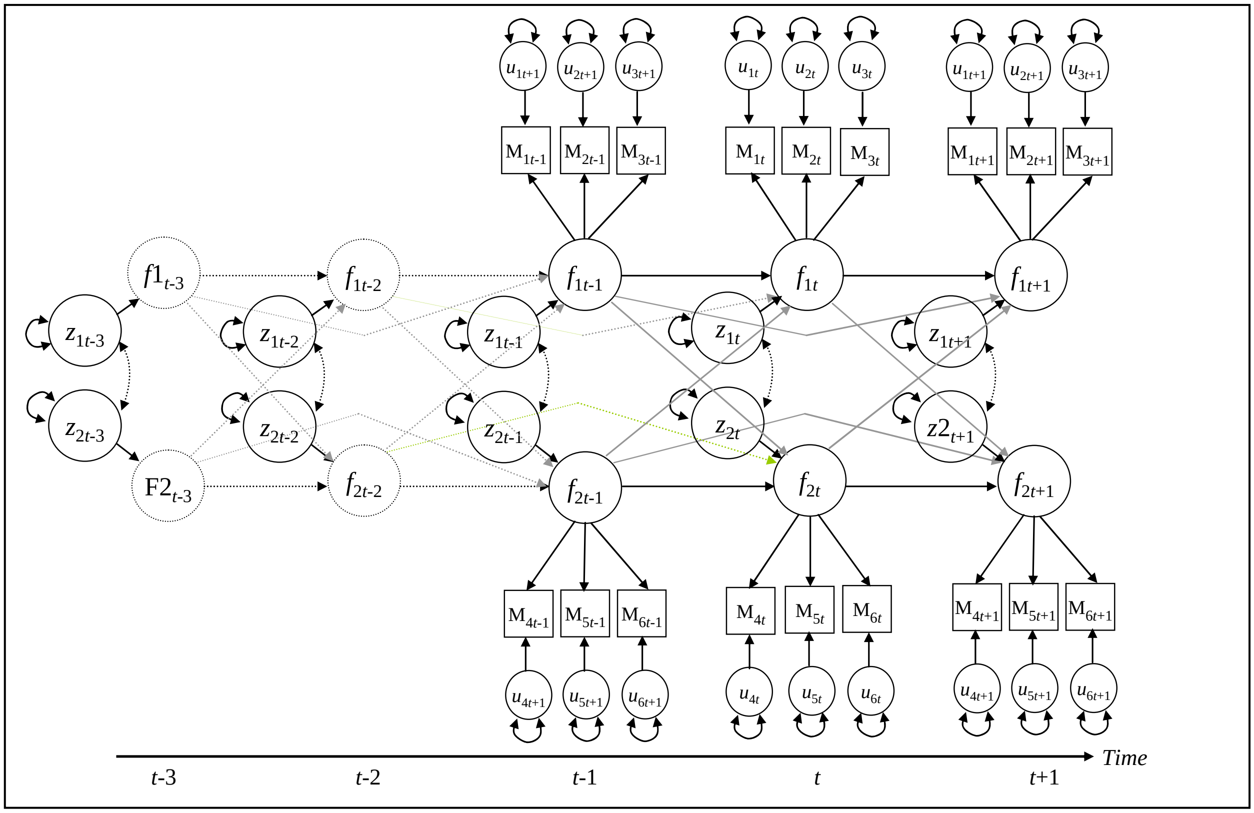
- Latent variable path models can be broken down into two related, but distinct parts:
    - Measurement model
        - relations among observed and latent variables
        - factor models
    - Structural model
        - relations among latent variables
- These two components become blurred with more complex models, but are useful when discussing SEMs



- **18. 6 Examples**




Learning Climate
Teaching Nondirectiveness
Performance Climate
Task Orientation
Perceived Competence
Self Determination
Ego Orientation
Intrinsic Motivation
Motivated Behaviors
γ 11
γ21
γ 32
γ 43
β51
β52
β53
β54
β65
ϕ 13
ϕ 12
ϕ 23
ζ 1
ζ 2
ζ 3
ζ4
ζ 5
ζ6


u1t+1 u2t+1 u3t+1
M1t-1 M2t-1 M3t-1
u1t u2t u3t
M1t M2t M3t
u1t+1 u2t+1 u3t+1
M1t+1 M2t+1 M3t+1
f1t-3
f1t-2
f1t-1
f1t
f1t+1
z1t-3
z1t-2
z1t-1
z1t
z1t+1
z2t-3
z2t-2
z2t-1
z2t
z2t+1
F2t-3
f2t-2
f2t-1
f2t
f2t+1
M4t-1 M5t-1 M6t-1
M4t M5t M6t
M4t+1 M5t+1 M6t+1
u4t+1 u5t+1 u6t+1
u4t u5t u6t
u4t+1 u5t+1 u6t+1
Time
t-3
t-2
t-1
t
t+1

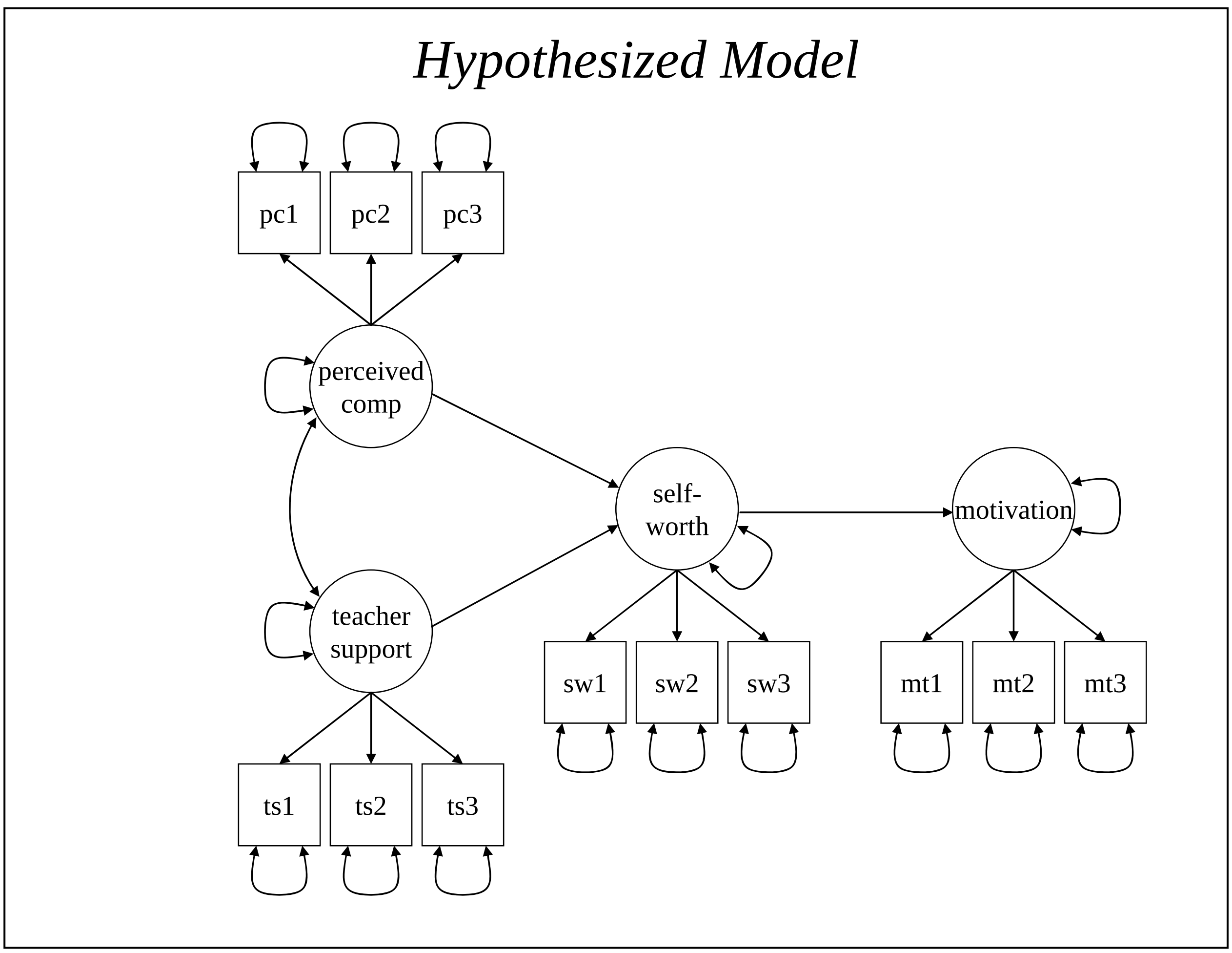
- **18. 7 Empirical Example**

- Relations among perceived competence, teacher support, self-worth, and motivation in high school students

    - Scales with multiple items for all variables

        - perceived competence = 4 items
        - teacher support = 5 items
        - self-worth = 4 items
        - motivation = 5 items

    - Students in high school (*N* = 228)



Results

```
MODEL RESULTS
                                                       Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
WORTH    ON
    PCOMP              0.770      0.079      9.721      0.000
    TEACH              0.212      0.075      2.837      0.005

MOTIV    ON
    WORTH              0.467      0.086      5.441      0.000

TEACH    WITH
    PCOMP              0.065      0.033      1.934      0.053

Variances
    PCOMP              0.537      0.088      6.075      0.000
    TEACH              0.282      0.063      4.458      0.000

Residual Variances
    WORTH              0.053      0.022      2.417      0.016
    MOTIV              0.093      0.024      3.877      0.000

!Standardized
WORTH    ON
    PCOMP              0.886      0.034     26.321      0.000
    TEACH              0.177      0.058      3.059      0.002

 MOTIV    ON
    WORTH              0.697      0.061     11.378      0.000

 TEACH    WITH
    PCOMP              0.166      0.082      2.030      0.042
```

• 18. 8 Steps in SEM

- Step 1: Specification: Formulation of a theoretical model in terms of structural equations

- Step 2: Identification: There must exist at least a unique solution for the model's parameters

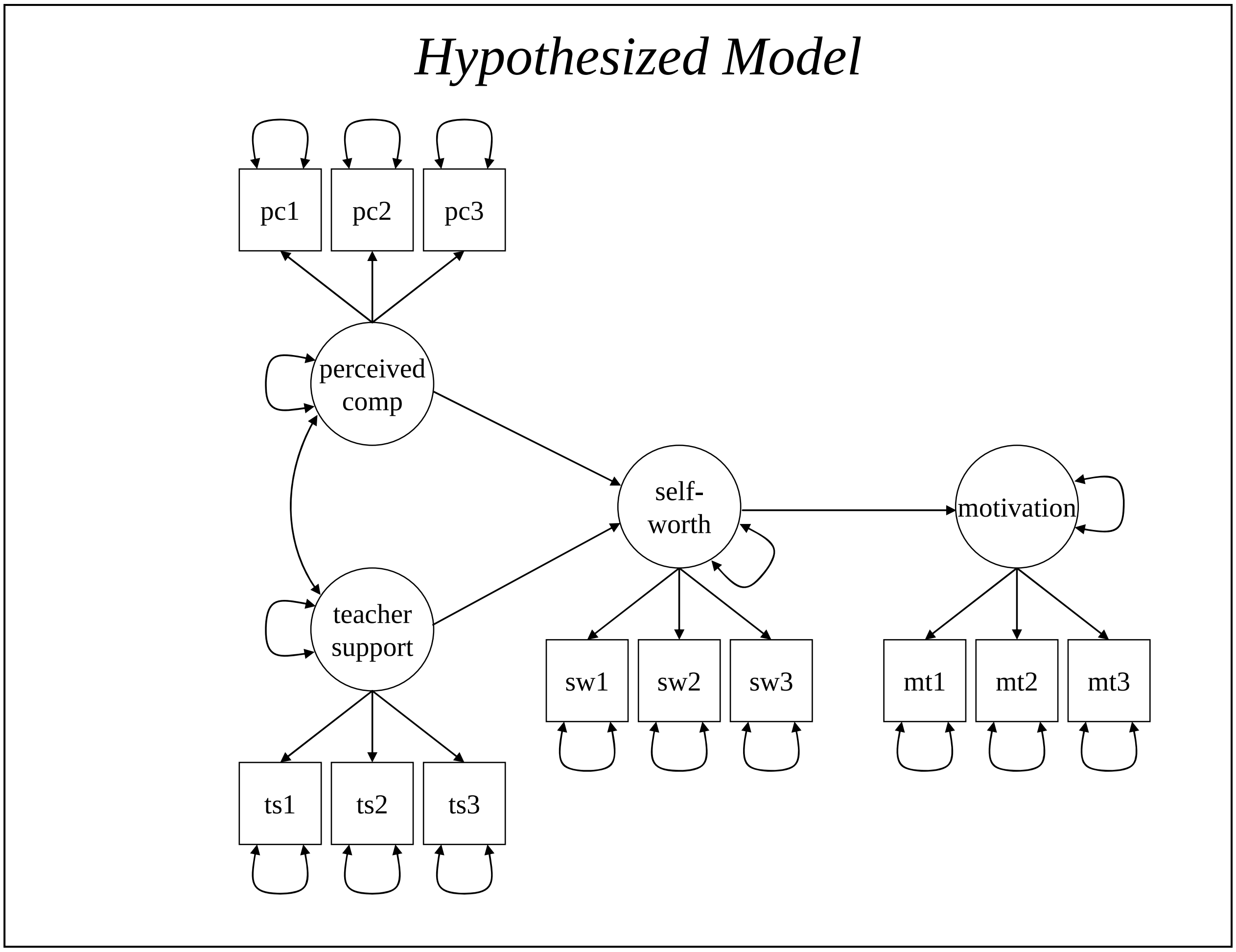
- Step 3: Estimation: The model's parameters are estimated from the observed data using an estimation method

- Step 4: Evaluation: The hypothesized covariances (based on the estimated parameters) are compared with the observed covariances to assess discrepancy

- Step 5: Reevaluation: Alternative models (hypotheses) are fitted to the same data

- **Step 1: Specification**

- Define hypothesis in algebraic terms



- **Step 2: Estimation of Model from Data**

- Specified structural model leads to a set of expectations (*E*)

- Estimation procedures attempt to minimize differences between observations *O* and *E*

- A common minimization function

$$f = (O - E)^2\, W^{-1}$$

- When weights $W = E$, approximate maximum likelihood estimates (ML) are obtained

- **Step 3: Model Evaluation**

- To examine how likely is the model, given the observed data

$$\boldsymbol{H_0}: \Sigma = S; (O = E)$$

$$\boldsymbol{H_a}: \Sigma \neq S; (O \neq E)$$

- Several indices of fit including a likelihood ratio, residuals, standard errors

- A rational goal is to separate those models with good fit (reasonable for the observed data) from those with poor fit (unreasonable for the observed data)

• 18. 9 Linear Regression Models

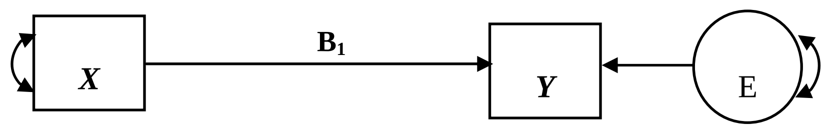
- A linear model is expressed (for n=1 to N) as

$$Y_n = B_0 + B_1 X_n + E_n$$

- where B_0 is the *intercept* term -- the predicted score of Y when X=0

- B_1 is the *coefficient* term -- the change in the predicted score of Y for a one unit change in X

- E is the *residual* score -- an unobserved and random score that is uncorrelated with X but forms part of the variance of Y



• 18. 10 Multiple Regression Models

- A linear model is expressed (for n=1 to N) as

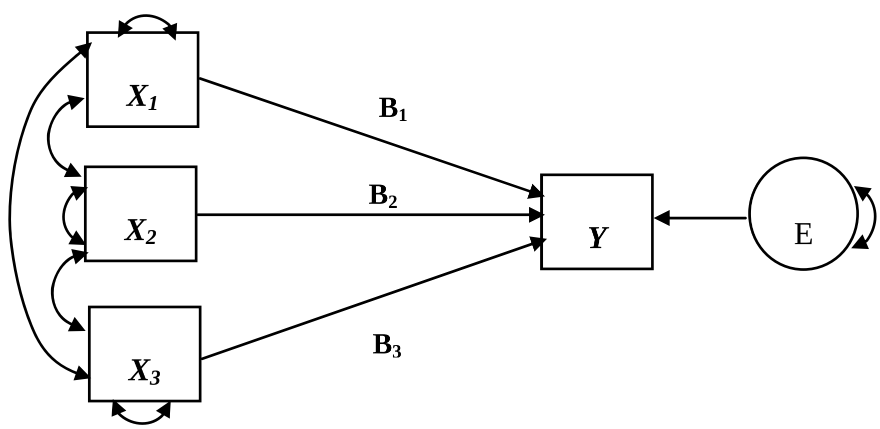
$$Y_n = B_0 + B_1 X_{1n} + B_2 X_{2n} + E_n$$

- B_0 is the *intercept* term -- the predicted score of Y when $X_1 = X_2 = 0$

- B_1 is the first *slope* term -- the change in the predicted score of Y for a one unit change in X_1 given "fixed" values of X_2

- B_2 is the second *slope* term -- the change in the predicted score of Y for a one unit change in X_2 given "fixed" values of X_1

- where E is the *residual* score when predicting Y from X_1 and X_2 -- an unobserved and random score that is uncorrelated with all predictors



• 18. 11 Model Assumptions

- In using multiple regression models, there are always some *testable* and *untestable* assumptions

- The testable assumptions deal with the *accuracy* (i.e., significance) of the prediction equation, as well as the *accuracy, size, and direction* of each regression coefficient (i.e., independent effect) within the general linear framework

- The untestable assumptions deal with the "deterministic" and "stochastic" features of general linear framework

- Any model is only valid to the degree the untestable assumptions are valid

- **18. 12 Benefits of SEM**

- Useful approach to test conceptual ideas

- Emphasis on confirmatory - and exploratory - approaches

- Useful tool to examine invariance and replication across samples

- Relatively easy to incorporate incomplete data and multiple groups

- **18. 13 Limitations of SEM**

- Common (false) belief of causal inferences; no causal inferences without a good design

- Some computer programs require relatively complex programming and, sometimes, they do not work

- Direct relations among variables become difficult to examine in large complex models

- An overall index of misfit is given for all parameters; difficulty to trace misfit when numerous parameters are involved

- SEM is still in development; its effective use is not yet well known