Factor Analysis

Qian-Li Xue

Biostatistics Program

Harvard Catalyst | The Harvard Clinical & Translational Science Center

Short course, October 27, 2016

Well-used latent variable models

Latent variable scale	Observed variable scale	
	Continuous	Discrete
Continuous	**Factor analysis** LISREL	Discrete FA IRT (item response)
Discrete	Latent profile Growth mixture	Latent class analysis, regression

General software: MPlus, Latent Gold, WinBugs (Bayesian), NLMIXED (SAS)

Objectives

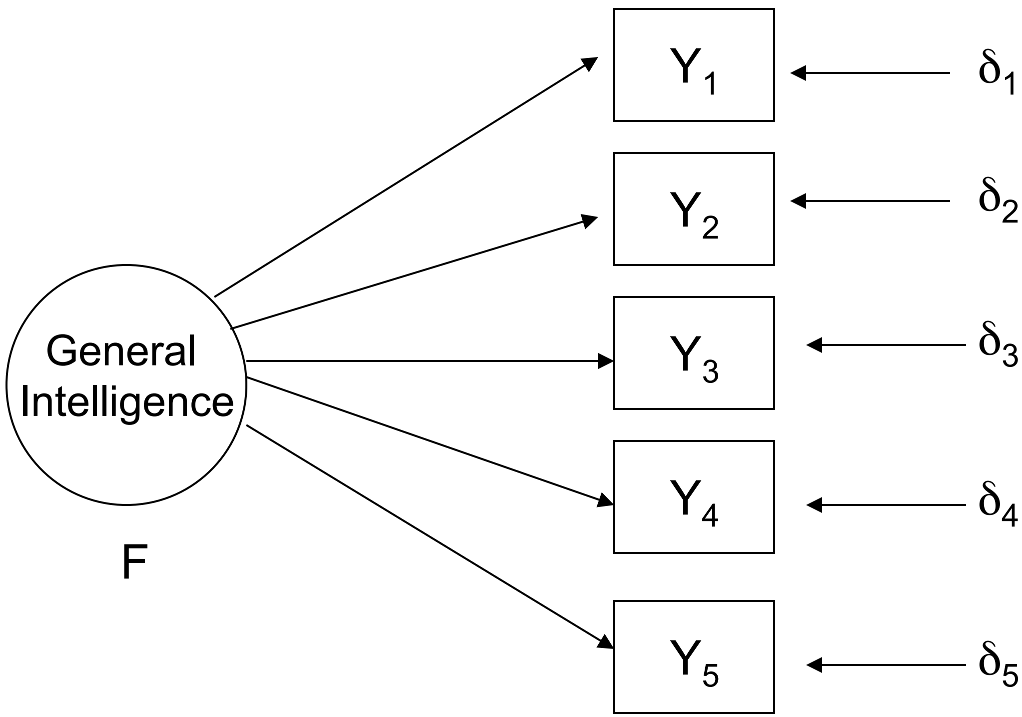
- What is factor analysis?
- What do we need factor analysis for?
- What are the modeling assumptions?
- How to specify, fit, and interpret factor models?
- What is the difference between exploratory and confirmatory factor analysis?
- What is and how to assess model identifiability?

What is factor analysis

- Factor analysis is a theory driven statistical data reduction technique used to explain **covariance** among observed random variables in terms of fewer unobserved random variables named factors

An Example: General Intelligence

(Charles Spearman, 1904)



Why Factor Analysis?

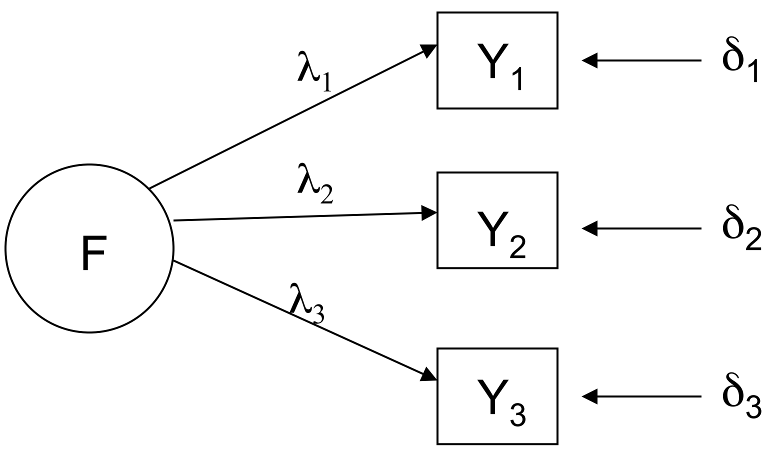
1. Testing of theory
   - Explain covariation among multiple observed variables by
   - Mapping variables to latent constructs (called “factors”)

2. Understanding the structure underlying a set of measures
   - Gain insight to dimensions
   - Construct validation (e.g., convergent validity)

3. Scale development
   - Exploit redundancy to improve scale’s validity and reliability

Part I. Exploratory Factor Analysis (EFA)

One Common Factor Model: Model Specification



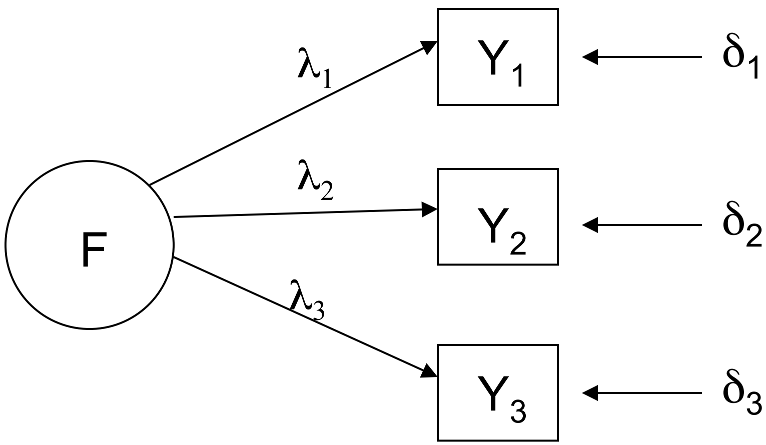
$$Y_1 = \lambda_1 F + \delta_1$$

$$Y_2 = \lambda_2 F + \delta_2$$

$$Y_3 = \lambda_3 F + \delta_3$$

- The factor F is <u>not observed</u>; only Y_1, Y_2, Y_3 are observed
- δ_i represent variability in the Y_i NOT explained by F
- Y_i is a <u>linear</u> function of F and δ_i

One Common Factor Model: Model Assumptions



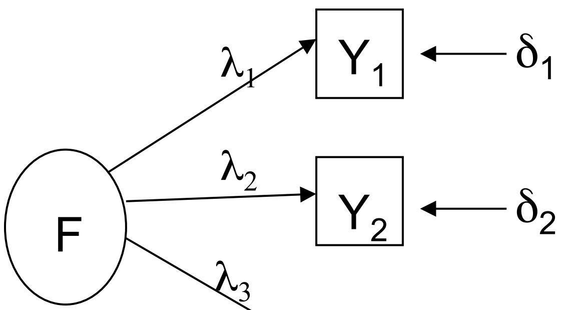
$$Y_1 = \lambda_1 F + \delta_1$$

$$Y_2 = \lambda_2 F + \delta_2$$

$$Y_3 = \lambda_3 F + \delta_3$$

- Factorial causation
- F is independent of δ_j, i.e. $\mathrm{cov}(F,\delta_j)=0$
- δ_i and δ_j are independent for $i \neq j$, i.e. $\mathrm{cov}(\delta_i,\delta_j)=0$
- Conditional independence: Given the factor, observed variables are independent of one another, i.e. $\mathrm{cov}(Y_i ,Y_j \mid F) = 0$

One Common Factor Model: Model Interpretation

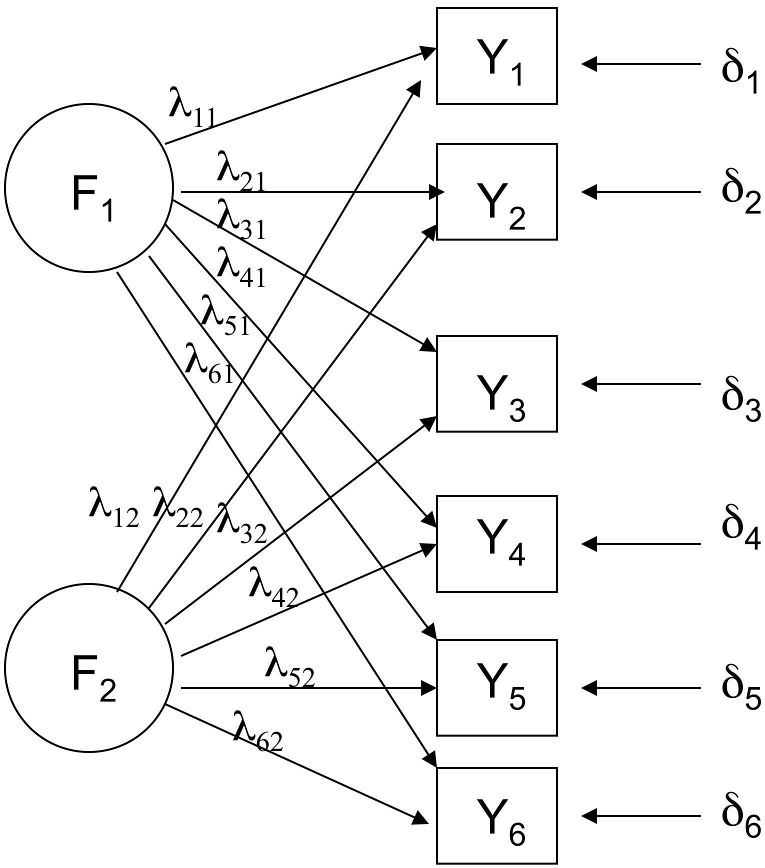


$$Y_1 = \lambda_1 F + \delta_1$$
$$Y_2 = \lambda_2 F + \delta_2$$
$$Y_3 = \lambda_3 F + \delta_3$$

Given all variables in standardized form, i.e. $var(Y_i)=var(F)=1$

- Factor loadings: λ_i
  $\lambda_i = corr(Y_i, F)$
- Communality of Y_i: h_i^2
  $h_i^2 = \lambda_i^2 = [corr(Y_i, F)]^2$
  =% variance of Y_i explained by F
- Uniqueness of Y_i: $1-h_i^2$
  = residual variance of Y_i
- Degree of factorial determination:
  $=\Sigma\, \lambda_i^2/n$, where n=# observed variables Y

Two-Common Factor Model (Orthogonal): Model Specification



$$Y_1 = \lambda_{11} F_1 + \lambda_{12} F_2 + \delta_1$$
$$Y_2 = \lambda_{21} F_1 + \lambda_{22} F_2 + \delta_2$$
$$Y_3 = \lambda_{31} F_1 + \lambda_{32} F_2 + \delta_3$$
$$Y_4 = \lambda_{41} F_1 + \lambda_{42} F_2 + \delta_4$$
$$Y_5 = \lambda_{51} F_1 + \lambda_{52} F_2 + \delta_5$$
$$Y_6 = \lambda_{61} F_1 + \lambda_{62} F_2 + \delta_6$$

F1 and F2 are common factors because they are shared by ≥2 variables !

Matrix Notation

with n variables and m factors

$$\mathbf{Y}_{\mathbf{nx1}} = \boldsymbol{\Lambda}_{\mathbf{nxm}}\mathbf{F}_{\mathbf{mx1}} + \boldsymbol{\delta}_{\mathbf{nx1}}$$

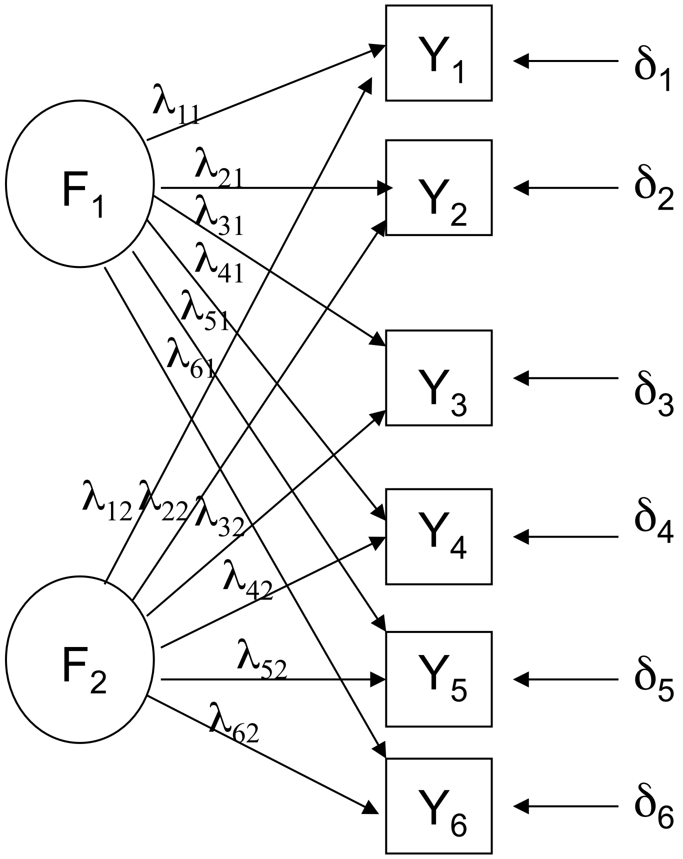
$$\begin{bmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \cdots & \lambda_{1m} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \lambda_{n1} & \cdots & \cdots & \lambda_{nm} \end{bmatrix}_{n\times m} \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}_{m\times 1} + \begin{bmatrix} \delta_1 \\ \vdots \\ \vdots \\ \delta_n \end{bmatrix}_{n\times 1}$$

Factor Pattern Matrix

- Columns represent derived factors
- Rows represent input variables
- Loadings represent degree to which each of the variables "correlates" with each of the factors
- Loadings range from -1 to 1
- Inspection of factor loadings reveals extent to which each of the variables contributes to the meaning of each of the factors.
- High loadings provide meaning and interpretation of factors (~ regression coefficients)

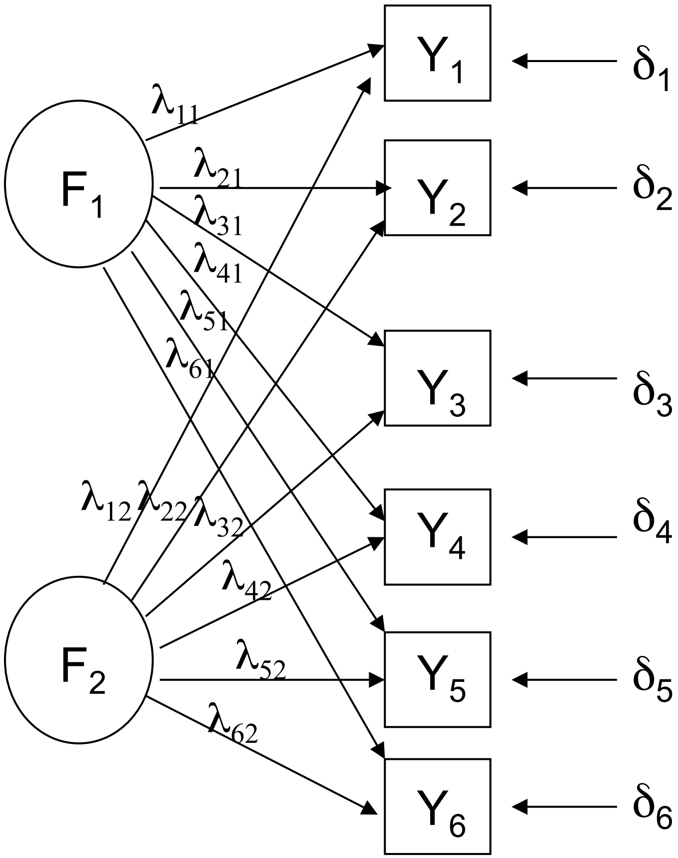
$$\begin{bmatrix} \lambda_{11} & \cdots & \cdots & \lambda_{1m} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \lambda_{n1} & \cdots & \cdots & \lambda_{nm} \end{bmatrix}_{n \times m}$$

Two-Common Factor Model (Orthogonal): Model Assumptions



- Factorial causation
- F_1 and F_2 are independent of δ_j, i.e. $cov(F_1,\delta_j)= cov(F_2,\delta_j)= 0$
- δ_i and δ_j are independent for $i \neq j$, i.e. $cov(\delta_i,\delta_j)=0$
- Conditional independence: Given factors F_1 and F_2, observed variables are independent of one another, i.e. $cov(Y_i ,Y_j \mid F_1, F_2) = 0$ for $i \neq j$
- Orthogonal (=independent): $cov(F_1,F_2)=0$

Two-Common Factor Model (Orthogonal): Model Interpretation



Given all variables in standardized form, i.e. $var(Y_i)=var(F_i)=1$;

AND **orthogonal** factors, i.e. $\text{cov}(F_1,F_2)=0$

- Factor loadings: λ_{ij}

  $\lambda_{ij} = corr(Y_i,F_j)$

- Communality of Y_i: h_i^2

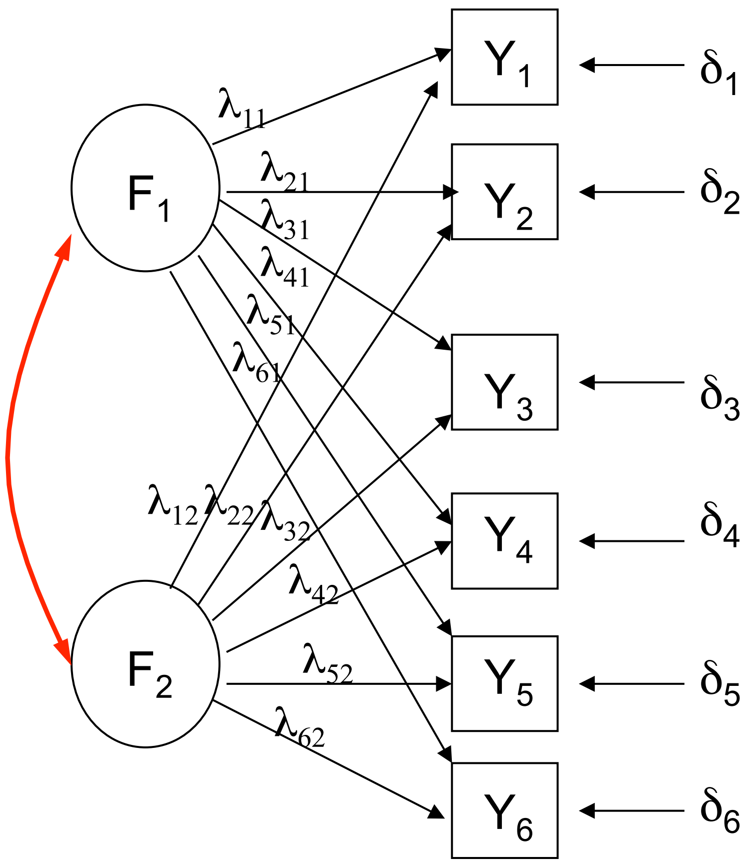
  $h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2$ = % variance of Y_i explained by F_1 AND F_2

- Uniqueness of Y_i: $1-h_i^2$

- Degree of factorial determination:

  $=\Sigma\, \lambda_{ij}^2/n$, n=# observed variables Y

Two-Common Factor Model : The Oblique Case



Given all variables in standardized form,
i.e. $var(Y_i)=var(F_i)=1$;
AND oblique factors (i.e. $cov(F_1,F_2)\neq 0$)

- The interpretation of factor loadings: λ_{ij} is no longer correlation between Y and F; it is direct effect of F on Y

- The calculation of communality of Yi (h_i^2) is more complex

Extracting initial factors

- Least-squares method (e.g. principal axis factoring with iterated communalities)
- Maximum likelihood method

Model Fitting: Extracting initial factors

Least-squares method (LS) (e.g. principal axis factoring with iterated communalities)

- Goal: minimize the sum of squared differences between observed and estimated corr. matrices
- Fitting steps:
  a) Obtain initial estimates of communalities (h^2)

     e.g. squared correlation between a variable and the remaining variables
  b) Solve objective function: $\det(R_{LS}-\eta I)=0$,

     where R_{LS} is the corr matrix with h^2 in the main diag. (also termed adjusted corr matrix), η is an eigenvalue
  c) Re-estimate h^2
  d) Repeat b) and c) until no improvement can be made

Model Fitting: Extracting initial factors

Maximum likelihood method (MLE)

- Goal: maximize the likelihood of producing the observed corr matrix
- Assumption: distribution of variables (Y and F) is multivariate normal
- Objective function: $\det(R_{MLE} - \eta I) = 0$,
  where $R_{MLE} = U^{-1}(R - U^2)U^{-1} = U^{-1}R_{LS}U^{-1}$, and U^2 is $\mathrm{diag}(1-h^2)$
- Iterative fitting algorithm similar to LS approach
- Exception: adjust R by giving greater weights to correlations with smaller unique variance, i.e. $1 - h^2$
- Advantage: availability of a large sample χ^2 significant test for goodness-of-fit (but tends to select more factors for large n!)

Choosing among Different Methods

- Between MLE and LS
  - LS is preferred with
    - few indicators per factor
    - Equeal loadings within factors
    - No large cross-loadings
    - No factor correlations
    - Recovering factors with low loadings (overextraction)
  - MLE if preferred with
    - Multivariate normality
    - unequal loadings within factors
  - Both MLE and LS may have convergence problems

Factor Rotation

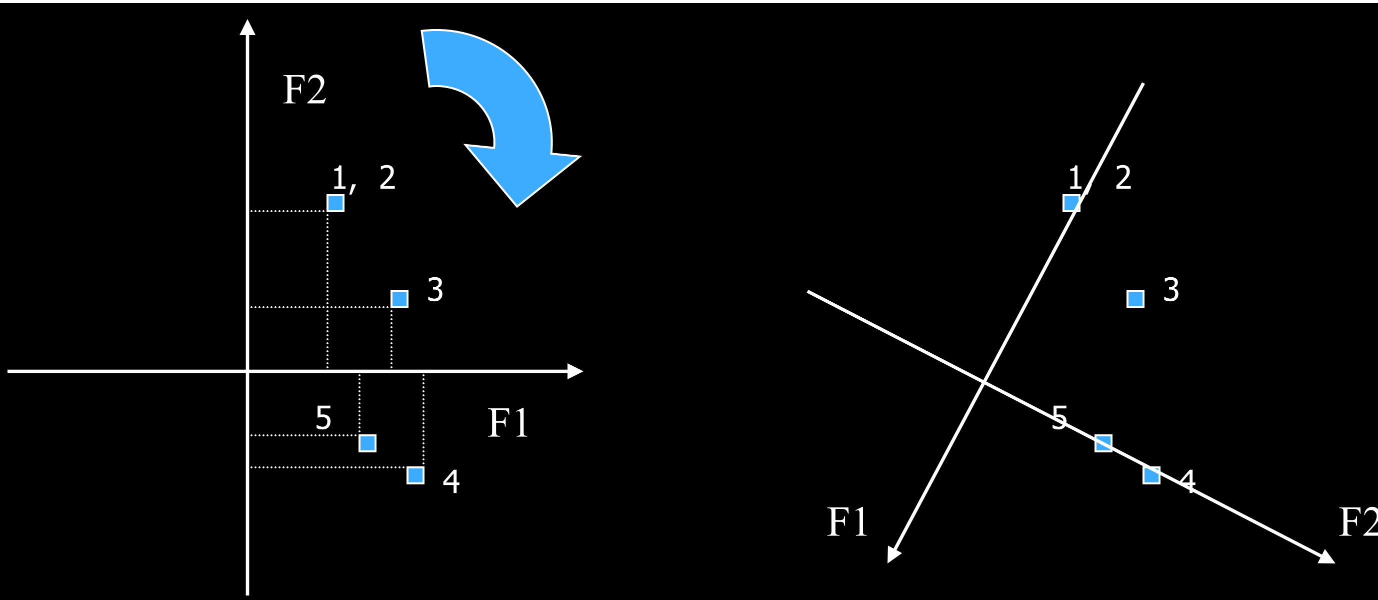
- Goal is simple structure
- Make factors more easily interpretable
  - While keeping the number of factors and communalities of Ys fixed!!!
- **Rotation does NOT improve fit!**

Factor Rotation

To do this we "rotate" factors:

- redefine factors such that 'loadings' (or pattern matrix coefficients) on various factors tend to be very high (-1 or 1) or very low (0)
- intuitively, it makes sharper distinctions in the meanings of the factors

Factor Rotation (Intuitively)



	Factor 1	Factor 2
x1	0.4	0.69
x2	0.4	0.69
x3	0.65	0.32
x4	0.69	-0.4
x5	0.61	-0.35

	Factor 1	Factor 2
x1	-0.8	0
x2	-0.8	0
x3	-0.6	0.4
x4	0	0.8
x5	0	0.7

Factor Rotation

- Uses "ambiguity" or non-uniqueness of solution to make interpretation more simple
- Where does ambiguity come in?
  - Unrotated solution is based on the idea that each factor tries to maximize variance explained, conditional on previous factors
  - What if we take that away?
  - Then, there is not one "best" solution

Factor Rotation: Orthogonal vs. Oblique Rotation

- Orthogonal: Factors are independent
  - **varimax**: maximize variance of squared loadings ***across variables*** (sum over factors)
    - Goal: the simplicity of interpretation of factors
  - quartimax: maximize variance of squared loadings ***across factors*** (sum over variables)
    - Goal: the simplicity of interpretation of variables
  - **Intuition: from previous picture, there is a right angle between axes**
- **Note: "Uniquenesses" remain the same!**

Factor Rotation: Orthogonal vs. Oblique Rotation

- Oblique: Factors are **NOT** independent. Change in "angle."
  - oblimin: minimize covariance of squared loadings between factors.
  - **promax**: simplify orthogonal rotation by making small loadings even closer to zero.
  - Target matrix: choose "simple structure" a priori.
  - **Intuition: from previous picture, angle between axes is not necessarily a right angle.**
- **Note: "Uniquenesses" remain the same!**

Pattern versus Structure Matrix

- In oblique rotation, one typically presents both a pattern matrix and a structure matrix
- Also need to report correlation between the factors
- The pattern matrix presents the usual factor loadings
- The structure matrix presents correlations between the variables and the factors
- For orthogonal factors, pattern matrix=structure matrix
- The pattern matrix is used to interpret the factors

Factor Rotation: Which to use?

- Choice is generally not critical
- Interpretation with orthogonal (varimax) is "simple" because factors are independent: "Loadings" are correlations.
- Configuration may appear more simple in oblique (promax), but correlation of factors can be difficult to reconcile.
- Theory? Are the conceptual meanings of the factors associated?

Factor Rotation: Unique Solution?

- The factor analysis solution is **NOT** unique!
- More than one solution will yield the same "result."

Derivation of Factor Scores

- Each object (e.g. each person) gets a factor score for each factor:
- The factors themselves are variables
- "Object's" score is weighted combination of scores on input variables $\hat{F} = \hat{W}Y$, where $\hat{W}$ is the weight matrix.
- These weights are **NOT** the factor loadings!
- Different approaches exist for estimating $\hat{W}$ (e.g. regression method)
- Factor scores are not unique
- Using factors scores instead of factor indicators can reduce measurement error, but does **NOT** remove it.
- Therefore, using factor scores as predictors in conventional regressions leads to inconsistent coefficient estimators!

Factor Analysis with Categorical Observed Variables

- Factor analysis hinges on the correlation matrix
- As long as you can get an interpretable correlation matrix, you can perform factor analysis
- Binary/ordinal items?
  - Pearson corrlation: Expect attenuation!
  - Tetrachoric correlation (binary)
  - Polychoric correlation (ordinal)

To obtain polychoric correlation in STATA:
polychoric var1 var2 var3 var4 var5 …
To run princial component analysis:
pcamat r(R), n(328)
To run factor analysis:
factormat r(R), fa(2) ipf n(328)

Criticisms of Factor Analysis

- Labels of factors can be arbitrary or lack scientific basis
- Derived factors often very obvious
  - defense: but we get a quantification
- "Garbage in, garbage out"
  - really a criticism of input variables
  - factor analysis reorganizes input matrix
- Correlation matrix is often poor measure of association of input variables.

Major steps in EFA

1. Data collection and preparation
2. Choose number of factors to extract
3. Extracting initial factors
4. Rotation to a final solution
5. Model diagnosis/refinement
6. Derivation of factor scales to be used in further analysis

Part II. Confirmatory Factor Analysis (CFA)

Exploratory vs. Confirmatory Factor Analysis

- Exploratory:
  - summarize data
  - describe correlation structure between variables
  - generate hypotheses
- Confirmatory
  - Testing correlated measurement errors
  - Redundancy test of one-factor vs. multi-factor models
  - Measurement invariance test comparing a model across groups
  - Orthogonality tests

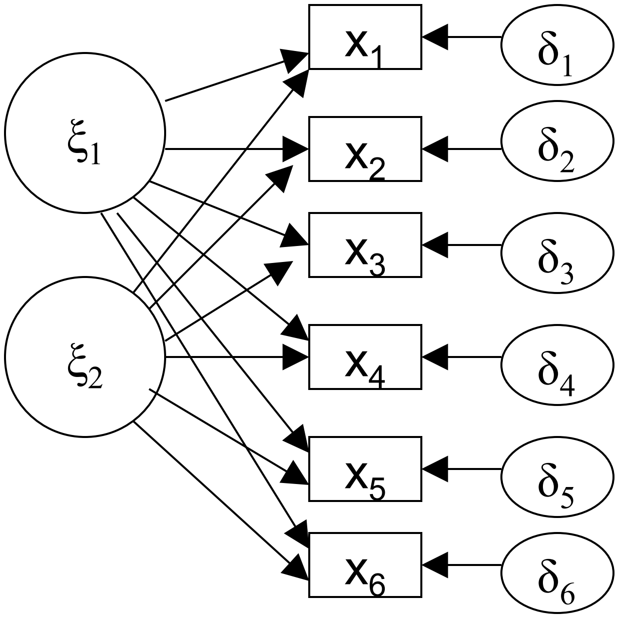
Confirmatory Factor Analysis (CFA)

- Takes factor analysis a step further.
- We can “test” or “confirm” or “implement” a “highly constrained a priori structure that meets conditions of model identification”
- But be careful, a model can never be confirmed!!
- CFA model is constructed in advance
- number of latent variables (“factors”) is pre-set by analyst (not part of the modeling usually)
- Whether latent variable influences observed is specified
- Measurement errors may correlate
- Difference between CFA and the usual SEM:
  - SEM assumes causally interrelated latent variables
  - CFA assumes interrelated latent variables (i.e. exogenous)

Exploratory Factor Analysis

Two factor model: $$x = \Lambda \xi + \delta$$

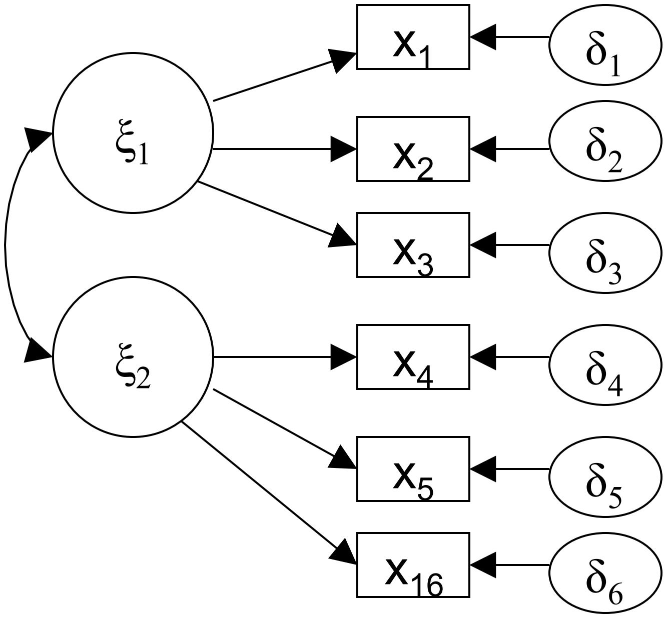
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$



CFA Notation

Two factor model: $$x = \Lambda \xi + \delta$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$



Difference between CFA and EFA

CFA

$$x_1 = \lambda_{11}\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$x_4 = \lambda_{42}\xi_2 + \delta_4$$
$$x_5 = \lambda_{52}\xi_2 + \delta_5$$
$$x_6 = \lambda_{62}\xi_2 + \delta_6$$
$$\mathrm{cov}(\xi_1, \xi_2) = \varphi_{12}$$

EFA

$$x_1 = \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \delta_3$$
$$x_4 = \lambda_{41}\xi_1 + \lambda_{42}\xi_2 + \delta_4$$
$$x_5 = \lambda_{51}\xi_1 + \lambda_{52}\xi_2 + \delta_5$$
$$x_6 = \lambda_{61}\xi_1 + \lambda_{62}\xi_2 + \delta_6$$
$$\mathrm{cov}(\xi_1, \xi_2) = 0$$

Model Constraints

- Hallmark of CFA
- Purposes for setting constraints:
  - Test a priori theory
  - Ensure identifiability
  - Test reliability of measures

Identifiability

- Let θ be a $t \times 1$ vector containing all unknown and unconstrained parameters in a model. The parameters θ are identified if $\Sigma(\theta_1) = \Sigma(\theta_2) \Leftrightarrow \theta_1 = \theta_2$
- Estimability ≠ Identifiability !!
- Identifiability – attribute of the model
- Estimability – attribute of the data

Model Constraints: Identifiability

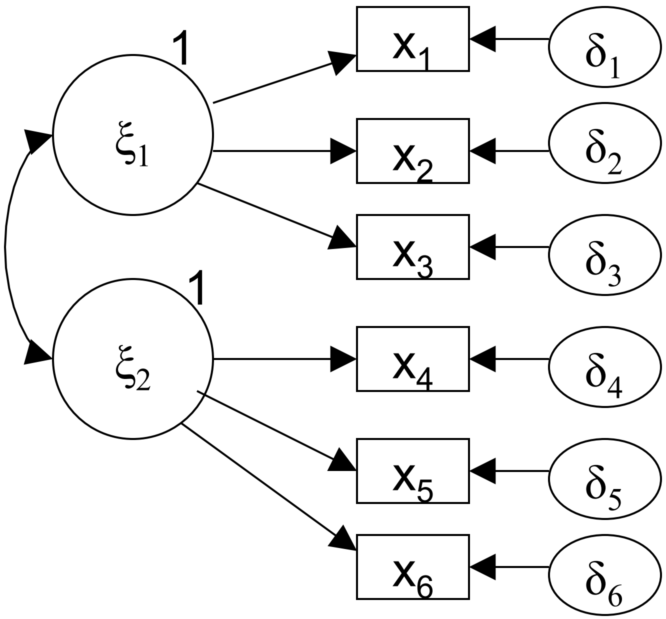
- Latent variables (LVs) need some constraints
- Because factors are unmeasured, their variances can take different values
- Recall EFA where we constrained factors:

$$F \sim N(0,1)$$

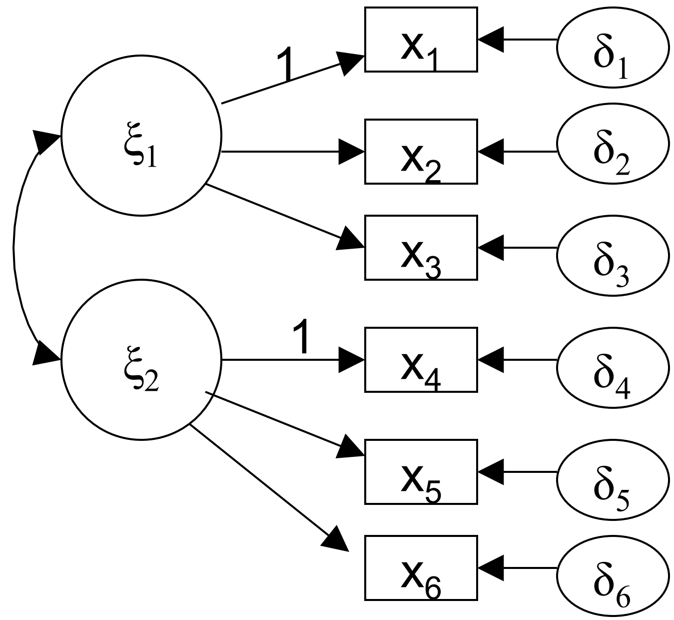
- Otherwise, model is not identifiable.
- Here we have two options:
  - Fix variance of latent variables (LV) to be 1 (or another constant)
  - Fix one path between LV and indicator

Necessary Constraints

Fix variances:



Fix path:



Model Parametrization

Fix variances:

$$x_1 = \lambda_{11}\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$x_4 = \lambda_{42}\xi_2 + \delta_4$$
$$x_5 = \lambda_{52}\xi_2 + \delta_5$$
$$x_6 = \lambda_{62}\xi_2 + \delta_6$$
$$\mathrm{cov}(\xi_1, \xi_2) = \varphi_{12}$$
$$\mathrm{var}(\xi_1) = 1$$
$$\mathrm{var}(\xi_2) = 1$$

Fix path:

$$x_1 = \xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$x_4 = \xi_2 + \delta_4$$
$$x_5 = \lambda_{52}\xi_2 + \delta_5$$
$$x_6 = \lambda_{62}\xi_2 + \delta_6$$
$$\mathrm{cov}(\xi_1, \xi_2) = \varphi_{12}$$
$$\mathrm{var}(\xi_1) = \varphi_{11}$$
$$\mathrm{var}(\xi_2) = \varphi_{22}$$

Identifiability Rules for CFA

(1) Two-indicator rule *(sufficient, not necessary)*

1) At least two factors
2) At least two indicators per factor
3) Exactly one non-zero element per row of Λ

   (translation: each x only is pointed at by one LV)

4) Non-correlated errors (Θ is diagonal)

   (translation: no double-header arrows between the δ's)

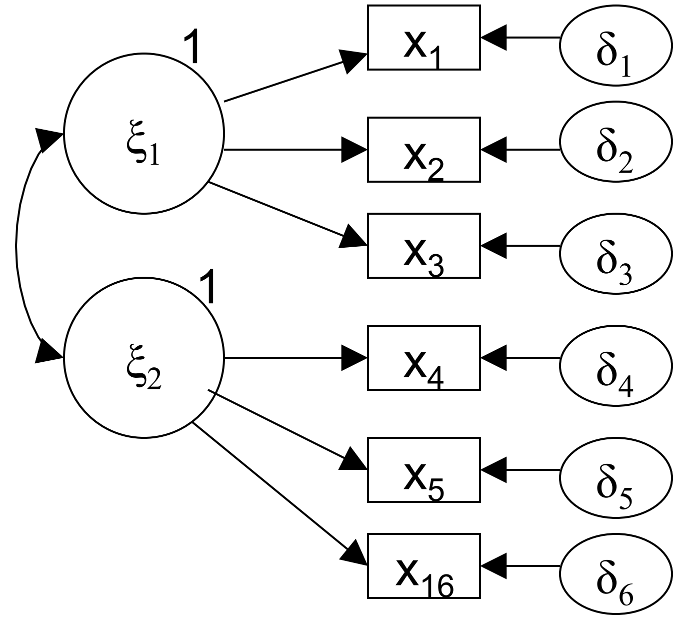
5) Factors are correlated (Φ has no zero elements)*

   (translation: there are double-header arrows between all of the LVs)

* Alternative less strict criteria: each factor is correlated with **at least** one other factor.

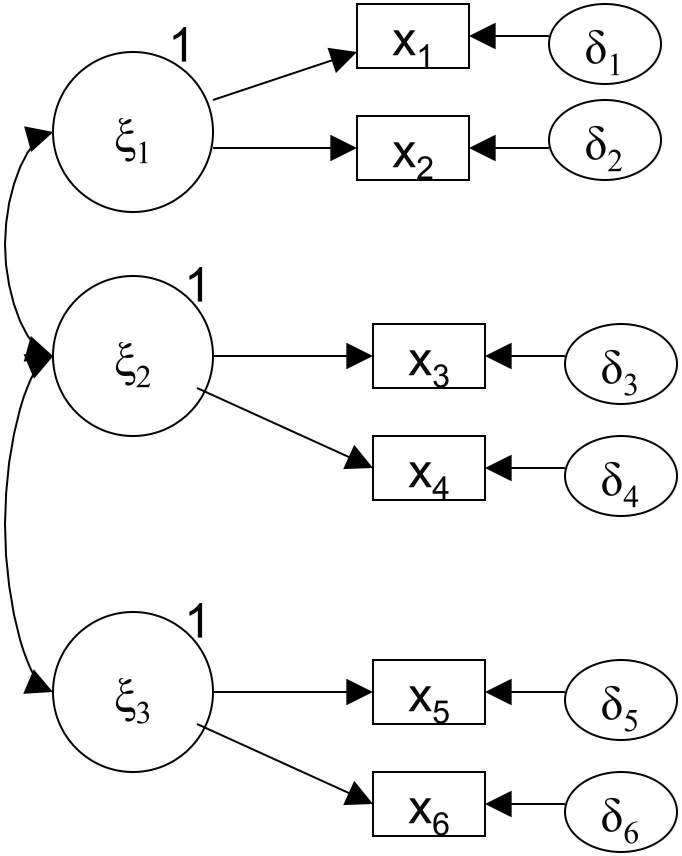
(see page 247 on Bollen)

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}
=
\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}
\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}
+
\begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}
$$

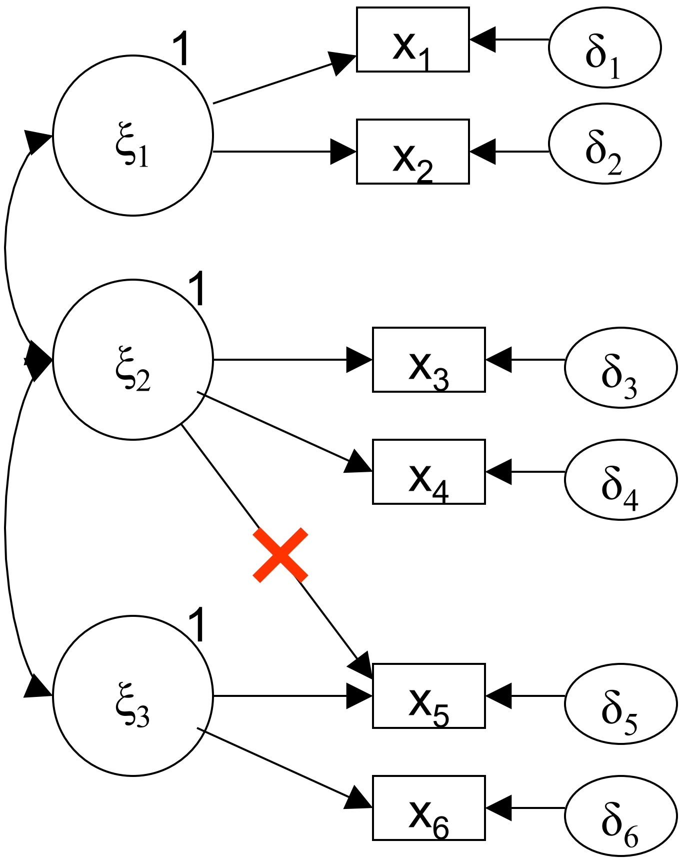


$$
\Theta = \mathrm{var}(\delta) =
\begin{bmatrix}
\theta_{11} & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_{22} & 0 & 0 & 0 & 0 \\
0 & 0 & \theta_{33} & 0 & 0 & 0 \\
0 & 0 & 0 & \theta_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & \theta_{55} & 0 \\
0 & 0 & 0 & 0 & 0 & \theta_{66}
\end{bmatrix}
\qquad
\Phi = \mathrm{var}(\xi) =
\begin{bmatrix} 1 & \varphi_{12} \\ \varphi_{12} & 1 \end{bmatrix}
$$

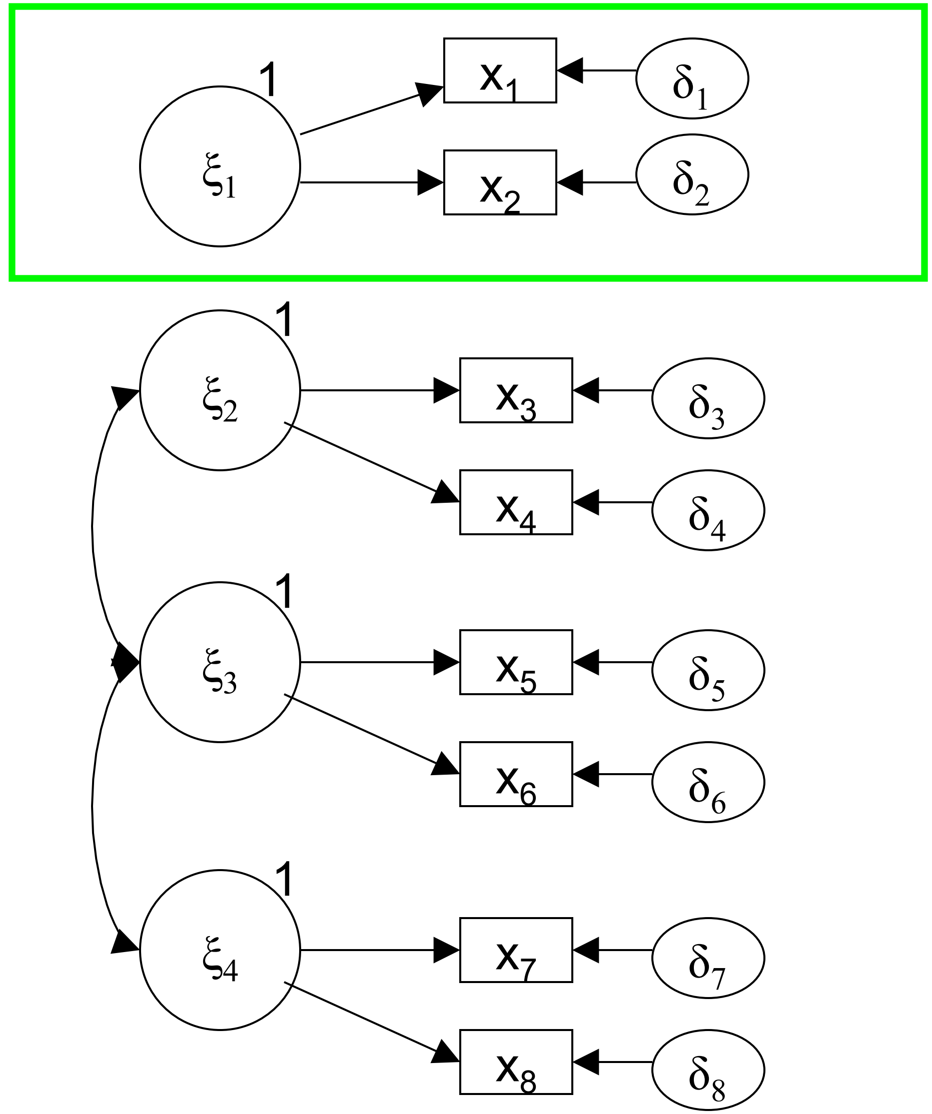
Example: Two-Indicator Rule



Example: Two-Indicator Rule



Example: Two-Indicator Rule



Identifiability Rules for CFA

(2) Three-indicator rule *(sufficient, not necessary)*

1) at least one factor
2) at least three indicators per factor
3) one non-zero element per row of Λ

   (translation: each x only is pointed at by one LV)

4) non-correlated errors (Θ is diagonal)

   (translation: no double-headed arrows between the δ's)

[Note: no condition about correlation of factors (no restrictions on Φ).]