The process of fitting structural equation models

1. Estimate

| Specify Model | → | Check identification | → | Prepare data | → | Estimate model |

2. Evaluate

Evaluate fit

Bad fit: Respecify model

3. Interpret

Good fit: Make inferences

| Interpret estimates | → | Consider other models | → | Report results |

SEM notation redux (RAM notation)

Exogenous (X)

Endogenous (Y)

Variance

Correlation

Observed variable

Regression path

Disturbance (Residual variance)

Residual Association (Disturbance Correlation)

X_1 X_2 Y_1 Y_2 Y_3 D_1 D_2 D_3
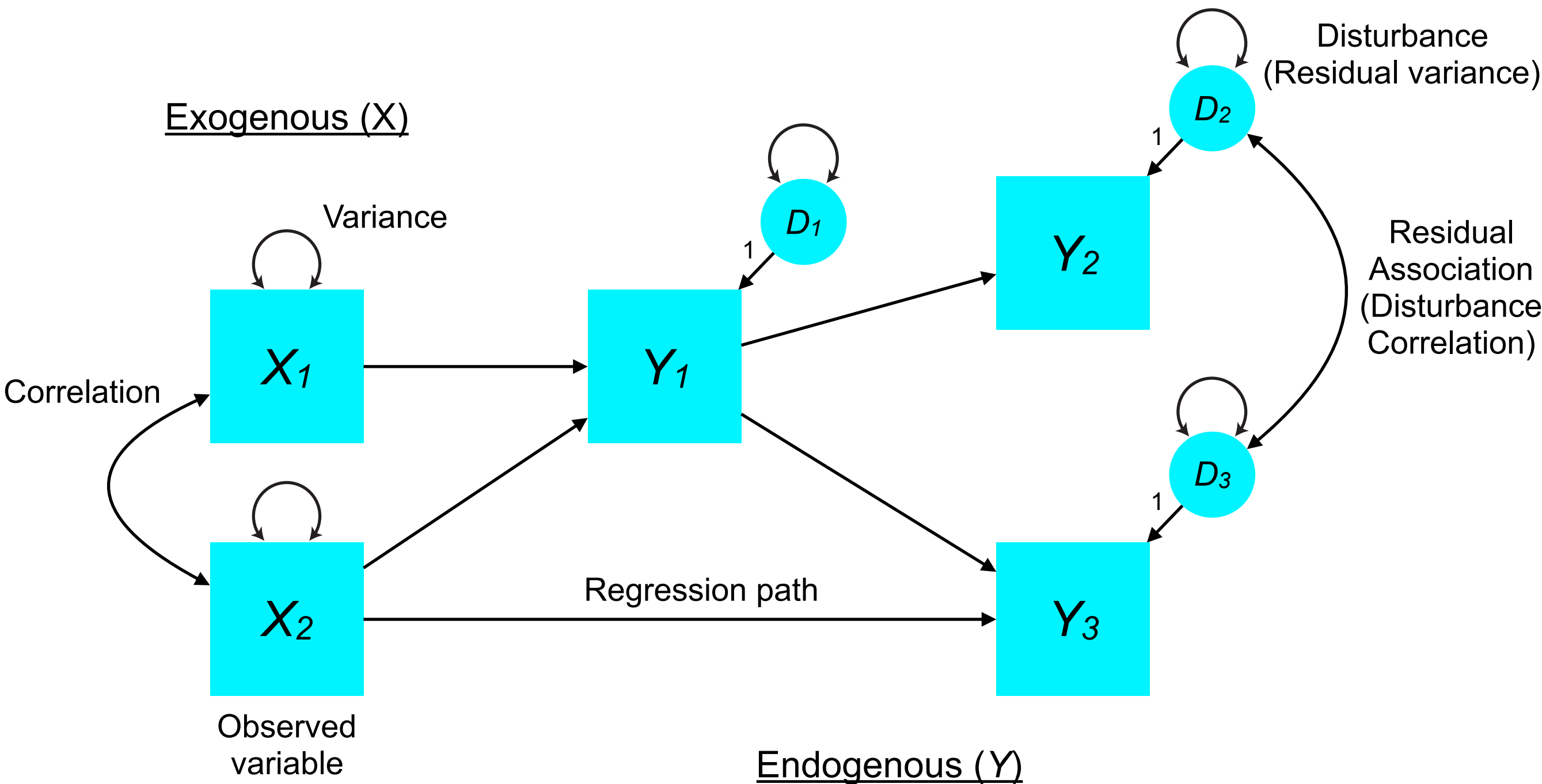
SEM notation redux (simplified RAM notation)

Disturbance
(Residual variance)

Exogenous (X)

Variance

Y_2

X_1

Y_1

Residual
Association
(Disturbance
Correlation)

Correlation

X_2

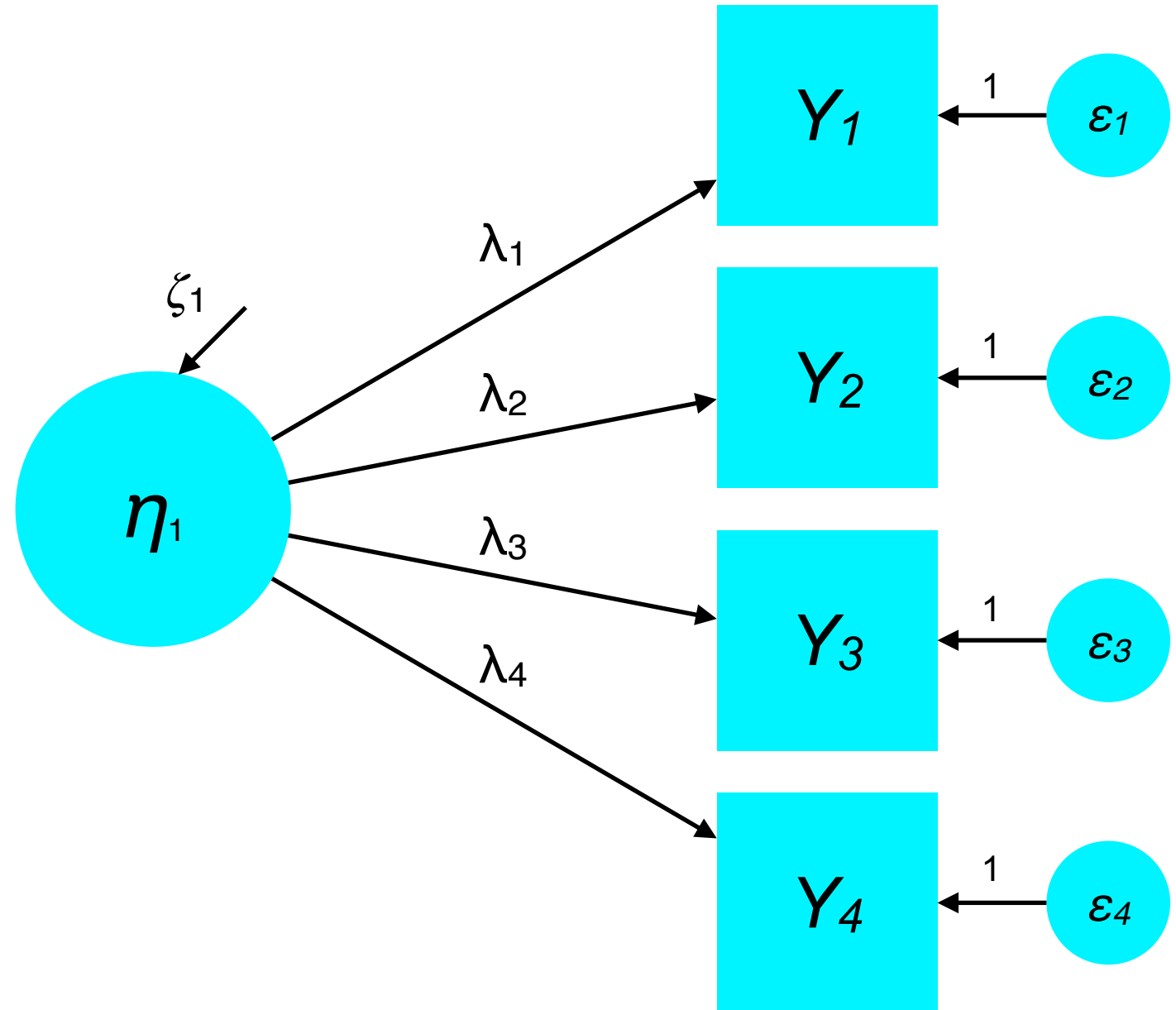Regression path

Y_3

Observed
variable

Endogenous (Y)
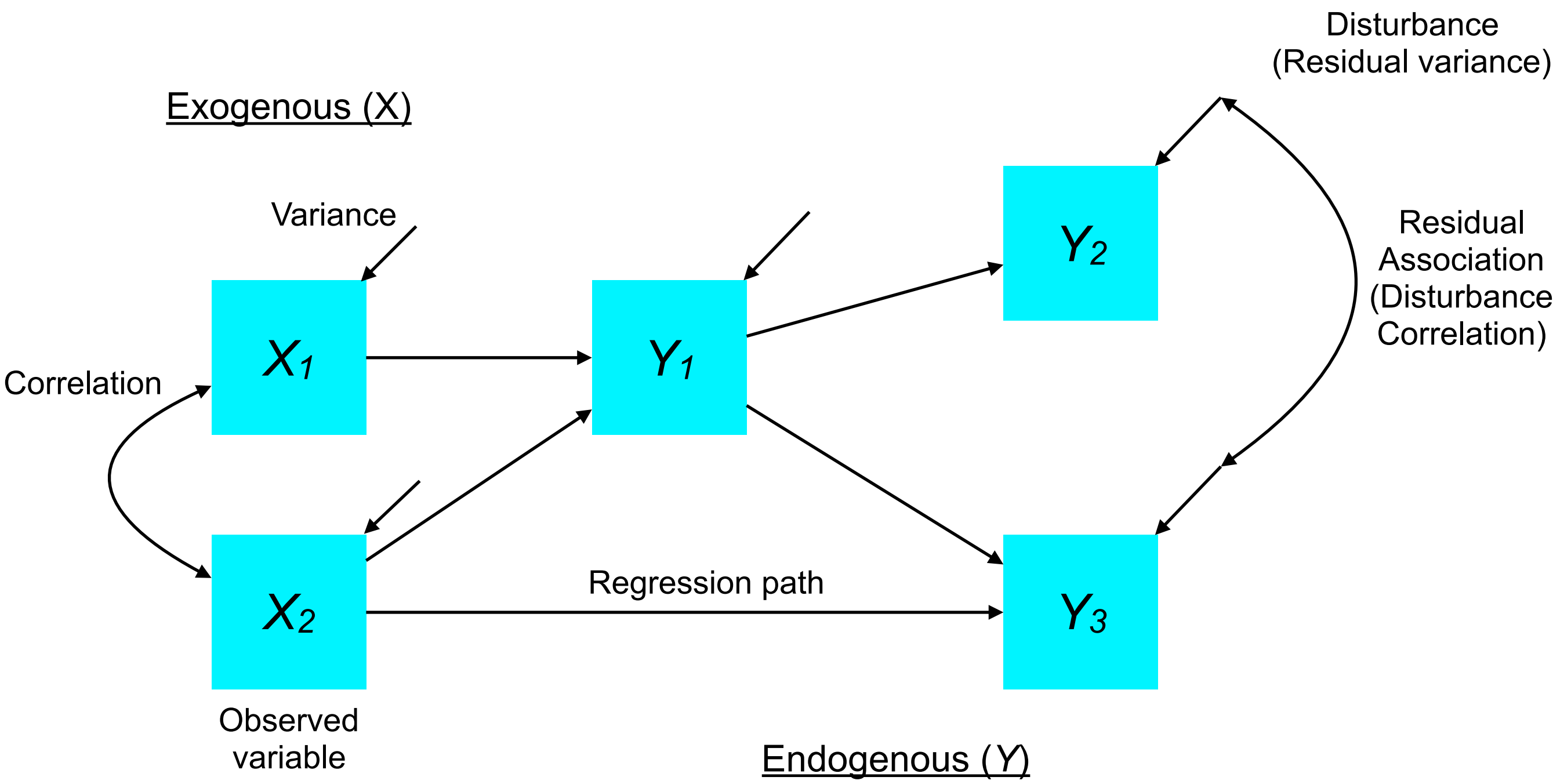
Common Factor Model

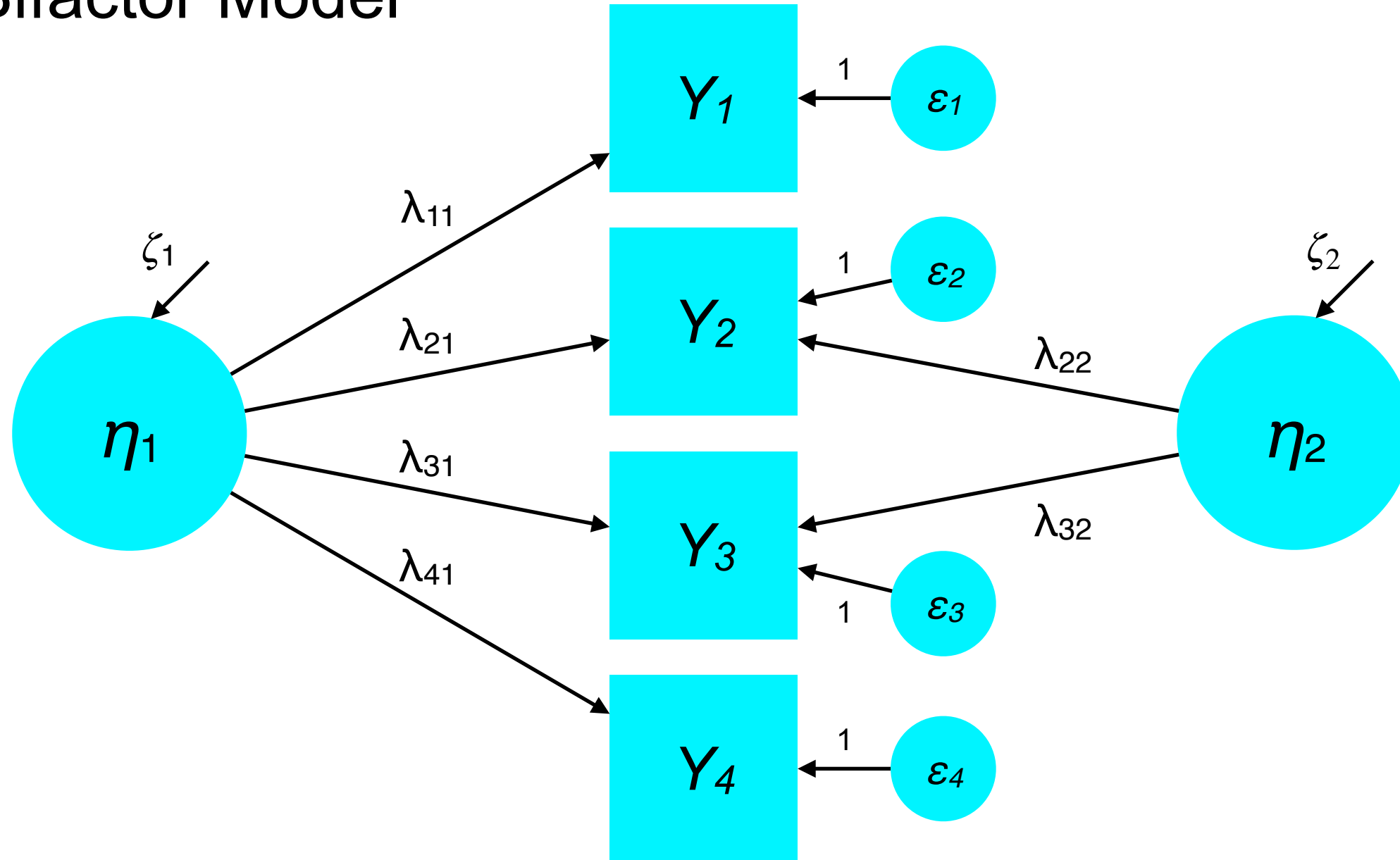$$y_1 = \alpha_1 + \lambda_1 \eta_1 + \varepsilon_1$$

$$y_2 = \alpha_2 + \lambda_2 \eta_1 + \varepsilon_2$$

$$y_3 = \alpha_3 + \lambda_3 \eta_1 + \varepsilon_3$$

$$y_4 = \alpha_4 + \lambda_4 \eta_1 + \varepsilon_4$$
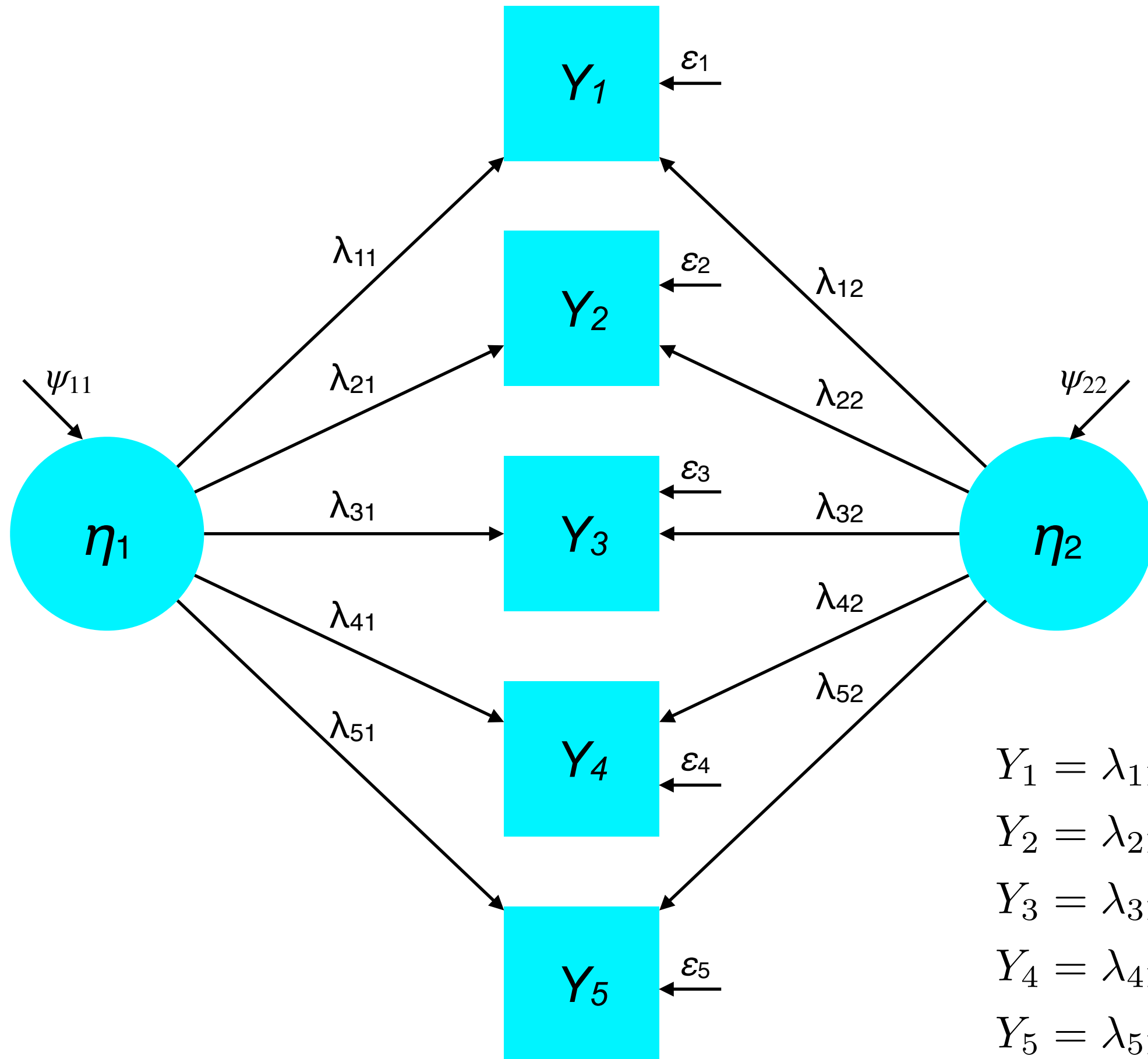
Bifactor Model

$$y_1 = \alpha_1 + \lambda_{11}\eta_1 + \varepsilon_1$$

$$y_2 = \alpha_2 + \lambda_{21}\eta_1 + \lambda_{22}\eta_2 + \varepsilon_2$$

$$y_3 = \alpha_3 + \lambda_{31}\eta_1 + \lambda_{32}\eta_2 + \varepsilon_3$$

$$y_4 = \alpha_4 + \lambda_{41}\eta_1 + \varepsilon_4$$

Two-factor model as typically instantiated in EFA

$$Y_1 = \lambda_{11}\boldsymbol{\eta}_1 + \lambda_{12}\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_1$$
$$Y_2 = \lambda_{21}\boldsymbol{\eta}_1 + \lambda_{22}\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_2$$
$$Y_3 = \lambda_{31}\boldsymbol{\eta}_1 + \lambda_{32}\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_3$$
$$Y_4 = \lambda_{41}\boldsymbol{\eta}_1 + \lambda_{42}\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_4$$
$$Y_5 = \lambda_{51}\boldsymbol{\eta}_1 + \lambda_{52}\boldsymbol{\eta}_2 + \boldsymbol{\varepsilon}_5$$

Estimating SEMs

Specify an estimable model, code into *lavaan* syntax

↓

Software estimates plausible starting values for all parameters

↓

Software estimates model-implied covariance matrix ($\mathbf{\Sigma}$) at current parameter values

↓

Compare model-implied covariance ($\mathbf{\Sigma}$) to observed covariance ($\mathbf{S}$) according to sample log-likelihood function

$$(\mathbf{Y}_i - \boldsymbol{\mu})' \mathbf{\Sigma}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu})$$

↓

Update model parameters to reduce discrepancies (using derivative matrices)

Repeat until there is minimal change in log-likelihood (e.g., 10^{-5}). This is called model convergence.

A book worth reading…

Latent Curve Models

A Structural Equation Perspective

KENNETH A. BOLLEN

University of North Carolina
Department of Sociology
Chapel Hill, North Carolina

PATRICK J. CURRAN

University of North Carolina
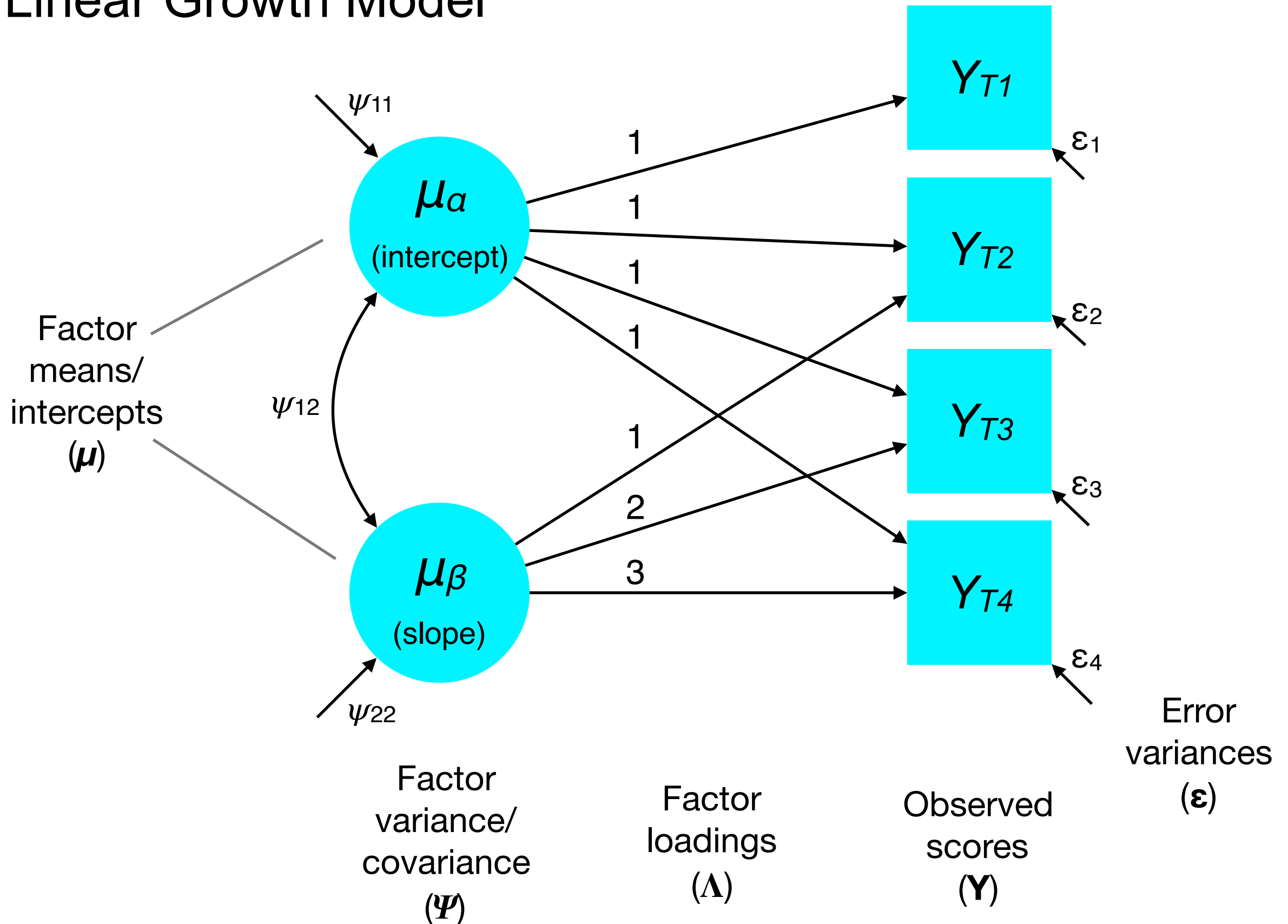Department of Psychology
Chapel Hill, North Carolina

- Forms of change
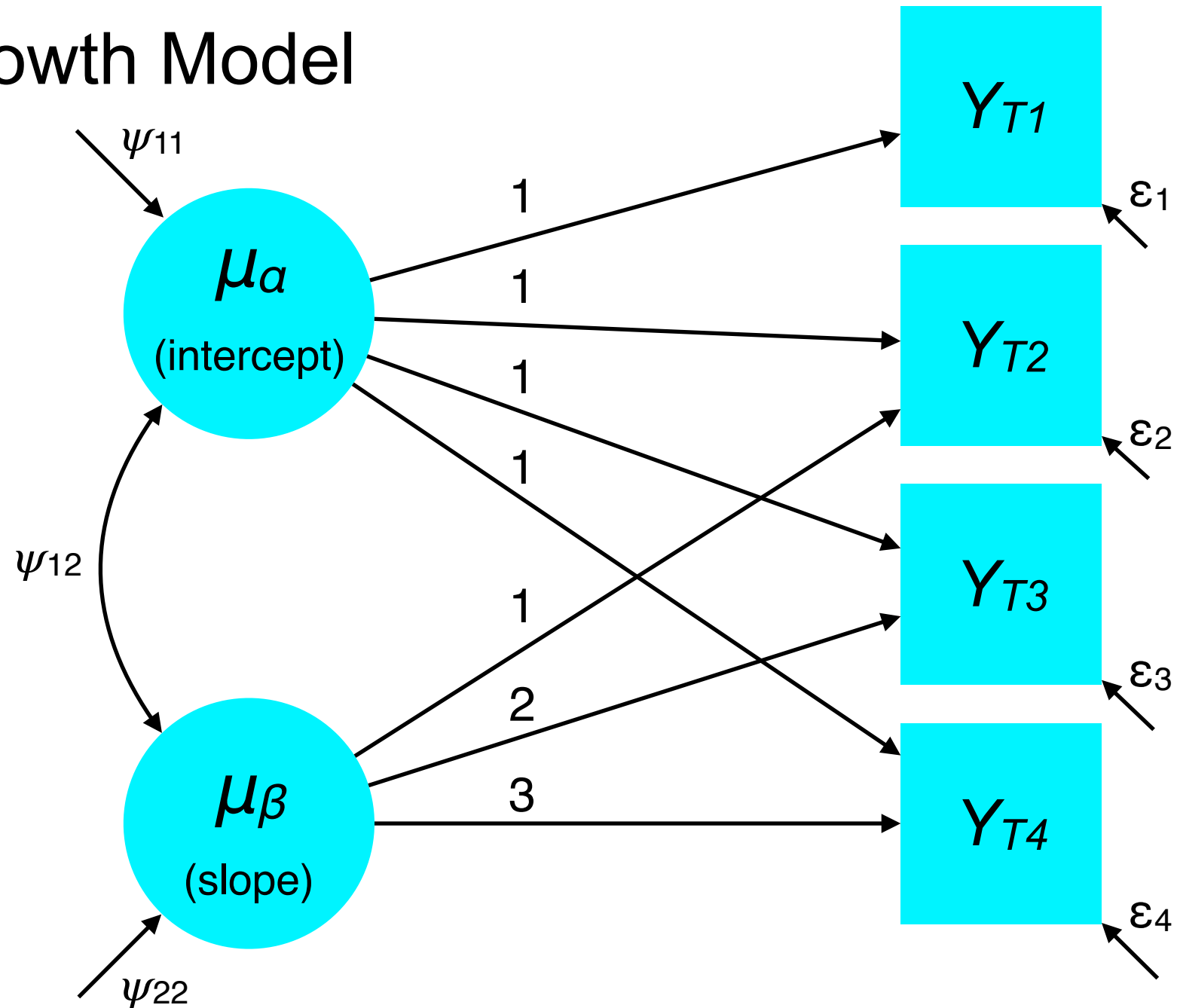
- Mean-level

- rank-order

- ipsative

Ergodicity

- Adapt Molenaar here: process of change within-individual may differ from mean-level change *between* individuals

Linear Growth Model

Linear Growth Model

$$y_{i,T1} = \alpha_i + \beta_i \lambda_{T1} + \varepsilon_{i,T1}$$
$$y_{i,T2} = \alpha_i + \beta_i \lambda_{T2} + \varepsilon_{i,T2}$$
$$y_{i,T3} = \alpha_i + \beta_i \lambda_{T3} + \varepsilon_{i,T3}$$
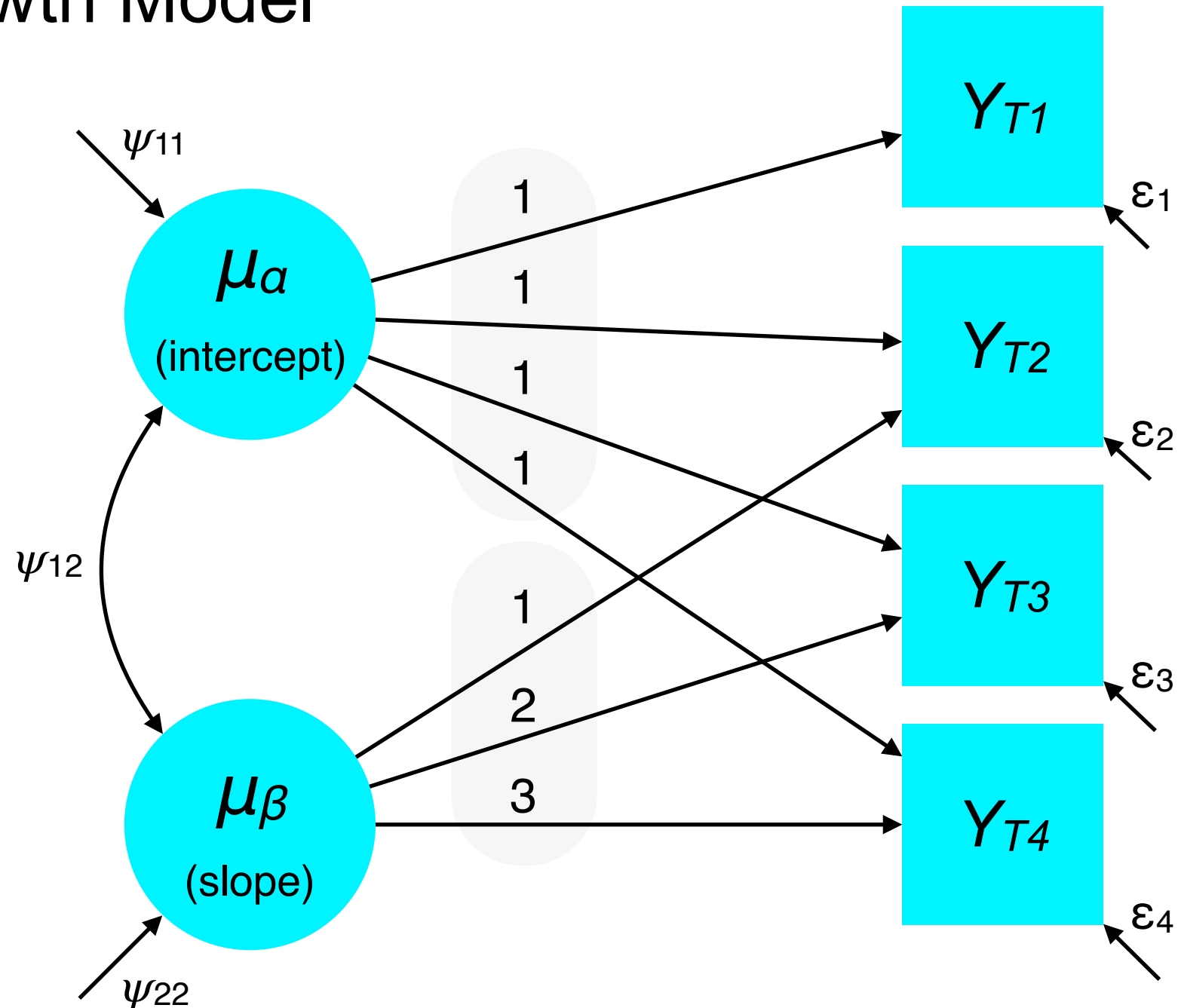$$y_{i,T4} = \alpha_i + \beta_i \lambda_{T4} + \varepsilon_{i,T4}$$

$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i}$$
$$\beta_i = \mu_\beta + \zeta_{\beta_i}$$
$$Var(\zeta_{\boldsymbol{\alpha}}) = \psi_{11}$$
$$Var(\zeta_{\boldsymbol{\beta}}) = \psi_{22}$$

Linear Growth Model

Quadratic Growth Model

Linear Growth Model

Mean structure

α_1

α_2

ψ_{11}

ψ_{12}

ψ_{22}

η_1 (intercept)

η_2 (slope)

1

1

1

1

1

1

2

3

Y_{T1}

Y_{T2}

Y_{T3}

Y_{T4}

ε_1

ε_2

ε_3

ε_4

λ_{11}

λ_{21}

λ_{31}

λ_{41}

Factor variance/ covariance (α)

Factor variance/ covariance (Ψ)

Factor loadings (Λ)

Observed scores (Y)

Error variances (ε)

Fixed loadings denote functional form of change

Fixed loadings denote functional form of change

$y_1 = \alpha_1$

$y_2 = \alpha_2$
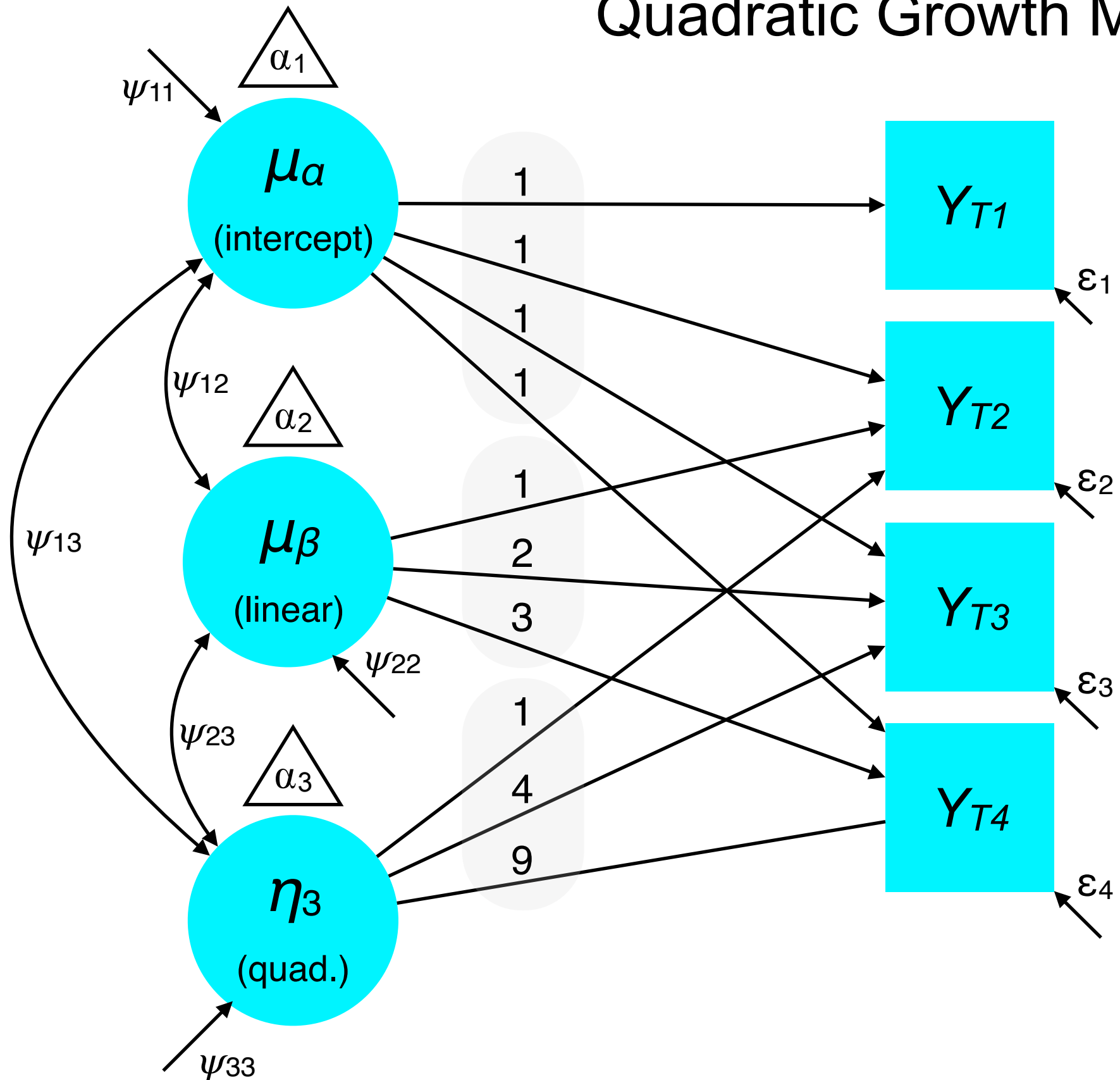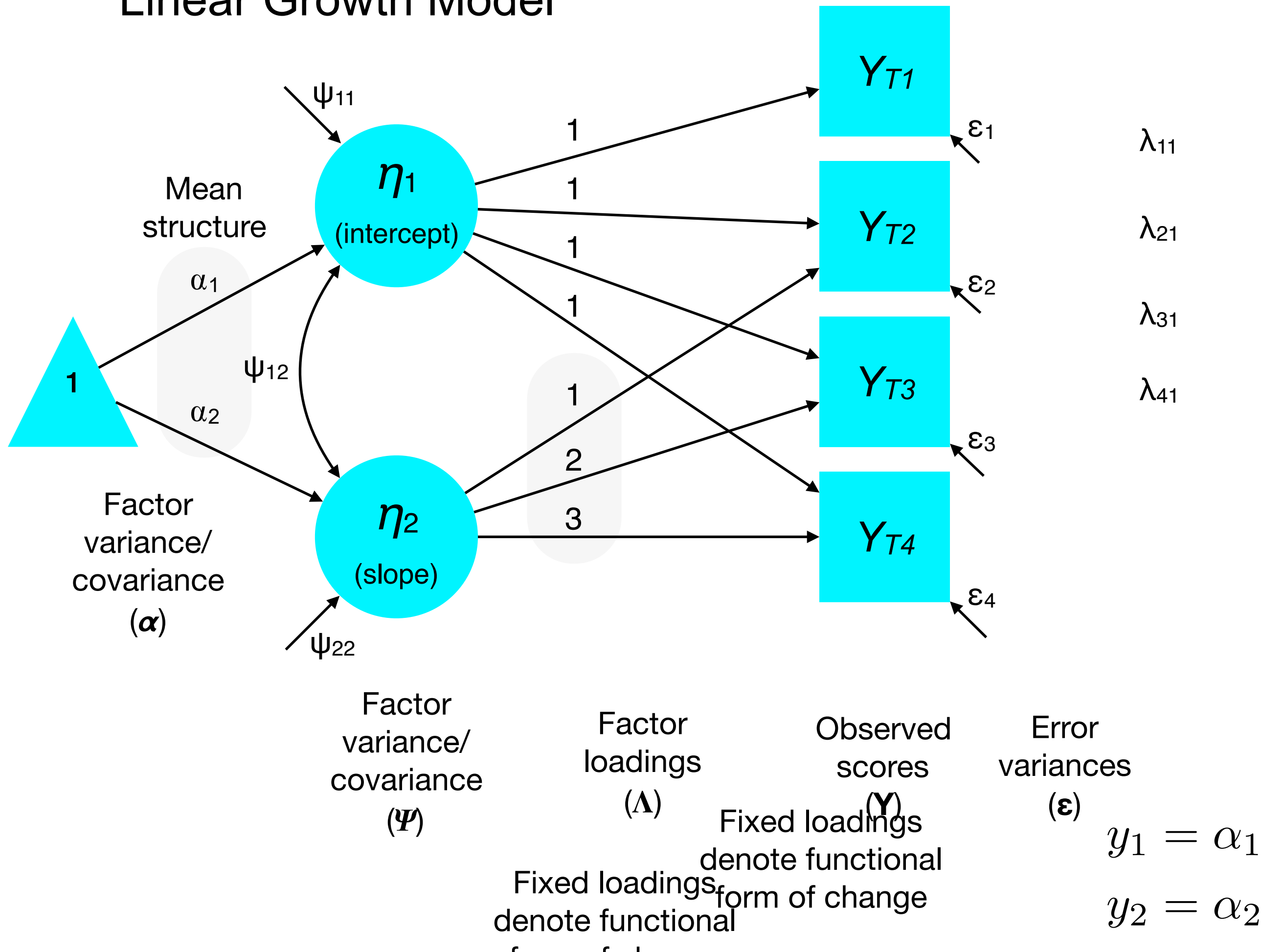
Organize wrt age, not wave

Points around missingness, covariance coverage
Accelerated longitudinal design, but only works if there are many more
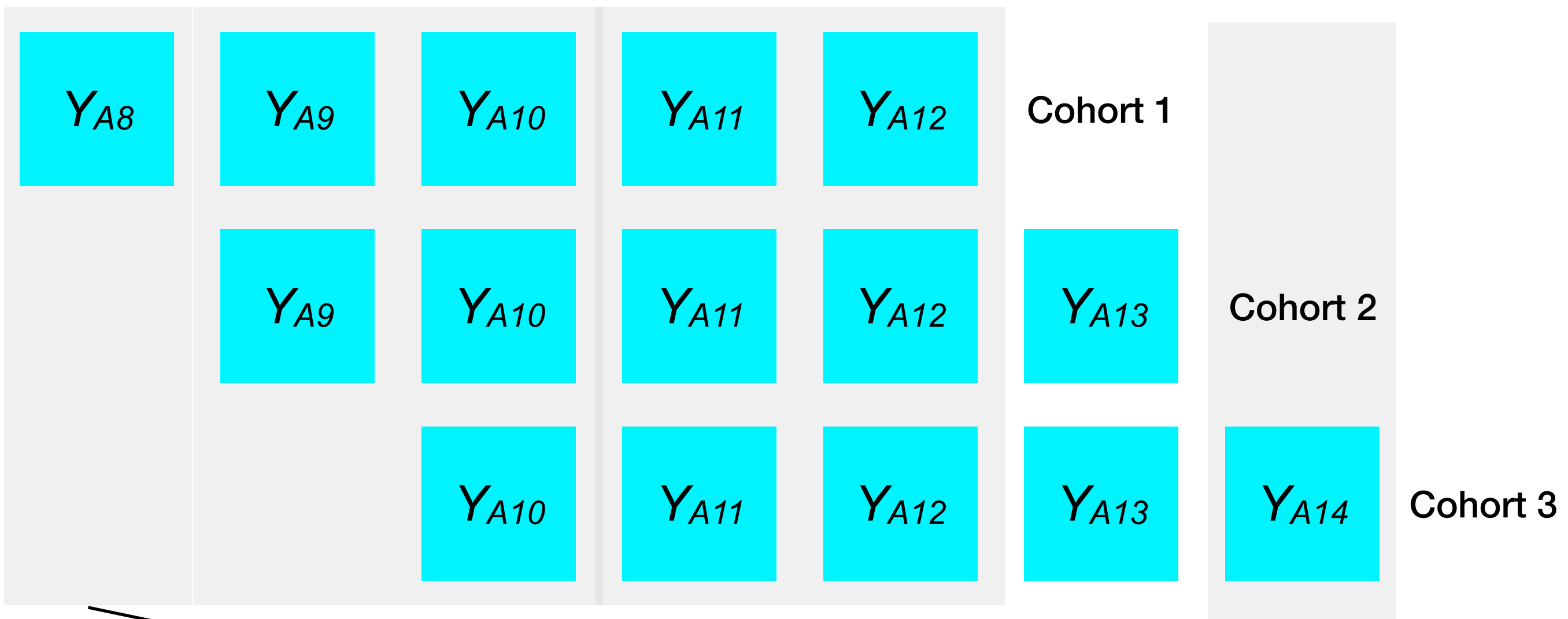waves than there are age *cohorts.*

Ad hoc accelerated longitudinal will not have covariance coverage! So if you are accelerating,
strongly recommend cohorts and not spacing the out too much

Organizing wrt wave means we are pooling estimates (e.g., means or variances) that
reflect different moments in developmental time (e.g., age)

Accelerated Longitudinal Cohort Design

aka Cohort-Sequential Design

Y_{A8} Y_{A9} Y_{A10} Y_{A11} Y_{A12} Cohort 1

Y_{A9} Y_{A10} Y_{A11} Y_{A12} Y_{A13} Cohort 2

Y_{A10} Y_{A11} Y_{A12} Y_{A13} Y_{A14} Cohort 3
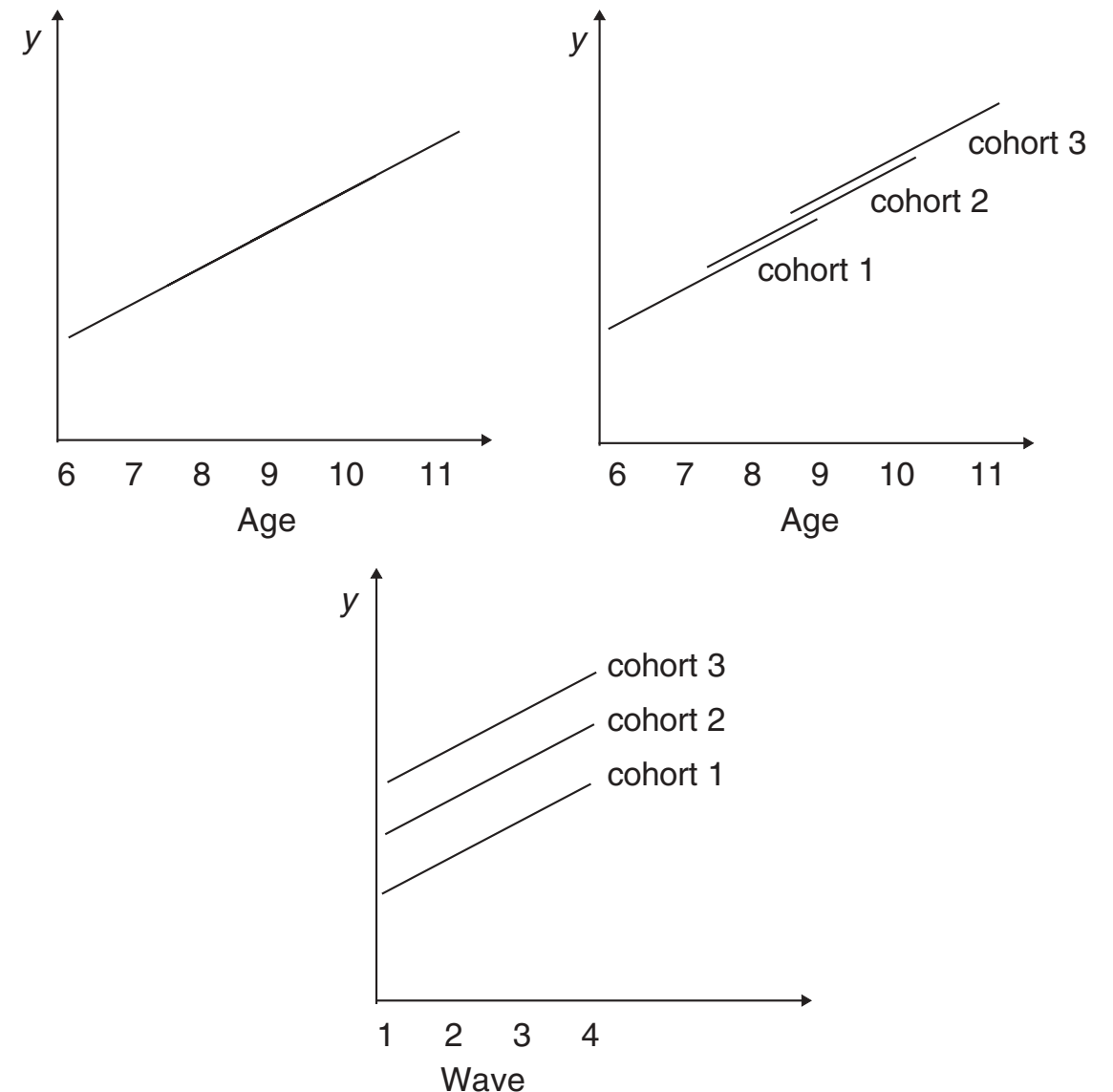
Coverage (A9, A10) = 0.66 *Coverage* (A10, A11) = 1.0

Coverage (A8, A14) = 0.0

(Assuming Equal n per cohort)

Estimation in cohort designs

- Risks of organizing data with respect to cohorts if there are meaningful developmental differences among them (esp. age)

- Missing by design (in sequential-cohort or ad hoc accelerated design) is MCAR by definition.

- Thus, FIML is an acceptable estimation method

- Based on the fact that the residual covariance between indicators is *not* a feature of the model — i.e., we assume conditional independence.

- If we thought there were some meaningful residual covariation (e.g., age 8 with age 14), we should worry about covariance coverage because the structural equations for **y** now depend on the theta $\boldsymbol{\Theta}$ matrix (residual covariance structure). (i.e., a diagonal matrix)

$$\boldsymbol{\theta} = \begin{bmatrix} \varepsilon_1 & & \\ 0 & \varepsilon \\ 0 & \\ 0 & \end{bmatrix}$$
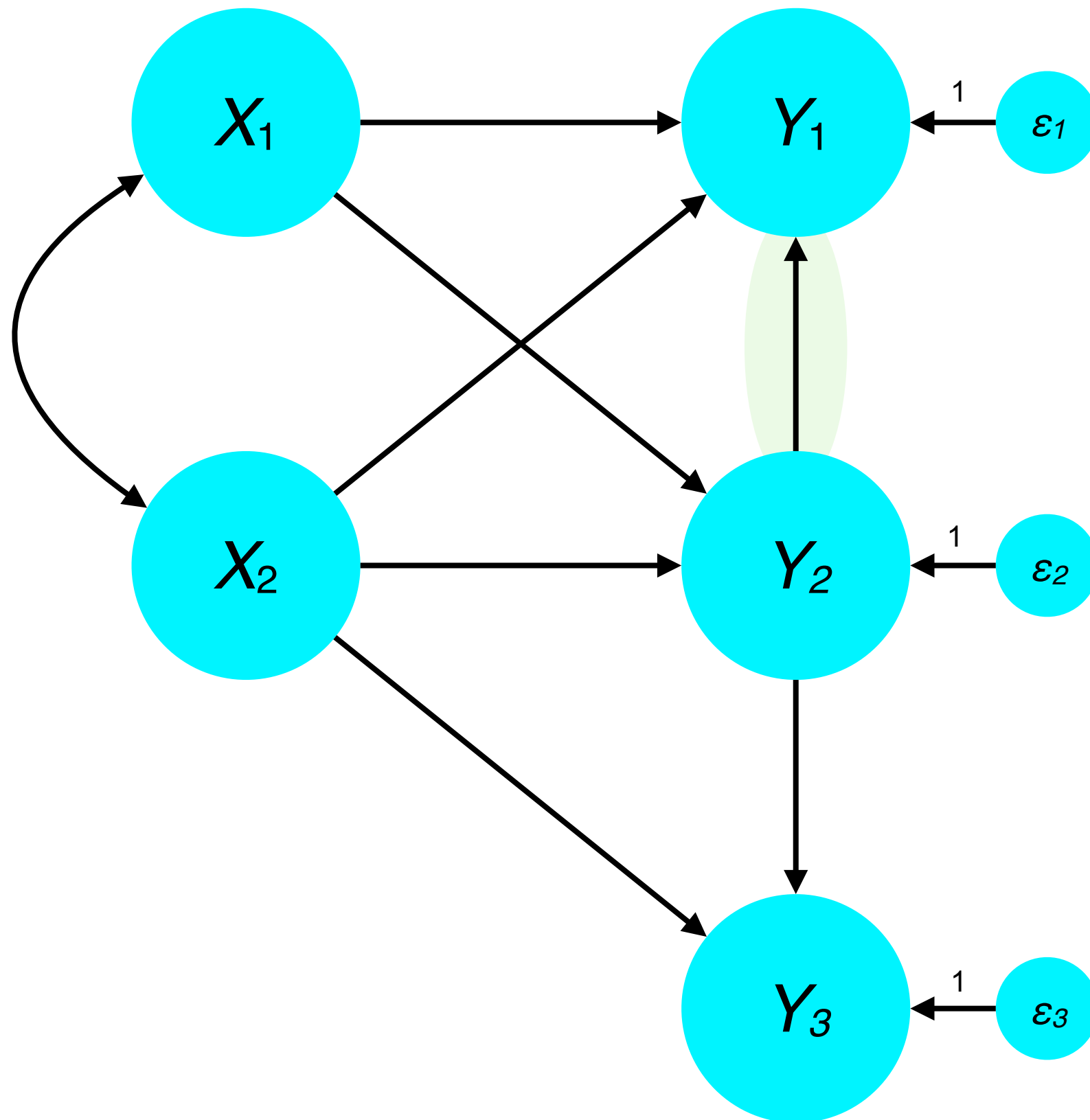
- d+2 rule

- Mean structure captures average mean level (when slope loading is 0) and average slope

- Variances capture individual differences in level and rate of change in growth process
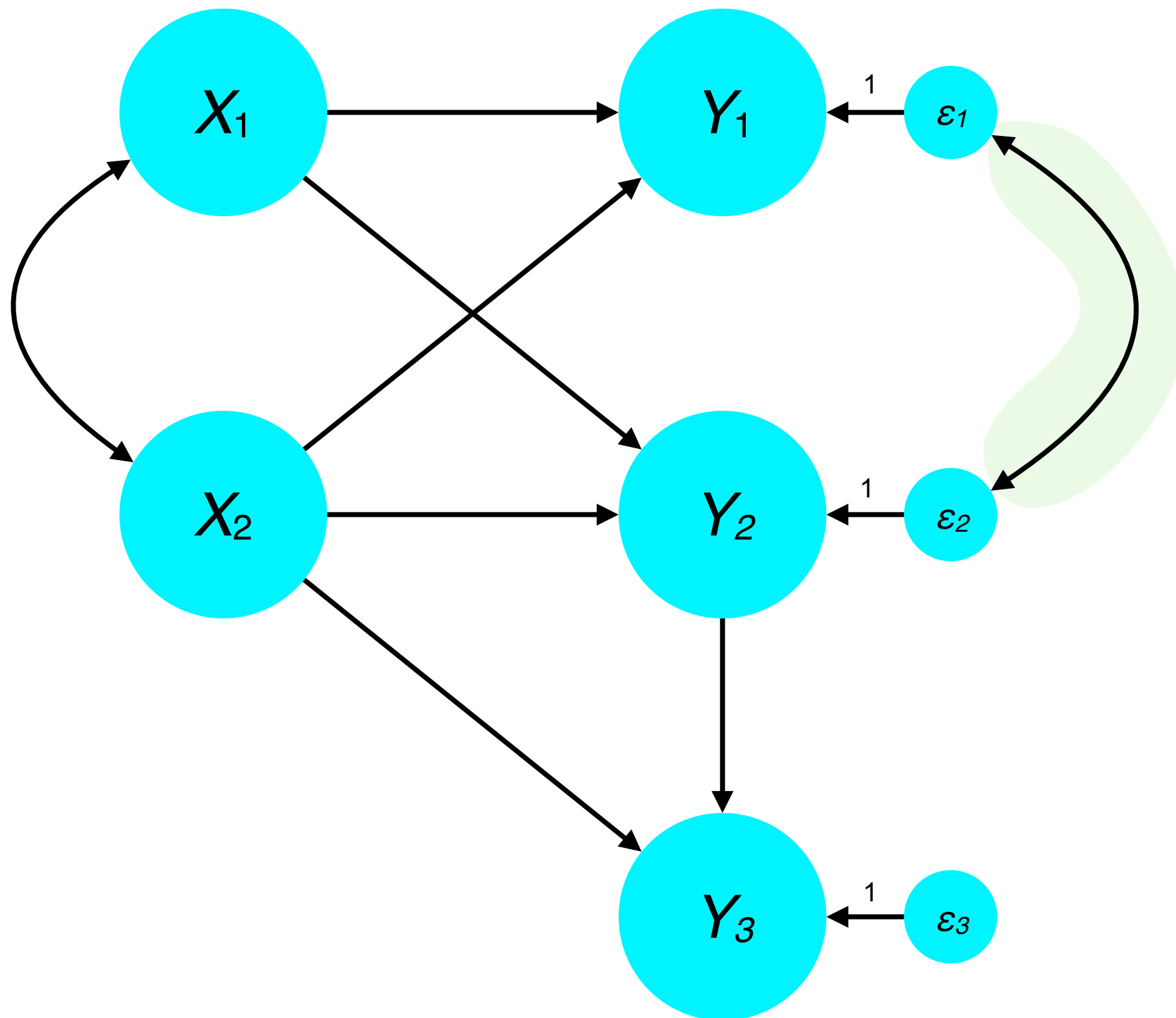
Equivalent Models: A
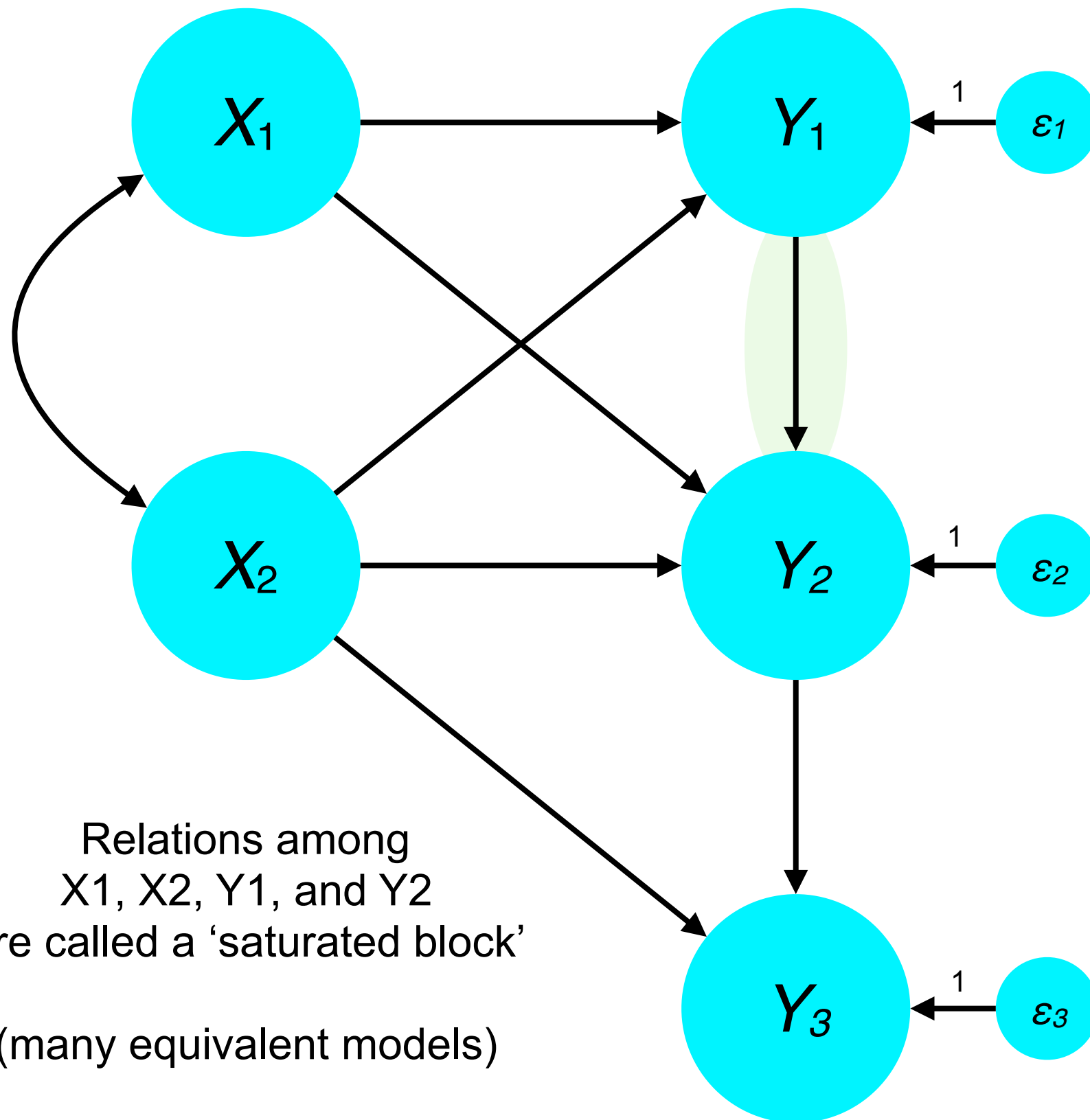
Relations among
X1, X2, Y1, and Y2
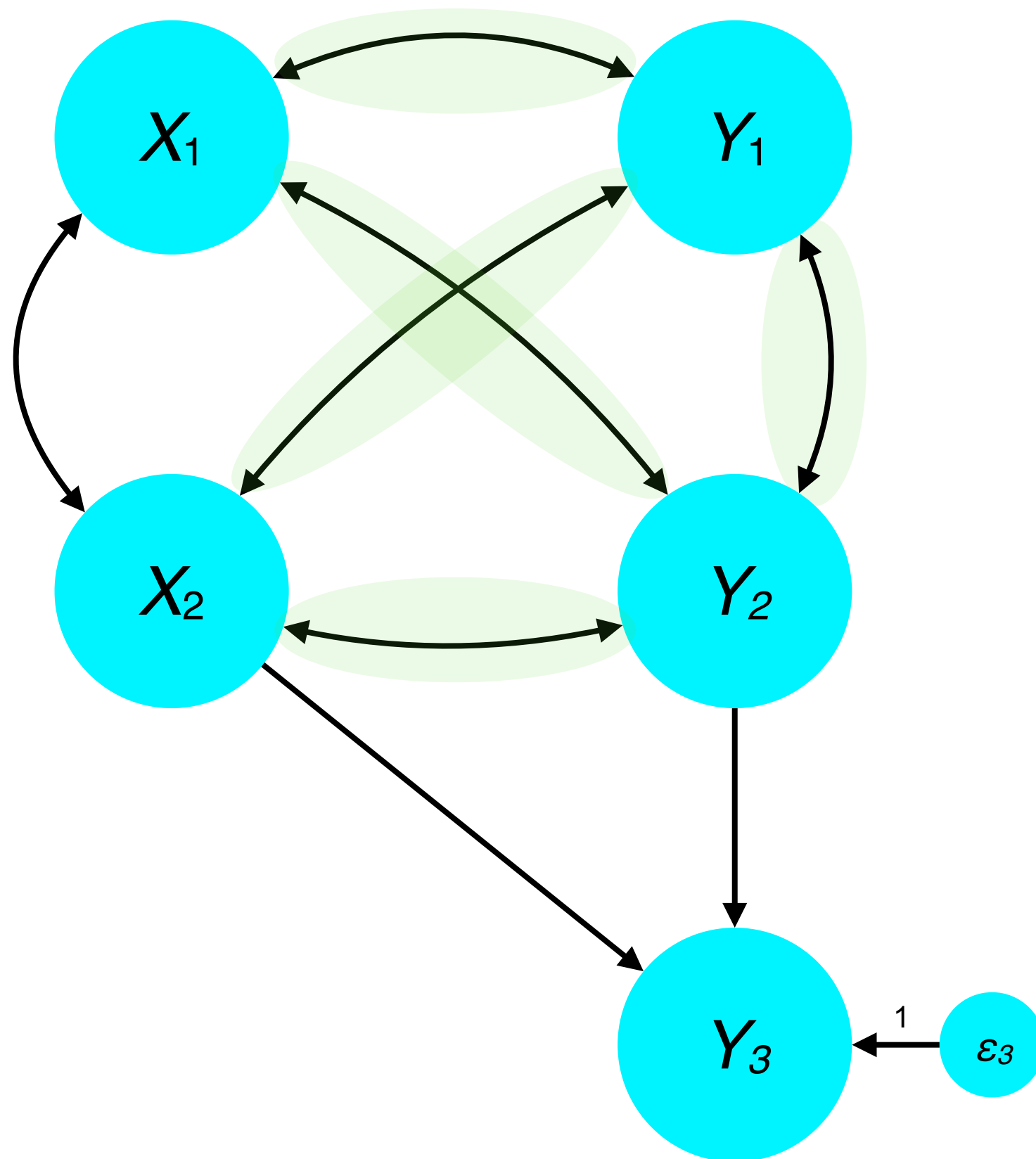are called a 'saturated block'
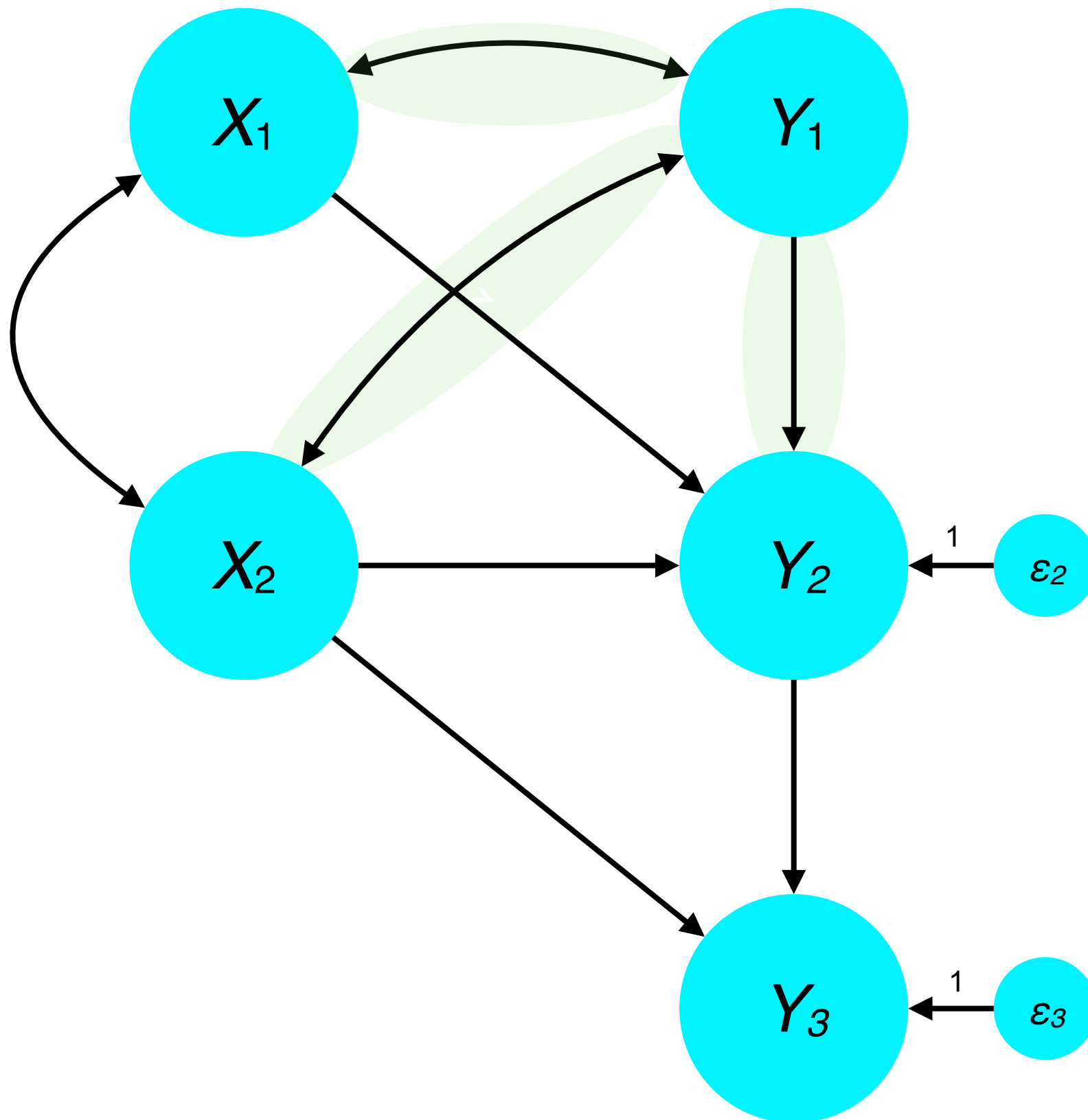
(many equivalent models)

Equivalent Models: B

Equivalent Models: C

Equivalent Models: D

Equivalent Models: E

Equivalent Models: F

Change in direction of Y - X effects