

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

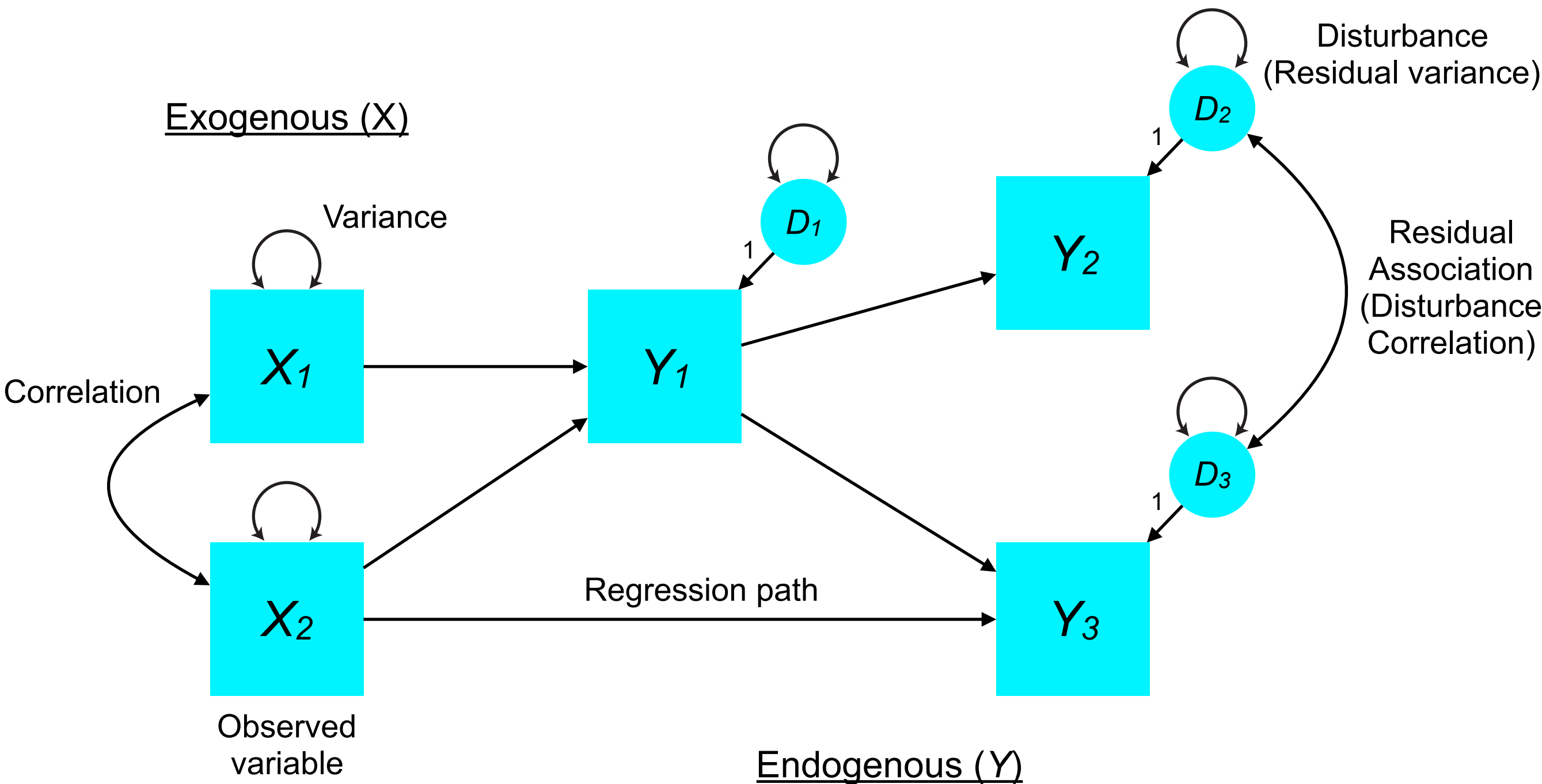
Interpret
estimates

Consider
other models

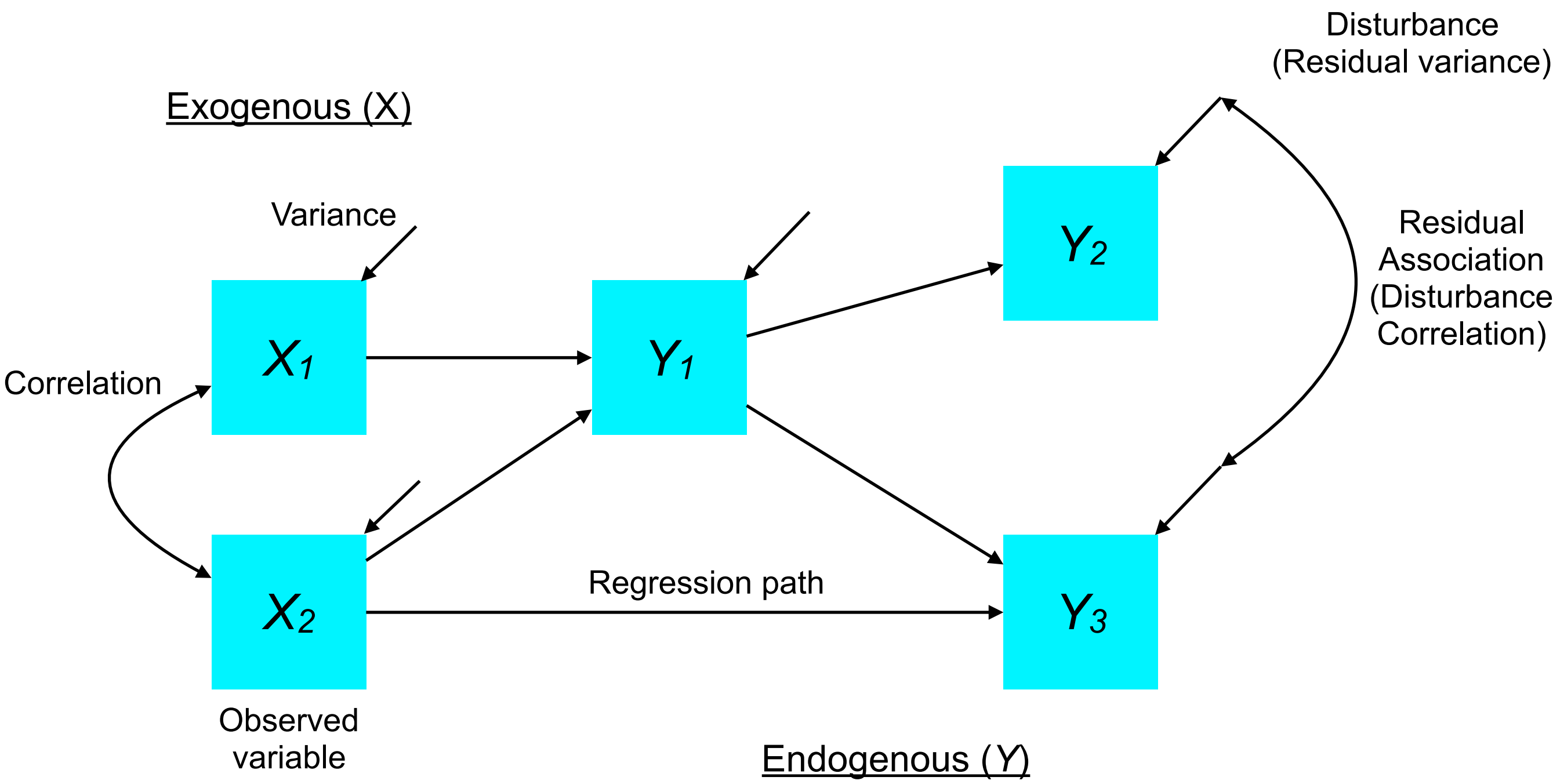
Report
results

Good fit:
Make inferences

SEM notation redux (RAM notation)



SEM notation redux (simplified RAM notation)



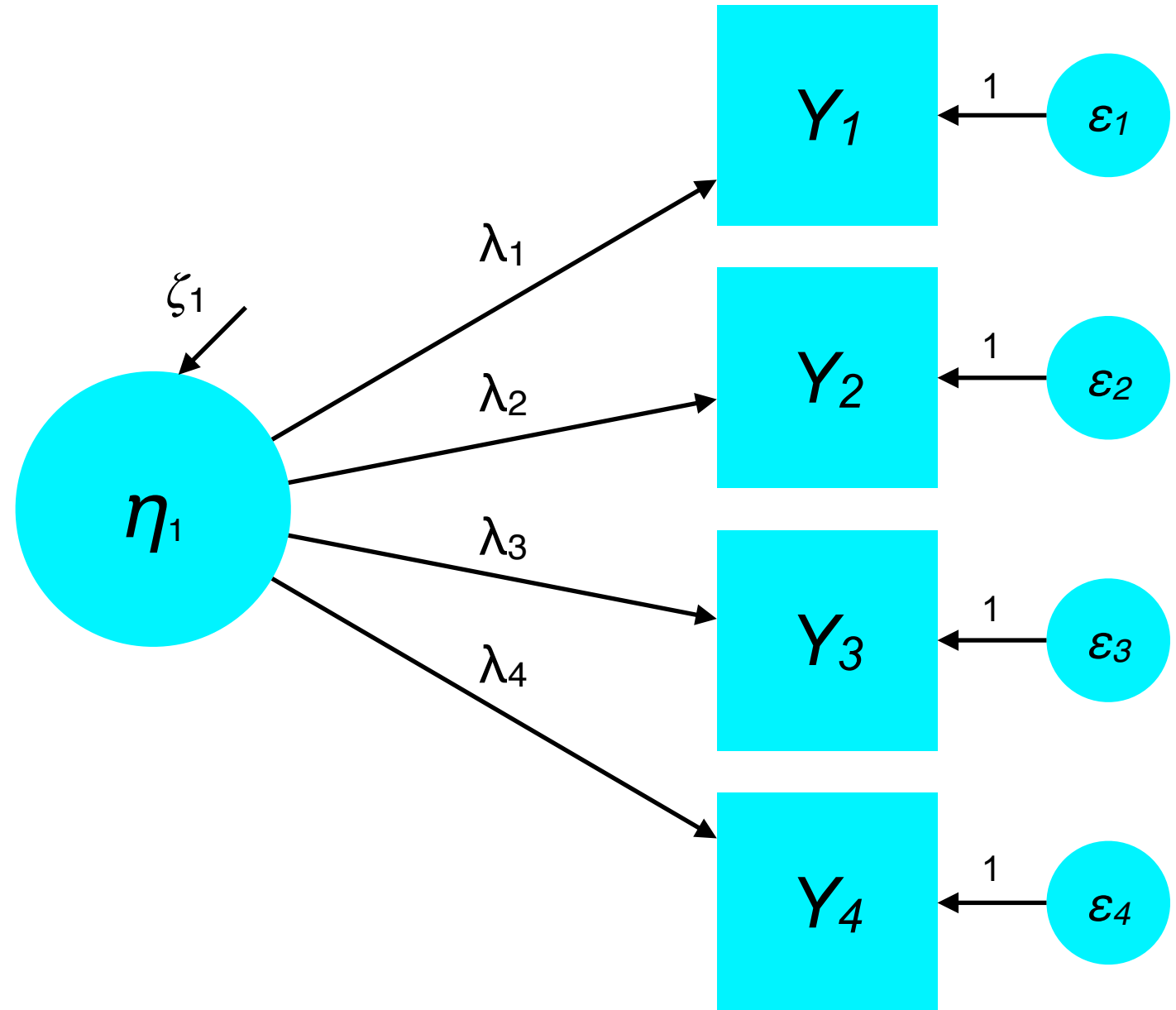
Common Factor Model

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

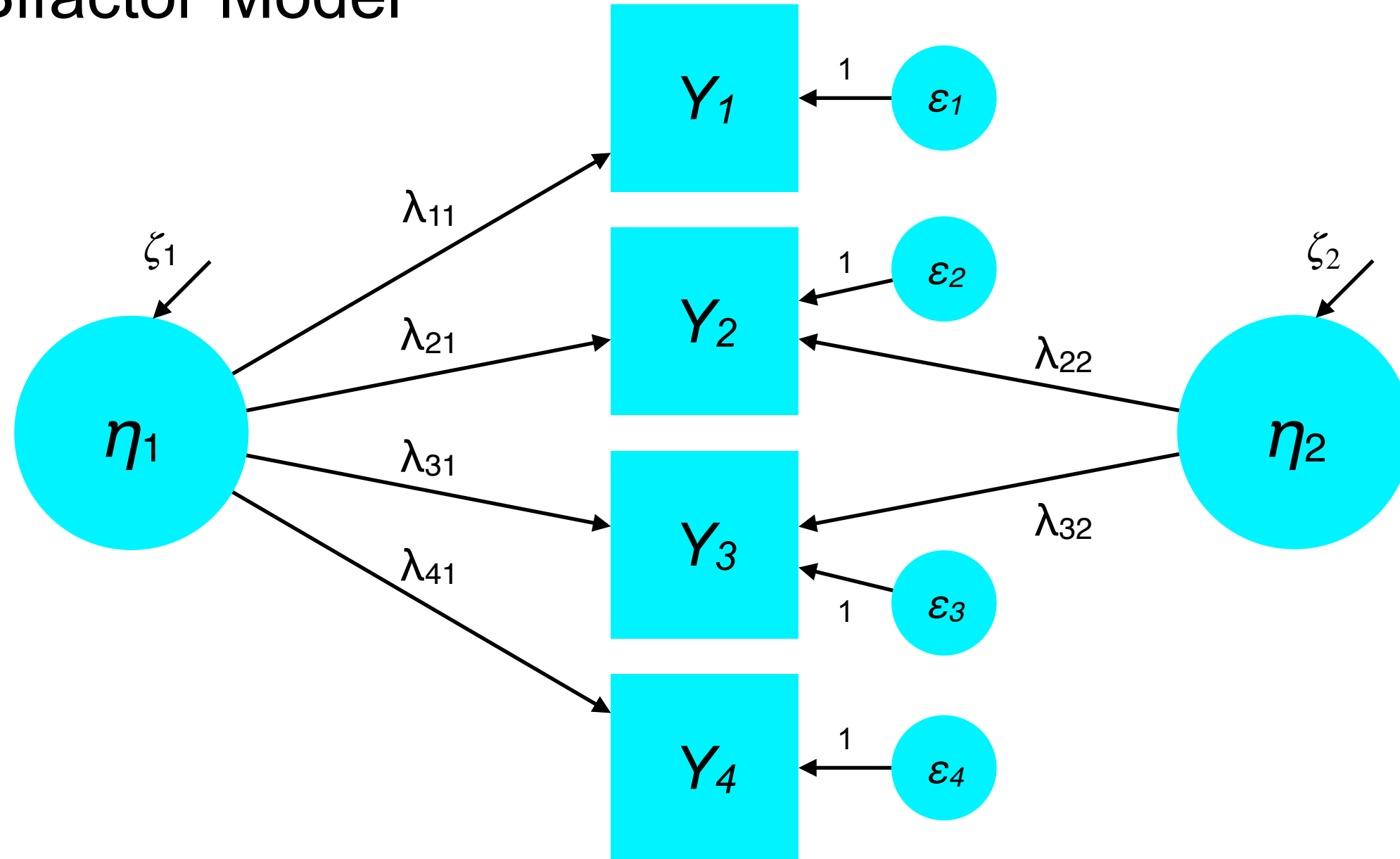
$$y_2 = \alpha_2 + \lambda_2 \eta_1 + \varepsilon_2$$

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

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



Bifactor Model



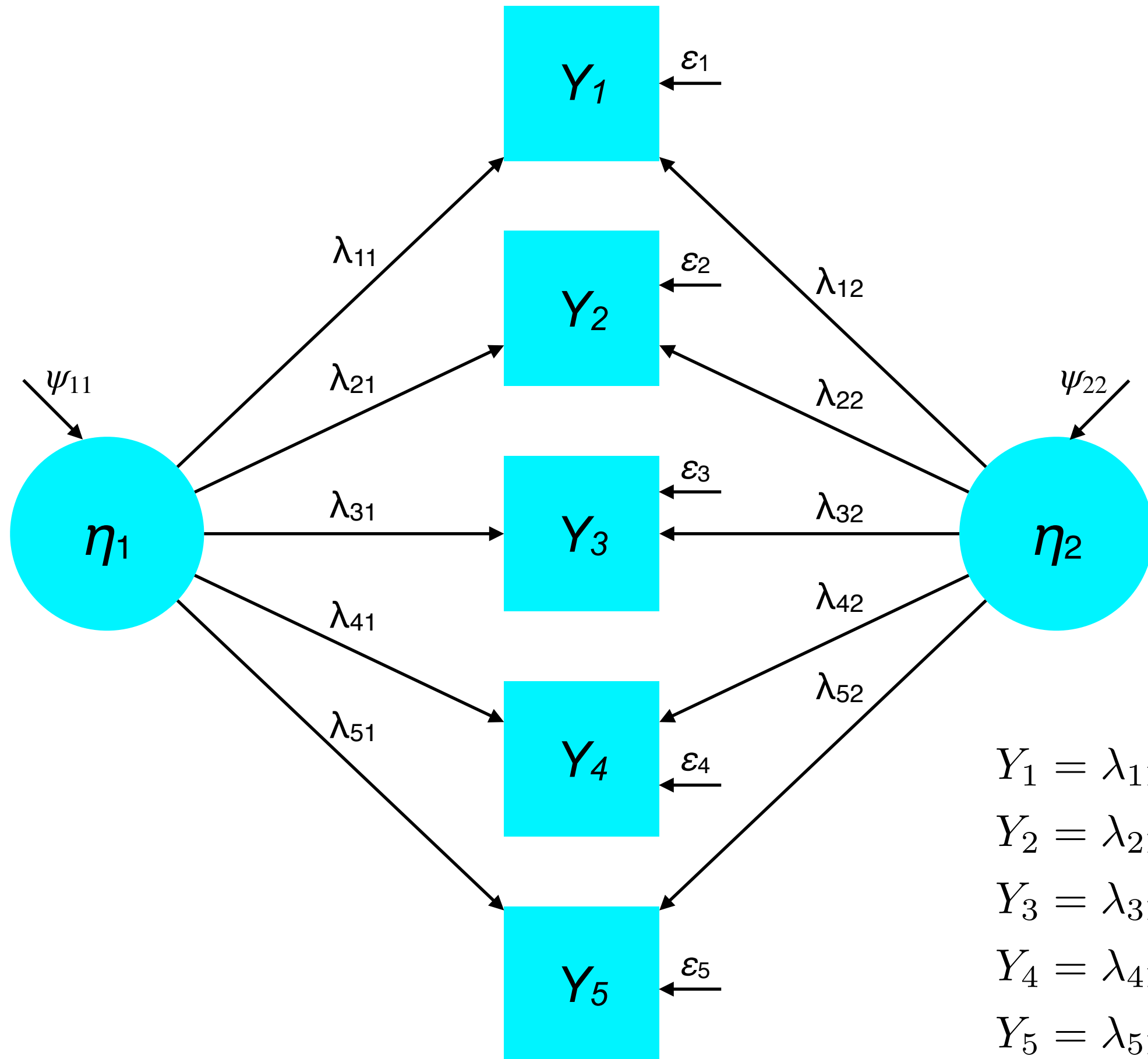
$$y_1 = \alpha_1 + \lambda_{11}\eta_1 + \varepsilon_1$$

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

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

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

Two-factor model as typically instantiated in EFA



Estimating SEMs

Specify an estimable model,
code into *lavaan* syntax

Software estimates plausible starting values
for all parameters

Software estimates model-implied covariance
matrix (Σ) at current parameter values

Compare model-implied covariance (Σ) to
observed covariance (S) according to sample
log-likelihood function

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

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

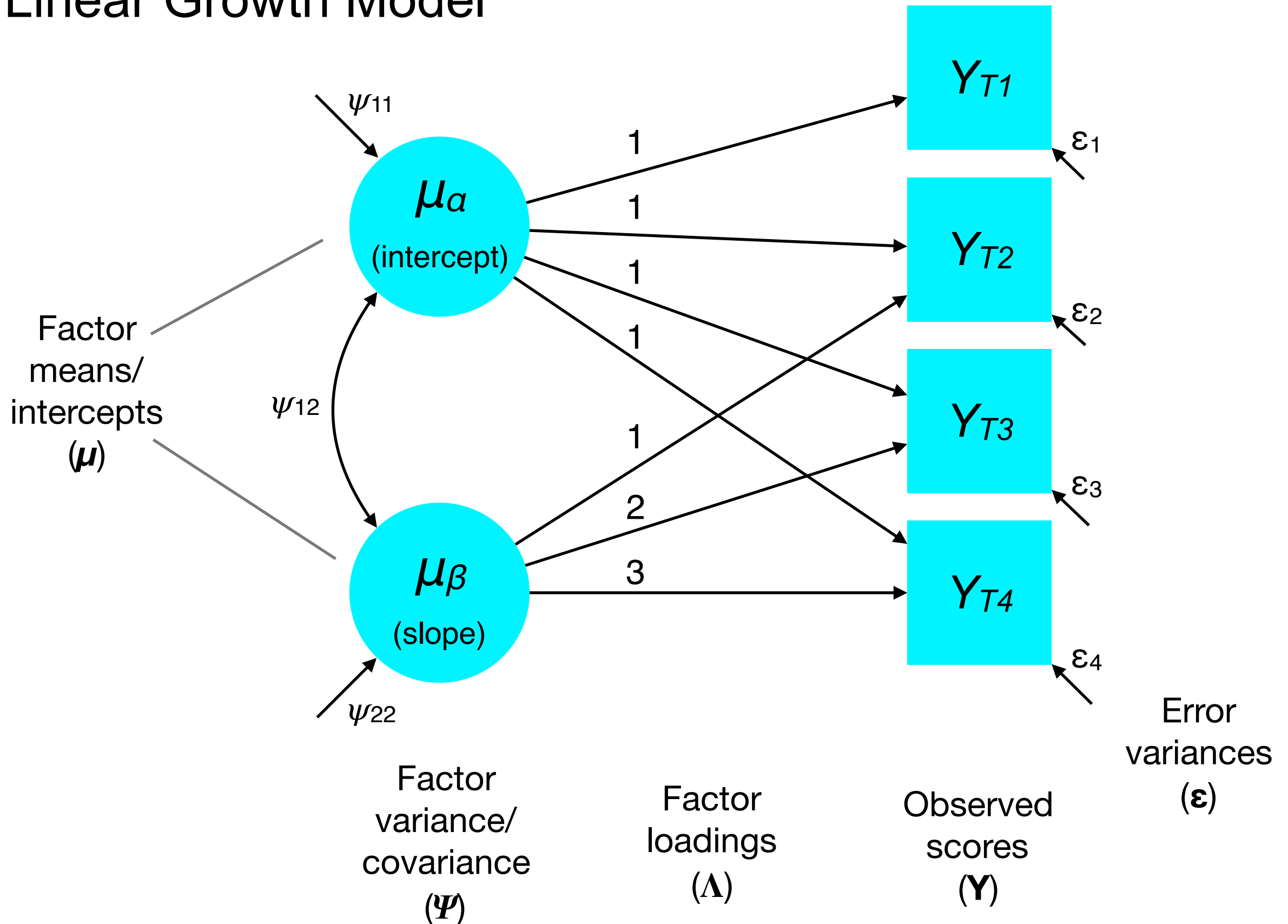
University of North Carolina
Department of Psychology
Chapel Hill, North Carolina

- Forms of change
- Mean-level
- rank-order
- ipsative

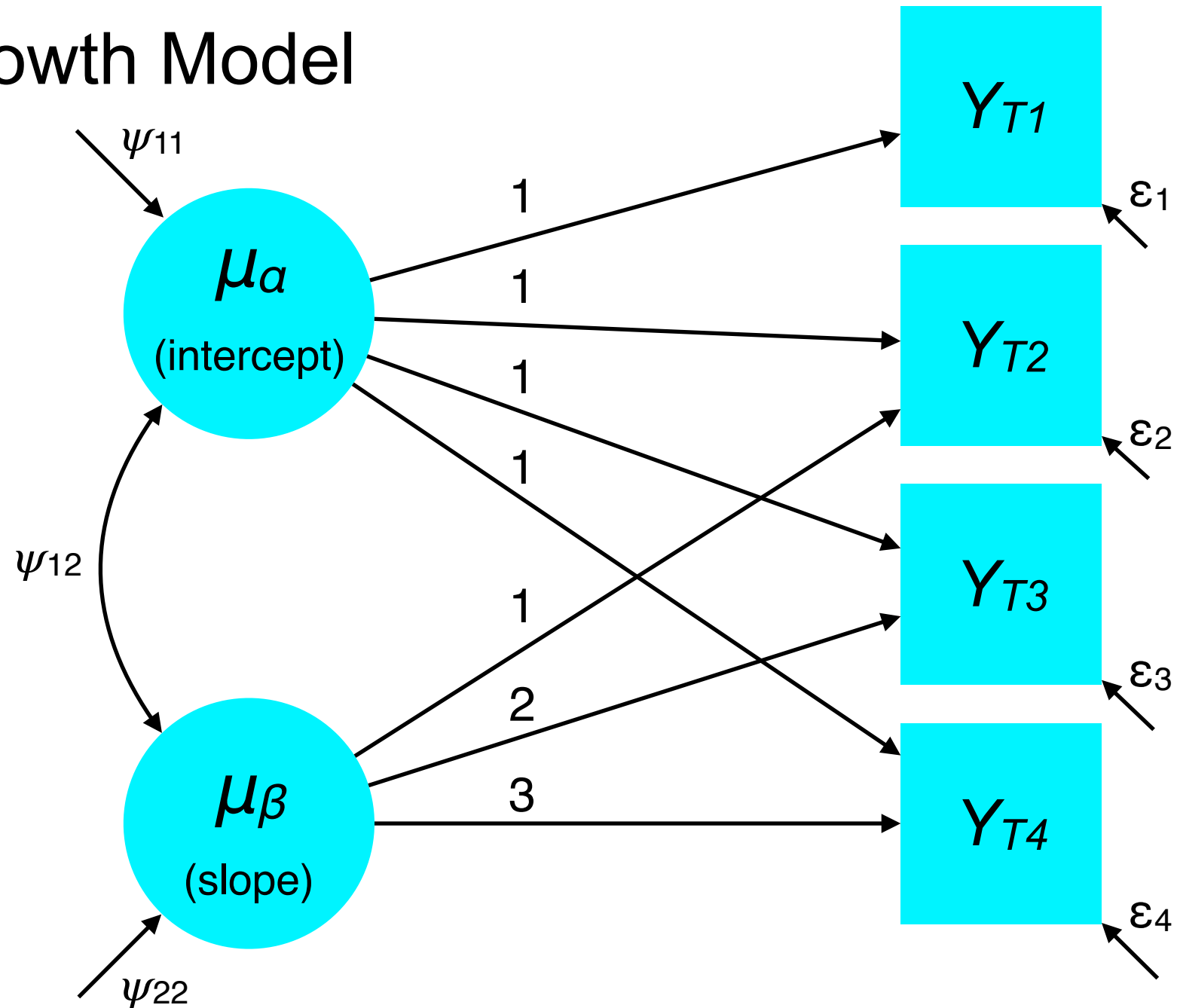
Ergodicity

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

Linear Growth Model



Linear Growth Model



$$y_{i,T1} = \alpha_i + \beta_i \lambda_{T1} + \epsilon_{i,T1}$$

$$y_{i,T2} = \alpha_i + \beta_i \lambda_{T2} + \epsilon_{i,T2}$$

$$y_{i,T3} = \alpha_i + \beta_i \lambda_{T3} + \epsilon_{i,T3}$$

$$y_{i,T4} = \alpha_i + \beta_i \lambda_{T4} + \epsilon_{i,T4}$$

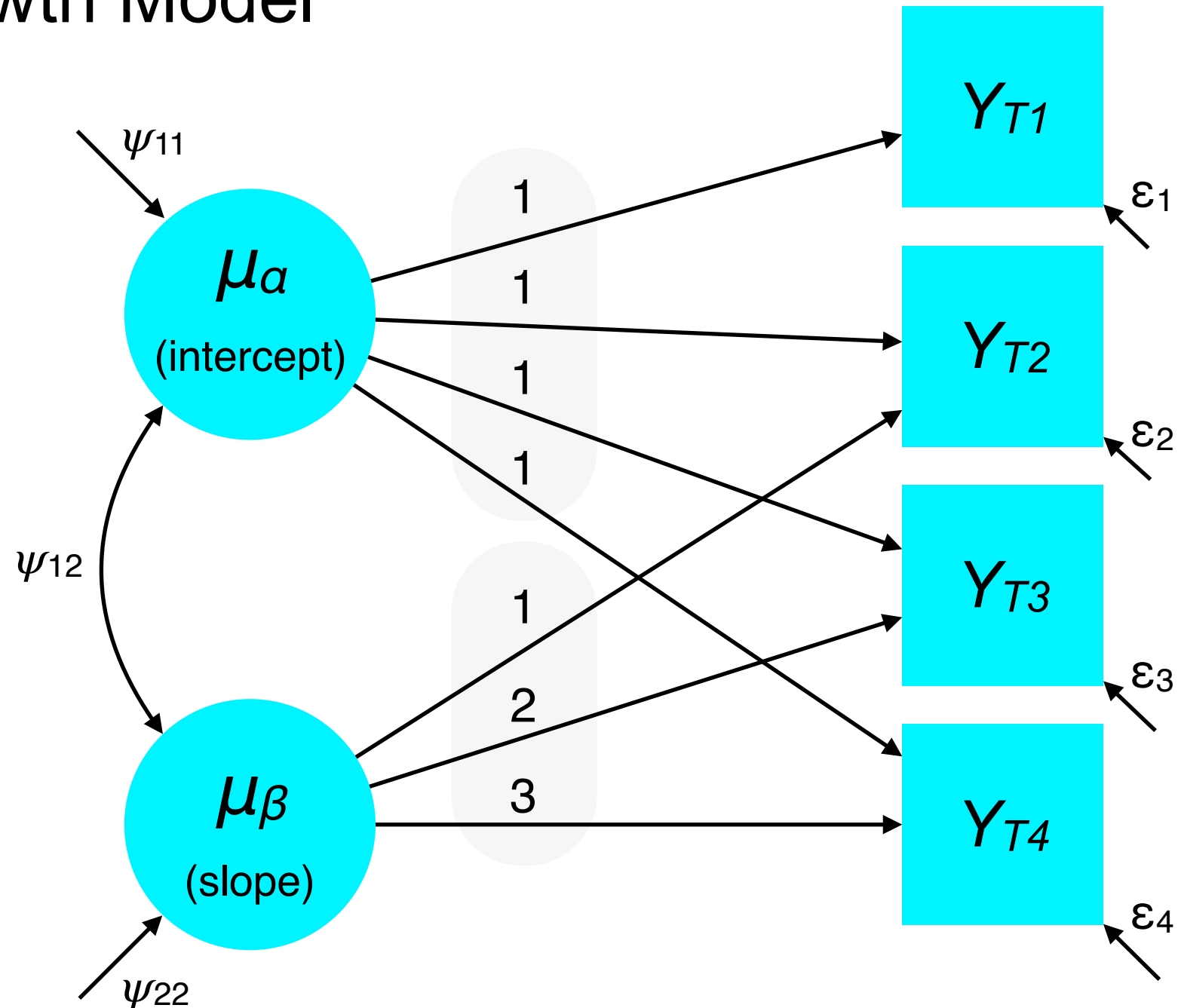
$$\alpha_i = \mu_\alpha + \zeta_{\alpha_i}$$

$$\beta_i = \mu_\beta + \zeta_{\beta_i}$$

$$Var(\zeta_\alpha) = \psi_{11}$$

$$Var(\zeta_\beta) = \psi_{22}$$

Linear Growth Model



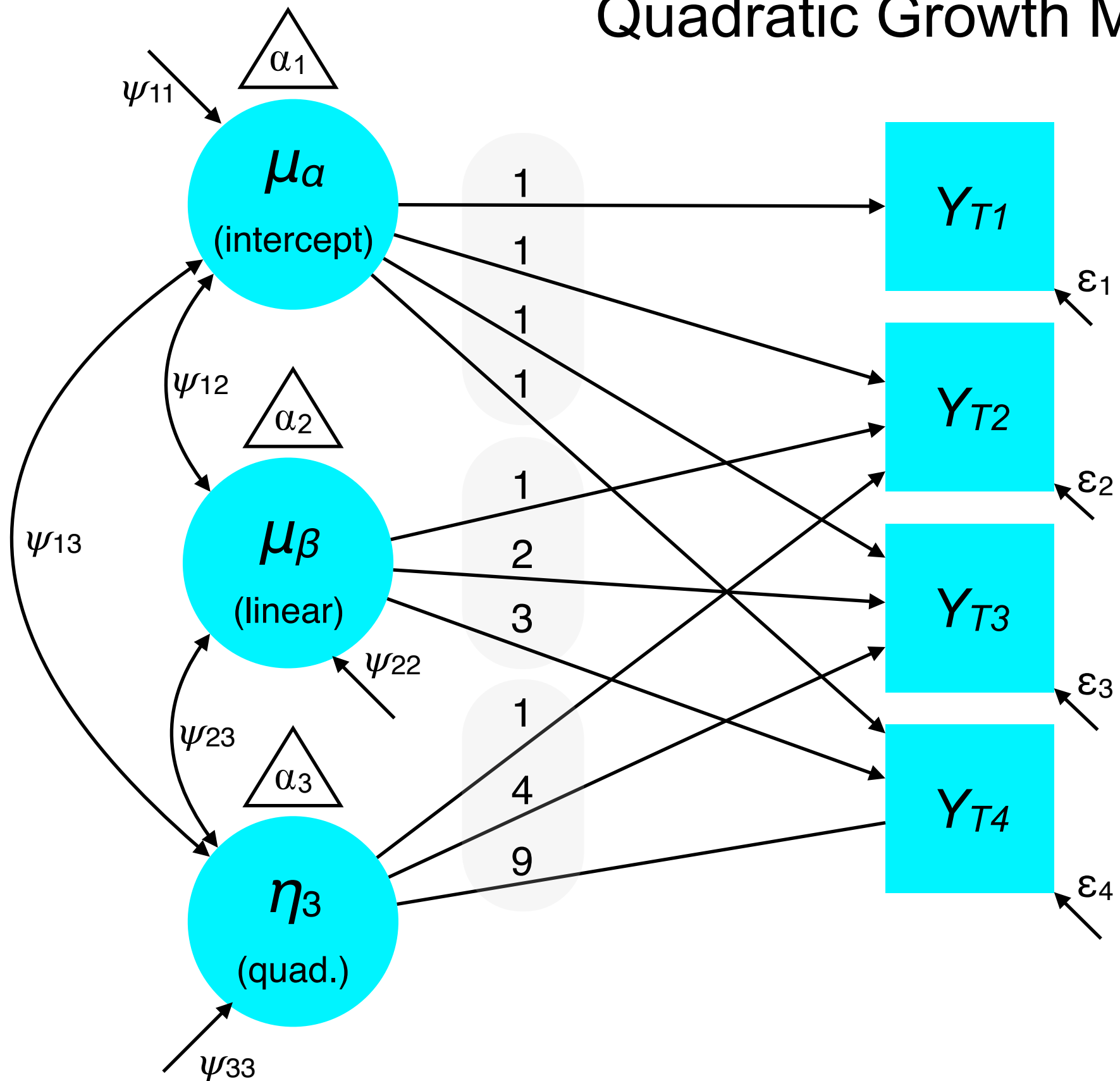
Fixed loadings denote:

a) functional form of change

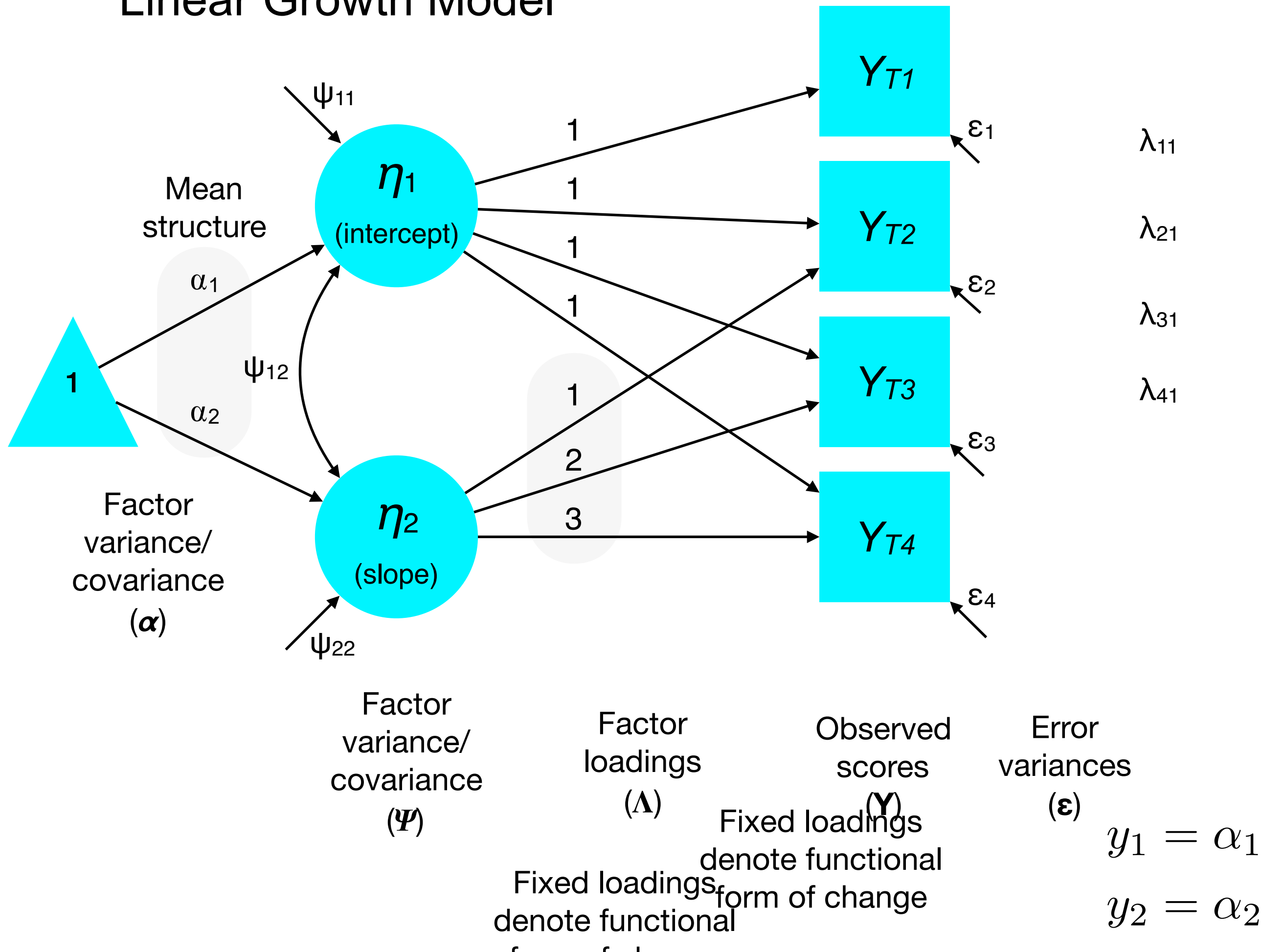
and

b) the spacing between observations

Quadratic Growth Model



Linear Growth Model



Organize wrt age, not wave

Points around missingness, covariance coverage

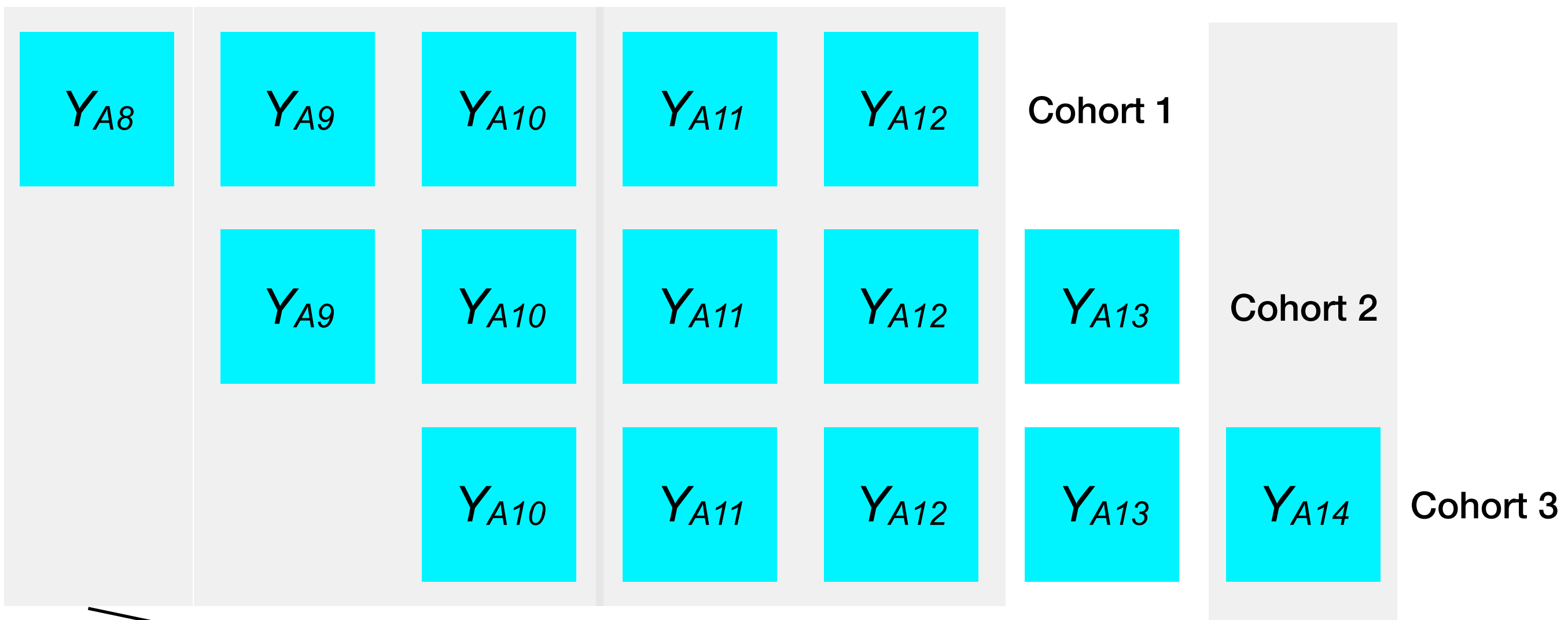
Accelerated longitudinal design, but only works if there are many more waves than there are age *cohorts*.

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



Coverage (A9, A10) = 0.66

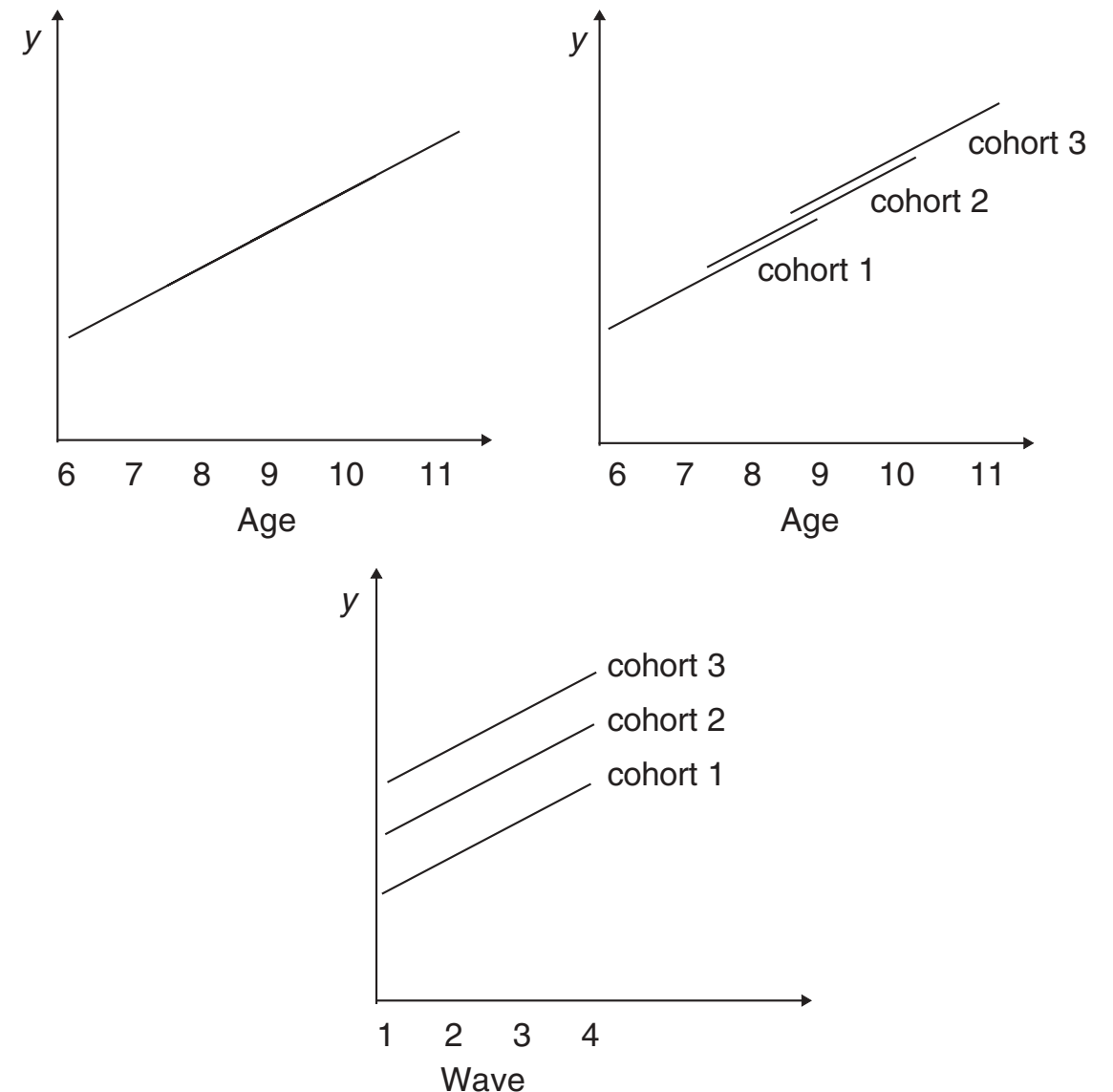
Coverage (A10, A11) = 1.0

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 **Θ** matrix (residual covariance structure). (i.e., a diagonal matrix)

$$\boldsymbol{\theta} = \begin{bmatrix} \varepsilon_1 & & & \\ 0 & \varepsilon_2 & & \\ 0 & 0 & \varepsilon_3 & \\ 0 & 0 & 0 & \varepsilon_4 \end{bmatrix}$$

- 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

Advantages of SEM

- Flexible latent basis
- Ability to constrain individual parameters to enforce a specific hypothesis about the form of developmental change (e.g., AR(1))
-

Disadvantages of SEM

- Conventionally, assume similar/identical spacing of observations.
- Can overcome this. Mplus TSCORES

