Latent Growth Curve Modeling

Neba Nfonsang
University of Denver

Introduction to Growth Curve Modeling (LGCM)

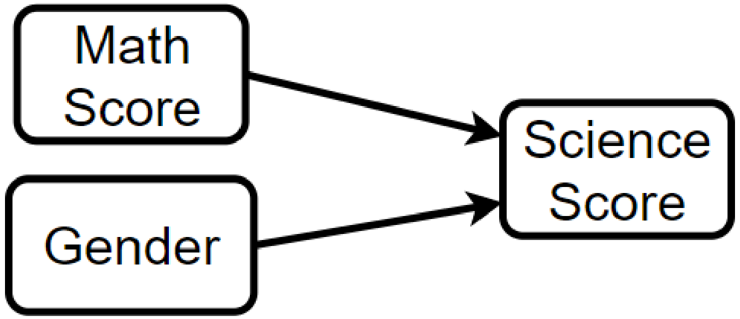
- There are lot of terms used to refer to the same underlying technique such as:
  - Growth curve modeling
  - Trajectory analysis
  - Latent curve modeling
  - Latent trajectory analysis
- How do we analyze repeated measure data?
- Traditional methods of analyzing repeated measure or time series data.
- First let's look at how two-time point data has been analyzed in the past.

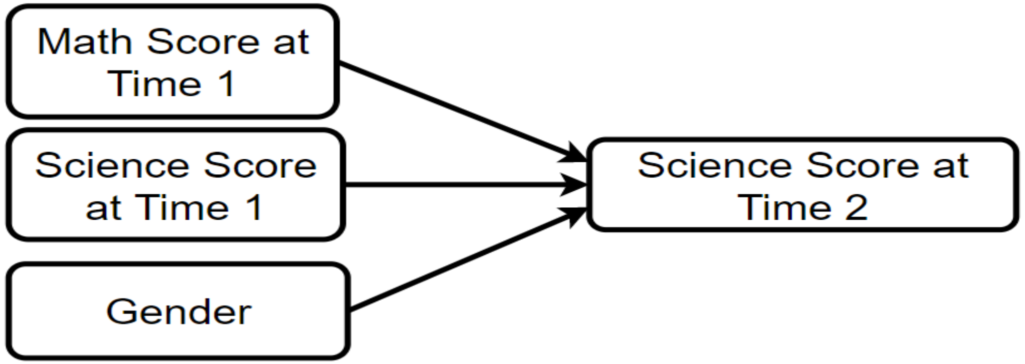
Traditional Approaches to LGCM

- Two approaches that have been used to analyze two-time point data (these methods can be extended to more than two-time point data):
  - **Raw change score:** if you have a pre and post test design, you find the difference between the pre and post test to become the unit of analysis. If there are two-time points, a t-test can be used. If there are more than two-time points, repeated ANOVA framework can be used where you are taking the difference between point 1 and 2, 2 and 3, 3 and 4.

Traditional Approaches to LGCM

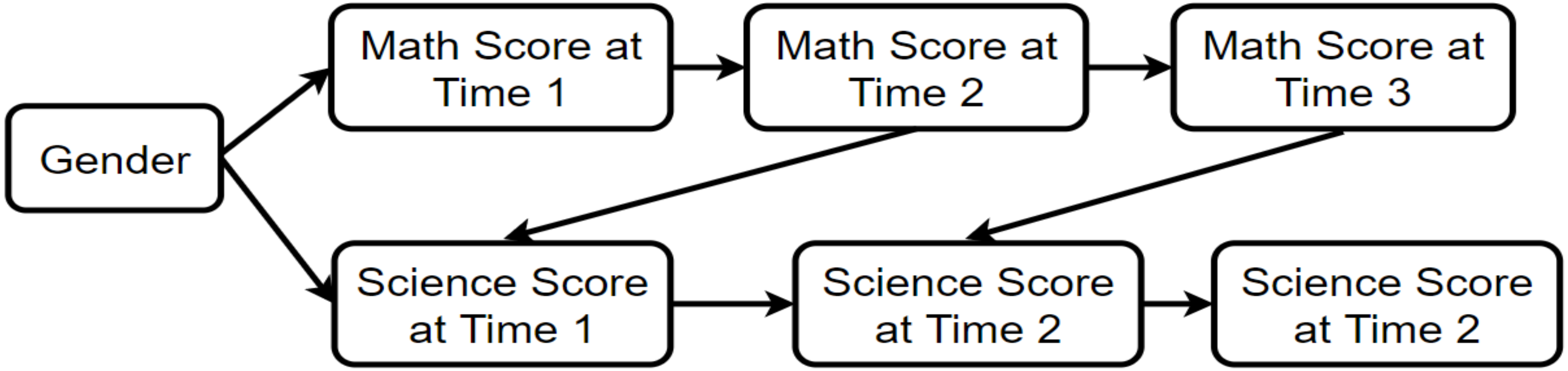
- **Residualized Change Model (or two time point regression model):** In this case, the pretest is used to predict the post test and what is left over which is the residual becomes the unit of analysis and you will see this in auto regressive models. In a cross-sectional regression, the variables are not time points, so this is limited in that it cannot establish temporal precedence.





Autoregressive Path Analytic Model

- Models time-adjacent observations with a set of regression models.
- Does not allow for the estimation of growth trajectories.



Autoregressive Path Models

- Advantages:
  - estimation of paths is straight forward
  - Overall fit is tested and direct and indirect effects as well
- Disadvantage:
  - Only handles change in two time point or a series of two-time points.
  - Assumes that observations are measured without error

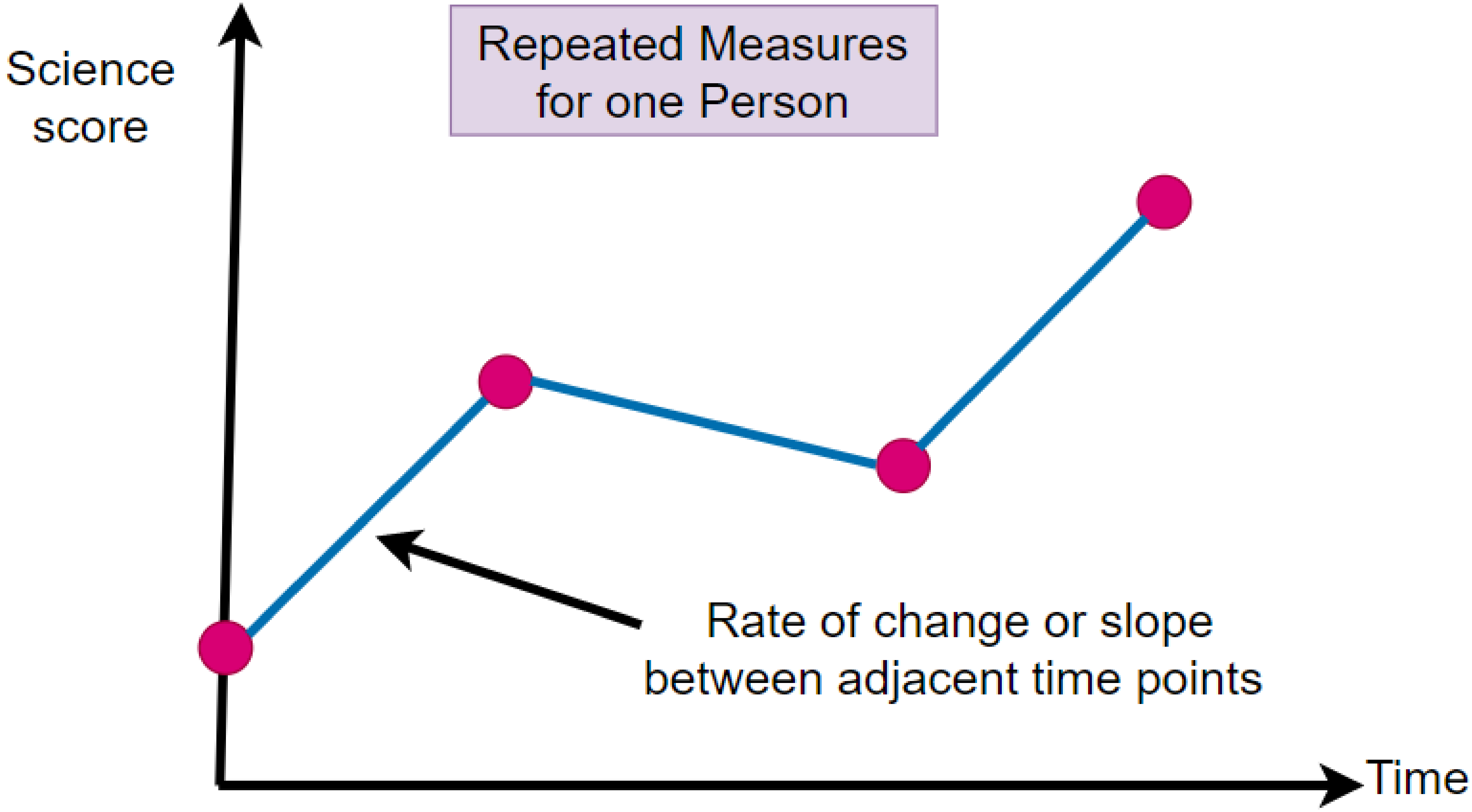
Repeated Measures ANOVA

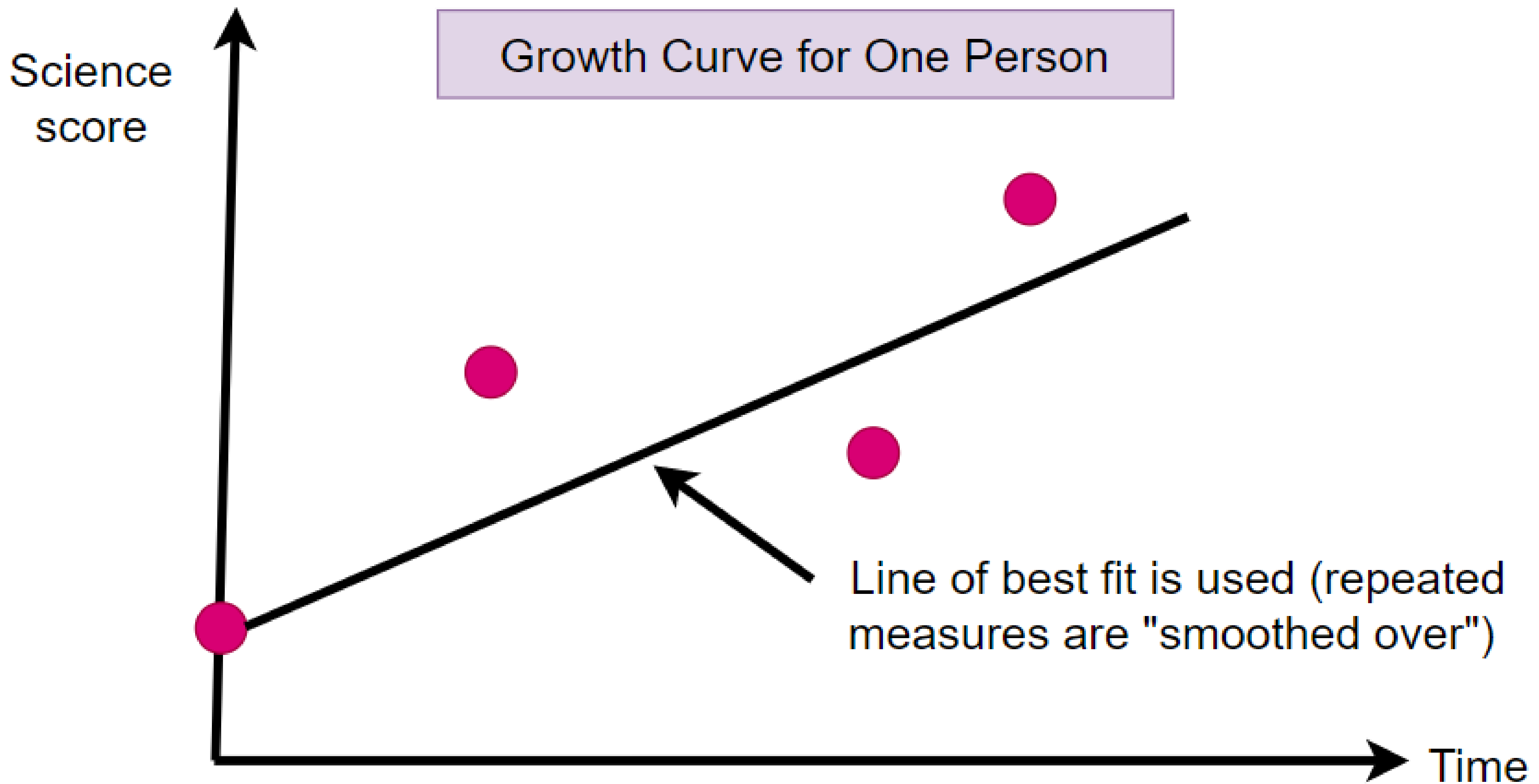
- Focuses on changes in means over time.
- Captures change in group or mean scores between adjacent time points.
- Not concerned with individual change.

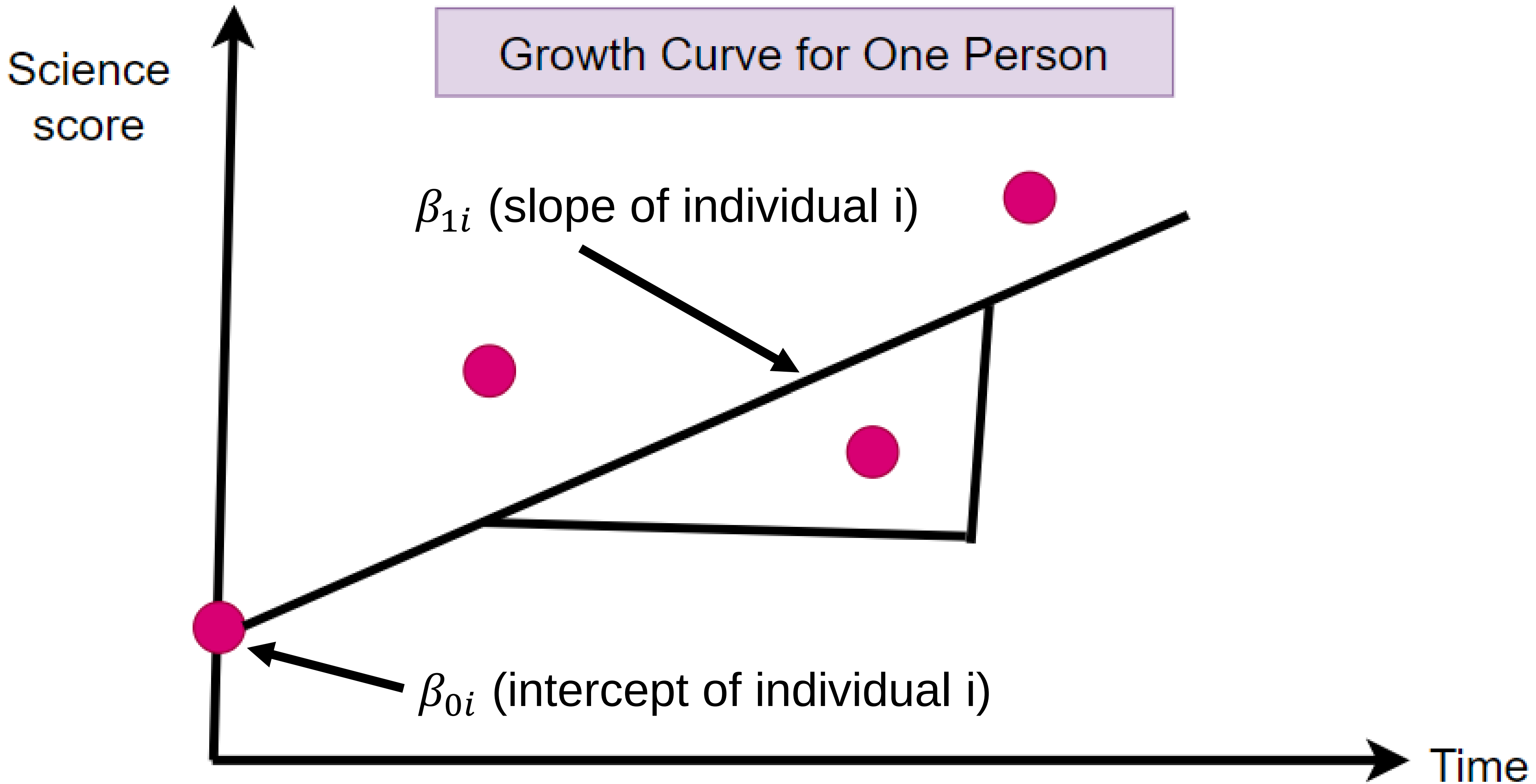
- Latent growth curve is interested in capturing:
  - change within an individual (intra-individual change)
  - differences between individuals (inter-individual change).

What is Growth Curve Modeling?

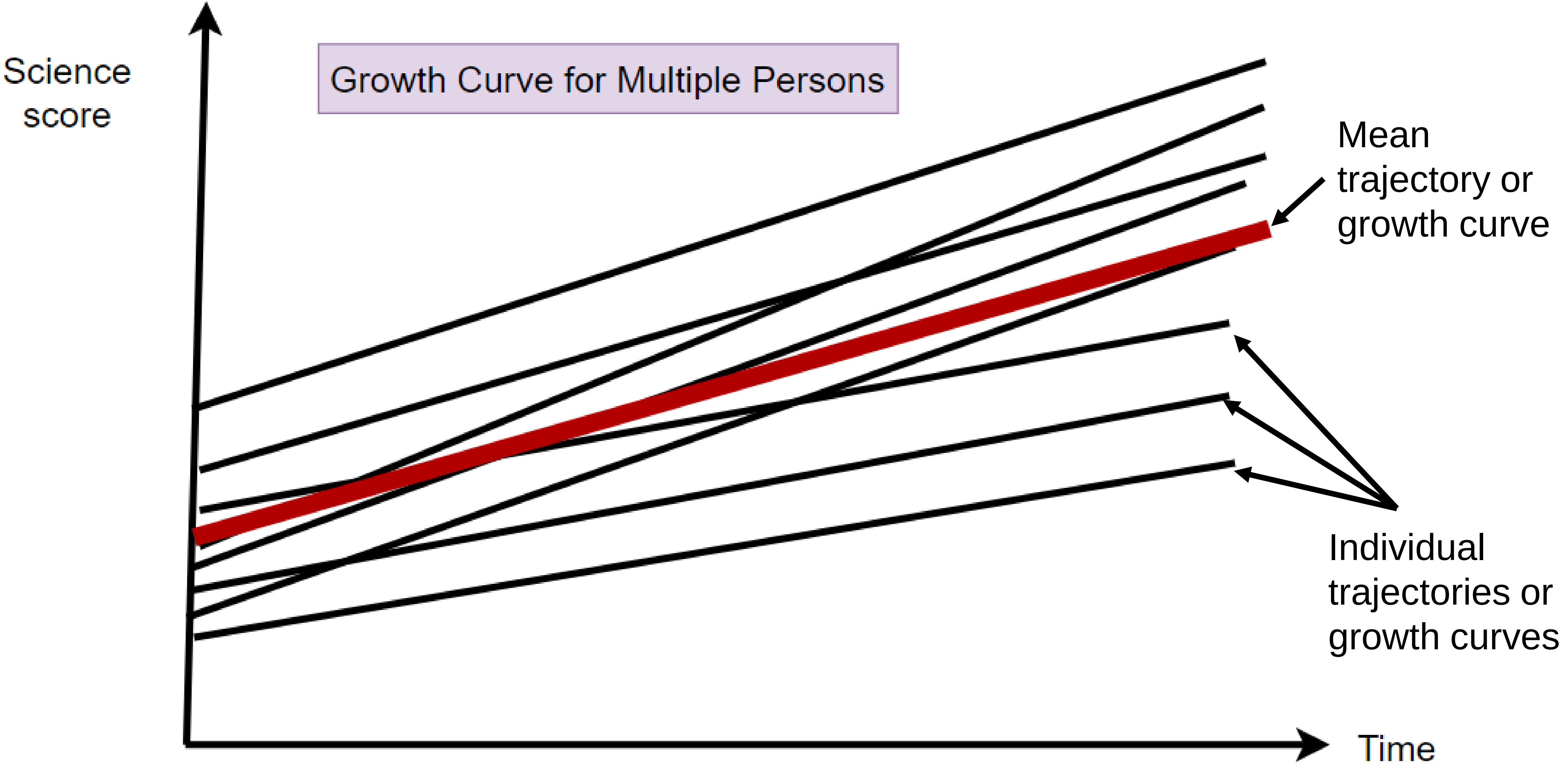
- A model for data that estimates change within each individual and compares change across individuals.
- That is, LGCM estimate inter-individual variability and intra-individual change (rate of change) over time.
- Latent growth curves capture continuous change over time.
- In repeated measure ANOVA, time is treated as a categorical variable but in latent growth curve modeling, time is treated as a continuous variable.


Science score
Repeated Measures for one Person
Rate of change or slope between adjacent time points
Time


Growth Curve for One Person
Science score
Line of best fit is used (repeated measures are "smoothed over")
Time

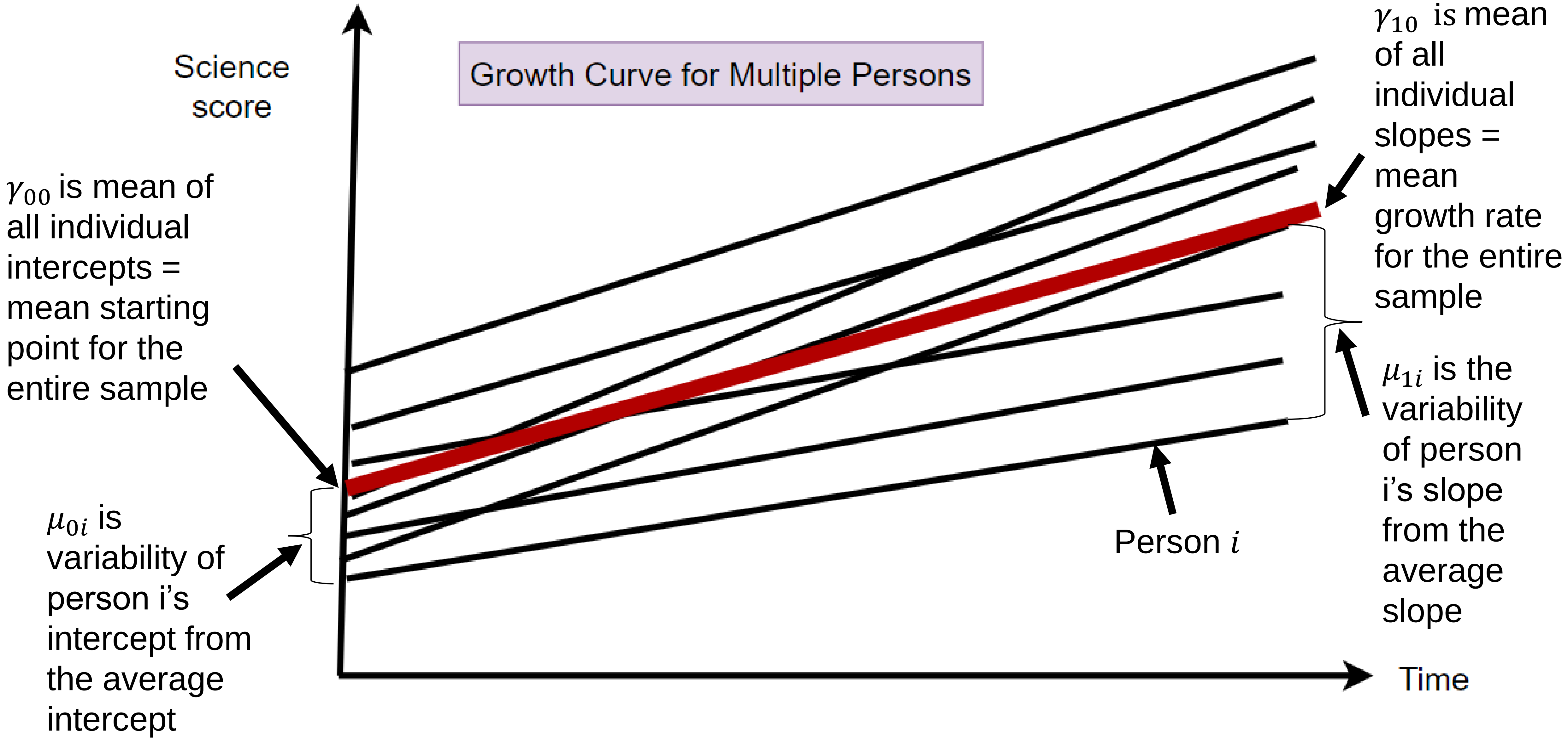


Growth Curve for One Person
Science score
β_{1i} (slope of individual i)
β_{0i} (intercept of individual i)
Time


Science score
Growth Curve for Multiple Persons
Mean trajectory or growth curve
Individual trajectories or growth curves
Time

The Latent Growth Curve

- Latent growth curves are characterized or can be summarized using trajectory means or trajectory variances
- Trajectory means:
  - Mean of intercepts
  - Mean of slopes
- Trajectory variance
  - Variance of trajectory intercepts
  - Variance of trajectory slopes
- A trajectory intercept indicates the starting score (at first time point or time point 0)
- A trajectory slope indicates change in score per unit change in time (growth rate).

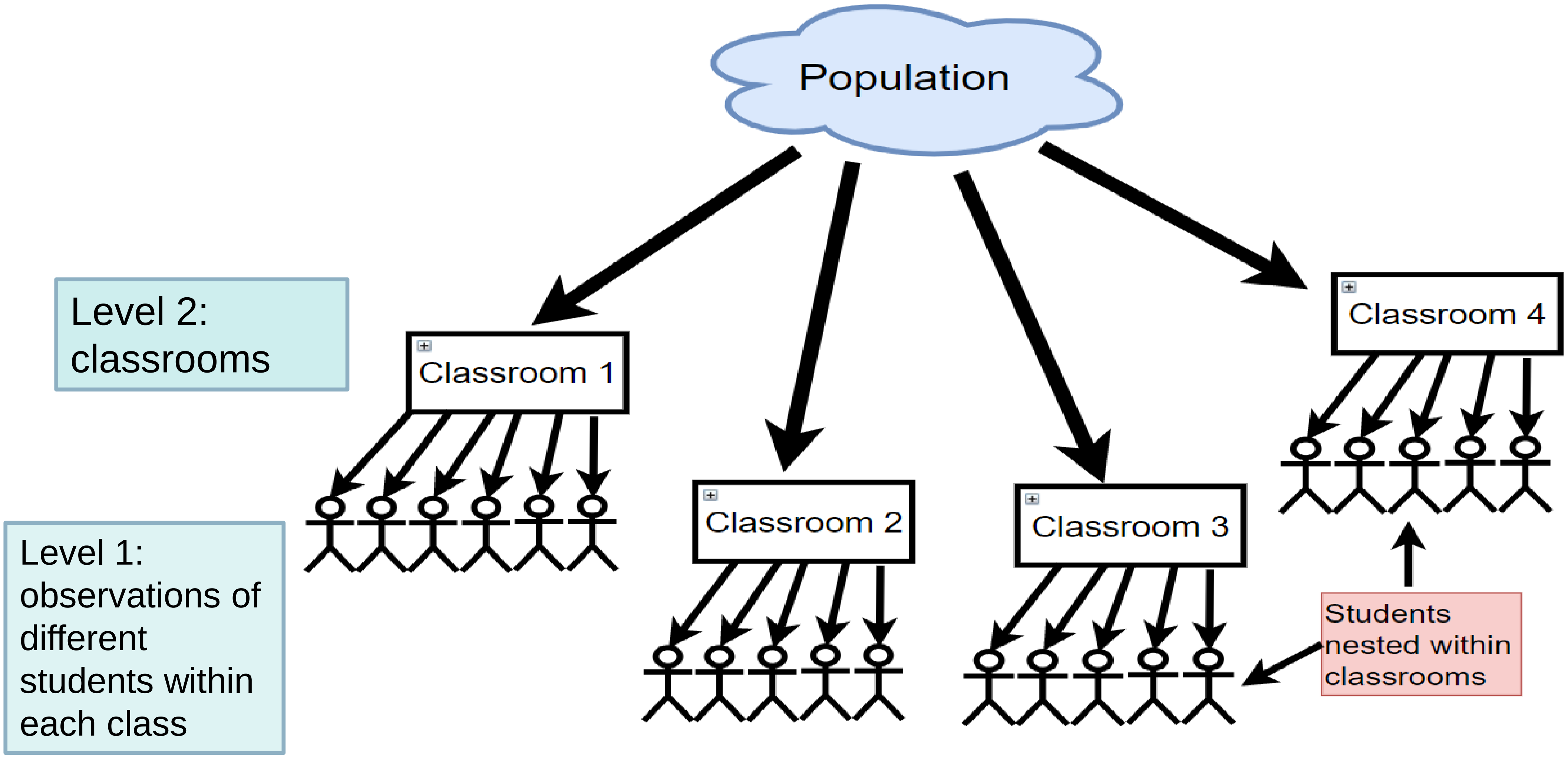


Growth Curve for Multiple Persons
Science score
Time
γ_{00} is mean of all individual intercepts = mean starting point for the entire sample
μ_{0i} is variability of person i’s intercept from the average intercept
γ_{10} is mean of all individual slopes = mean growth rate for the entire sample
μ_{1i} is the variability of person i’s slope from the average slope
Person i

Multilevel Modeling Framework

- Latent growth curves can be modeled in a multilevel modeling framework.
- Multilevel modeling involves data that is nested within units.
  - Students nested within a classroom
  - Repeated measures data nested within individuals

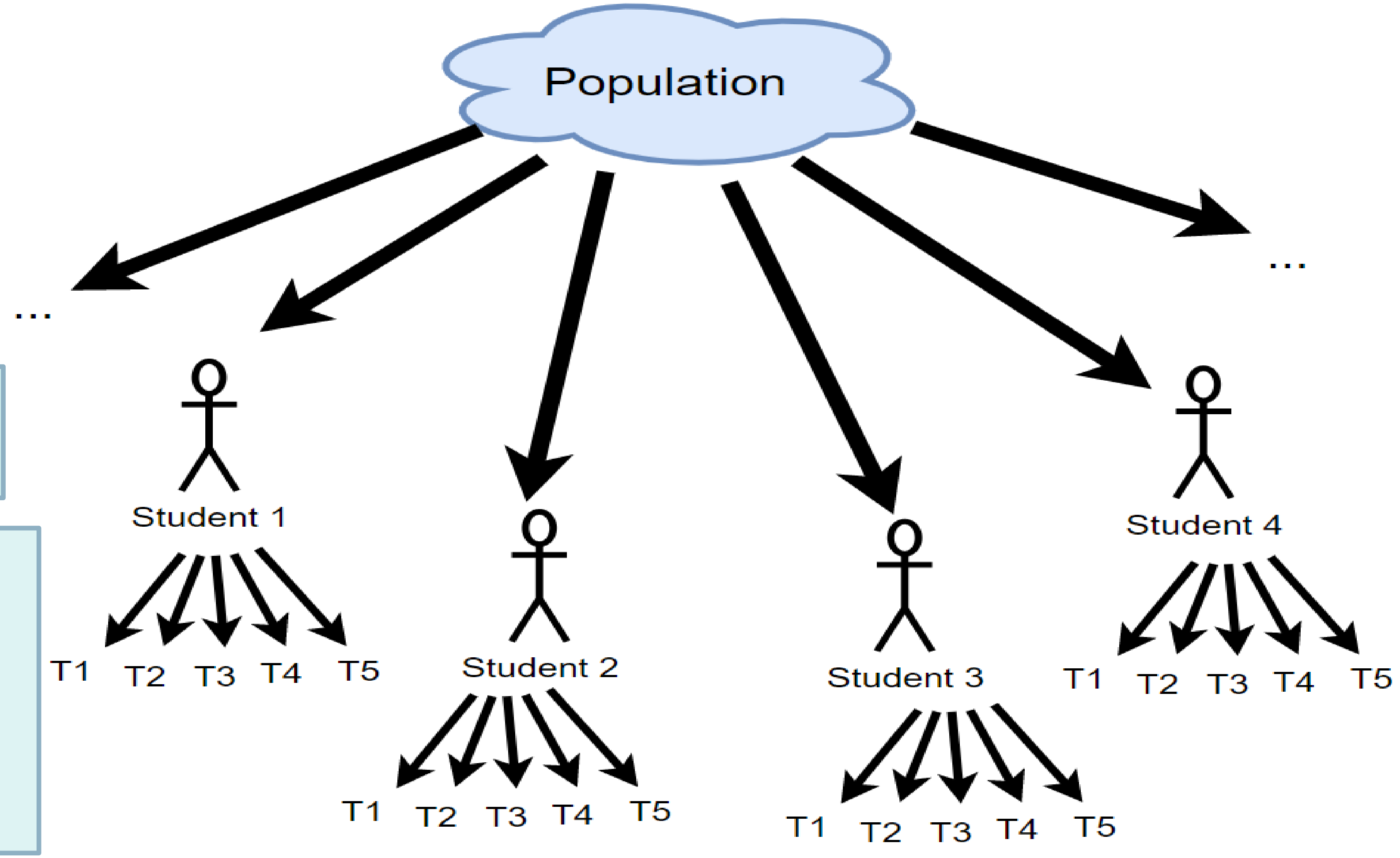
Multilevel Modeling Framework



Multilevel Modeling Framework

Level 1:
students

Level 1:
Observations
at different
time points
within each
student



Scores at
different time
points nested
within each
student

Multilevel Modeling Framework

- For students nested within classroom, the level 1 and 2 equations are as follows:
- Level 1:
  - $y_{ij} = \beta_{Oj} + \beta_{1j}x_{ij} + r_{ij}$
- Level 2
  - $\beta_{0j} = \gamma_{00} + u_{0j}$
  - $\beta_{1j} = \gamma_{10} + u_{1j}$

A level 2 or class variable could be added with its slope or impact on the class intercept or slope

- y_{ij} = observed outcome of individual i in class j.
- x_{ij} = attribute of student i in class j.
- β_{Oj} = intercept of class j or expected outcome of class j given x = 0.
- β_{1j} = slope of class j or change y_{ij} when x_{ij} changes by 1 unit.
- r_{ij} = unique effect of individual i on y_{ij}.
- γ_{00} = mean of intercept for all classes.
- γ_{10} = mean of slopes across classes.
- u_{0j} = unique effect of class j on β_{0j}.
- u_{1j} = unique effect of class j on β_{1j}.

Multilevel Modeling Framework

- For time series data nested within each student, level 1 and 2 equations are:
- Level 1 (within-person):
  - $y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + r_{ti}$
- Level 2 (between persons)
  - $\beta_{0i} = \gamma_{00} + u_{0i}$
  - $\beta_{1i} = \gamma_{10} + u_{1i}$

β_{0i} and β_{0i} could be allow to vary as a function the person characteristics

- y_{ti} = observed outcome at time point x_{ti} for individual i.
- x_{ti} = time point t
- β_{0i} = intercept or starting point for an individual i.
- β_{1i} = slope or growth rate of individual i (on average).
- r_{ti} = residual of individual i at time t.
- γ_{00} = mean of all intercepts or mean starting point across all individuals.
- γ_{10} = mean of all slopes or mean growth rate across all individuals.
- μ_{0i} = random (unique) effect of individual i on β_{0i}.
- μ_{1i} = random (unique) effect of individual i on β_{1i}.

Multilevel Modeling Framework

- Level 1 (within-person):
  - $y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + r_{ti}$
- Level 2 (between persons)
  - $\beta_{0i} = \gamma_{00} + u_{0i}$
  - $\beta_{1i} = \gamma_{10} + u_{1i}$
- Combined Equation
  - $y_{ti} = \gamma_{00} + u_{0i} + (\gamma_{10} + u_{1i})x_{ti} + r_{ti}$
  - $y_{ti} = \underbrace{\gamma_{00} + \gamma_{10} * x_{ti}}_{\text{Fixed effects}} + \underbrace{u_{0i} + + u_{1i} * x_{ti}}_{\text{Random effects}} + \underbrace{r_{ti}}_{\text{Residual or random error}}$

This is an unconditional model where there is no predictor

Variance Component

- The random effects are assumed to be normally distributed with mean of 0 and variances tau's as shown :
- $r_{ti} \sim N(0, \delta^2)$
- $\begin{pmatrix} u_{oi} \\ u_{1i} \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$

- δ^2 the population residual variance
- τ_{00} is the variance of the intercepts β_{0i} around $the\ mean\ intercept\ \gamma_{00}$
- τ_{10} is the variance of the slope β_{1i} of individual trajectories around $the\ mean\ slope\ \gamma_{10}$.
- τ_{11} is the covariance between the intercepts and slopes of individuals.

The Goal of Growth Curve Modeling

- **To estimate the parameters of the trajectories:**
  - β_{0i} = intercept or starting point for an individual i.
  - β_{1i} = slope or growth rate of individual i (on average).
  - γ_{00} = mean of all intercepts or mean starting point across all individuals.
  - δ^2 the population error variance
  - τ_{00} is the variance of the intercepts β_{0i} around $the\ mean\ intercept\ \gamma_{00}$
  - τ_{10} is the variance of the slope β_{1i} of individual trajectories around $the\ mean\ slope\ \gamma_{10}$.
  - τ_{11} is the covariance between the intercepts and slopes of individuals.

Fixed Effects and Random Effects

- The fixed effects are constant in the population and capture the mean structure (mean of intercepts and mean of slopes across units, denoted by gamma)

- Random effects vary across units in the population, denoted by u.
- Random effect is the deviation of individuals from the mean of the sample.

Research Questions For LGCM

- **Research questions that find estimate of intercept and slopes:**
  - What is the mean growth rate of an individual? β_{0i}
  - What is the initial status of an individual? β_{1i}
- **Research questions that find estimates of fixed effects (mean intercept or mean slope):**
  - What is the mean initial status of all the individuals γ_{00} ?
  - What is the mean growth rate across all individual trajectories, γ_{10} ?

Research Questions For LGCM

- **Research questions that find estimates of random effects (variance of intercept and slope):**
  - Is there a significant variation in the initial status or starting point of individuals, $\tau_{00} \; = 0$?
  - Is there a significant variation in the growth rate across individuals, $\tau_{10} \; = 0$?
  - Do the individuals with higher growth rate have a higher starting point, $\tau_{11} = 0$? (Is there a significant correlation between the intercepts and the slopes)

Structural Equation Modeling Framework

- The structural equation framework is a flexible and powerful multivariate methodology where we can run different type of test including t-test, ANOVA, multiple regression, and complex mediation, multiple group analysis, factor analysis where we can find relationship between latent factors.
- The emphasis is usually on the model as a whole. So there can be more than one outcome in a structural equation framework.

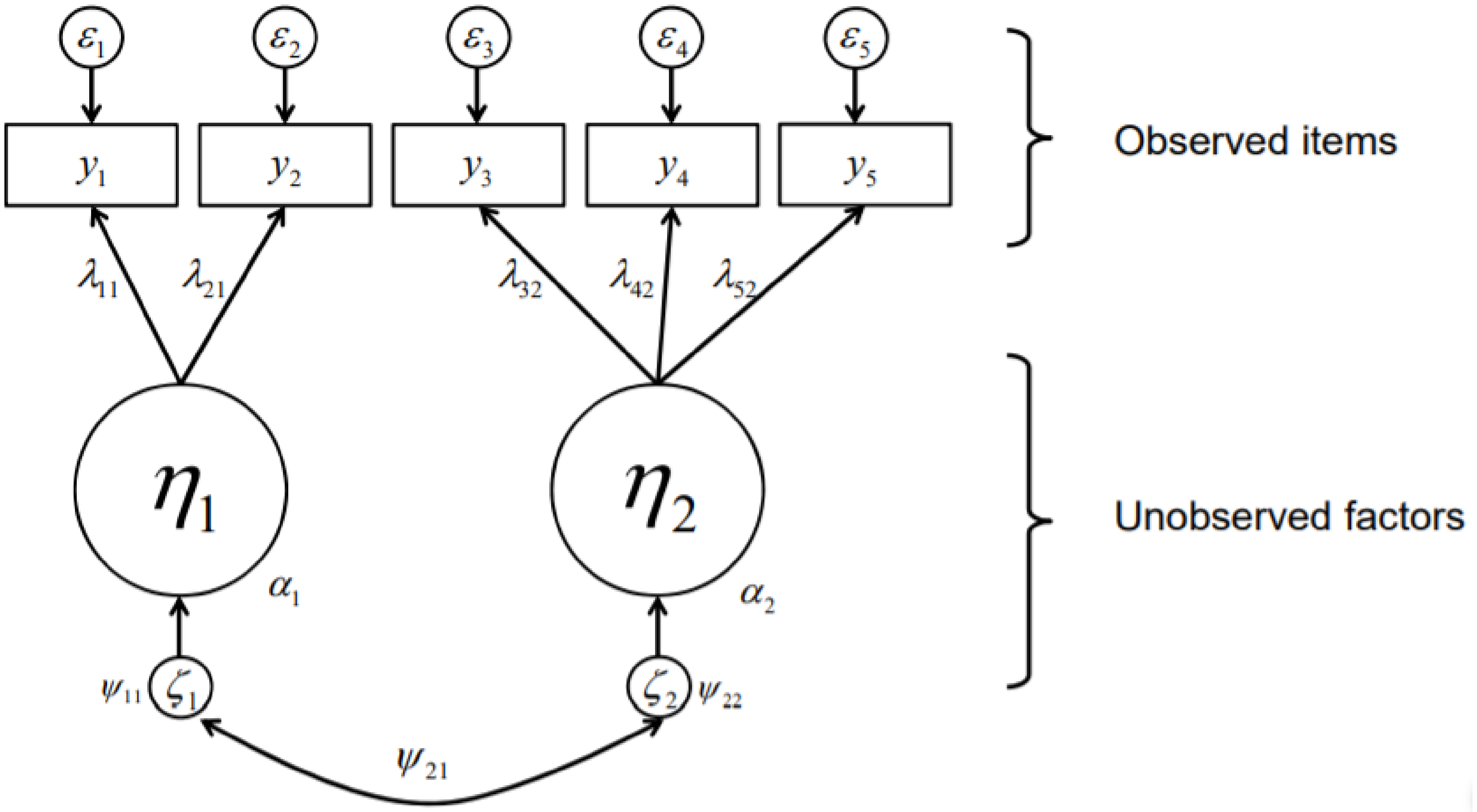
Structural Equation Modeling Framework

- LCM is a special case of confirmatory factor analysis (CFA).
- Factor analysis assume that the covariance among a set of observed variables is due to an underlying Factor.
- That means, one or more latent factors explain the covariance among a set of variables or influence a set of observed variables (items).

Structural Equation Modeling Framework

- Each item has a mean, and a residual. The residual variance is the variance of the item.
- Each factor has a factor mean and a factor variance.
- CFA test a model that specifies the number and structure of latent factors behind a set of observed measures.
- The model specification is determined by theory
- Local and global fit of model to data is given attention.

Structural Equation Modeling Framework



The general structure of a CFA model.

Structural Equation Modeling Framework

- Growth can be captured as a latent factor:
  - The growth model is unobserved and we can model that using observed repeated observations.
  - There is a latent trajectory for each individual that can be modeled or inferred or estimated using the observed repeated measures.
  - The time points are fixed in the model and serve as observed variables (items).
  - Note in the multilevel modeling, we had: $y_{ti} = \beta_{Oi} + \beta_{1i}x_{ti} + r_{ti}$
  - So, the mean intercept and mean slope would be factors that affect all the observations at different time points.

Structural Equation Modeling Framework

Linear LCM

Level 2: Latent Growth Factor

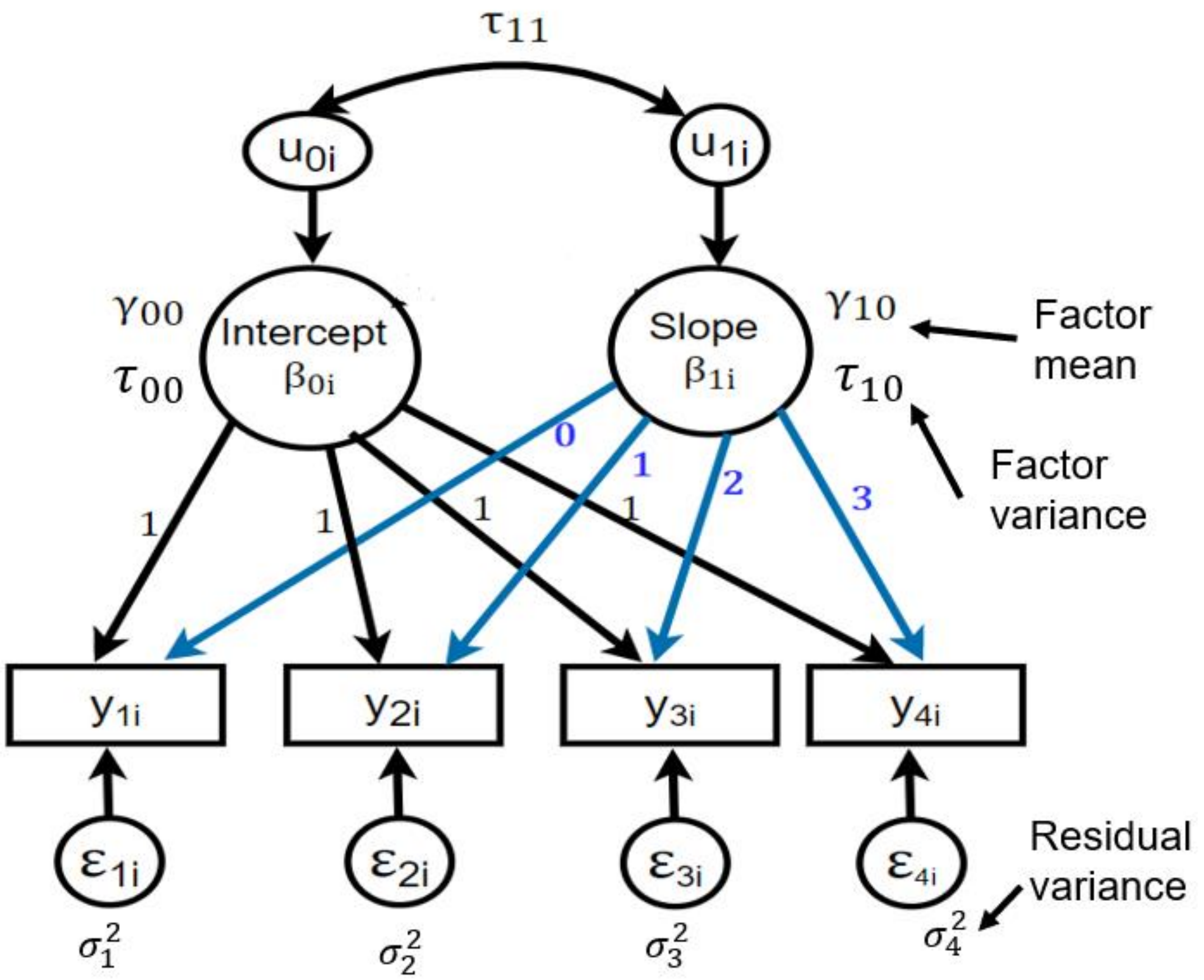
$\beta_{0i} = \gamma_{00} + u_{0i}$

$\beta_{1i} = \gamma_{10} + u_{1i}$

Level 1: Repeated Measures

$y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + \varepsilon_{ti}$

Note: ε_{ti} $(error\ term)$ is same as $r_{ti}(residual)$



Structural Equation Modeling Framework

Linear LCM

Level 2:

$$\beta_{0i} = \gamma_{00} + u_{0i}$$

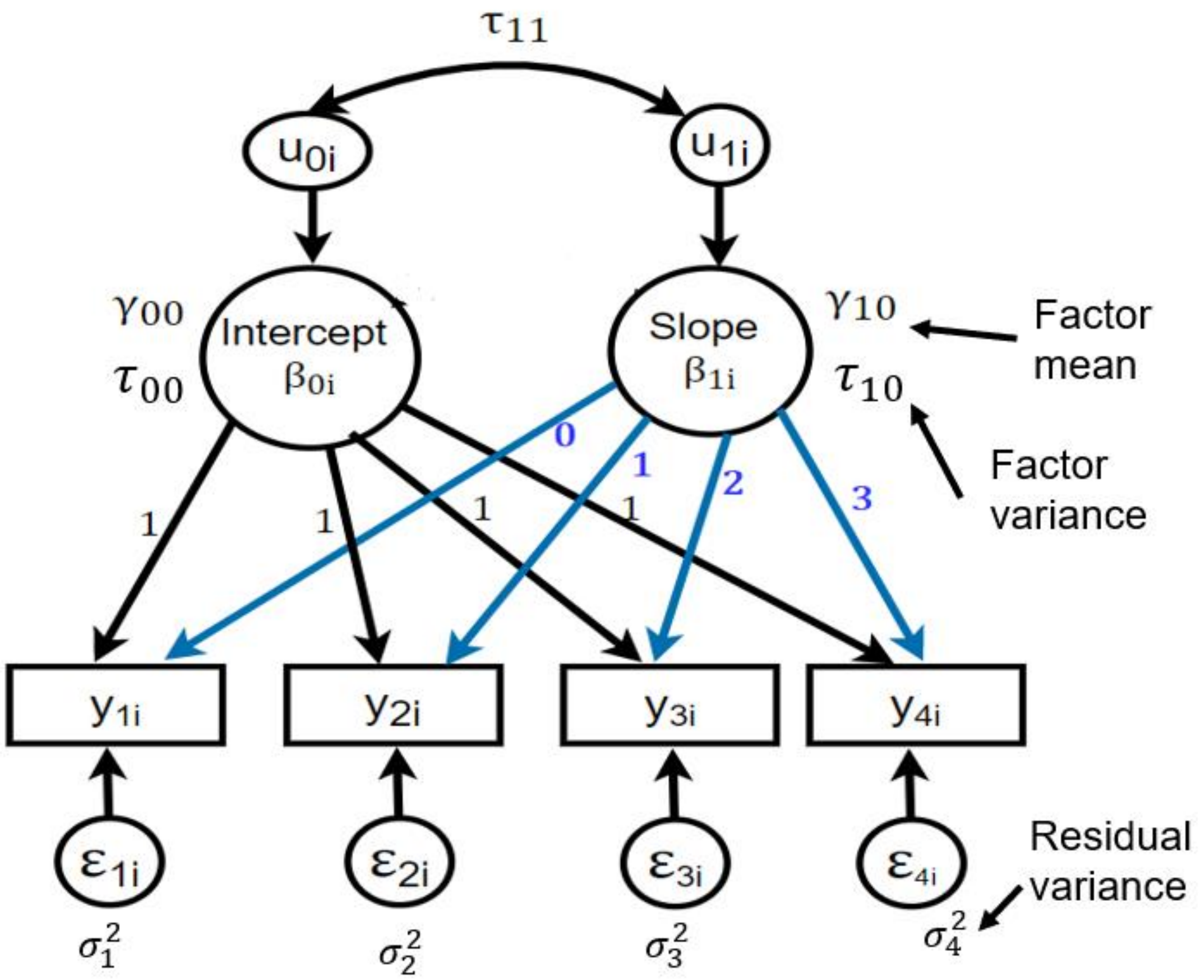
$$\begin{bmatrix}\beta_{0i}\\ \beta_{1i}\end{bmatrix} = \begin{bmatrix}\gamma_{00}\\ \gamma_{10}\end{bmatrix} + \begin{bmatrix}u_{0i}\\ u_{1i}\end{bmatrix}$$

Factor loading matrix

Level 1:

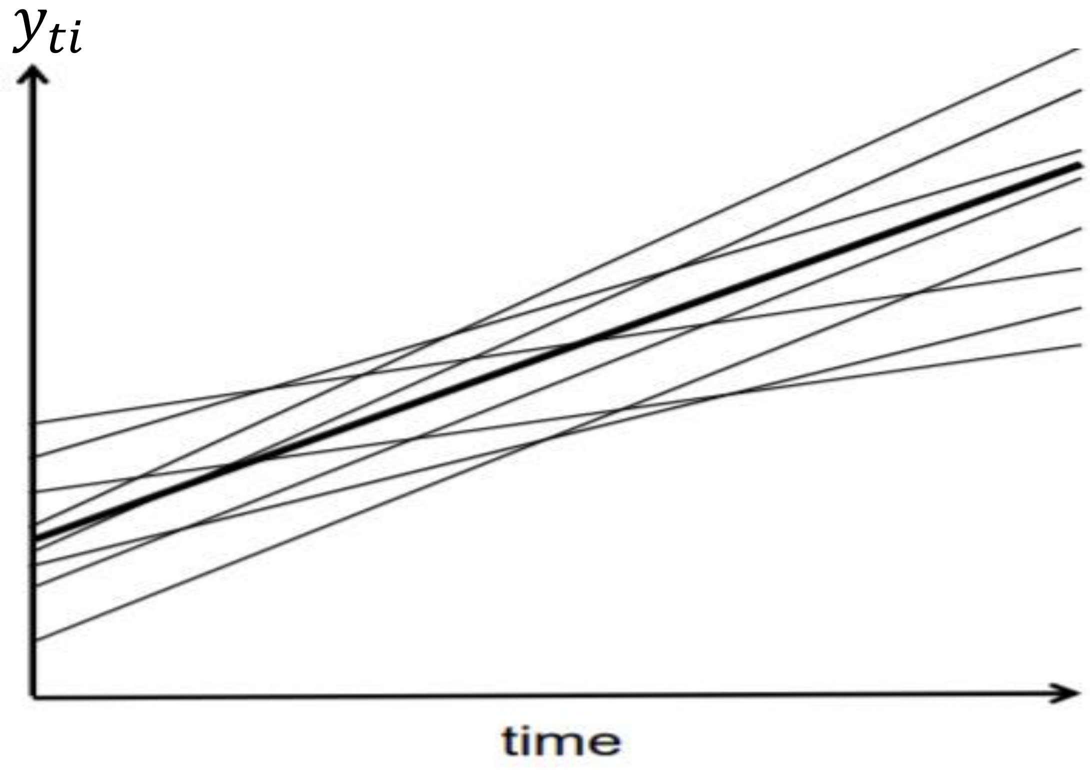
$$y_{ti} = \Lambda\beta_{0i} + \varepsilon_i$$

$$\begin{bmatrix}y_{1i}\\ y_{2i}\\ y_{3i}\\ y_{4i}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix}\begin{bmatrix}\beta_{0i}\\ \beta_{1i}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\\ \varepsilon_{3i}\\ \varepsilon_{4i}\end{bmatrix}$$



Structural Equation Modeling Framework

Linear LCM (Intercept and Linear Slope Model)

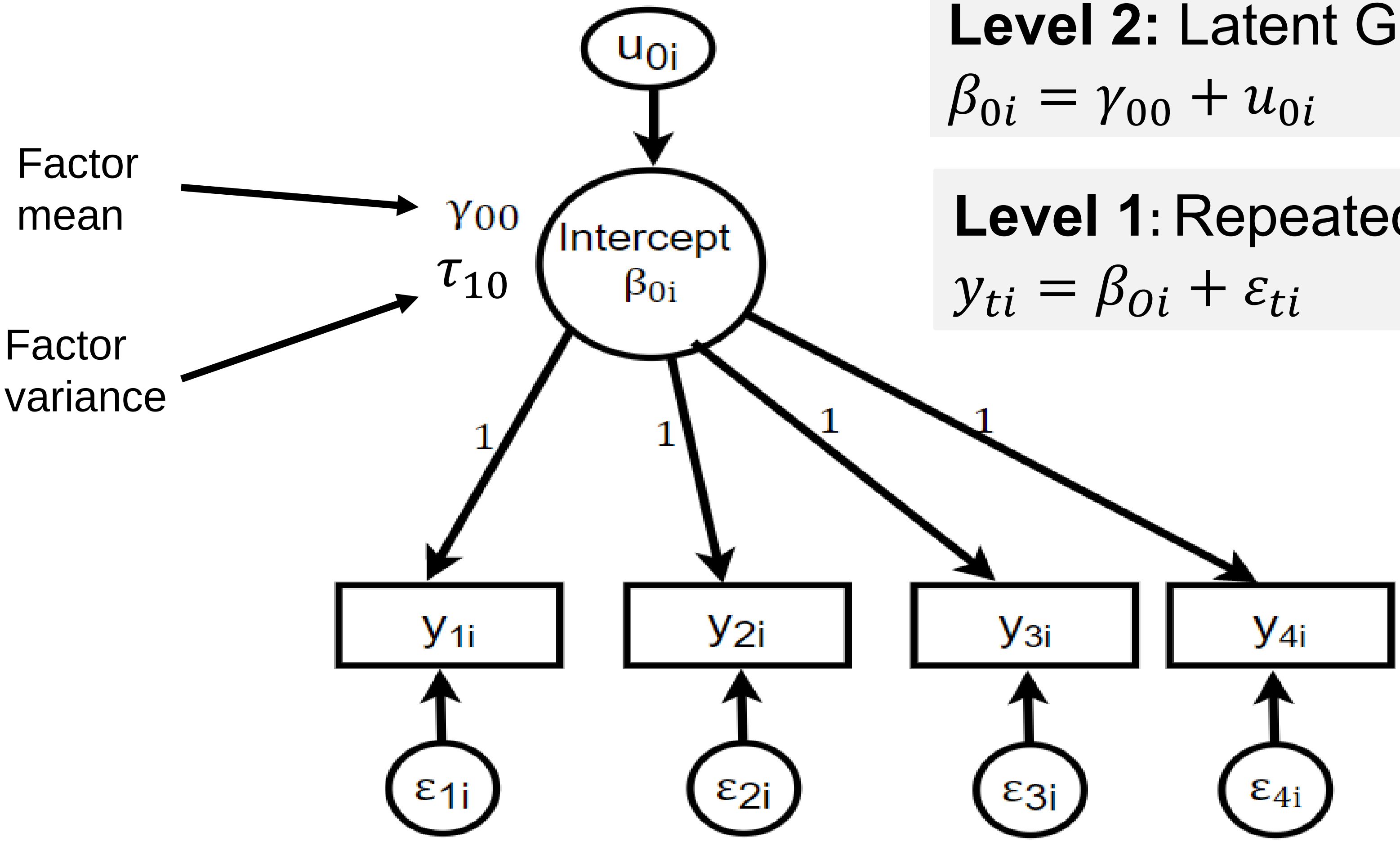


Structural Equation Modeling Framework

- Linear LCM Mean Structure:
  - Mean of intercepts and mean of slopes
- Variance Components:
  - Variance-covariance matrix of intercept and slopes
  - Covariance structure among time specific residuals σ^2. Covariance of errors between time points is 0 and variances of residuals across time points are $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$ (each time point is assume to have a unique residual variance. Sometimes each time point could be assumed to have the same residual variance).

Structural Equation Modeling Framework

Intercept-only LCM



Level 2: Latent Growth Factor

$$\beta_{0i} = \gamma_{00} + u_{0i}$$

Level 1: Repeated Measures

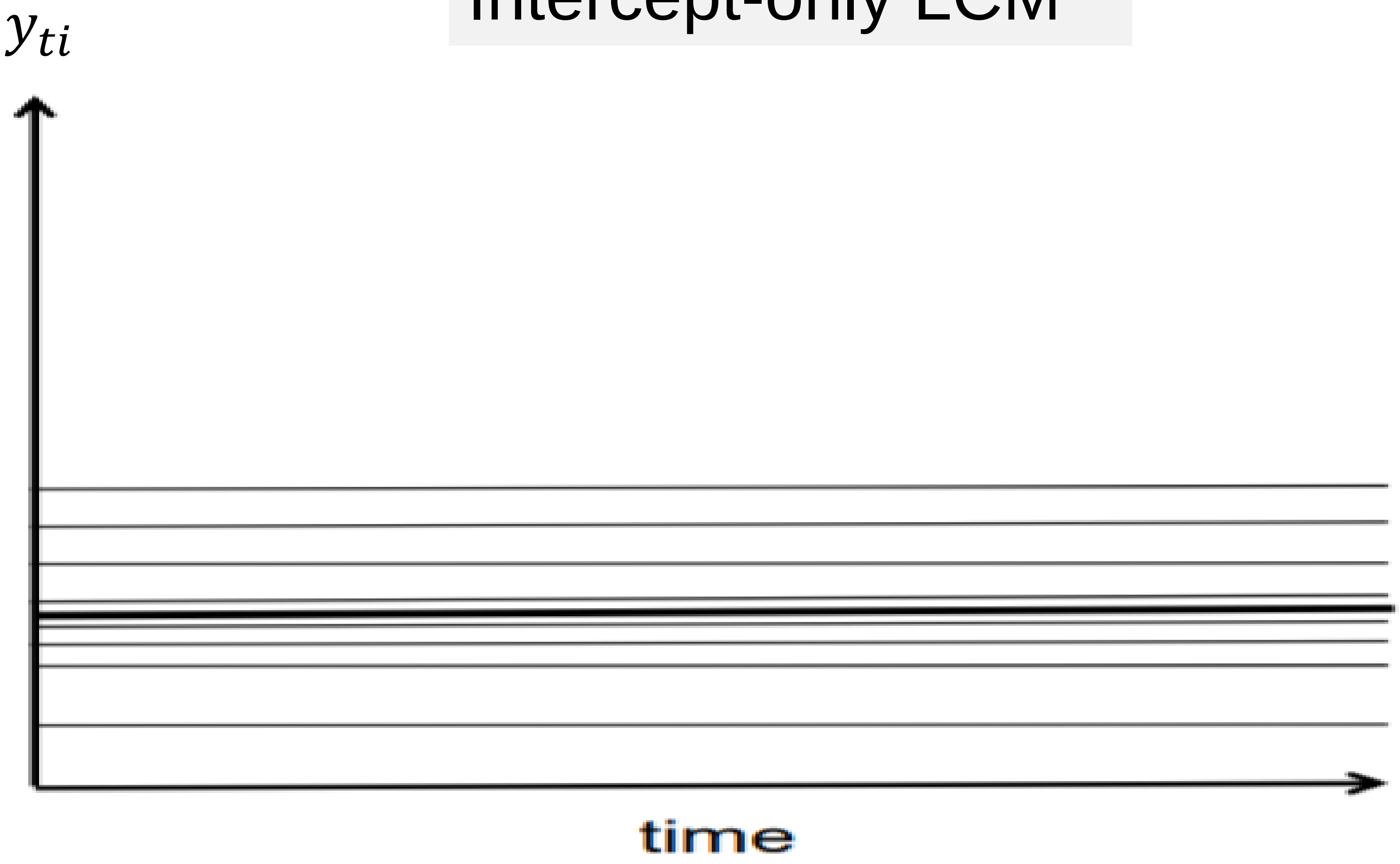
$$y_{ti} = \beta_{0i} + \varepsilon_{ti}$$

Note:

$$\beta_{1i} = 0$$
$$\gamma_{10} = 0$$

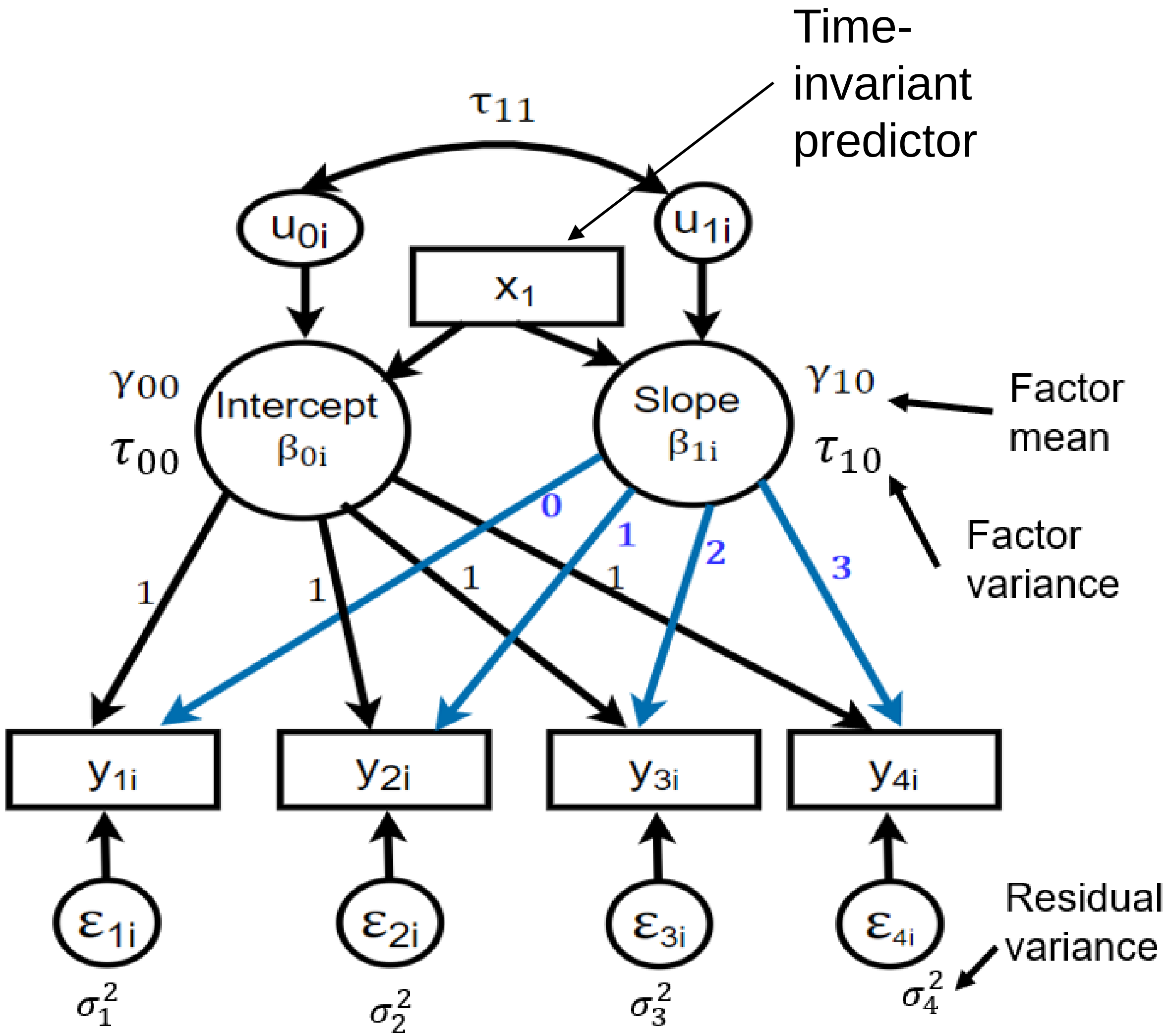
Structural Equation Modeling Framework

Intercept-only LCM



Structural Equation Modeling Framework

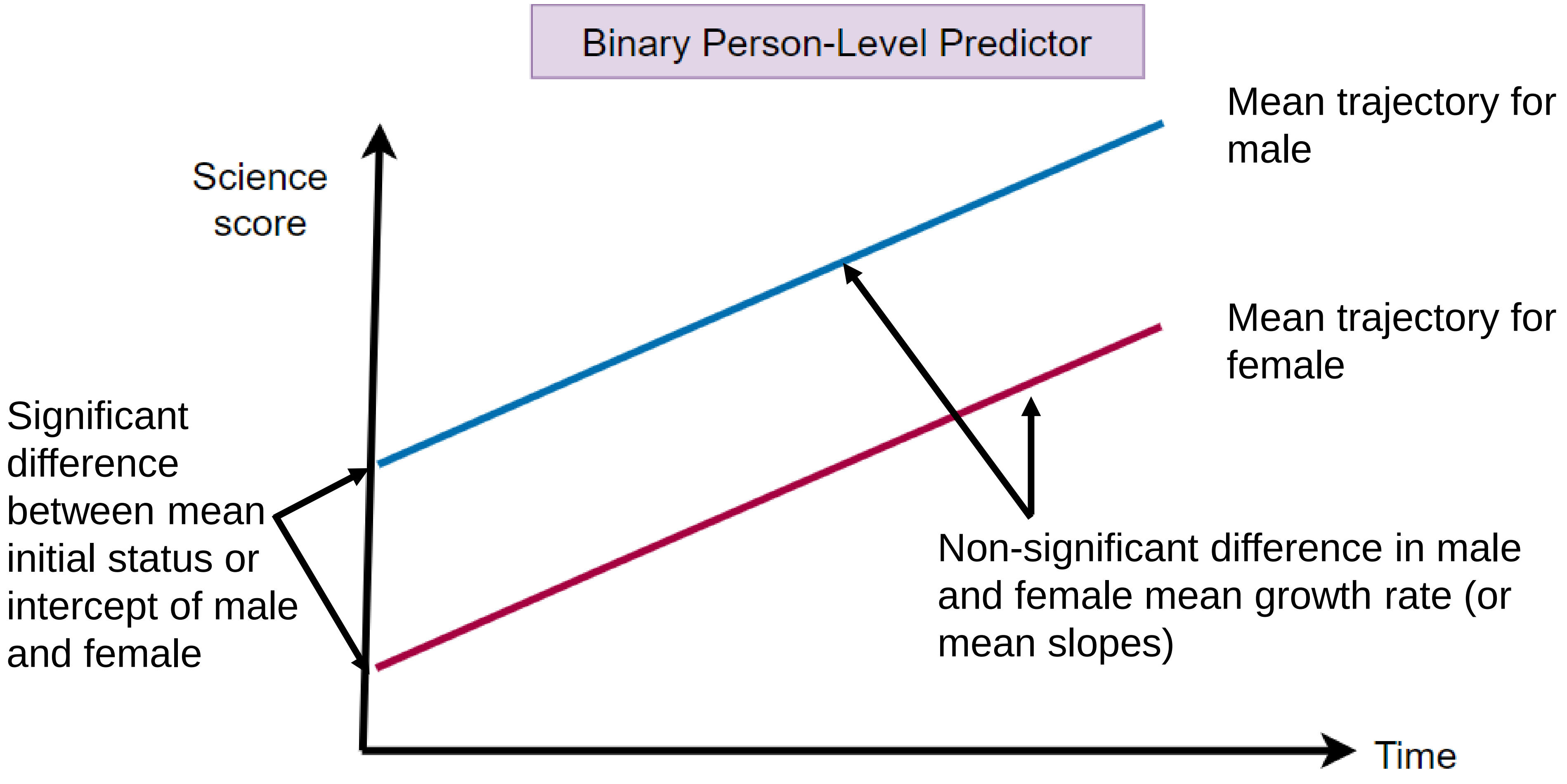
- **Time-invariant predictors:**
  - Time invariant predictors can be added at level 2 to explain the variance in slopes and intercepts.
- **Level 1:** Repeated Measures
  - $y_{ti} = \beta_{0i} + \beta_{1i} x_{ti} + \varepsilon_{ti}$
- **Level 2:** Latent Growth Factor
  - $\beta_{0i} = \gamma_{00} + \gamma_{01} * x_1 + u_{0i}$
  - $\beta_{1i} = \gamma_{10} + \gamma_{11} * x_1 + u_{1i}$



Structural Equation Modeling Framework

- Time-Invariant Predictor:
  - With time-invariant predictors, the model becomes conditional. That implies outcomes can be estimated given predictor values.
  - If the time-invariant predictor is binary, then the mean trajectories can be developed for each binary value. This is similar to conducting an ANOVA when the mean trajectory intercepts and slopes are compared.

Structural Equation Modeling Framework



Non-Linear Trajectories

- We can add polynomial terms to any degree to the linear model to obtain nonlinear trajectories
- Generally, a polynomial trajectory can be written as:
  - $y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + \beta_{2i}x_{ti}^2 + \cdots + \beta_{ki}x_{ti}^k + \varepsilon_{ti}$
  - For k degrees of polynomial

- **The Quadratic Trajectory**

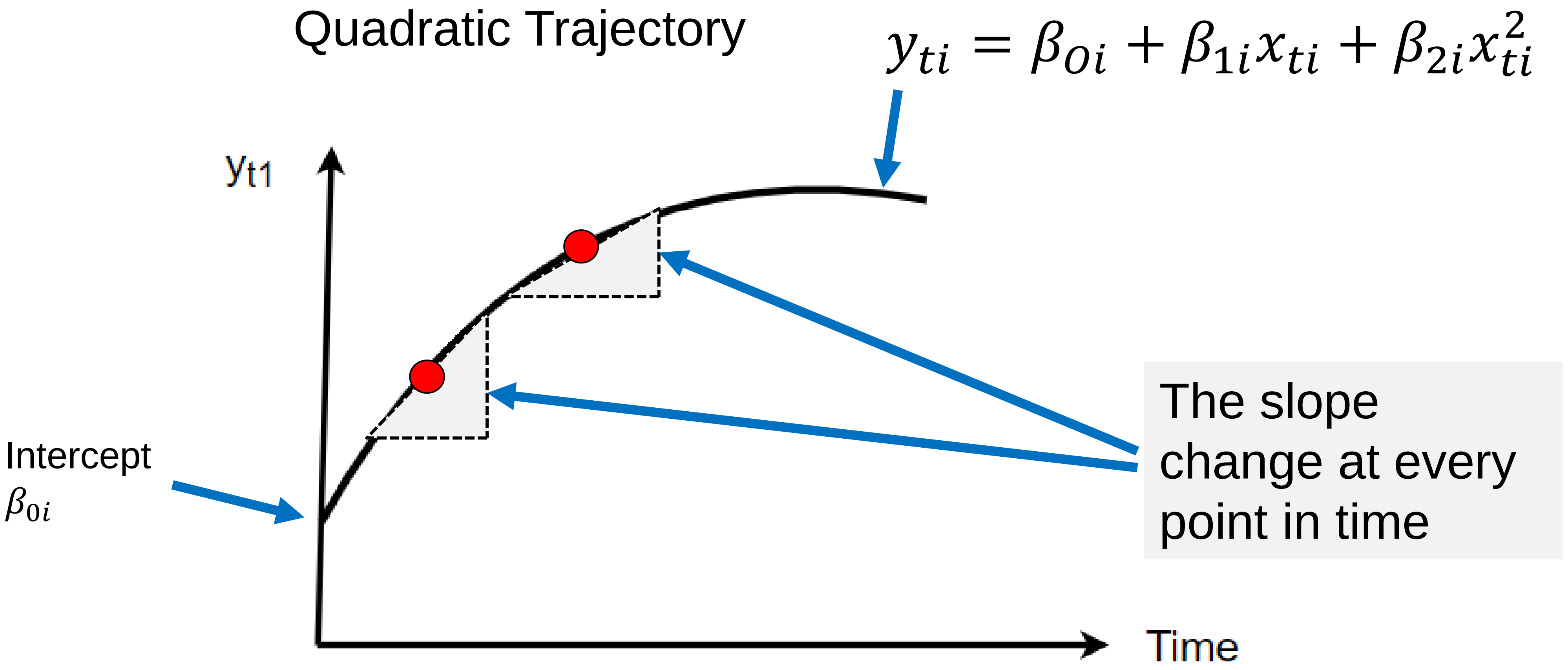
Level 1: Repeated Measures

$$y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + \beta_{2i}x_{ti}^2 + \varepsilon_{ti}$$

Level 2: Latent Growth Factor

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$
$$\beta_{2i} = \gamma_{20} + u_{2i}$$

Non-Linear Trajectories



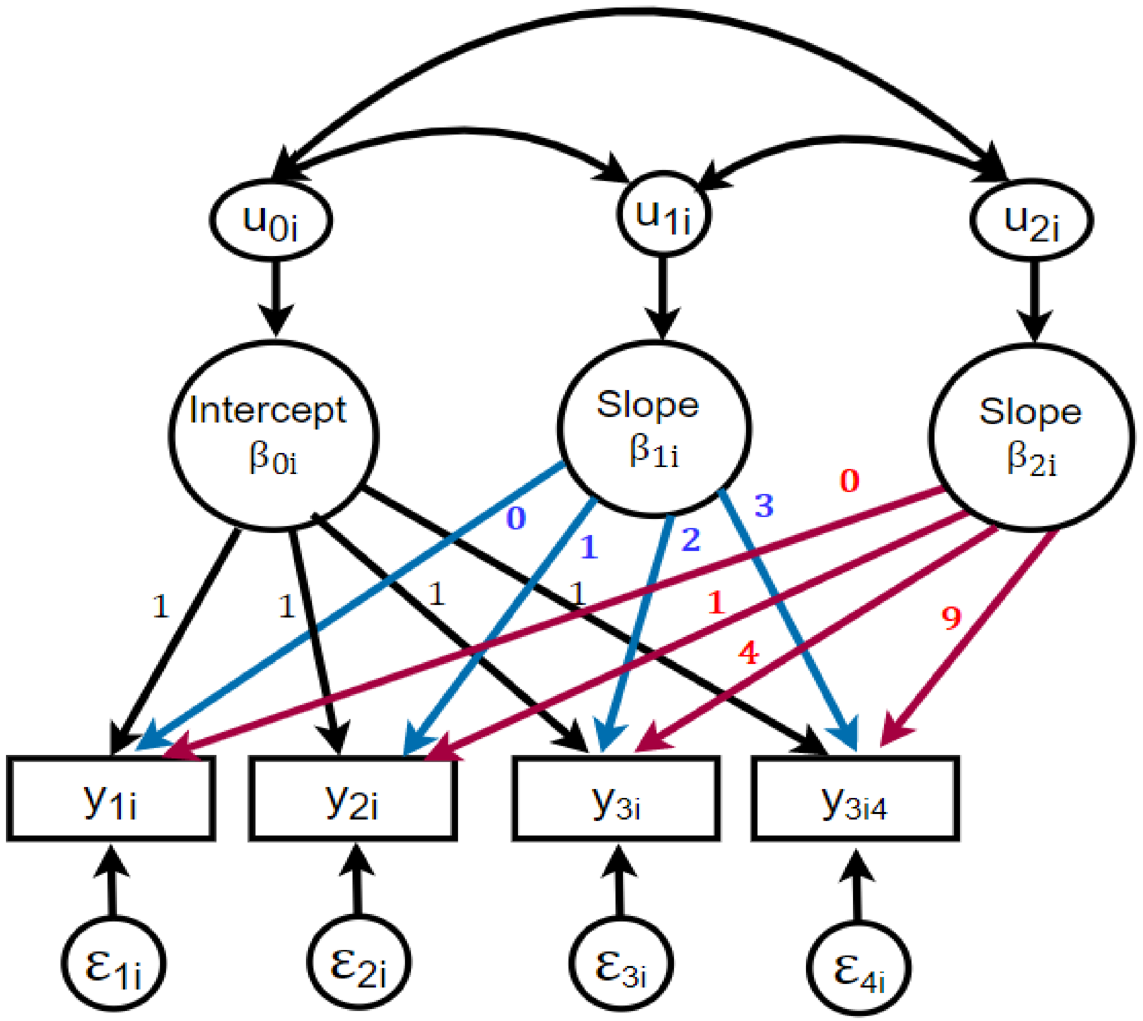
Non-Linear Trajectories

Level 2: Latent Growth Factor

$\beta_{0i} = \gamma_{00} + u_{0i}$
$\beta_{1i} = \gamma_{10} + u_{1i}$
$\beta_{2i} = \gamma_{20} + u_{2i}$

Level 1: Repeated Measures

$y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + \beta_{2i}x_{ti}^2 + \varepsilon_{ti}$

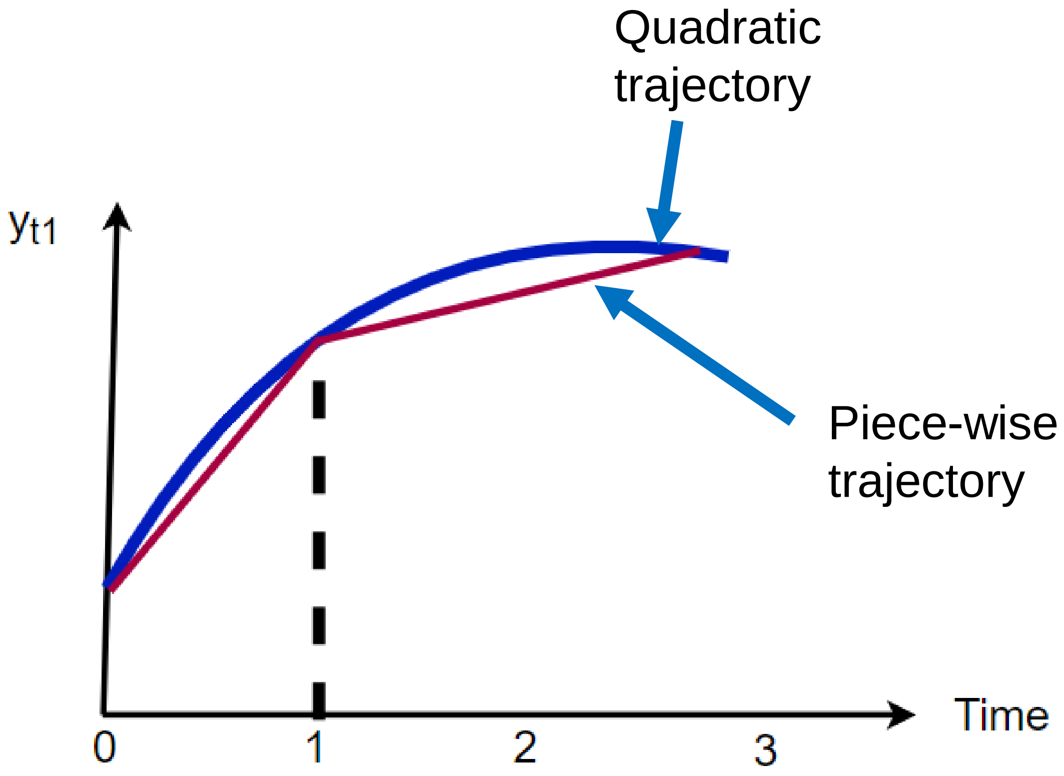


Non-Linear Trajectories

A non-linear trajectory such as a quadratic trajectory can be approximated using a combination of linear functions.

$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{bmatrix}$$

This example consist of a linear function from time 0 to 1 and another linear function from time 1 to 3



For the second linear function, the time start ti

Data Structure

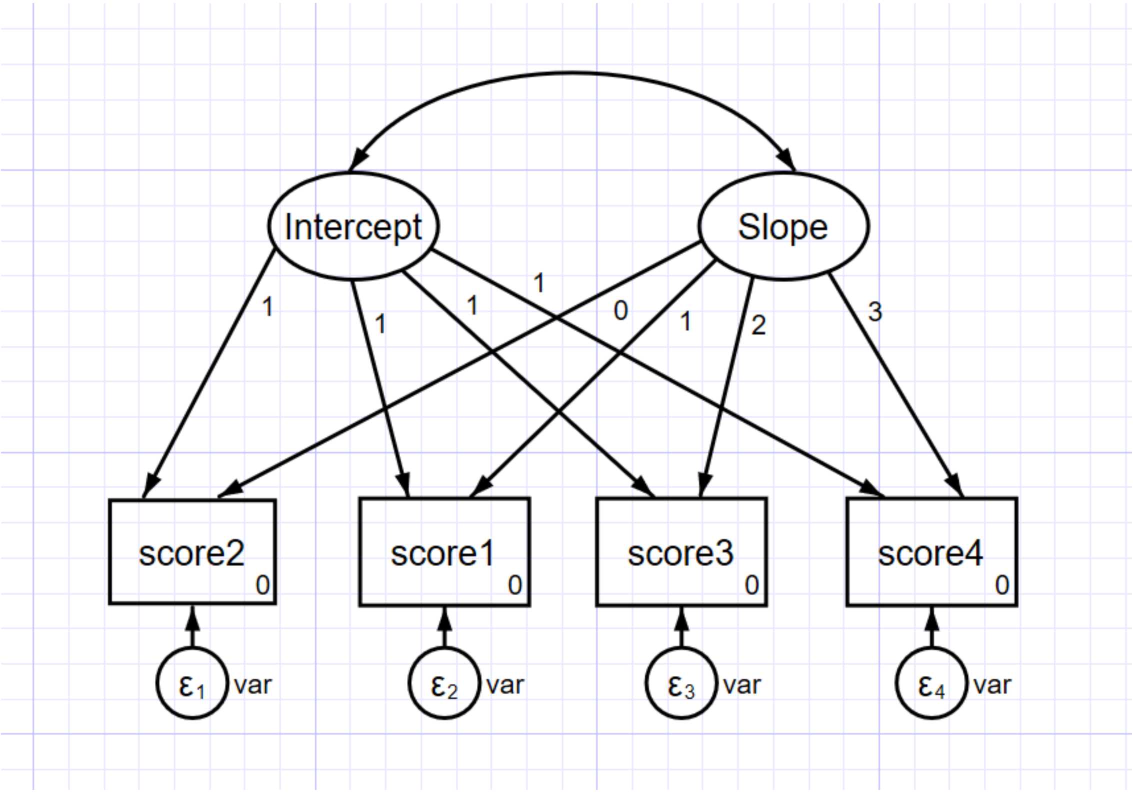
ID	Gender	Science Score	Grade
1	1	80	9
1	1	90	10
1	1	87	11
2	0	85	9
2	0	92	10
2	0	90	11
3	1	85	9
3	1	80	10
3	1	90	11

ID	Gender	Science9	Science10	Science11
1	1	80	90	87
2	0	85	92	90
3	1	85	80	90

A wide format data structure is typically used in structural equation modeling (SEM)

A long format data structure is used for multilevel modeling (HLM)

How to Run Latent Growth Model in Stata



- **Use the SEM Builder**
  - Create the scores at different time points
  - Create the latent variables and name them as intercept and slope
  - Create the paths and their constraints
  - Make the errors vary by adding “var”

Stata Syntax

Syntax in a do file

```
sem (Score0 <- Intercept@1 Slope@0 _cons@0) ///
        (Score1 <- Intercept@1 Slope@1 _cons@0) ///
        (Score3 <- Intercept@1 Slope@2 _cons@0) ///
        (Score4 <- Intercept@1 Slope@3 _cons@0),
        covstruct(_lexogenous, diagonal) ///
        method(mlmv) ///
        latent(Intercept Slope) ///
        var(e.Score0@var e.Score1@var e.Score2@var e.Score3@var)///
        cov(Intercept*Slope) ///
        means(Intercept Slope) noncapslatent
```