PSY 5939: Longitudinal Data Analysis

Contents

1 From mixed models to LGM

1.1 Model building

1.1.1 Homework 2

In homework 2, I walked you through "model building" for this dataset and research question

The idea behind model building is that you don't start at the final, most complex model you're interested in

- You build up to it
- Final model is your goal, you have a **plan**

The simpler models provide information about

- What your data looks like
- What works in the model
- What *doesn't* work in the model

1.1.2 Listening to your data: non-model based

- Plots of means across time
- Spaghetti plots
- Sample size at different time points

1.1.3 Listening to your data: model-based

Unconditional model:

- How much variance is due to repeated observations?

Linear growth model:

- How much variance is accounted for by a linear trend?
- What is the average linear trend?
- What is its variability?

Linear growth with L2 predictor:

- How much variance in intercept variability is due to gender?
- How much variance in linear trend is due to gender?
- How are average trajectories different across gender?

1.1.4 What questions does this model answer?

. . . and what questions does it **NOT** answer?

For example, in Homework 2, I asked "**What can you say** about variability in how students' behavioral problems change over time?"

- A model that does not include random slopes will not allow you to answer this question

In preparing for your final project:

- **You** will be in charge of model building (but use the logic in the assignment(s) as a guide)
- Think about the questions you want to answer and how the models you choose can help you answer them

1.1.5 Model building strategies

> "Essentially, all models are wrong, but some are useful"
>
> – George Box

Start with simpler models, see what they tell you
Work up to (what you hope is) your final model

- Sometimes you won't get there

A few websites with good discussions of model building:

- http://andrewgelman.com/2009/05/advice_on_stati/
- http://www.theanalysisfactor.com/7-guidelines-model-building/
- Also see the Preacher et al. book on Latent Growth Curve Models

1.2 SEM latent growth model vs mixed model

1.2.1 SEM latent growth model vs mixed model

The SEM latent growth model and the mixed model are **very similar**

- The main growth parameters are the same (*conceptually*, and if no missing data, *numerically*)
    - Mean intercept and slope, variance of intercept and slope, etc.
- Individual, "latent" (unobserved, estimated) trajectories for each individual

But the models are **conceptualized** differently

SEM latent growth model:

- Not a multilevel model like Mplus, so no level 1 and level 2
- Residual variance is conceptually different, which gives more flexibility
- Uses the "wide" or "multivariate" data format

1.2.2 Compare results

Reisby data from Don Hedeker's website:

- https://hedeker.people.uic.edu/long.html
- Hamilton Depression Inventory (HamD) is the outcome
    - 24+: severe depression
    - 17 to 23: moderate depression
    - 8 to 16: mild depression
    - 7 or less: normal
- Measured at 6 time points (weeks - coded 0, 1, 2, 3, 4, 5)
- No missing data

1.2.3 Compare results

Mixed model:

- Linear growth model with no predictors
- Random intercept
- Random slope

Latent growth model:

- Linear growth model with no predictors
- Random intercept
- Random slope
- *Constraint*: Equal residual variances across time points (more on this soon)

Equivalent when **no missing data** and **same assessment schedule** for all

1.2.4 Mixed model results - SPSS

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	23.369565	.629097	46	37.148	.000	22.103259	24.635871
Week	-2.330435	.228187	46	-10.213	.000	-2.789752	-1.871117

a. Dependent Variable: HamD.

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		11.751656	1.225195	9.592	.000	9.579772	14.415940
Intercept + Week [subject = ID]	UN (1,1)	12.049456	3.849890	3.130	.002	6.441726	22.538896
	UN (2,1)	-.903925	1.059988	-.853	.394	-2.981464	1.173614
	UN (2,2)	1.723673	.504316	3.418	.001	.971423	3.058450

a. Dependent Variable: HamD.

1.2.5 Mixed model results - R

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: HamD ~ 1 + Week + (1 + Week | ID)
##    Data: reisby_tall
##
##      AIC      BIC   logLik deviance df.resid
##   1628.7   1650.5   -808.4   1616.7      270
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8965 -0.4665  0.0237  0.4568  3.8181
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 12.050   3.471
##           Week         1.724   1.313    -0.20
##  Residual             11.752   3.428
## Number of obs: 276, groups:  ID, 46
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  23.3696     0.6291   37.15
## Week         -2.3304     0.2282  -10.21
##
## Correlation of Fixed Effects:
##      (Intr)
## Week -0.391
```

1.2.6 Latent growth model results - Mplus

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 INT      |
    HAMD0              1.000      0.000    999.000    999.000
    HAMD1              1.000      0.000    999.000    999.000
    HAMD2              1.000      0.000    999.000    999.000
    HAMD3              1.000      0.000    999.000    999.000
    HAMD4              1.000      0.000    999.000    999.000
    HAMD5              1.000      0.000    999.000    999.000

 LIN      |
    HAMD0              0.000      0.000    999.000    999.000
    HAMD1              1.000      0.000    999.000    999.000
    HAMD2              2.000      0.000    999.000    999.000
    HAMD3              3.000      0.000    999.000    999.000
    HAMD4              4.000      0.000    999.000    999.000
    HAMD5              5.000      0.000    999.000    999.000

 LIN      WITH
    INT               -0.904      1.060     -0.853      0.394

 Means
    INT               23.370      0.629     37.148      0.000
    LIN               -2.330      0.228    -10.213      0.000

 Intercepts
    HAMD0              0.000      0.000    999.000    999.000
    HAMD1              0.000      0.000    999.000    999.000
    HAMD2              0.000      0.000    999.000    999.000
    HAMD3              0.000      0.000    999.000    999.000
    HAMD4              0.000      0.000    999.000    999.000
    HAMD5              0.000      0.000    999.000    999.000

 Variances
    INT               12.049      3.850      3.130      0.002
    LIN                1.724      0.504      3.418      0.001

 Residual Variances
    HAMD0             11.752      1.225      9.592      0.000
    HAMD1             11.752      1.225      9.592      0.000
    HAMD2             11.752      1.225      9.592      0.000
    HAMD3             11.752      1.225      9.592      0.000
    HAMD4             11.752      1.225      9.592      0.000
    HAMD5             11.752      1.225      9.592      0.000
```

1.2.7 Latent growth model input - Mplus

```
TITLE:
Reisby data - linear growth model to match mixed model
```

```
DATA:
FILE IS reisby_wide.dat;

VARIABLE:
NAMES ARE ID HamD0 HamD1 HamD2 HamD3 HamD4 HamD5;
USEVARIABLES ARE HamD0 HamD1 HamD2 HamD3 HamD4 HamD5;

MODEL:
! new style specification of linear growth model;
int lin | HamD0@0 HamD1@1 HamD2@2 HamD3@3 HamD4@4 HamD5@5;
! fix residual variances to be equal across all 6 time points;
HamD0 (1);
HamD1 (1);
HamD2 (1);
HamD3 (1);
HamD4 (1);
HamD5 (1);
```

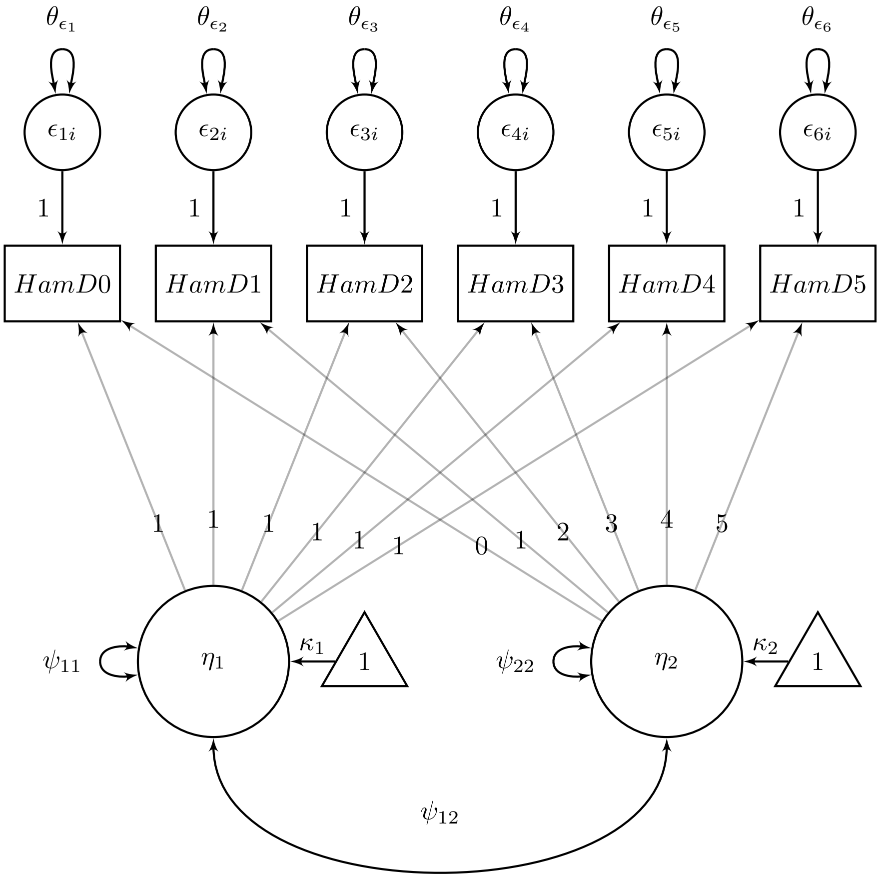
1.2.8 Latent growth model - lavaan

```
model <- 'i =~ 1*HamD0 + 1*HamD1 + 1*HamD2 + 1*HamD3 + 1*HamD4 + 1*HamD5
s =~ 0*HamD0 + 1*HamD1 + 2*HamD2 + 3*HamD3 + 4*HamD4 + 5*HamD5
HamD0 ~~ b*HamD0
HamD1 ~~ b*HamD1
HamD2 ~~ b*HamD2
HamD3 ~~ b*HamD3
HamD4 ~~ b*HamD4
HamD5 ~~ b*HamD5'
fit <- growth(model, data = reisby_wide)
summary(fit)
```

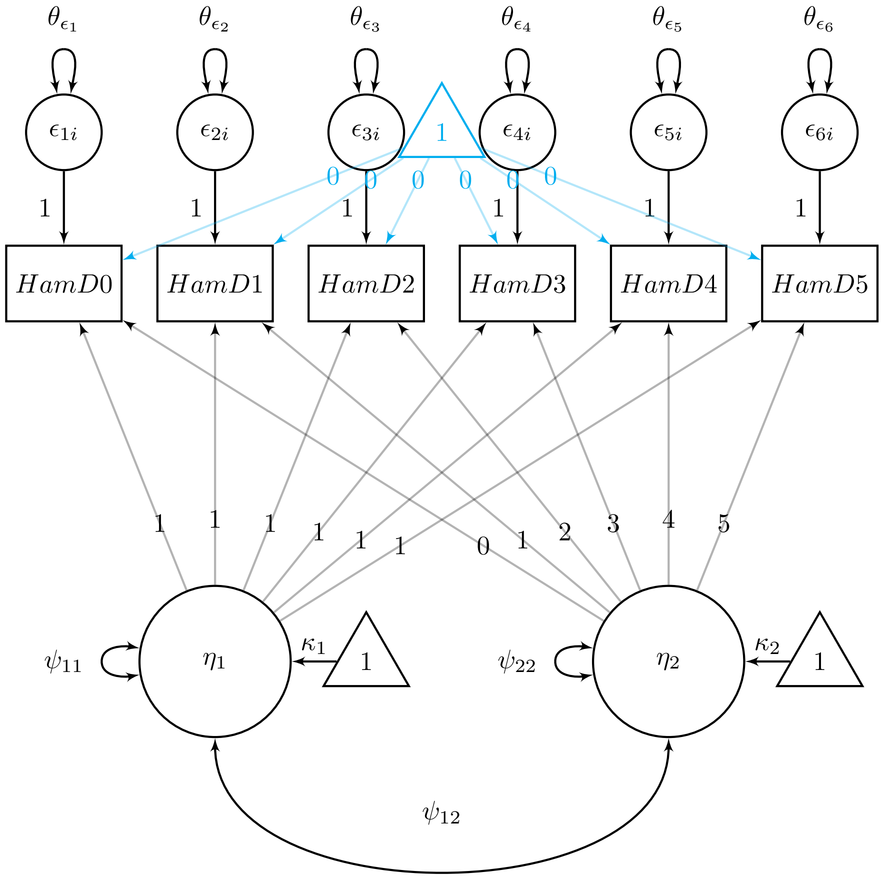
```
## lavaan 0.6-10 ended normally after 48 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        11
##   Number of equality constraints                     5
##
##   Number of observations                            46
##
## Model Test User Model:
##
##   Test statistic                                24.087
##   Degrees of freedom                                21
##   P-value (Chi-square)                           0.289
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
```

```
##                    Estimate  Std.Err  z-value  P(>|z|)
##   i =~
##     HamD0             1.000
##     HamD1             1.000
##     HamD2             1.000
##     HamD3             1.000
##     HamD4             1.000
##     HamD5             1.000
##   s =~
##     HamD0             0.000
##     HamD1             1.000
##     HamD2             2.000
##     HamD3             3.000
##     HamD4             4.000
##     HamD5             5.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   i ~~
##     s                -0.904    1.060   -0.853    0.394
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .HamD0             0.000
##    .HamD1             0.000
##    .HamD2             0.000
##    .HamD3             0.000
##    .HamD4             0.000
##    .HamD5             0.000
##     i                23.370    0.629   37.148    0.000
##     s                -2.330    0.228  -10.213    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .HamD0      (b)   11.752    1.225    9.592    0.000
##    .HamD1      (b)   11.752    1.225    9.592    0.000
##    .HamD2      (b)   11.752    1.225    9.592    0.000
##    .HamD3      (b)   11.752    1.225    9.592    0.000
##    .HamD4      (b)   11.752    1.225    9.592    0.000
##    .HamD5      (b)   11.752    1.225    9.592    0.000
##     i                12.049    3.850    3.130    0.002
##     s                 1.724    0.504    3.418    0.001
```

1.2.9 SEM latent growth model



1.2.10 SEM latent growth model



1.2.11 SEM latent growth model

The intercept and slope are **latent variables**: η_1 and η_2

- The **means** of the latent intercept and slope (κ_1 and κ_2) are analogous to
  - β_{00} and β_{01} from mixed models
- The **variances** of the latent intercept and slope (ψ_{11} and ψ_{22}) are analogous to
  - $\sigma^2_{r_{0i}}$ and $\sigma^2_{r_{1i}}$
- The **covariance** between the latent intercept and slope (ψ_{12}) is analogous to
  - $\sigma_{r_{0i}r_{1i}}$

In the SEM framework, a latent growth model is a *confirmatory factor analysis (CFA)* with a specialized set of *constraints*

- More on this later

1.2.12 Linear latent growth model with 6 time points

6 parameters are estimated

- Matches "number of free parameters" in Mplus output

Be able to count them all!

- **Two latent factor means**: κ_1 is the mean intercept, κ_2 is the mean linear slope

- **Two factor variances**: ψ_{11} is the intercept variance, ψ_{22} is the slope variance
- **One factor covariance**: ψ_{21} is the covariance between intercept and slope
- **Six residual variances**
  - but **we only estimate one value** because we *constrained* $\theta_{\epsilon_{0i}} = \theta_{\epsilon_{1i}} = \theta_{\epsilon_{2i}} = \theta_{\epsilon_{3i}} = \theta_{\epsilon_{4i}} = \theta_{\epsilon_{5i}}$

$\rightarrow 2 + 2 + 1 + 1 = 6$

1.3 Results

1.3.1 What does the LGM tell us?

- Mean intercept (and its variance)
- Mean linear slope (and its variance)
- Covariance between intercept and slope (can convert to correlation)
- Residual variance *at each time point*

All of these things except the residual variances are interpretable

- Residual variances may be of *substantive* interest – for example, does the model with different residual variances at each time point fit better than one with fixed residual variances

1.3.2 Intercept interpretation

Mean intercept is the average value of the outcome at time = 0

- Put time = 0 at a substantively interesting point: here, it's at the start
- Mean intercept = 23.370, p<.001
- At the beginning of the study, the average HamD score was 23.370, p<.001

Intercept variance tells you how the intercept varies *across people*

- About 95% of *individuals* have an intercept in [mean intercept ± 1.96*$\sqrt{variance}$]
- Intercept variance = 12.049, p<.01
- SD = $\sqrt{12.049}$ = 3.471
- At the beginning of the study, about 95% of **participants** had HamD scores between 16.567 and 30.173

1.3.3 Slope interpretation

Mean linear slope is the average increase in the outcome for a 1-unit increase in time

- Mean slope = -2.330, p<.001
- Each week, HamD scores *decreased* an average of 2.330 points, p<.001

Slope variance tells you how the slope varies *across people*

- About 95% of *individuals* have a slope in [mean slope ± 1.96*$\sqrt{variance}$]
- Slope variance = 1.724, p<.01
- SD = $\sqrt{1.724}$ = 1.313
- About 95% of **participants** had slopes between -4.903 and 0.243

1.3.4 Intercept-slope covariance / correlation interpretation

Positive covariance

- People with *high* intercepts have *high* slopes, people with *low* intercepts have *low* slopes

Negative covariance

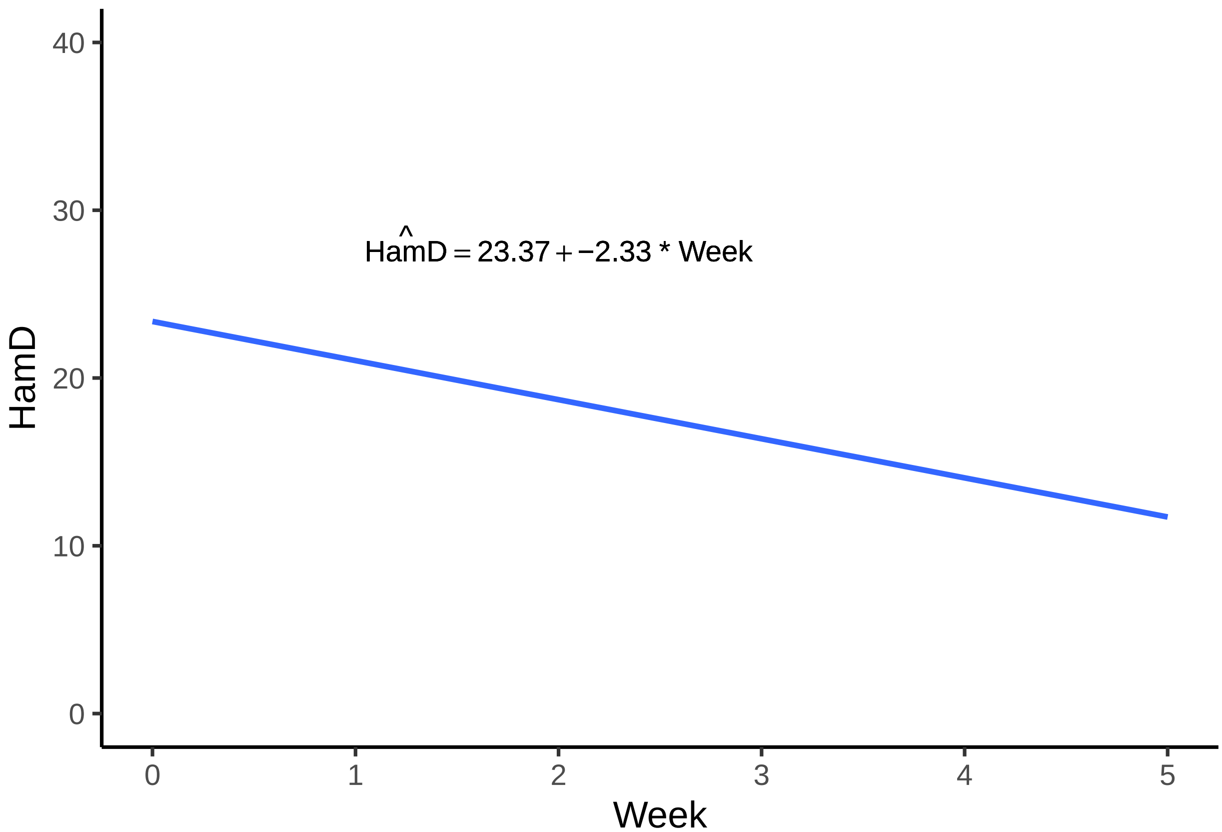
- People with *high* intercepts have *low* slopes, people with *low* intercepts have *high* slopes

Covariance between intercepts and slopes is -0.904, NS

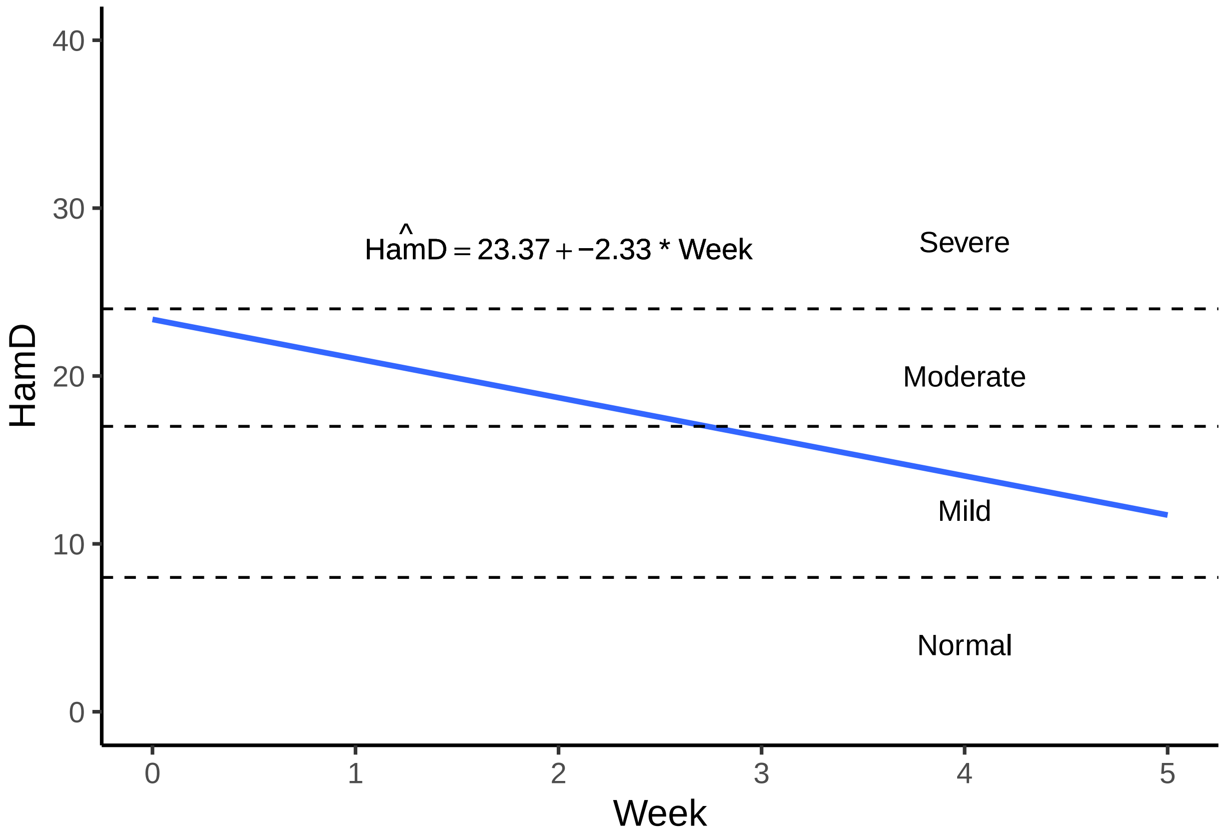
- (Start with high HamD scores → lower (more negative) slope)
- (Start with low HamD scores → higher (closer to zero) slope)

corr(I, S) = $\frac{cov(I,S)}{SD(I)SD(S)} = \frac{-0.904}{3.471 \times 1.313} = -0.198$

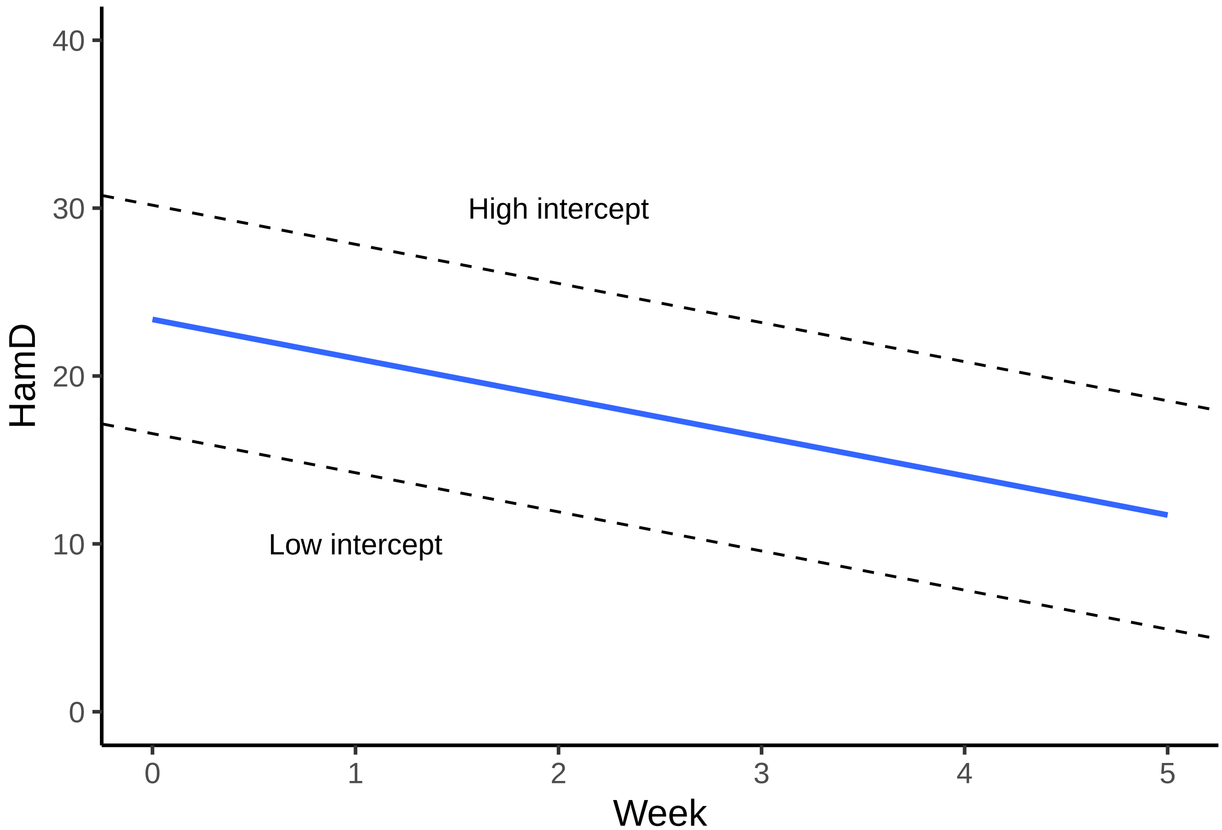
1.3.5 Average intercept and slope



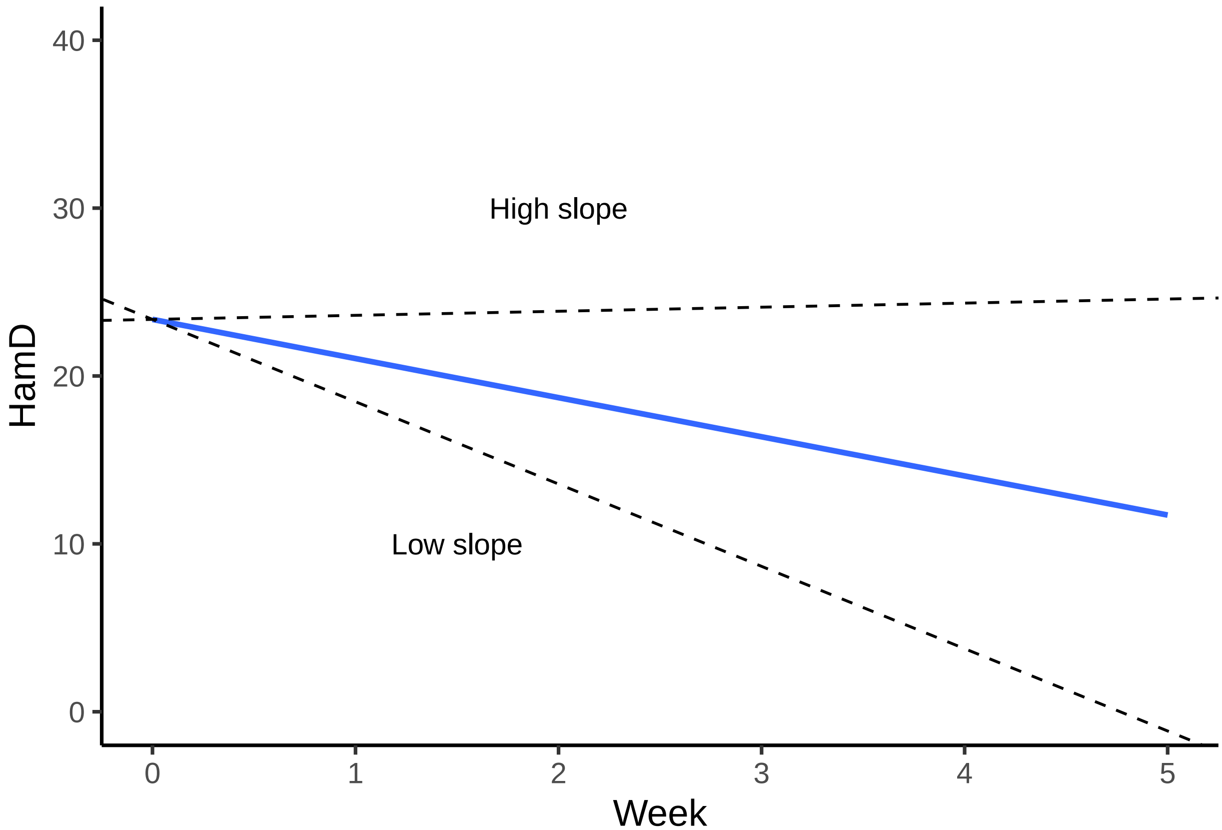
1.3.6 Average intercept and slope



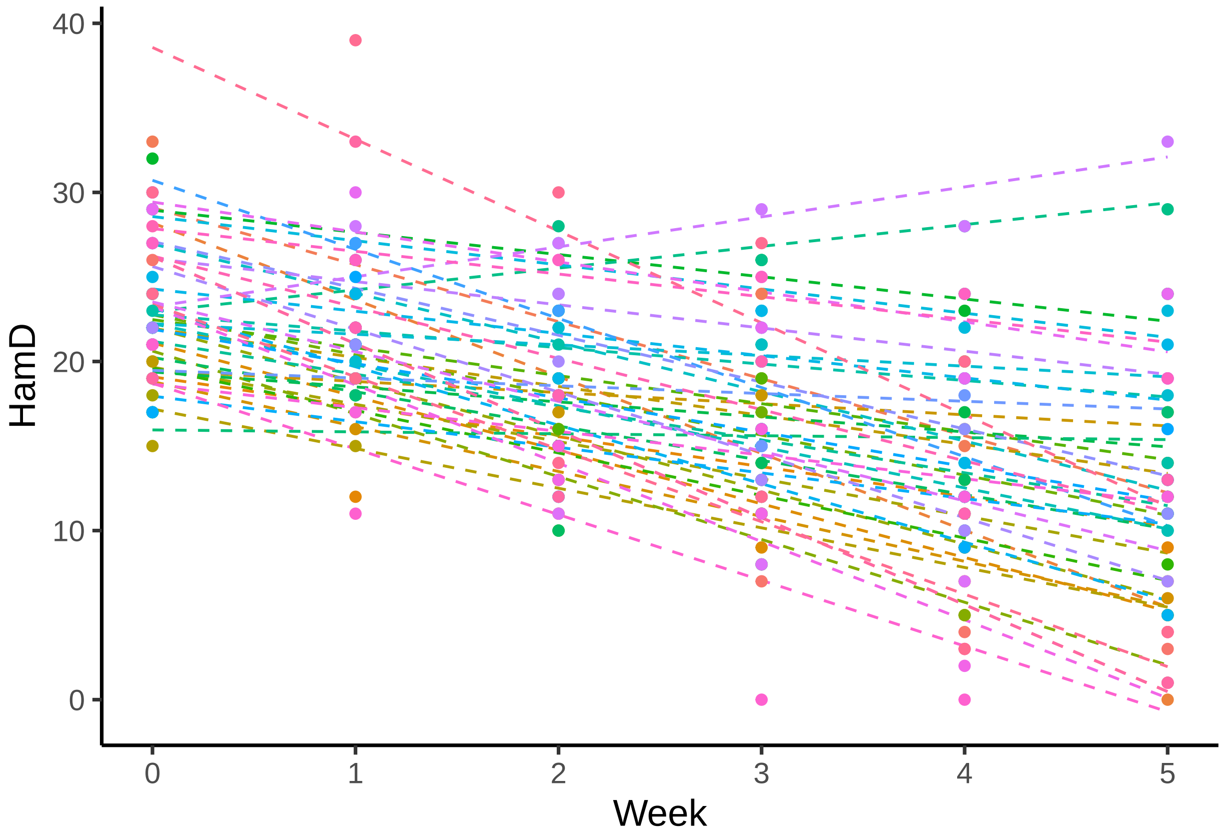
1.3.7 95% of individuals' intercept values



1.3.8 95% of individuals' slope values



1.3.9 Spaghetti plot



2 LGM as special case of factor analysis

2.1 Exploratory vs confirmatory factor analysis

2.1.1 Factor analysis

Factor analysis is a general statistical procedure for **data reduction**

- You can "reduce" the pieces of information you have from many observed variables to fewer latent variables

Factor analysis is also a way to identify **shared variance**

- One or more unobserved variables that explain what the many observed variables have in common
- For example, you give people several tests that measure math ability, reading ability, mental rotation, analogies, etc.
- These test scores are observed, but they reflect an unobserved variable (IQ or intelligence)

2.1.2 Exploratory factor analysis (EFA)

Exploratory factor analysis, as the name suggests, **explores** the factor structure of a set of observed variables

Often done with new measures, before much is known about them

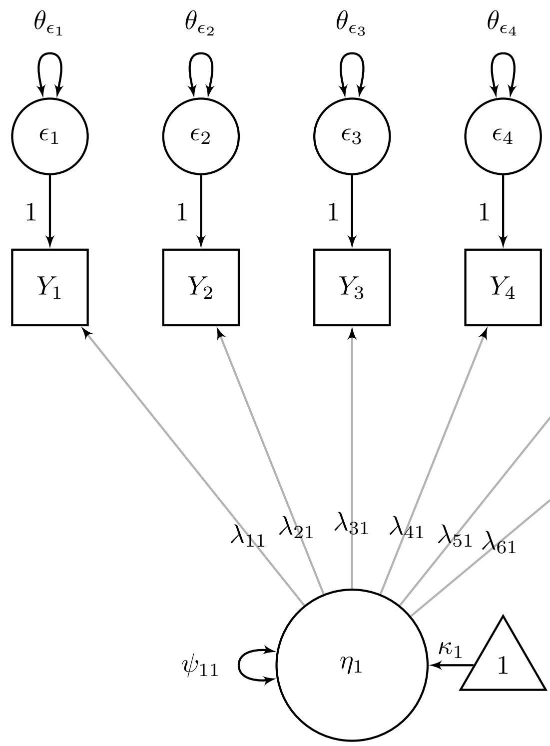
Try several different models with different numbers of factors

- Which model is best? Compare them

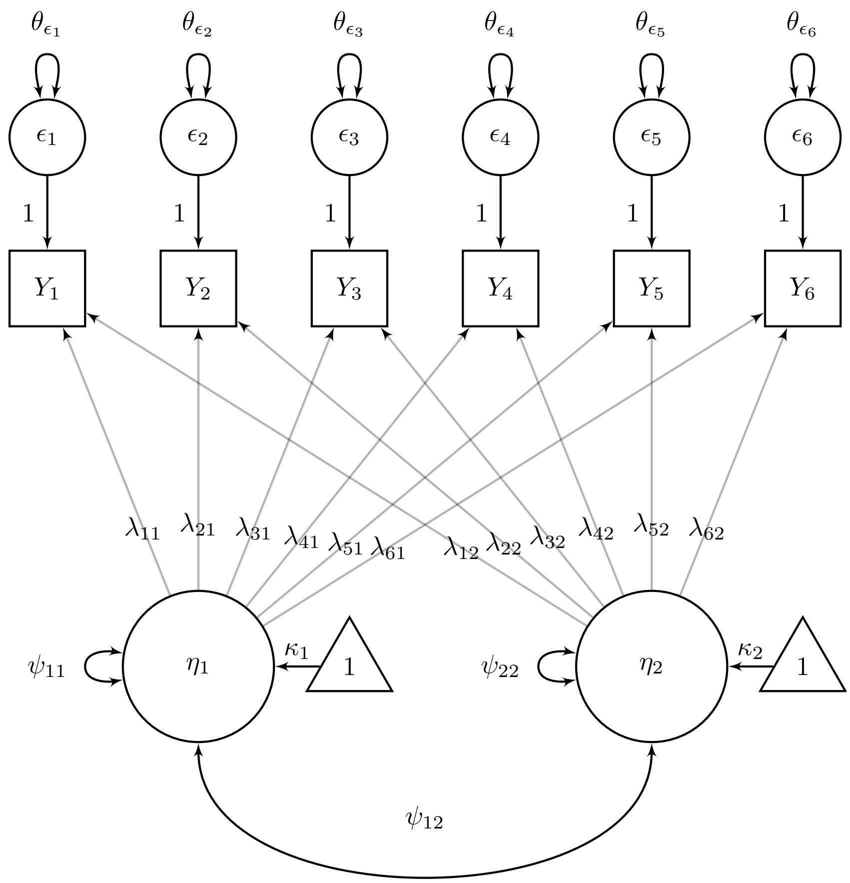
These models are relatively "free" (in the statistical sense)

- Estimate a lot of things
- All items to "load" on all factors
- (Some constraints to identify the model, that's all)

2.1.3 Exploratory factor analysis: 1 factor EFA



2.1.4 Exploratory factor analysis: 2 factor EFA



2.1.5 Confirmatory factor analysis (CFA)

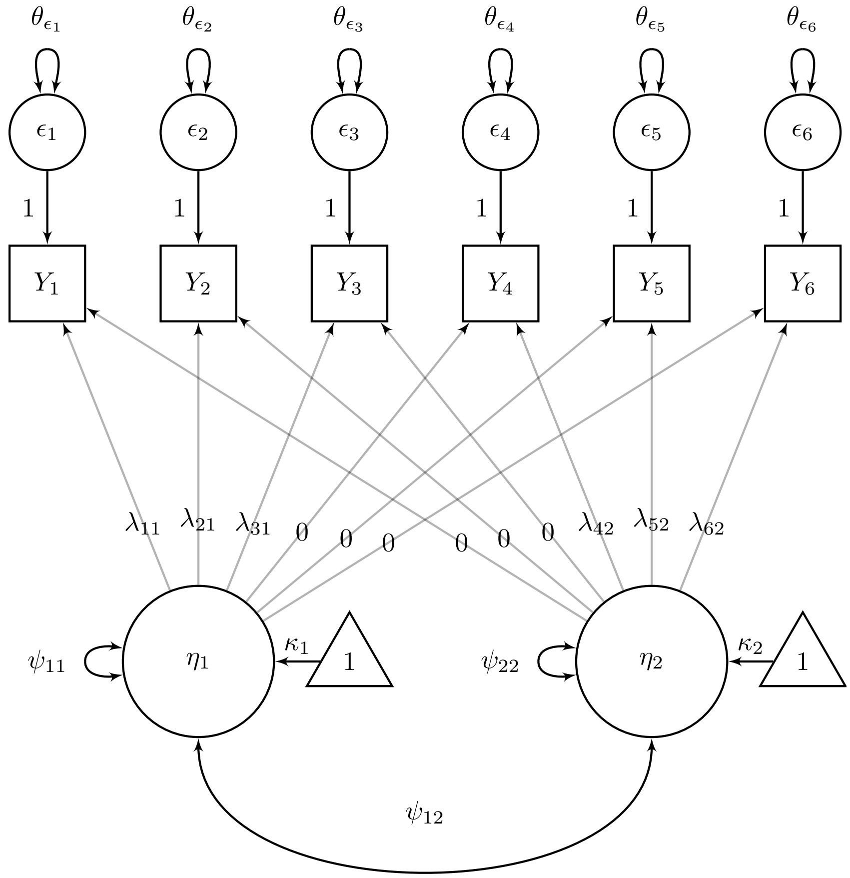
Confirmatory factor analysis is used when you know

- How many factors there are
- Which items load on which factors

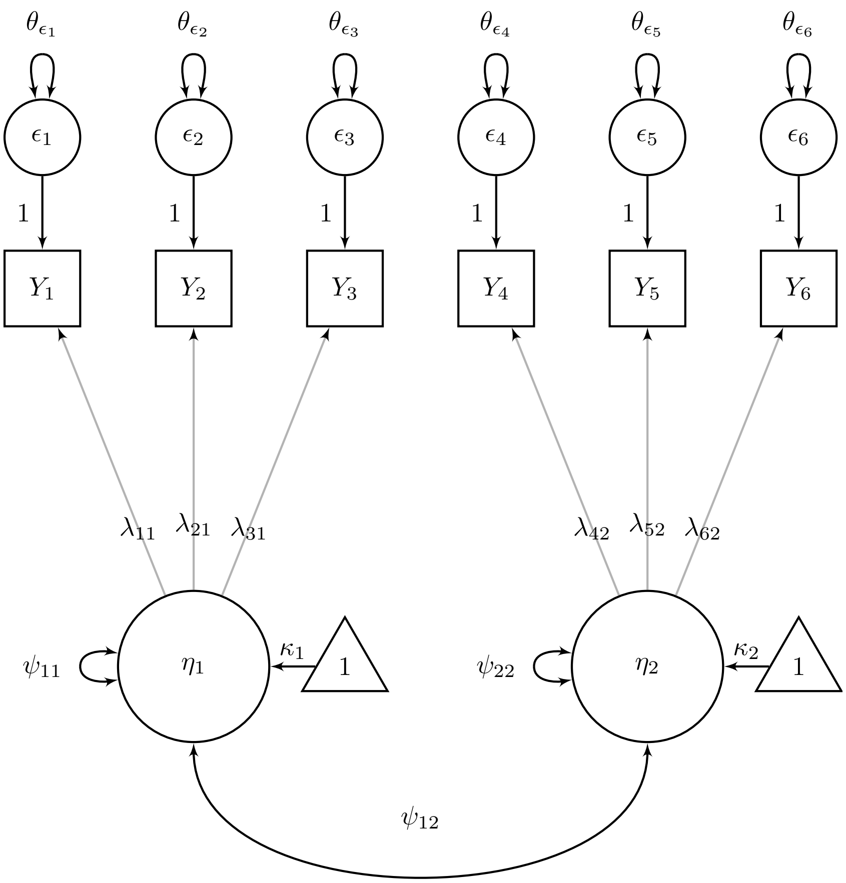
CFAs have many **more constraints** than EFAs – not as “free”

- For example, by saying that an item loads on Factor 1 but not Factor 2, you are *constraining* or *fixing* the loading of that item for Factor 2 to a value of 0

2.1.6 Confirmatory factor analysis: 2 factor CFA



2.1.7 Confirmatory factor analysis: 2 factor CFA



2.2 CFA notation

2.2.1 Confirmatory factor analysis

There are two parts to a confirmatory factor analysis

Measurement portion

- Defines the latent variables in your model
- Usually the **main focus** of a CFA

Structural portion

- Defines relationships between the latent variables
- Also defines relationships between the latent variables and other observed variables
- Often **not** emphasized in CFA

2.2.2 CFA equations: measurement portion

Defining the latent variables in the model

We can write an equation to predict person i's score on a single observed indicator k (Y_{ki}):

$$Y_{ki} = \nu_k + \lambda_{k1}\eta_{1i} + \lambda_{k2}\eta_{2i} + \epsilon_{ki}$$

This is basically a regression equation with

- An intercept (ν_k) (nu)
- Two (latent) predictor variables (η_{1i} and η_{2i}) (eta)
- Regression coefficients (λ_{k1} and λ_{k2}) (lambda)
- Residual or error term (ϵ_{ki}) (epsilon)

2.2.3 CFA equations for entire model

Six indicator variables: first 3 load on factor 1, last 3 load on factor 2

$$\begin{aligned}
Y_{1i} &= \nu_1 + \lambda_{11}\eta_{1i} + (0)\eta_{2i} + \epsilon_{1i} \\
Y_{2i} &= \nu_2 + \lambda_{21}\eta_{1i} + (0)\eta_{2i} + \epsilon_{2i} \\
Y_{3i} &= \nu_3 + \lambda_{31}\eta_{1i} + (0)\eta_{2i} + \epsilon_{3i} \\
Y_{4i} &= \nu_4 + (0)\eta_{1i} + \lambda_{42}\eta_{2i} + \epsilon_{4i} \\
Y_{5i} &= \nu_5 + (0)\eta_{1i} + \lambda_{52}\eta_{2i} + \epsilon_{5i} \\
Y_{6i} &= \nu_6 + (0)\eta_{1i} + \lambda_{62}\eta_{2i} + \epsilon_{6i}
\end{aligned}$$

2.2.4 Matrix notation: measurement portion

$$Y_i = \nu + \Lambda\eta_i + \epsilon_i$$

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \\ Y_{4i} \\ Y_{5i} \\ Y_{6i} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \\ \varepsilon_{6i} \end{bmatrix}$$

2.2.5 CFA equations: structural portion

Relationships between latent variables
(also between observed variables and latent variables)

Now that the latent variables (η_1 and η_2) have been defined, we can talk about them

$$\begin{aligned}
\eta_{1i} &= \kappa_1 + \zeta_{1i} \\
\eta_{2i} &= \kappa_2 + \zeta_{2i}
\end{aligned}$$

η_{1i} and η_{2i} are **individual** factor score values on latent variables η_1 and η_2
κ_1 and κ_2 are the latent variable means (kappa)
ζ_{1i} and ζ_{2i} are the **individual** deviations around those means (zeta)

2.2.6 Matrix notation: structural portion

$$\eta_i = \kappa + \zeta_i$$

$$\begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \begin{bmatrix} \zeta_{1i} \\ \zeta_{2i} \end{bmatrix}$$

2.2.7 Latent variable covariance matrix

The CFA model equations include **individual** residuals

For example, ζ_{1i} is the difference between individual i's value on latent variable 1 and the mean of latent variable 1

- You do not actually get estimates of those values
- Instead, you get a **variance-covariance matrix of the residuals**

$$\Psi = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix} = \begin{bmatrix} {\sigma_{\eta_1}}^2 & \\ \sigma_{\eta_1 \eta_2} & {\sigma_{\eta_2}}^2 \end{bmatrix}$$

2.2.8 Residual covariance matrix

The residuals (ε_{ki}) have a covariance matrix

A starting point is to assume different residual variances for each variable, but that the covariances are all 0

$$\Theta = \begin{bmatrix} \theta_{\varepsilon_1} & & & & & \\ 0 & \theta_{\varepsilon_2} & & & & \\ 0 & 0 & \theta_{\varepsilon_3} & & & \\ 0 & 0 & 0 & \theta_{\varepsilon_4} & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_5} & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_6} \end{bmatrix}$$

2.3 CFA identification

2.3.1 Identifying the CFA model

Identification refers to *constraining* a model enough to get **unique estimates**

Known quantities: observed correlations
Unknown quantities: loadings, coefficients, etc.

For a model to be *identified*, we need to have fewer unknowns than knowns

In the CFA, the latent variable scores are completely missing (latent, unobserved, unknown), so we have to impose some *constraints* to get the number of unknowns $\leq$ the number of knowns

- This is called "identifying the model"

2.3.2 Methods of identifying the CFA model

Need to impose constraints in **2 parts** of the model

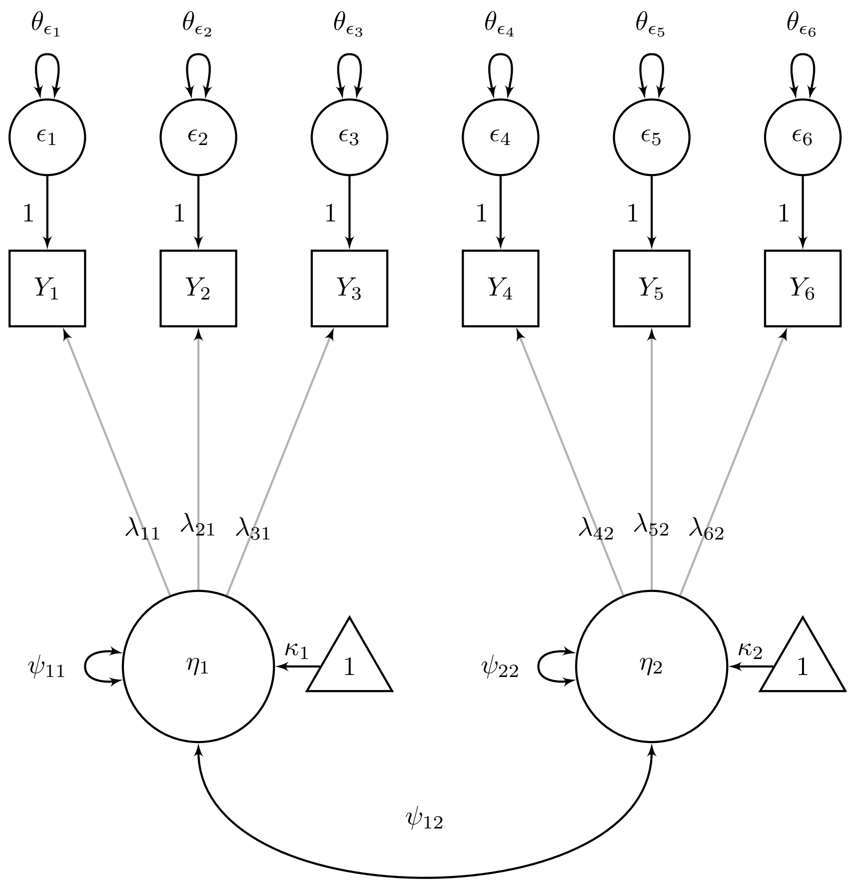
For the **measurement** portion:

- Fix (at least) one *loading* for each factor to some value (usually 1), OR
- Fix the *variance* of each of the latent variables to 1

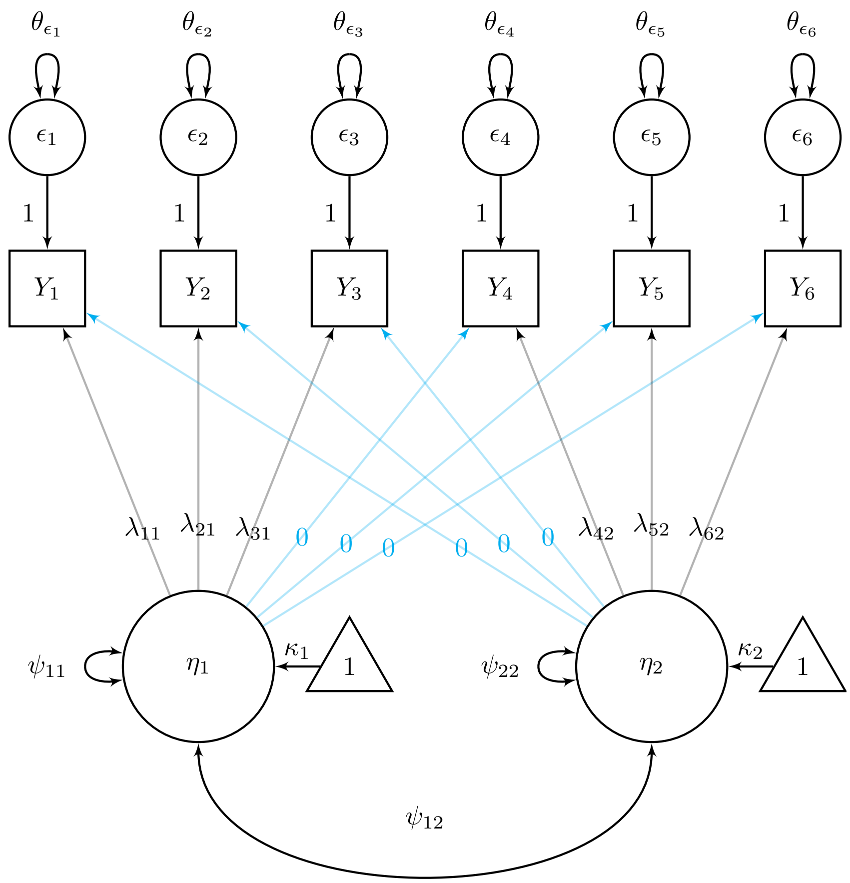
For the **structural** portion:

- Fix one item *intercept* from each factor to a value, OR
- Fix the latent variable *means* to 0

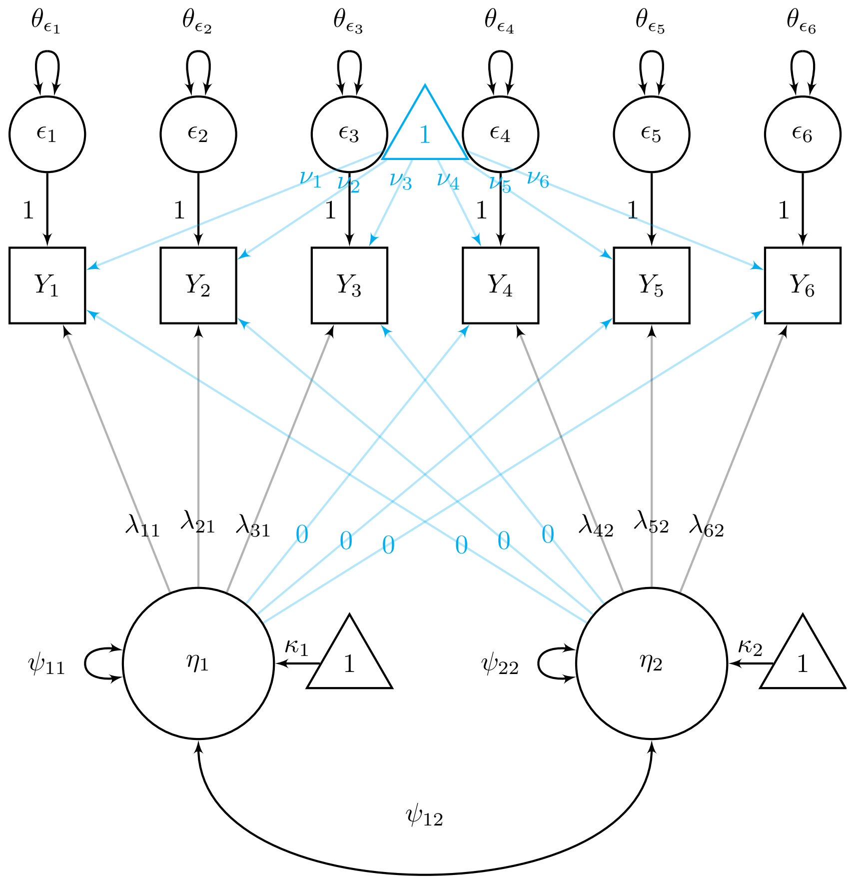
2.3.3 Full CFA figure



2.3.4 Full CFA figure



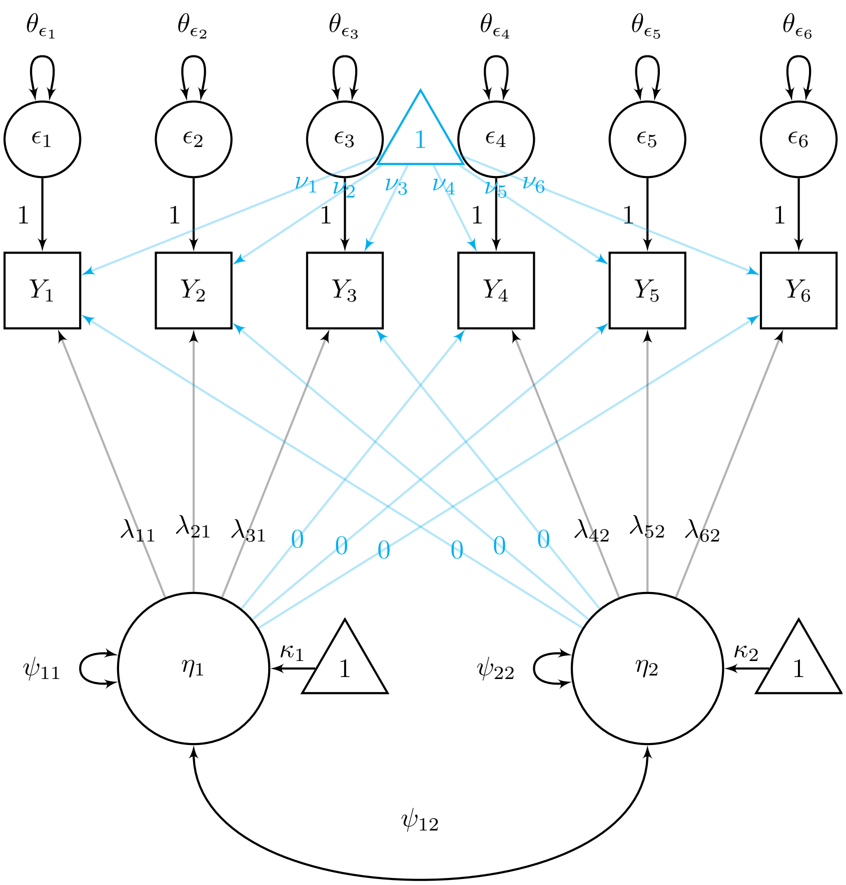
2.3.5 Full CFA figure



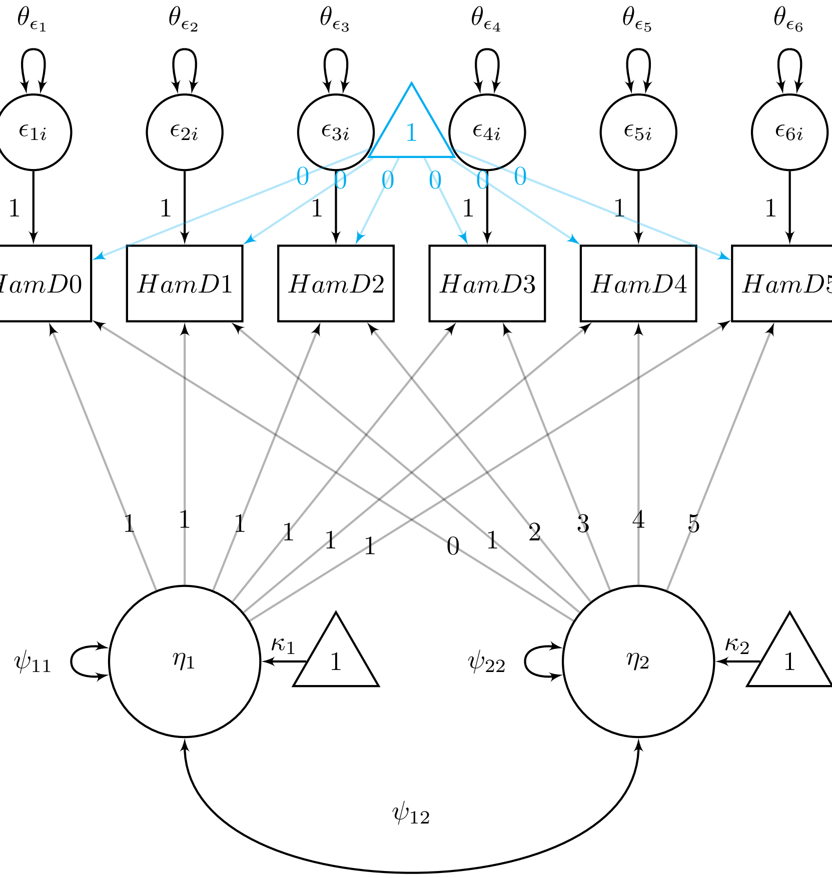
3 LGM related to CFA and mixed models

3.1 LGM as special CFA

3.1.1 Confirmatory factor analysis



3.1.2 SEM latent growth model



3.1.3 SEM latent growth model

The SEM linear latent growth model is actually a two-factor CFA with **very specialized** set of constraints

The measurement portion of CFA is *somewhat* analogous to level 1

The structural (mean) portion is *somewhat* analogous to level 2

BUT the SEM growth model doesn't really use the multilevel framework

- No level 1 and level 2

3.1.4 How is a LGM different from a CFA?

Confirmatory factor analysis

- Typically fixes means of each latent factor to 0
- Focus of CFA is usually **not** on the means
- Focus of CFA is usually on covariances

Latent growth models

- Intercept and slope are latent factors
- **Latent variable means** are what we want to know about
- Instead, we will fix the **item means** to 0

- This seems really weird, but it "pushes" all the information about means to the latent variables

3.1.5 Methods of identifying the LGM

Need to impose contraints in **2 parts** of the model

For the **measurement** portion

- Fix (at least) one *loading* for each factor to **some value**, OR
- ~~Fix the variance of the latent variables to 1~~

For the **mean** portion

- Fix one item *intercept* from each factor to a value, OR
- ~~Fix the latent variable means to 0~~

3.2 LGM as mixed model

3.2.1 Return to mixed models

In the longitudinal mixed model, linear growth for an individual is:

$$Y_{ij} = \pi_{0i} + \pi_{1i}(WEEK) + e_{ij}$$

where

- Y_{ij} is the value of person i's outcome at week j
- π_{0i} is person i's intercept
- π_{1i} is person i's slope
- e_{ij} is person i's error at week j

3.2.2 Equations at different time points

The outcome was measured at 6 different weeks
(Everyone was measured at the same time)
Center at first week, so WEEK is now coded 0, 1, 2, 3, 4, 5

$$Y_{0i} = \pi_{0i} + \pi_{1i}(0) + e_{ij}$$
$$Y_{1i} = \pi_{0i} + \pi_{1i}(1) + e_{ij}$$
$$Y_{2i} = \pi_{0i} + \pi_{1i}(2) + e_{ij}$$
$$Y_{3i} = \pi_{0i} + \pi_{1i}(3) + e_{ij}$$
$$Y_{4i} = \pi_{0i} + \pi_{1i}(4) + e_{ij}$$

$$Y_{5i} = \pi_{0i} + \pi_{1i}(5) + e_{ij}$$

3.2.3 Equations at different time points

The outcome was measured at 6 different weeks
(Everyone was measured at the same time)
Center at first week, so WEEK is now coded 0, 1, 2, 3, 4, 5

$$Y_{0i} = \pi_{0i}(1) + \pi_{1i}(0) + e_{ij}(1)$$
$$Y_{1i} = \pi_{0i}(1) + \pi_{1i}(1) + e_{ij}(1)$$
$$Y_{2i} = \pi_{0i}(1) + \pi_{1i}(2) + e_{ij}(1)$$
$$Y_{3i} = \pi_{0i}(1) + \pi_{1i}(3) + e_{ij}(1)$$
$$Y_{4i} = \pi_{0i}(1) + \pi_{1i}(4) + e_{ij}(1)$$

$$Y_{5i} = \pi_{0i}(1) + \pi_{1i}(5) + e_{ij}(1)$$

3.2.4 π_{0i} and π_{1i} as latent variables

In the mixed model, the individual slopes and intercepts are **unobserved**

- We talked about the mini-regressions, but we didn't see them

So π_{0i} and π_{1i} were basically **latent variables**

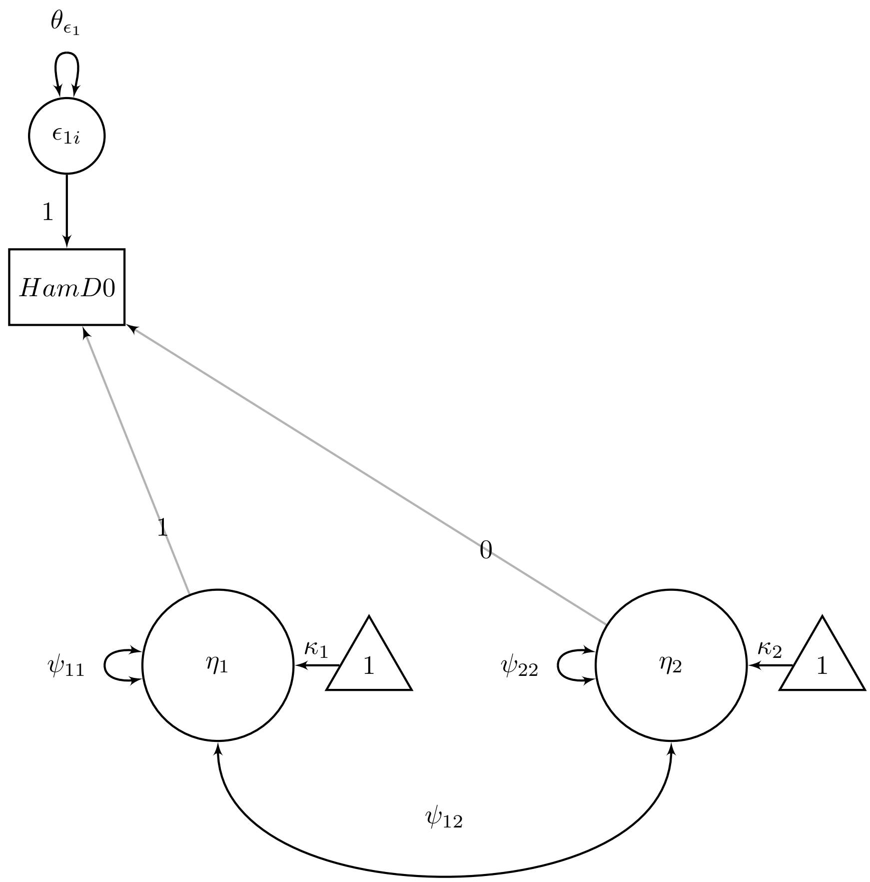
- We infer a person's intercept and slope from their observed values on the Y variables

Mixed models for longitudinal data and SEM latent growth models are, for the most basic models at least, virtually identical

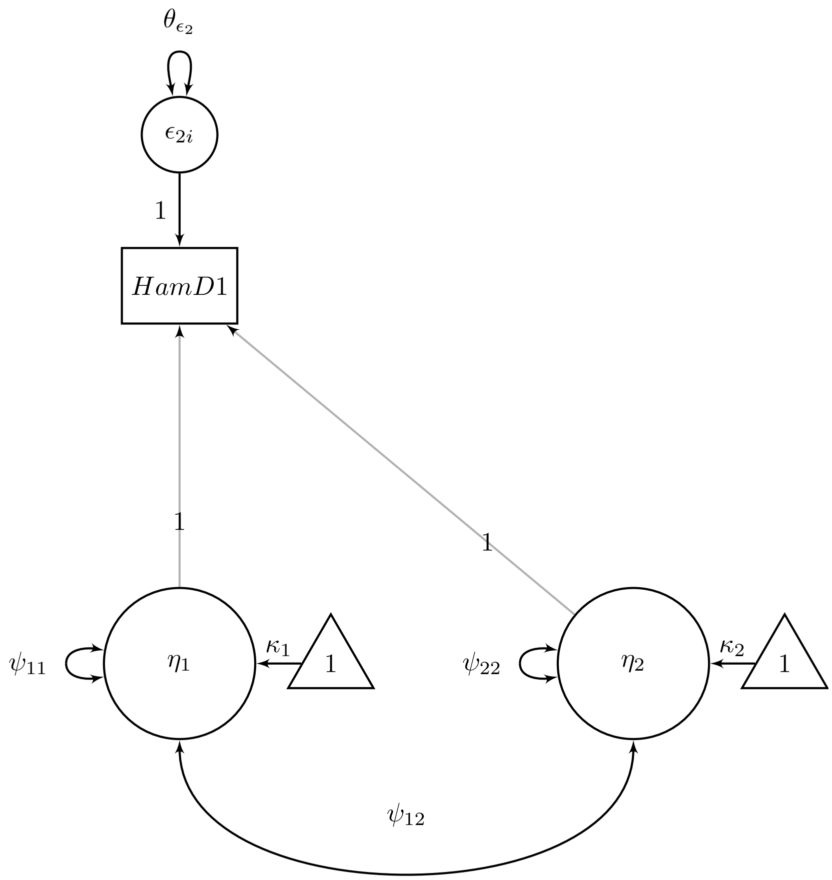
- The SEM growth model explicitly conceives of the intercept and slope as latent variables that are inferred from the observed variables (Ys)

3.3 Measurement model / level 1 equations

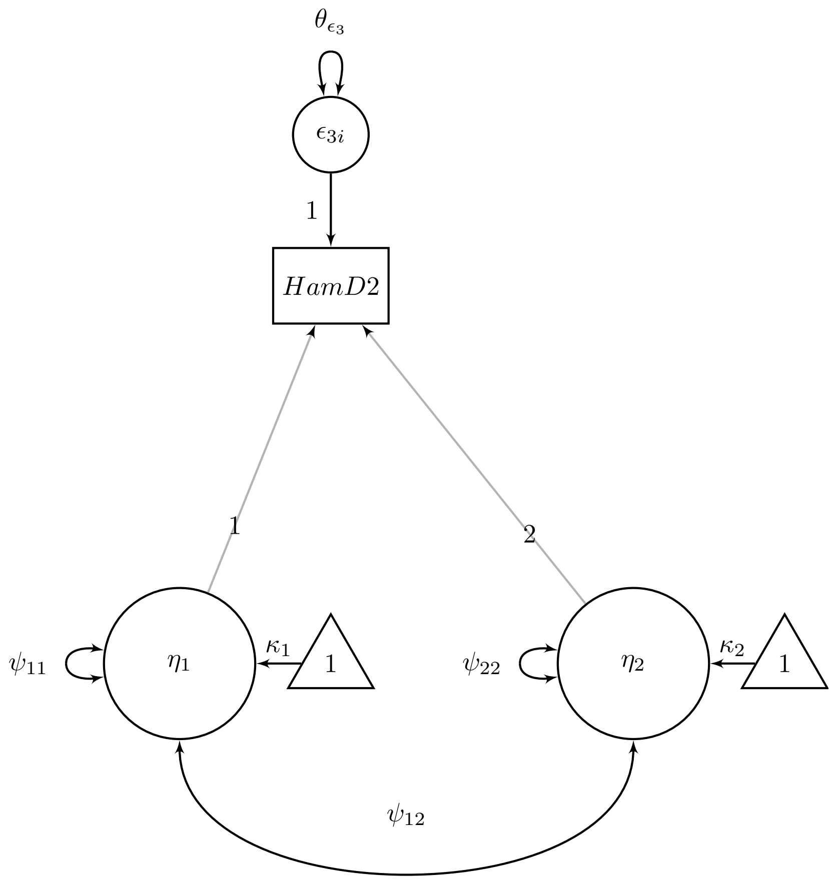
3.3.1 Week 0 equation: $Y_{0i} = \pi_{0i}(1) + \pi_{1i}(0) + e_{ij}(1)$



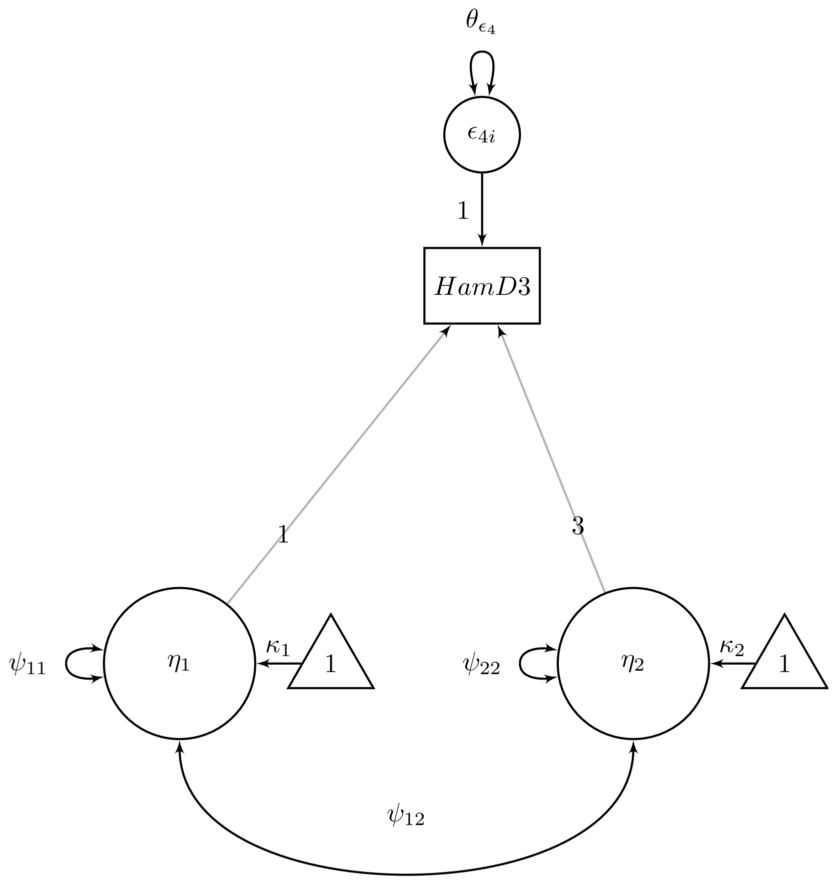
3.3.2 Week 1 equation: $Y_{1i} = \pi_{0i}(1) + \pi_{1i}(1) + e_{ij}(1)$



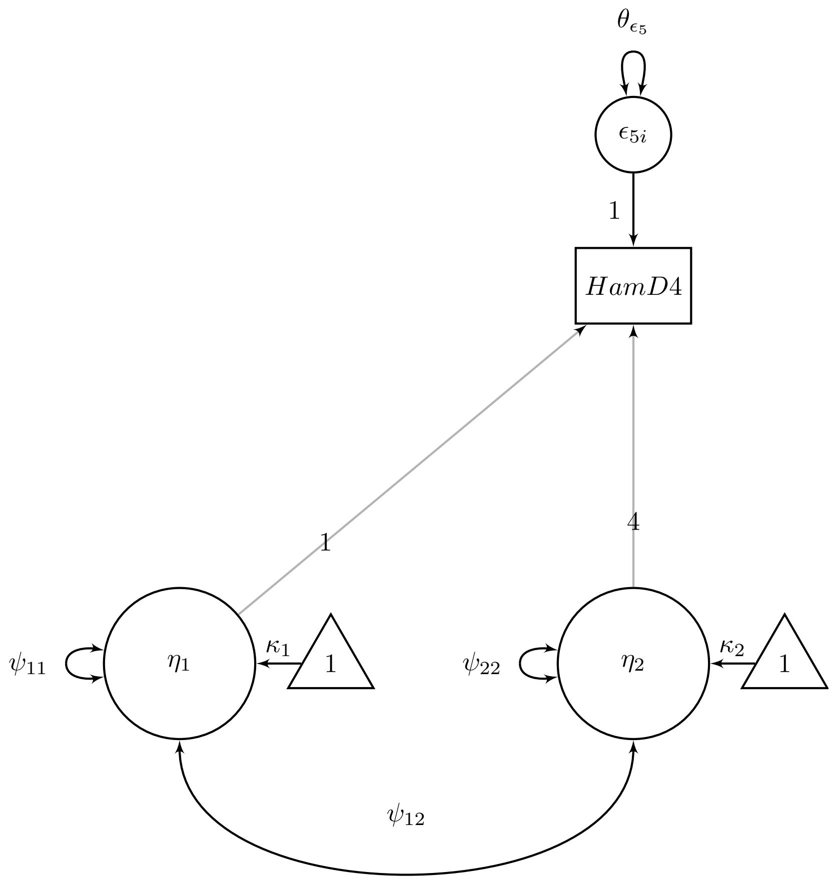
3.3.3 Week 2 equation: $Y_{2i} = \pi_{0i}(1) + \pi_{1i}(2) + e_{ij}(1)$



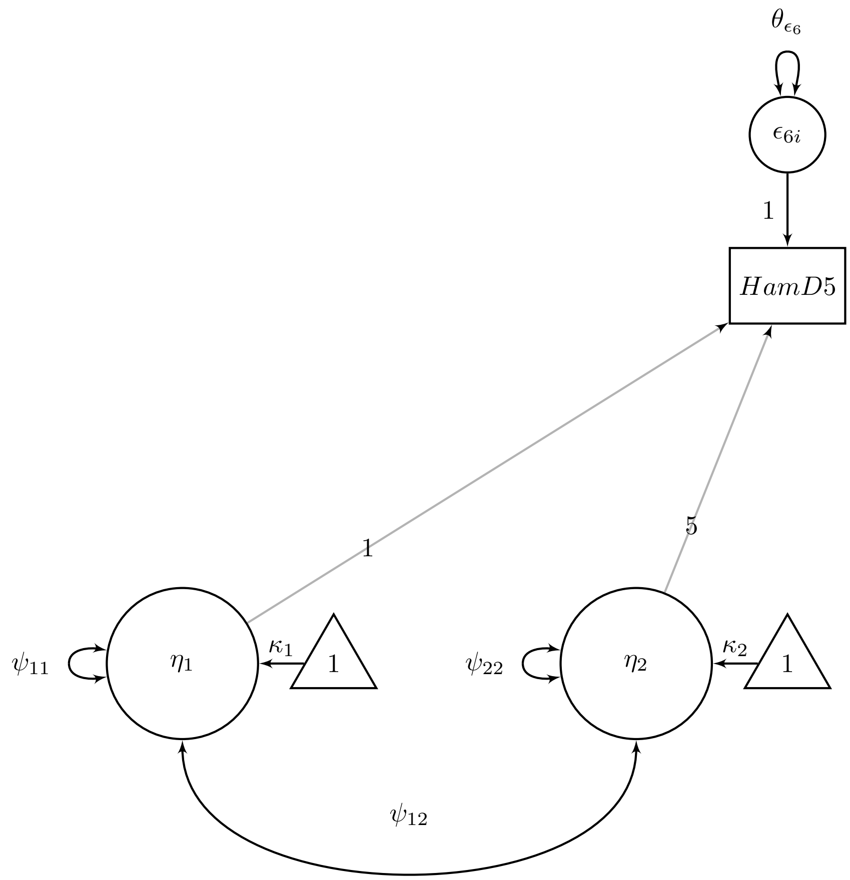
3.3.4 Week 3 equation: $Y_{3i} = \pi_{0i}(1) + \pi_{1i}(3) + e_{ij}(1)$



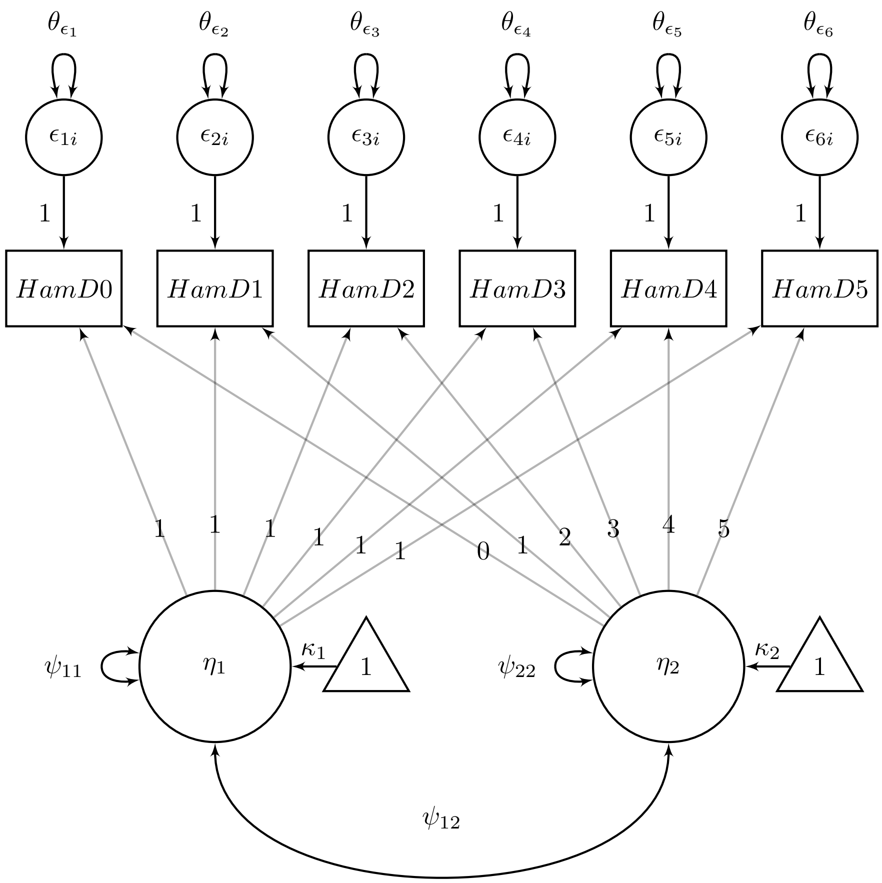
3.3.5 Week 4 equation: $Y_{4i} = \pi_{0i}(1) + \pi_{1i}(4) + e_{ij}(1)$



3.3.6 Week 5 equation: $Y_{5i} = \pi_{0i}(1) + \pi_{1i}(5) + e_{ij}(1)$



3.3.7 SEM latent growth model: measurement



3.4 Structural model / level 2 equations

3.4.1 Mixed model versus SEM LGM

Compare the **equations** for the SEM growth model to the mixed models equations

Compare the **results** from earlier in class today

- Virtually the same, except for different notation
- With no missing data and same assessment schedule for all, they are *numerically* the same

What is different:

SEM growth model does **not** view the data as nested – all variables are at the same "level"

BUT

Remember that we had to *constrain* the LGM to make it match the mixed model

- Constrain the residual variances at each time point to be equal
- Resulted in a *single* value for the residual variance instead of *six* values

3.4.2 SEM growth model: structural

The SEM LGM is a highly constrained CFA **with a mean structure**

The mean structure is represented by two equations:

$$\eta_{1i} = \kappa_1 + \zeta_{1i}$$

$$\eta_{2i} = \kappa_2 + \zeta_{2i}$$

where

- η_{1i} and η_{2i} are person i's latent intercept and slope
- κ_1 and κ_2 are the **means** for the latent intercept and slope
- ζs are residuals capturing individual differences

3.4.3 SEM growth model: latent variable covariance matrix

The SEM growth model (like the mixed model version) doesn't produce intercept and slope residuals for every individual

- The model produces a **covariance matrix** of the residuals

The SEM growth model version is basically the same as MLM

The ψ matrix gives the variances and covariances of the latent variables (i.e., the latent intercept and latent slope)

$$\Psi = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

3.4.4 Mixed model: residual covariance matrix

Longitudinal mixed models produce a **single value** for the level 1 residual

- Error is defined as residuals for each person's trajectory
- We assume that every individual has *about the same error* in their trajectory
- We assume level 1 variance is the same for all individuals

Error is per person, across all time points

We don't (and can't) examine residual error at each time point separately – but we can in SEM

3.4.5 SEM growth model: residual covariance matrix

The SEM growth model produces a *residual covariance matrix* with (default) a unique variance estimate for each time point

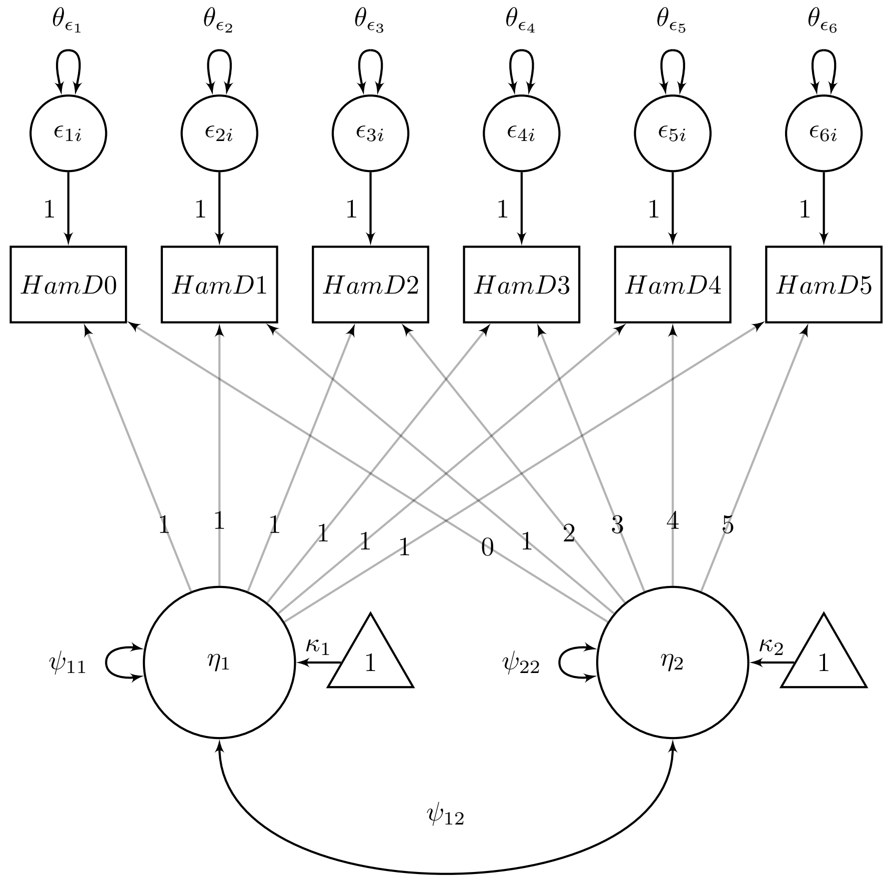
- More flexibility in modeling change and variability over time
- Can have different residual variance at each time, or simplify model and have a single value that's the same for all time points (this is equivalent to mixed model)

$$\Theta = \begin{bmatrix} \theta_{\varepsilon_1} & & & & & \\ 0 & \theta_{\varepsilon_2} & & & & \\ 0 & 0 & \theta_{\varepsilon_3} & & & \\ 0 & 0 & 0 & \theta_{\varepsilon_4} & & \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_5} & \\ 0 & 0 & 0 & 0 & 0 & \theta_{\varepsilon_6} \end{bmatrix} = \begin{bmatrix} {\sigma_{\varepsilon_1}}^2 & & & & & \\ 0 & {\sigma_{\varepsilon_2}}^2 & & & & \\ 0 & 0 & {\sigma_{\varepsilon_3}}^2 & & & \\ 0 & 0 & 0 & {\sigma_{\varepsilon_4}}^2 & & \\ 0 & 0 & 0 & 0 & {\sigma_{\varepsilon_5}}^2 & \\ 0 & 0 & 0 & 0 & 0 & {\sigma_{\varepsilon_6}}^2 \end{bmatrix}$$

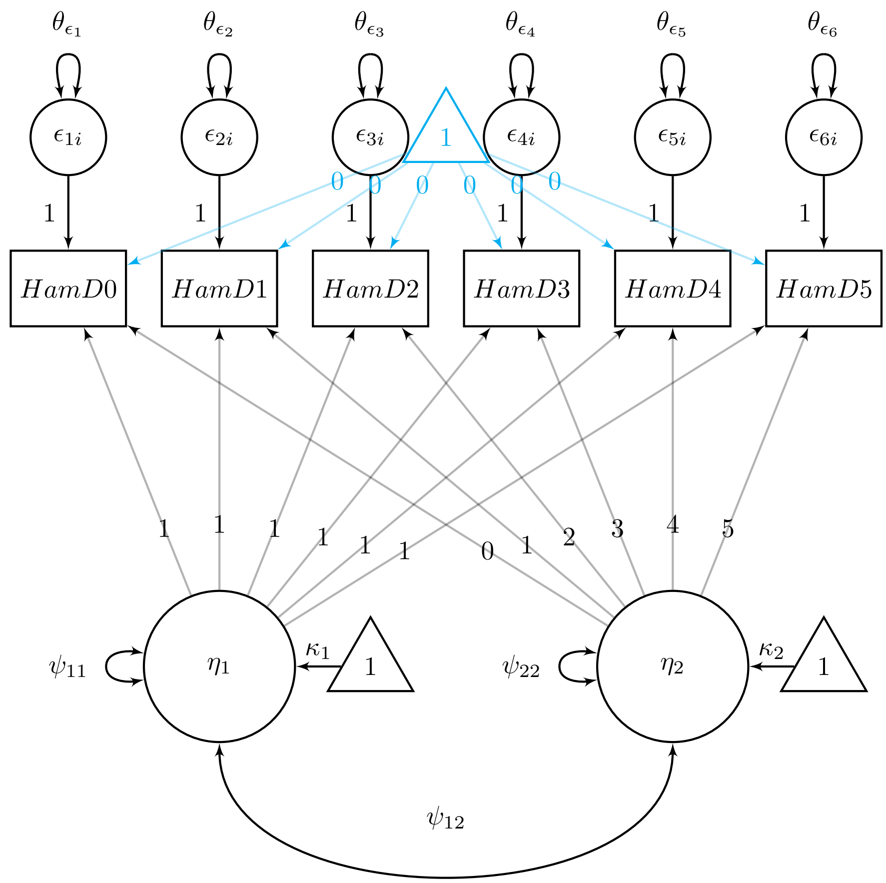
Error is per time point, across all people

3.5 Summary

3.5.1 Full SEM growth model



3.5.2 Full SEM growth model



3.5.3 Full SEM growth model: optional equal residual variances