

SC705: Advanced Statistics
Instructor: Natasha Sarkisian
Class notes: Latent Growth Models in SEM

In order to understand the implementation of latent growth models in SEM, we need to first consider the issue of SEM with mean structures.

Mean structures

So far in using SEM we were only dealing with covariances. Oftentimes, however, we are also interested in means – either their absolute value or how they differ by group (especially means of latent variables).

This type of analysis requires both the covariance matrix and the means. Essentially, what it does is it introduces intercepts into the measurement models and the structural model:

That is, so far we used:

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \varepsilon$$

$$\eta = B\eta + \Gamma\xi + \zeta$$

Now we add the intercepts:

$$X = \tau_x + \Lambda_x \xi + \delta$$

$$Y = \tau_y + \Lambda_y \eta + \varepsilon$$

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta$$

So we have four extra vectors now:

τ_x is the vector of means for indicators x

τ_y is the vector of means for indicators y

α is the vector of means (really, intercepts) of endogenous latent variables

κ is the vector of means of exogenous latent variables

See handout, pp.306-307 from Byrne

The way we can represent that graphically is by introducing the constant into the diagram: See Kline, p.266, p. 270

Identification of models with means:

In models with means we need to take into account whether the mean structure is identified. The rule is that the total number of means and intercepts cannot exceed the total number of means of observed variables. We can also count the total number of data point and total number of parameters by counting means and intercepts as parameters and the number of data points as $n*(n+3)/2$.

Means of latent factors: Comparing means across groups

Note that the identification constraints do not allow us to have a model with means of all indicators evaluated alongside the mean for the latent factor – we have to assume the mean of the latent factor to be zero. However, if we compare two or more groups, we can evaluate the relative size of factor means. To do that, we set the mean in one group to zero and freely evaluate the means of other groups. The parameters for these other groups will indicate the differences of latent variable means in these groups from group 1, and significance tests for these parameters will serve as significance tests for those differences.

So if we have two groups, then in group 1, we would specify vectors TX and TY as free, and KA and AL as fixed to zero. Then in group 2, we would specify vectors TX and TY as invariant (as well as LX and LY – this kind of comparison only makes sense if the factors are identically defined in both groups), and KA and AL as freely estimated.

Latent growth models

The idea of growth models in SEM is the same as in HLM: we allow starting values and the trajectories to vary from person to person, and calculate average trajectory as well as the amount of variance around it; then we try to explain that variance. So the intercept and the slope (effect of time) in HLM were random variables. But in SEM we conceptualize both the intercept and the growth slope as latent variables.

Example: figure 10.4 on p.282 in Kline.

Note that the factor loadings for the intercept should all be set to 1. Factor loadings for the slope, however, can be specified differently, depending on the time intervals between the observations. In this example, all time intervals are equal, therefore the distances between the values of factor loadings are also equal. The factor loadings also depend on which time point we want to become the intercept. For instance, in this example, the first time point is selected to be the intercept, but in the example that we'll do below, third time point will be the intercept.

Note that we also need to specify the mean structure for those latent variables in order to be able to get the mean values for them (like in HLM, where we had fixed effects and variance components, here too we want to have the mean value and the variance estimate for intercept and slope).

One advantage of doing this model in LISREL rather than in HLM is that in LISREL we can allow for correlated measurement errors (typically, serially correlated, like in the picture). A disadvantage, however, is that we have to have equal number of observations per person, and they have to be done at the same time – this stems from the way the data have to be structured for this type of analysis.

LISREL example

For an example of doing this in LISREL, we'll use the same data we used with HLM: NYS2 in Chapter 9 of HLM6. But, here we need to structure it differently. To prepare the data, I merged Nys21.sav and Nys22.sav into a single file (matched on id), that has the following variables:

attit
expo
age
ages
age14
age16
age145
age14s
age16s
age145s
id
female
minority
income

I transferred it to Stata using StatTransfer program, and then did the following:

```
drop ages-age145s  
reshape wide attit expo, i(id) j(age)
```

The resulting dataset contains:

id
attit14
expo14
attit15
expo15
attit16
expo16
attit17
expo17
attit18
expo18
female
minority
income
minfem

I transferred it back to SPSS to import it into LISREL. This file (nys2.sav) is available on the course website. Upon importing the data, we should define variables and obtain the covariance matrix and the means – these will be in files nys.cov and meansnys.meas.

```
!Preliis syntax  
SY='C:\nys2.PSF'  
OU MA=CM SM=nys.cov ME=meansnys.meas
```

Just to remind you, here's how the basic model of change looked in HLM:

Level-1 Model

$$Y = B0 + B1*(AGE16) + R$$

Level-2 Model

$$B0 = G00 + U0$$

$$B1 = G10 + U1$$

Sigma_squared = 0.02873

Tau

INTRCPT1,B0	0.04572	-0.00093
AGE16,B1	-0.00093	0.00313

Tau (as correlations)

INTRCPT1,B0	1.000	-0.078
AGE16,B1	-0.078	1.000

The outcome variable is ATTIT

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.493325	0.014864	33.189	240	0.000
For AGE16 slope, B1					
INTRCPT2, G10	0.032357	0.005350	6.048	240	0.000

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.21383	0.04572	235	1754.38522	0.000
AGE16 slope, U1	0.05595	0.00313	235	446.20764	0.000
level-1, R	0.16949	0.02873			

Like in HLM, first we want to start with the basic change model, without any explanatory variables.

TI Change only (random intercept and slope) model for attitude

DA NI=15 NO=241 MA=CM

LA

ID ATTIT14 EXPO14 ATTIT15 EXPO15 ATTIT16 EXPO16 ATTIT17

EXPO17 ATTIT18 EXPO18 FEMALE MINORITY INCOME MINFEM

CM=C:\nys.cov

ME =C:\meansnys.me

SE

2 4 6 8 10/

MO NX=5 NK=2 LX=FU, FI PH=SY,FR TD=SY, FI TX=FI KA=FR

LK

INTERCPT SLOPE

FR TD 1 1 TD 2 2 TD 3 3 TD 4 4 TD 5 5 TD 2 1 TD 3 2 TD 4 3 TD 5 4

VA 1.0 LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1

VA -2.0 LX 1 2

VA -1.0 LX 2 2

VA 0.0 LX 3 2

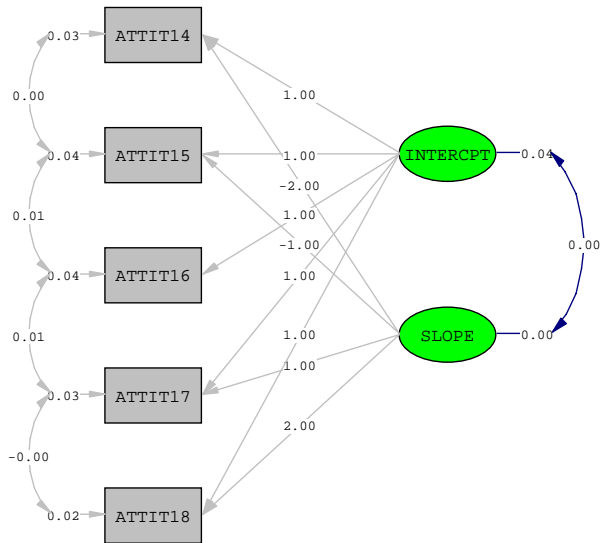
VA 1.0 LX 4 2

VA 2.0 LX 5 2

PD

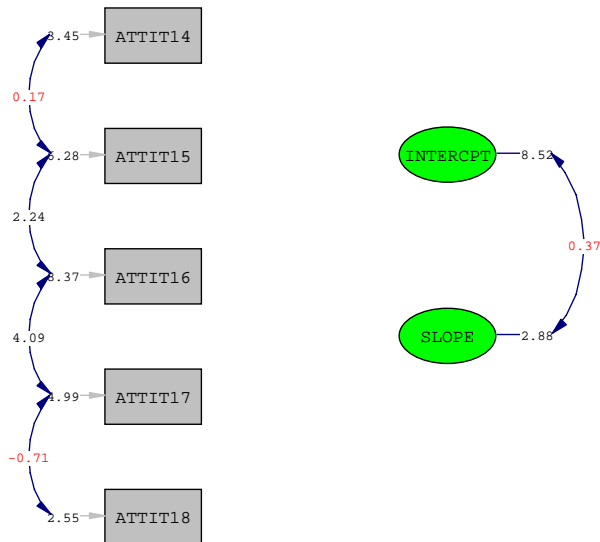
OU

Estimates:

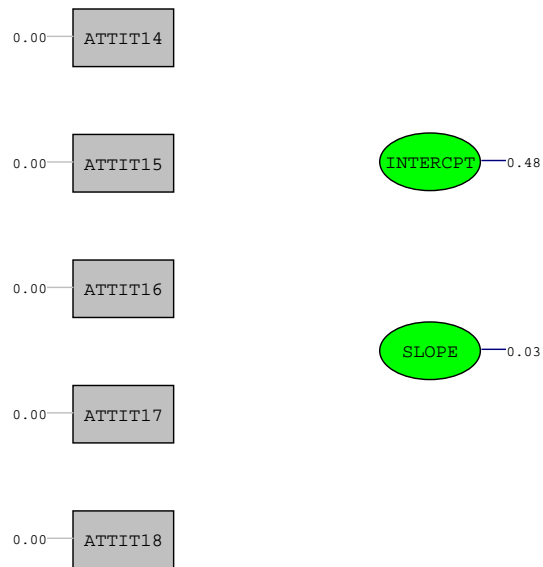


Chi-Square=21.86, df=6, P-value=0.00129, RMSEA=0.105

Significances:



Means:



Now let's estimate the same change model but with a quadratic term:

TI Change only (random intercept and slope) model for attitude, with quadratic term

DA NI=15 NO=241 MA=CM

LA

ID ATTIT14 EXPO14 ATTIT15 EXPO15 ATTIT16 EXPO16 ATTIT17

EXPO17 ATTIT18 EXPO18 FEMALE MINORITY INCOME MINFEM

CM=C:\nys.cov

ME =C:\meansnys.me

SE

2 4 6 8 10/

MO NX=5 NK=3 LX=FU, FI PH=SY,FR TD=SY, FI TX=FI KA=FR

LK

INTERCPT SLOPE SLOPE2

FR TD 1 1 TD 2 2 TD 3 3 TD 4 4 TD 5 5 TD 2 1 TD 3 2 TD 4 3 TD 5 4

VA 1.0 LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1

VA -2.0 LX 1 2

VA -1.0 LX 2 2

VA 0.0 LX 3 2

VA 1.0 LX 4 2

VA 2.0 LX 5 2

VA 4.0 LX 1 3

VA 1.0 LX 2 3

VA 0.0 LX 3 3

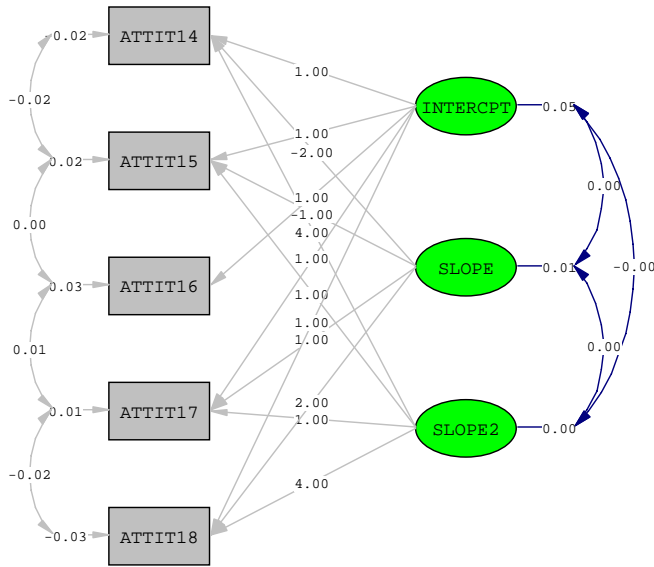
VA 1.0 LX 4 3

VA 4.0 LX 5 3

PD

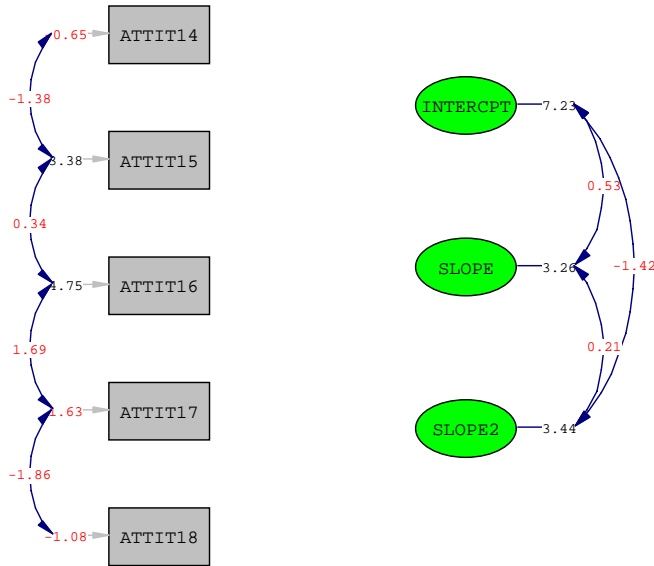
OU

Estimates:

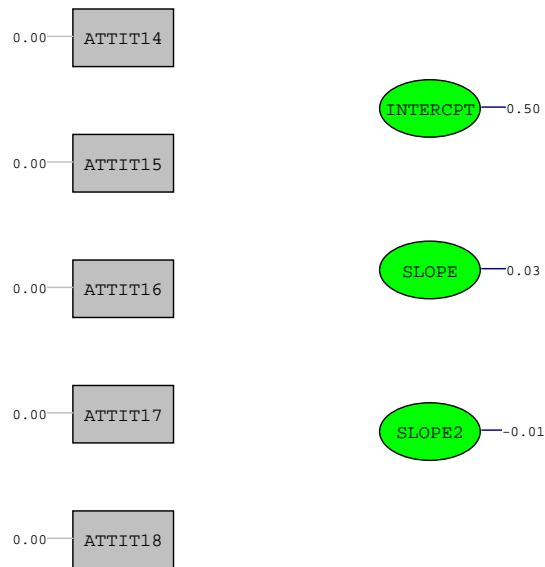


Chi-Square=3.07, df=2, P-value=0.21554, RMSEA=0.047

T-values:



Means:



Check whether there is a significant improvement in chi-square:

$$21.86 - 3.07 = 18.79, \text{ df} = 6 - 2 = 4$$

Alpha = .01 critical value for df = 4 is 13.28, so it's a significant improvement. We can also see that in RMSEA and chi-square significance.

The second step of this process is to predict change. Here, we will predict change using time-invariant (i.e. level 2) variables, GENDER, MINORITY, and INCOME:

TI Predicting change in the random intercept and slope for attitude, with quadratic term

DA NI=15 NO=241 MA=CM

LA

ID ATTIT14 EXPO14 ATTIT15 EXPO15 ATTIT16 EXPO16 ATTIT17

EXPO17 ATTIT18 EXPO18 FEMALE MINORITY INCOME MINFEM

CM=nys.cov

ME =meansnys.mea

SE

2 4 6 8 10 12 13 14/

MO NY=5 NE=3 NX=3 NK=3 LX=ID LY=FU,FI PH=SY,FR PS=SY,FR TD=ZE TE=SY, FI

TY=FI TX=FI KA=FR AL=FR GA=FR

LK

FEMALE MINORITY INCOME

LE

INTERCPT SLOPE SLOPE2

FR TE 1 1 TE 2 2 TE 3 3 TE 4 4 TE 5 5 TE 2 1 TE 3 2 TE 4 3 TE 5 4

VA 1.0 LY 1 1 LY 2 1 LY 3 1 LY 4 1 LY 5 1

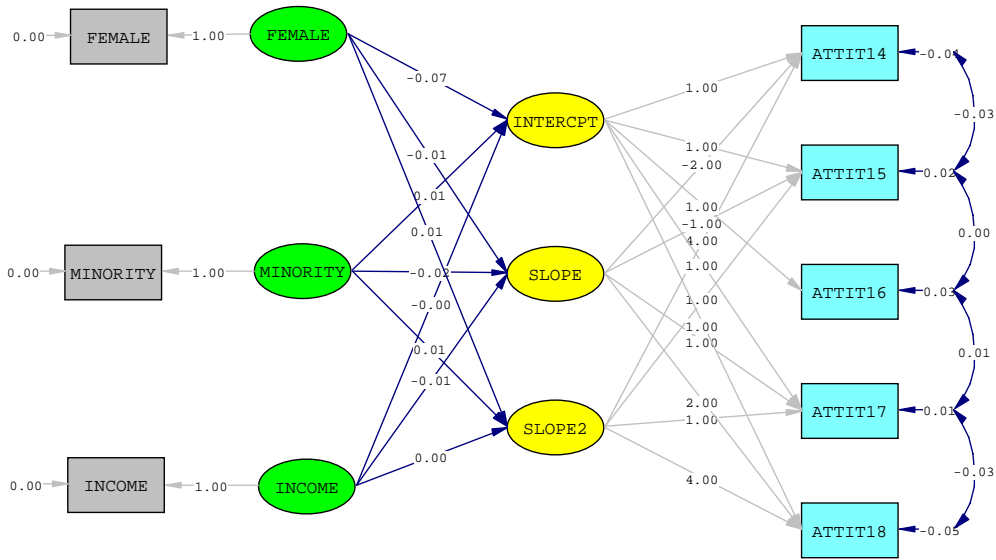
VA -2.0 LY 1 2

VA -1.0 LY 2 2

VA 0.0 LY 3 2

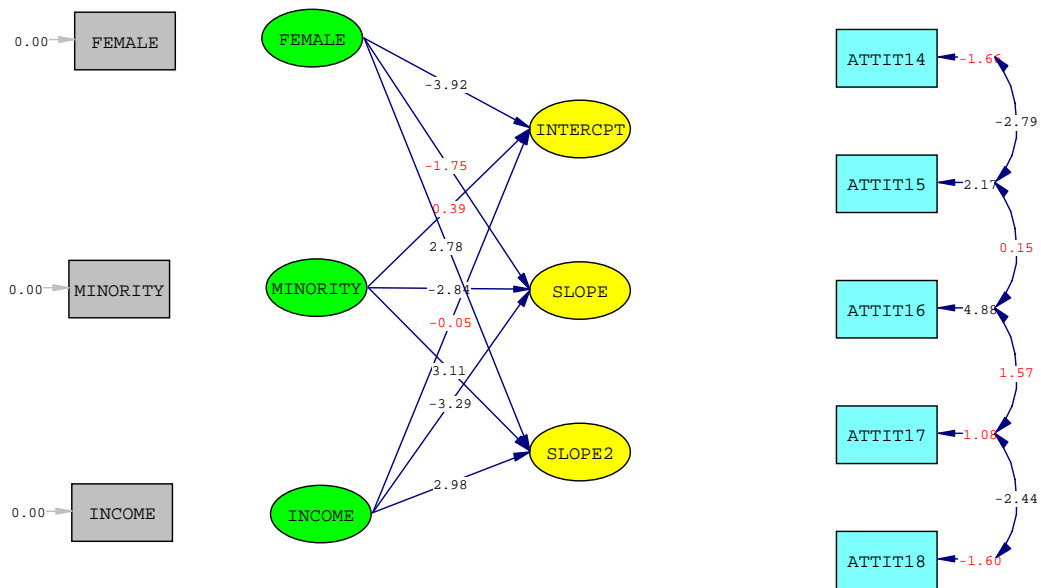
VA 1.0 LY 4 2
 VA 2.0 LY 5 2
 VA 4.0 LY 1 3
 VA 1.0 LY 2 3
 VA 0.0 LY 3 3
 VA 1.0 LY 4 3
 VA 4.0 LY 5 3
 PD
 OU

Estimates:

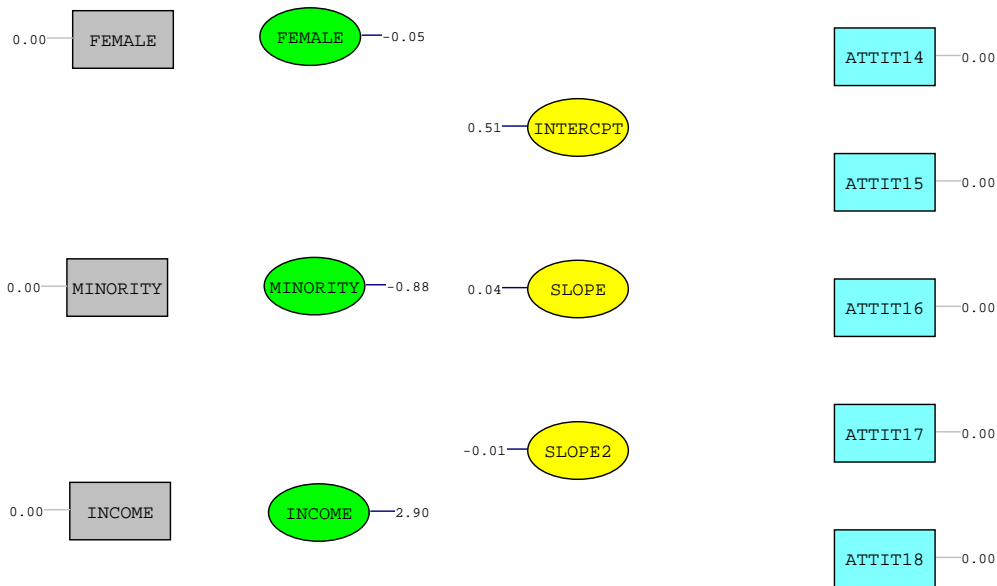


Chi-Square=52.27, df=8, P-value=0.00000, RMSEA=0.152

T-values:



Means:



Example:

Wright, John Paul, David E. Carter, and Francis T. Cullen. 2005. "A Life-Course Analysis of Military Service in Vietnam." *Journal of Research in Crime and Delinquency*, 42(1), 55-83.