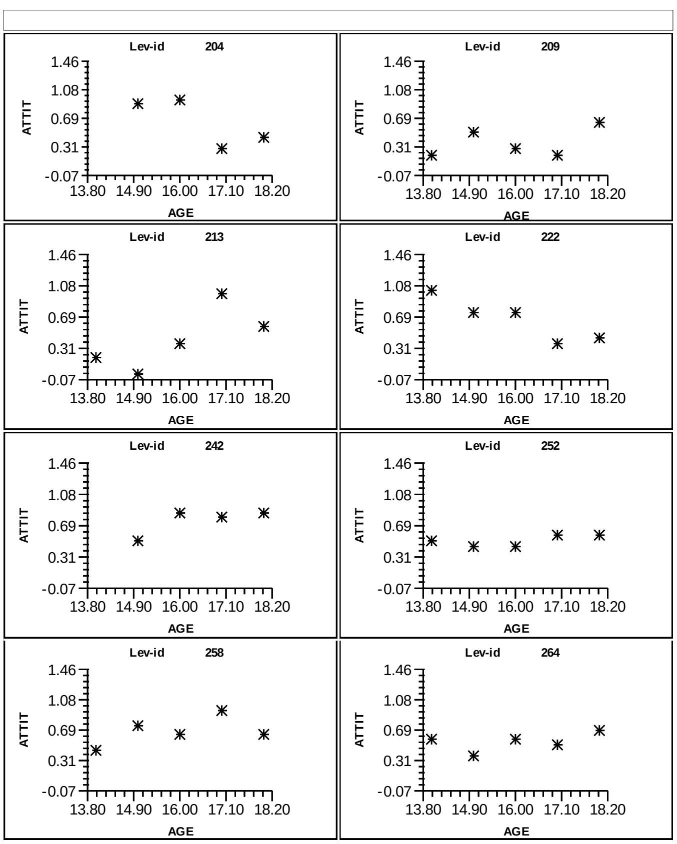
SC705: Advanced Statistics
Instructor: Natasha Sarkisian
Class notes: Longitudinal Data Analysis in HLM and SEM

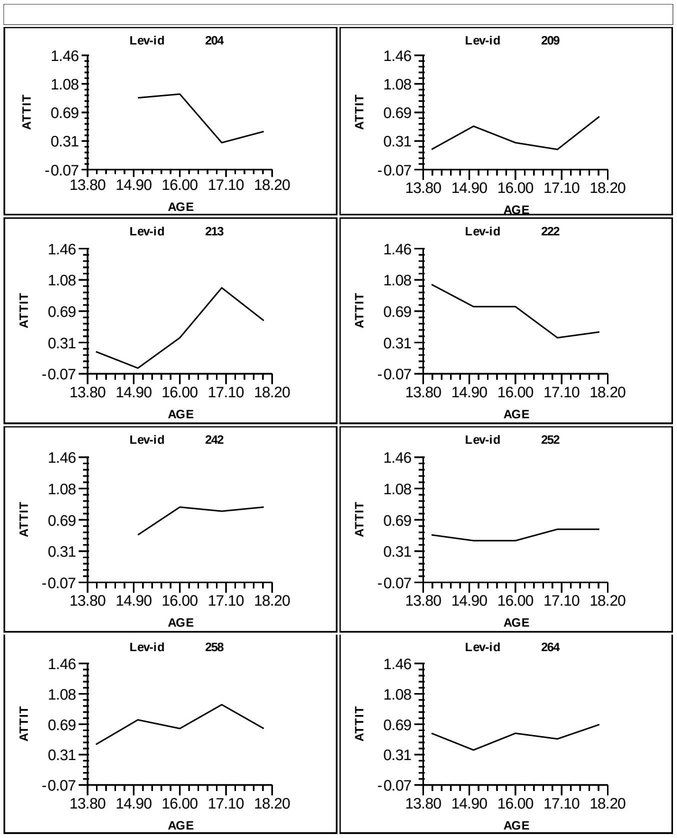
Growth Curve Models in HLM

So far, when using HLM, we've worked with one type of hierarchical data – students nested within schools. HLM can also be used to model longitudinal data where multiple observations over time are nested within one person.

We will use NYS2.MDM from Chapter 9 folder. This file contains data for a cohort of adolescents in the National Youth Survey, ages 14 to 18. The dependent variable ATTIT is a 9-item scale assessing attitudes favorable to deviant behavior (property damage, drug and alcohol use, stealing, etc.). The level-1 independent variables include: EXPO measuring exposure to deviant peers (students were asked how many of their friends engaged in the 9 deviant behaviors), AGE (age in years), AGES (age in years squared), AGE14 (age minus 14), AGE16 (age minus 16), AGE145 (age minus 14.5), and the three corresponding squared variables. Level 2 include person-level variables: FEMALE, MINORITY, INCOME, and an interaction term for MINFEM.

What we will study is how attitudes change over time, and what shapes that change. First, let's examine individual trajectories.





Now let's try to model these trajectories. First, we will assume that we can model them using a linear model. Therefore, we'll estimate an unconditional linear growth model:

```
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

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.837
     AGE16, B1                        0.453
 ----------------------------------------------------

The outcome variable is    ATTIT
```

```
Final estimation of fixed effects:
----------------------------------------------------------------------------
                                       Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   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
----------------------------------------------------------------------------

The outcome variable is    ATTIT

Final estimation of fixed effects
(with robust standard errors)
----------------------------------------------------------------------------
                                       Standard             Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00          0.493325   0.014833    33.259       240    0.000
For    AGE16 slope, B1
   INTRCPT2, G10          0.032357   0.005338     6.061       240    0.000
----------------------------------------------------------------------------

Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
-----------------------------------------------------------------------------
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
-----------------------------------------------------------------------------

Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = -99.676230
 Number of estimated parameters = 4
```

The mean growth trajectory is:
Attitude=.493 + .032*Age16

Now let’s estimate an unconditional quadratic growth model and compare the fit:

```
Level-1 Model

       Y = B0 + B1*(AGE16) + B2*(AGE16S) + R

Level-2 Model
       B0 = G00 + U0
       B1 = G10 + U1
       B2 = G20 + U2

 Sigma_squared =      0.02291


Tau
 INTRCPT1,B0      0.05825      -0.00033      -0.00416
    AGE16,B1     -0.00033       0.00369      -0.00033
```

```
  AGE16S,B2     -0.00416      -0.00033       0.00118


Tau (as correlations)
 INTRCPT1,B0  1.000 -0.022 -0.502
    AGE16,B1 -0.022  1.000 -0.160
   AGE16S,B2 -0.502 -0.160  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.822
     AGE16, B1                        0.530
    AGE16S, B2                        0.358
 ----------------------------------------------------

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          0.514018   0.017307    29.700       240    0.000
 For    AGE16 slope, B1
    INTRCPT2, G10          0.031463   0.005333     5.900       240    0.000
 For   AGE16S slope, B2
    INTRCPT2, G20         -0.010696   0.003652    -2.929       240    0.004
 ----------------------------------------------------------------------------

 Final estimation of fixed effects
 (with robust standard errors)
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          0.514018   0.017270    29.764       240    0.000
 For    AGE16 slope, B1
    INTRCPT2, G10          0.031463   0.005320     5.914       240    0.000
 For   AGE16S slope, B2
    INTRCPT2, G20         -0.010696   0.003643    -2.936       240    0.004
 ----------------------------------------------------------------------------

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        0.24135       0.05825   222    1247.17000    0.000
    AGE16 slope, U1        0.06075       0.00369   222     503.78215    0.000
   AGE16S slope, U2        0.03437       0.00118   222     347.59593    0.000
  level-1,       R         0.15136       0.02291
 -----------------------------------------------------------------------------
 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = -129.616127
 Number of estimated parameters = 7
```

The average growth trajectory becomes:
Attitude = 0.514+.031*Age16 – 0.011*Age16S

Our quadratic model does have smaller deviance value, but let's test the quadratic model against the linear model:

```
Variance-Covariance components test
 -----------------------------------
 Chi-square statistic         =     29.93990
 Number of degrees of freedom =    3
 P-value                      = 0.000
```

We conclude that quadratic model is a better fit, and proceed to estimating conditional models using person-level (time-invariant) predictors at first.

```
  The model specified for the fixed effects was:
 ----------------------------------------------------

   Level-1                   Level-2
   Coefficients             Predictors
 ----------------------   ---------------
         INTRCPT1, B0      INTRCPT2, G00
                             FEMALE, G01
                           MINORITY, G02
$                            INCOME, G03
      AGE16 slope, B1      INTRCPT2, G10
                             FEMALE, G11
                           MINORITY, G12
$                            INCOME, G13
     AGE16S slope, B2      INTRCPT2, G20
                             FEMALE, G21
                           MINORITY, G22
$                            INCOME, G23

'$' - This level-2 predictor has been centered around its grand mean.


Level-1 Model

       Y = B0 + B1*(AGE16) + B2*(AGE16S) + R

Level-2 Model
       B0 = G00 + G01*(FEMALE) + G02*(MINORITY) + G03*(INCOME) + U0
       B1 = G10 + G11*(FEMALE) + G12*(MINORITY) + G13*(INCOME) + U1
       B2 = G20 + G21*(FEMALE) + G22*(MINORITY) + G23*(INCOME) + U2

Sigma_squared =      0.02291

 Tau
 INTRCPT1,B0      0.05662      -0.00042      -0.00391
    AGE16,B1     -0.00042       0.00364      -0.00025
   AGE16S,B2     -0.00391      -0.00025       0.00112

Tau (as correlations)
 INTRCPT1,B0  1.000 -0.029 -0.492
    AGE16,B1 -0.029  1.000 -0.122
   AGE16S,B2 -0.492 -0.122  1.000

  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.818
     AGE16, B1                        0.527
    AGE16S, B2                        0.346
 ----------------------------------------------------
```

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.562491	0.025856	21.754	237	0.000
FEMALE, G01	-0.100283	0.034929	-2.871	237	0.005
MINORITY, G02	-0.019852	0.044100	-0.450	237	0.653
INCOME, G03	0.003602	0.007755	0.464	237	0.642
For AGE16 slope, B1					
INTRCPT2, G10	0.039149	0.008110	4.827	237	0.000
FEMALE, G11	-0.003239	0.010823	-0.299	237	0.765
MINORITY, G12	-0.028441	0.013824	-2.057	237	0.040
INCOME, G13	-0.003963	0.002373	-1.670	237	0.096
For AGE16S slope, B2					
INTRCPT2, G20	-0.019852	0.005501	-3.609	237	0.001
FEMALE, G21	0.014754	0.007364	2.003	237	0.046
MINORITY, G22	0.012461	0.009468	1.316	237	0.190
INCOME, G23	0.002798	0.001620	1.727	237	0.085

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.562491	0.029658	18.966	237	0.000
FEMALE, G01	-0.100283	0.034379	-2.917	237	0.004
MINORITY, G02	-0.019852	0.039082	-0.508	237	0.611
INCOME, G03	0.003602	0.006930	0.520	237	0.603
For AGE16 slope, B1					
INTRCPT2, G10	0.039149	0.007686	5.094	237	0.000
FEMALE, G11	-0.003239	0.010304	-0.314	237	0.753
MINORITY, G12	-0.028441	0.014088	-2.019	237	0.044
INCOME, G13	-0.003963	0.002075	-1.910	237	0.057
For AGE16S slope, B2					
INTRCPT2, G20	-0.019852	0.006129	-3.239	237	0.002
FEMALE, G21	0.014754	0.007121	2.072	237	0.039
MINORITY, G22	0.012461	0.009555	1.304	237	0.194
INCOME, G23	0.002798	0.001383	2.023	237	0.044

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.23795	0.05662	219	1196.00045	0.000
AGE16 slope,	U1	0.06030	0.00364	219	495.23926	0.000
AGE16S slope,	U2	0.03342	0.00112	219	336.68827	0.000
level-1,	R	0.15135	0.02291			

Finally, let’s estimate a quadratic growth model with a time-varying covariate (EXPO). Here, we will use EXPO grand-centered. If we wanted to take this analysis one step further, we could have created a mean exposure variable on person level (level 2) and then used EXPO group centered on level 1 and mean of EXPO on level 2.

```
   Level-1                  Level-2
   Coefficients             Predictors
 ----------------------   ---------------
        INTRCPT1, B0      INTRCPT2, G00
                            FEMALE, G01
                          MINORITY, G02
$                           INCOME, G03
 %     EXPO slope, B1     INTRCPT2, G10
                            FEMALE, G11
                          MINORITY, G12
$                           INCOME, G13
      AGE16 slope, B2     INTRCPT2, G20
                            FEMALE, G21
                          MINORITY, G22
$                           INCOME, G23
     AGE16S slope, B3     INTRCPT2, G30
                            FEMALE, G31
                          MINORITY, G32
$                           INCOME, G33

'%' - This level-1 predictor has been centered around its grand mean.
'$' - This level-2 predictor has been centered around its grand mean.


Level-1 Model

       Y = B0 + B1*(EXPO) + B2*(AGE16) + B3*(AGE16S) + R

Level-2 Model
       B0 = G00 + G01*(FEMALE) + G02*(MINORITY) + G03*(INCOME) + U0
       B1 = G10 + G11*(FEMALE) + G12*(MINORITY) + G13*(INCOME) + U1
       B2 = G20 + G21*(FEMALE) + G22*(MINORITY) + G23*(INCOME) + U2
       B3 = G30 + G31*(FEMALE) + G32*(MINORITY) + G33*(INCOME) + U3

 Sigma_squared =      0.02030

 Tau
 INTRCPT1,B0      0.02273      -0.00288       0.00068      -0.00147
     EXPO,B1     -0.00288       0.03327      -0.00273       0.00185
    AGE16,B2      0.00068      -0.00273       0.00276      -0.00034
   AGE16S,B3     -0.00147       0.00185      -0.00034       0.00058


Tau (as correlations)
 INTRCPT1,B0  1.000 -0.105  0.086 -0.405
     EXPO,B1 -0.105  1.000 -0.285  0.420
    AGE16,B2  0.086 -0.285  1.000 -0.270
   AGE16S,B3 -0.405  0.420 -0.270  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.342
      EXPO, B1                        0.062
     AGE16, B2                        0.330
    AGE16S, B3                        0.166
 ----------------------------------------------------

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                     Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
```

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.536548	0.018864	28.443	237	0.000
FEMALE, G01	-0.087195	0.025319	-3.444	237	0.001
MINORITY, G02	-0.003917	0.032033	-0.122	237	0.903
INCOME, G03	0.006434	0.005601	1.149	237	0.252
For EXPO slope, B1					
INTRCPT2, G10	0.551921	0.041454	13.314	237	0.000
FEMALE, G11	-0.048549	0.058298	-0.833	237	0.406
MINORITY, G12	-0.404139	0.071710	-5.636	237	0.000
INCOME, G13	-0.042315	0.013089	-3.233	237	0.002
For AGE16 slope, B2					
INTRCPT2, G20	0.018852	0.007483	2.519	237	0.013
FEMALE, G21	0.008663	0.009922	0.873	237	0.384
MINORITY, G22	-0.008015	0.012682	-0.632	237	0.528
INCOME, G23	-0.001653	0.002179	-0.759	237	0.449
For AGE16S slope, B3					
INTRCPT2, G30	-0.011305	0.004845	-2.333	237	0.021
FEMALE, G31	0.014522	0.006476	2.242	237	0.026
MINORITY, G32	0.003959	0.008345	0.474	237	0.635
INCOME, G33	0.002238	0.001416	1.580	237	0.115

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.536548	0.019993	26.837	237	0.000
FEMALE, G01	-0.087195	0.024535	-3.554	237	0.001
MINORITY, G02	-0.003917	0.032658	-0.120	237	0.905
INCOME, G03	0.006434	0.005623	1.144	237	0.254
For EXPO slope, B1					
INTRCPT2, G10	0.551921	0.038407	14.370	237	0.000
FEMALE, G11	-0.048549	0.057455	-0.845	237	0.399
MINORITY, G12	-0.404139	0.072271	-5.592	237	0.000
INCOME, G13	-0.042315	0.014359	-2.947	237	0.004
For AGE16 slope, B2					
INTRCPT2, G20	0.018852	0.007315	2.577	237	0.011
FEMALE, G21	0.008663	0.009423	0.919	237	0.359
MINORITY, G22	-0.008015	0.013110	-0.611	237	0.541
INCOME, G23	-0.001653	0.002002	-0.826	237	0.410
For AGE16S slope, B3					
INTRCPT2, G30	-0.011305	0.004920	-2.298	237	0.022
FEMALE, G31	0.014522	0.006166	2.355	237	0.019
MINORITY, G32	0.003959	0.008755	0.452	237	0.651
INCOME, G33	0.002238	0.001269	1.763	237	0.079

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.15078	0.02273	197	402.26089	0.000
EXPO slope,	U1	0.18240	0.03327	197	244.37700	0.012
AGE16 slope,	U2	0.05252	0.00276	197	307.13303	0.000
AGE16S slope,	U3	0.02415	0.00058	197	250.50418	0.006
level-1,	R	0.14248	0.02030			

<u>Example</u>: Baldwin, Scott A., and John P. Hoffmann. 2002. The Dynamics of Self-Esteem: A Growth-Curve Analysis. *Journal of Youth and Adolescence*, *31*, 2, 101–113.

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:

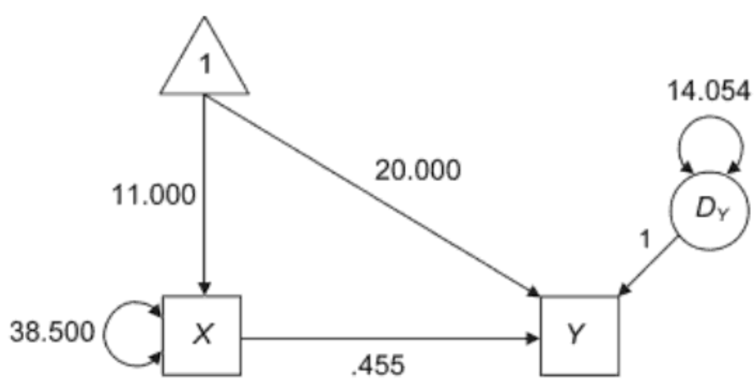


FIGURE 11.1. A path model with a mean structure.

(From Kline, 3rd ed, p. 301)

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 points and total number of parameters by counting means and intercepts as parameters and the number of data points as n*(n+3)/2. Note that the identification constraints do not allow us to have a model with constants for measurement equations of all indicators evaluated alongside the mean for the latent factor – we have to either assume the mean of the latent factor to be zero or intercepts for indicators are zero. So we could specify vectors TX and TY as free and KA and AL as fixed to zero, or KA and AL as free and TX and TY as 0.

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.

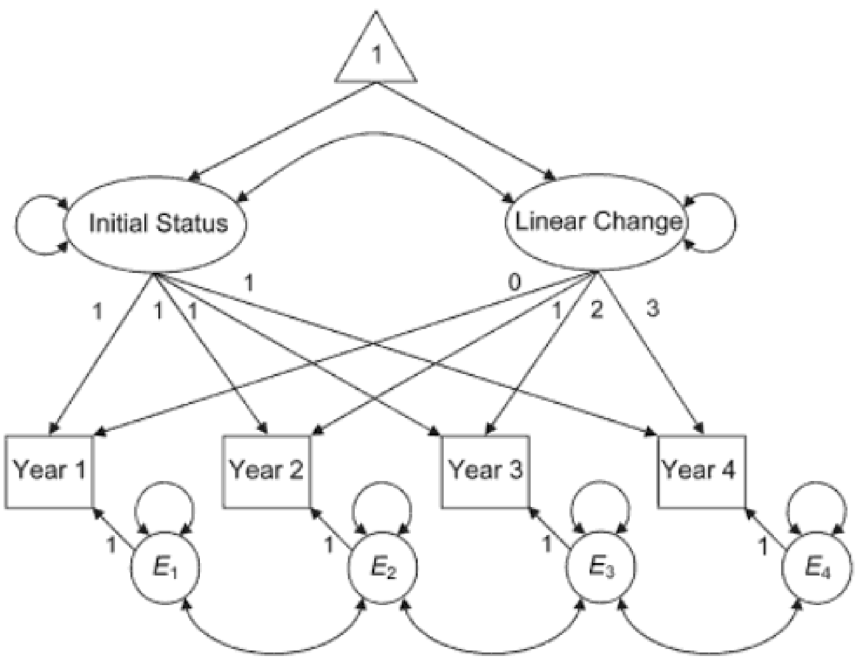


FIGURE 11.2. Latent growth model of change in level of alcohol use over 4 years.

(Kline, 3rd ed, p. 307)

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 diagram). 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.

<u>LISREL example</u>

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

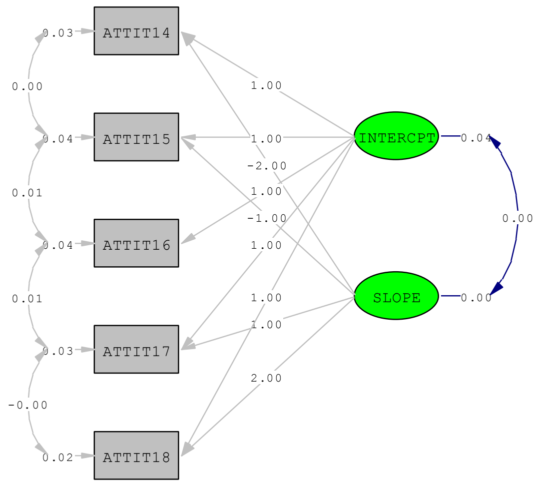
I transferred it back to SPSS to import it into LISREL. This file (nys2.sav) is available on the course website. Upon importing the data, we should define variables and obtain the covariance matrix and the means – these will be in files nys.cov and meansnys.mea.

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

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

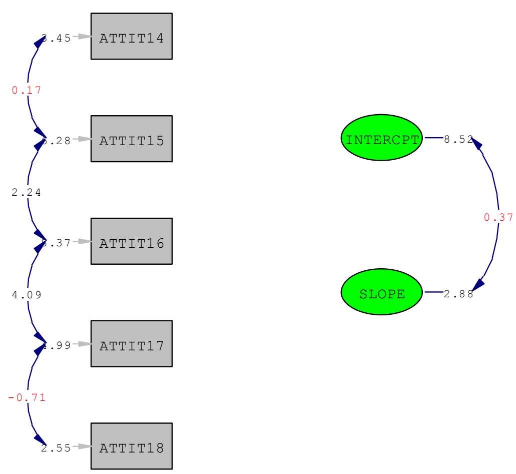
```
TI  Change only (random intercept and slope) model for attitude
DA NI=15 NO=241 MA=CM
LA
ID    ATTIT14    EXPO14    ATTIT15    EXPO15    ATTIT16    EXPO16    ATTIT17
EXPO17   ATTIT18    EXPO18    FEMALE MINORITY INCOME MINFEM
CM=C:\nys.cov
ME =C:\meansnys.mea
SE
2 4 6 8 10/
MO NX=5 NK=2 LX=FU, FI 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:

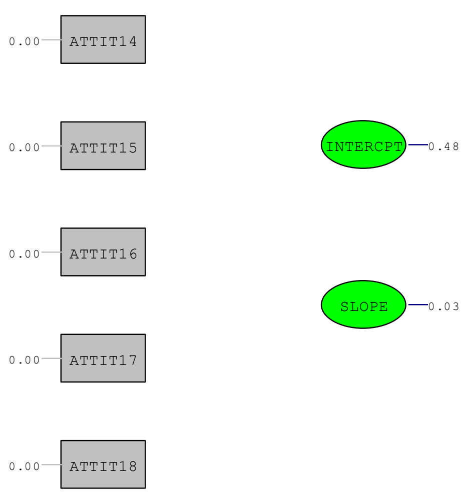


Chi-Square=21.86, df=6, P-value=0.00129, RMSEA=0.105

Significances:



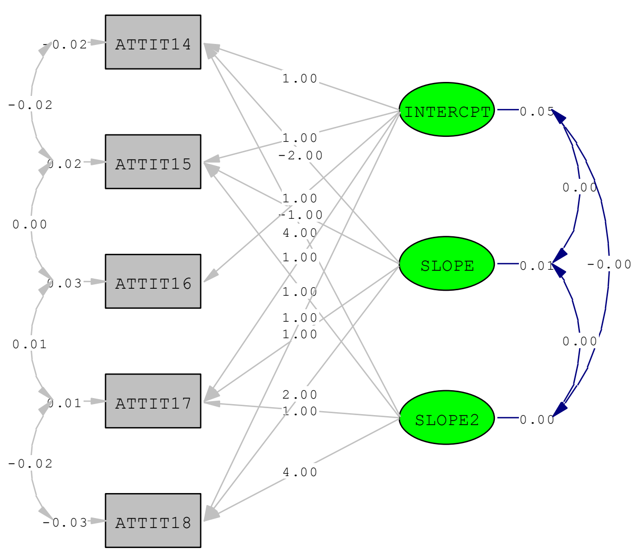
Means:



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

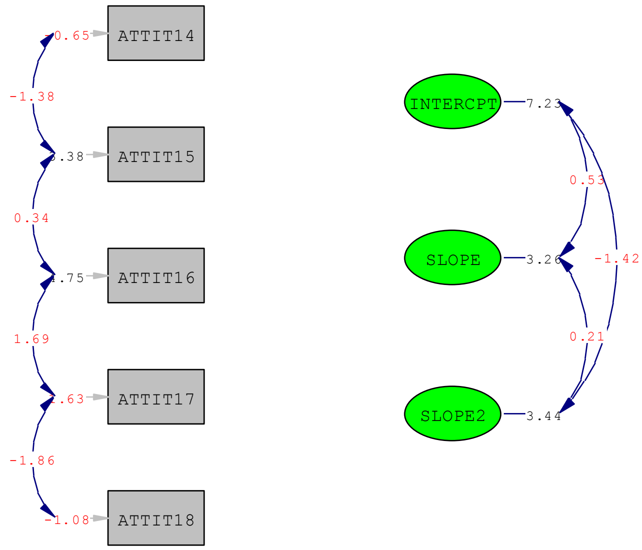
```
TI  Change only (random intercept and slope) model for attitude, with quadratic term
DA NI=15 NO=241 MA=CM
LA
ID    ATTIT14     EXPO14    ATTIT15     EXPO15    ATTIT16     EXPO16    ATTIT17
EXPO17   ATTIT18     EXPO18     FEMALE MINORITY INCOME MINFEM
CM=C:\nys.cov
ME =C:\meansnys.mea
SE
2 4 6 8 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:

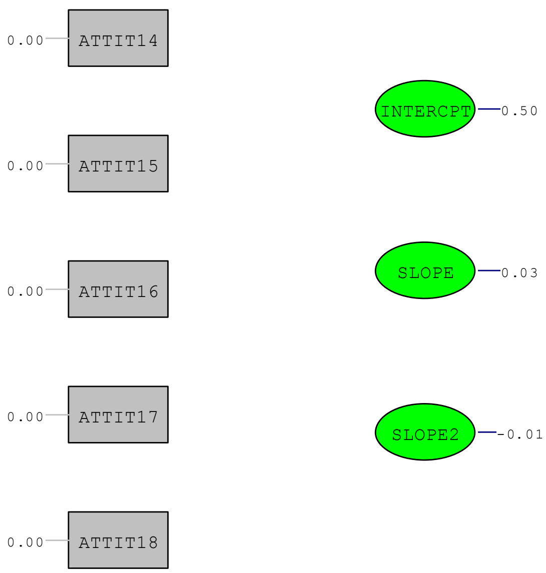


Chi-Square=3.07, df=2, P-value=0.21554, RMSEA=0.047

T-values:



Means:



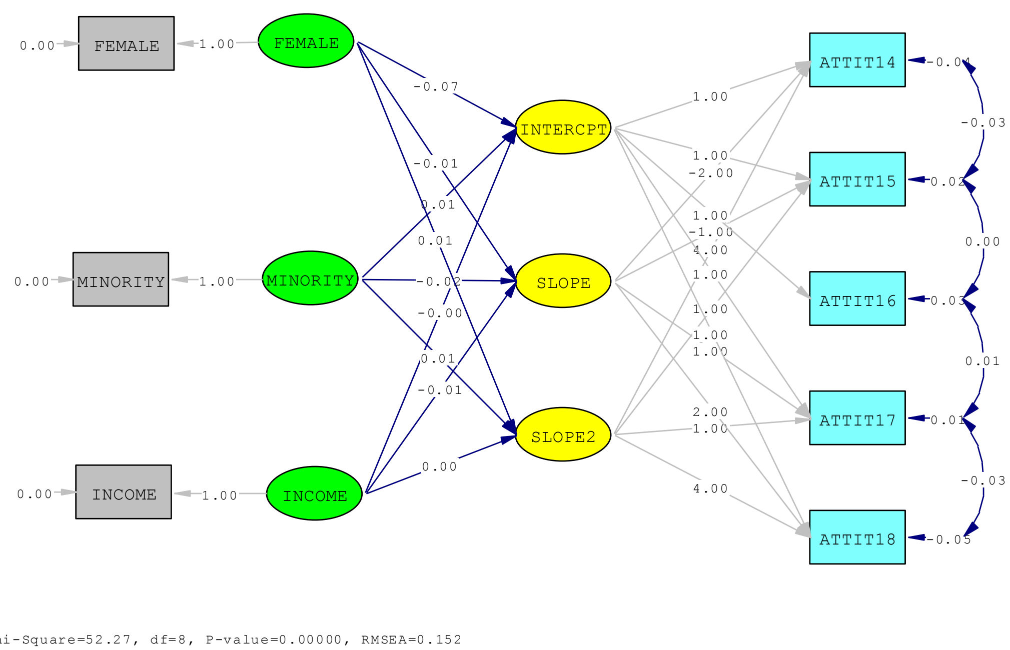
Check whether there is a significant improvement in chi-square:
21.86-3.07=18.79, df=6-2=4
Alpha=.01 critical value for df=4 is 13.28, so it's a significant improvement. We can also see that in RMSEA and chi-square significance.

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:

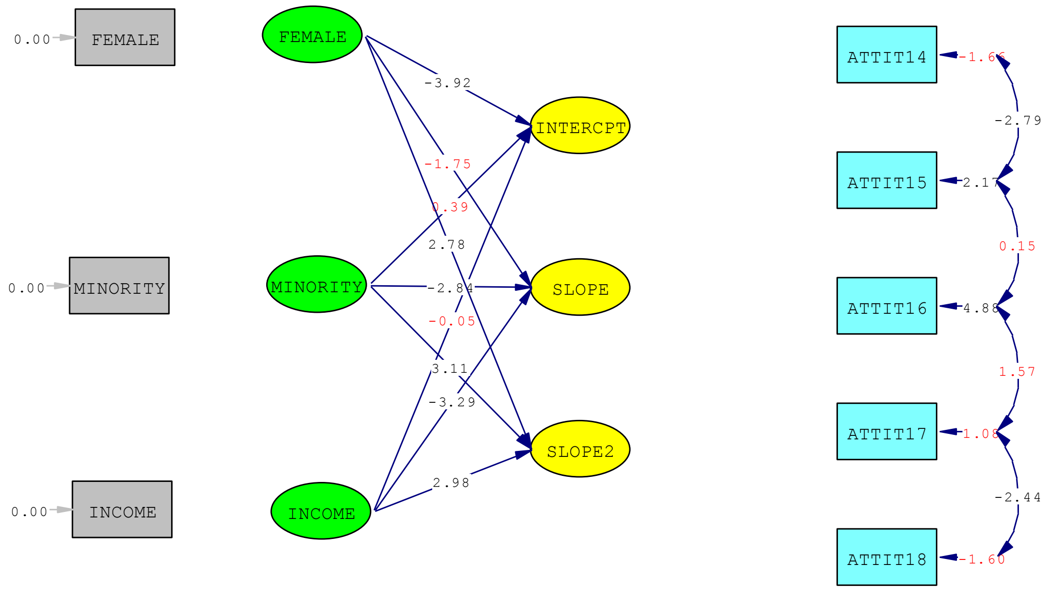
TI Predicting change in the random intercept and slope for attitude, with quadratic term
DA NI=15 NO=241 MA=CM
LA

```
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
```

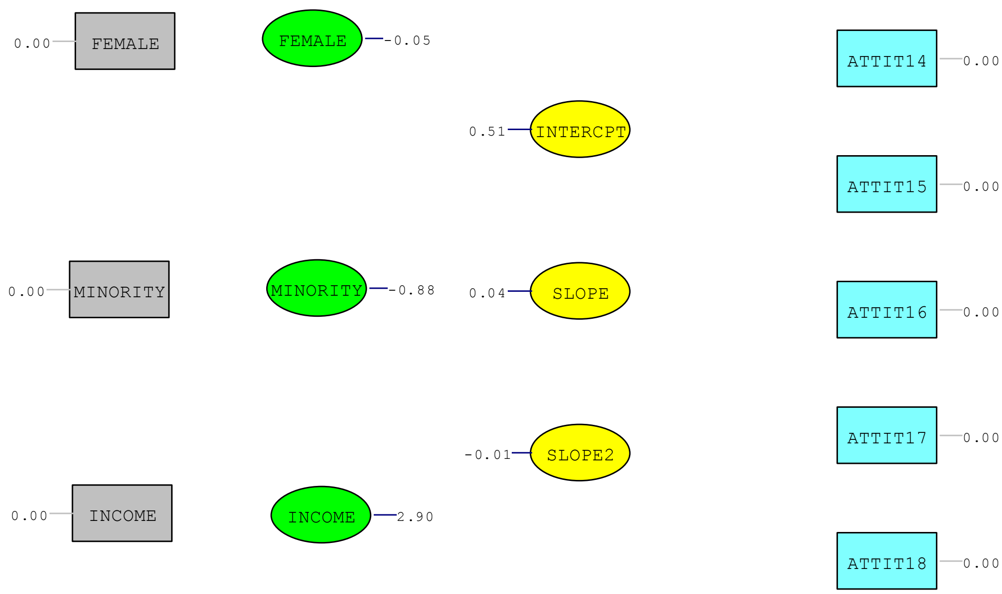
Estimates:



T-values:



Means:



<u>Example:</u>

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.

Other Types of Longitudinal models Using SEM

Longitudinal models are also very useful when we are interested in reciprocal relationships. Their value lies in the ability to examine both stability and change of variables (and relationships between variables) over time. Panel data are especially useful when we have repeat measures of the same variables (if they do not, then these data are analyzed the same way cross-sectional data would be).

Types of relationships in panel models:
1. Correlation between X1 and Y1 = synchronous correlation
2. Correlation between X1 and X2 and between Y1 and Y2 = autocorrelations, or stabilities.
3. Correlation between X1 and Y2 and between Y1 and X2 = cross-lagged correlations
4. The paths between measurement errors = autocorrelated error terms.

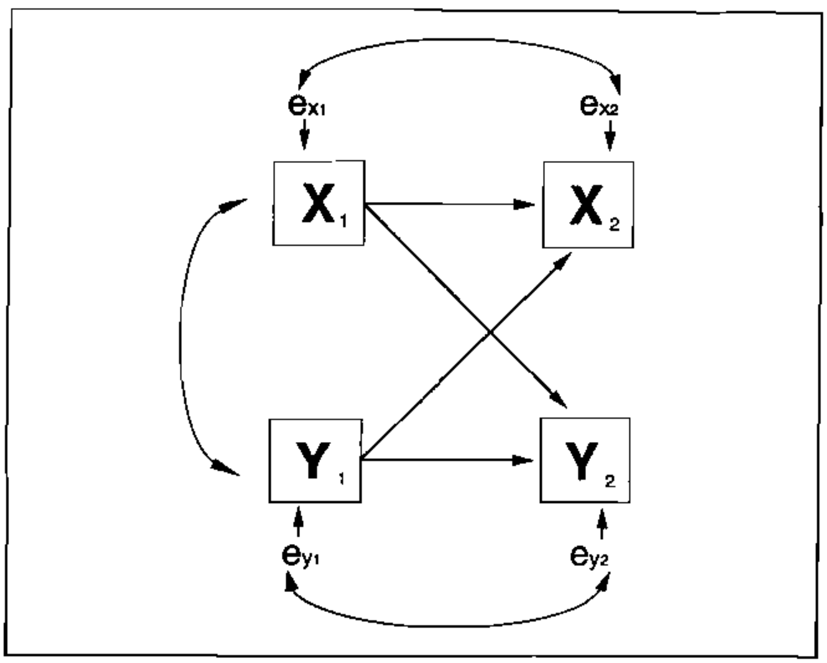


Figure 6.3. Two-Variable, Two-Wave Panel Model

<u>Stability of measures</u>

Stability is the most important concept added by panel models. If a variable is perfectly stable, that means that Y2 is perfectly determined by Y1 and has no other causes but itself. In this context, if we add some predictors at time 1, e.g. X1, we will find no causal relationship between X1 and Y2. Note that, in this situation, we would omit Y1 (or the relationship between Y1 and Y2) from the model, we would probably observe a relationship between X1 and Y2, but it would probably be erroneous to assume that X1 caused Y2 even though X1 happened prior to Y2 – the reason for their correlation lies in the correlation between X1 and the omitted Y1, and there may be many possible reasons for that correlation. So such a model can be misspecified, and, of course, if we don't have data on Y1, such a misspecification will likely go undetected.

E.g. if school achievement at time 2 is strongly related to school achievement at time 1, we cannot omit that relationship – if we do, we will witness many time1 predictors of time 2 school achievement, but they all may be misleading.

Note, that high stability for a variable means we will find very little in terms of causal antecedents for this variable. Low stability, in contrast, suggests that a variable is changing rapidly, and although this gives us an opportunity to find the causes for that change, it also may indicate low reliability of the measure or possibly even a change in that variable's meaning.

Note, that when working with longitudinal SEM models, you should use covariances and at all costs avoid using correlations as these remove differences in variability across time, and therefore ignore growth/change.

Autocorrelated error terms

These reflect the fact that when a measure is administered at different times, a substantial amount of variance may be shared because same method of data collection is used, or because respondents remember their earlier answers. We can only include these in the models if we have more than one indicator of X1 and X2, and Y1 and Y2 – otherwise, the model will not be identified. So if we suspect autocorrelated measurement errors, we need multiple-indicator models. Otherwise, to keep the model identified, we drop these paths, but by doing so, we incorporate any measurement-specific correlations into our measure of stability.

Note that in order to model these in LISREL, we need to be able to correlate measurement errors corresponding to exogenous variables' indicators with those of endogenous variables' indicators. This is done using an additional matrix – Theta Delta Epsilon, $\Theta\delta\varepsilon$ (TH). By default, this matrix is a fixed matrix (all zeros) and we cannot free the entire matrix on MO line, but we can free its elements (usually want we want to free is its diagonal elements) using FR command; it is a square matrix with both dimensions = number of X indicators + number of Y indicators.

Stability of causal processes

Stability of causal processes is different from stability of measures – it means that the effects of X on Y is stable over time – i.e., is the same for every time interval of the same length. Typically, if we are interested in the effect of X on Y, it would be desirable for that effect to be stable, unless we predict that it varies over time for a certain reason. We can check such stability if we have more than two time points.

Also, we need to consider the issue of temporal lag – i.e., how long of a time interval do we have between time 1 and time 2. If that interval is too short, we might have not observed the effect of X on Y yet; if it's too long, that effect might have decayed from its maximum. This is even more complicated if we think that the optimum time lag would be different for the relationship X→Y vs. Y→X. This is important to consider if one is collecting data; with secondary data, we usually have no choice.

Causal predominance

When examining reciprocity using panel data, we are often interested in evaluating causal predominance – i.e., which causal relationship is stronger, X→Y or Y→X. To evaluate that, we need to first evaluate a model that estimates both freely, then constrain them to be equal (using EQ command, e.g., EQ GA 2 1 GA 1 2 or EQ BE 4 1 BE 3 2), and see if there was a significant decrease in fit by evaluating chi-square change between the unconstrained and constrained model. If there was no statistically decrease, none of the causal relationships dominates. If the fit decreases significantly, the relationships are different, and the one with the larger standardized coefficient indicates the causally predominant relationship. Note that if the two latent variables have different units (which is based on the units of the reference indicator), you have to standardize them first by setting their variance to 1 and estimating all the lambdas freely – otherwise, their coefficients will be different because their units are different.

Example: Maruyama, Geoffrey, Norman Miller, and Rolf Holtz. 1986. "The relation between popularity and achievement: a longitudinal test of the lateral transmission of value hypothesis." Journal of Personality and Social Psychology 51(4): 730-741.

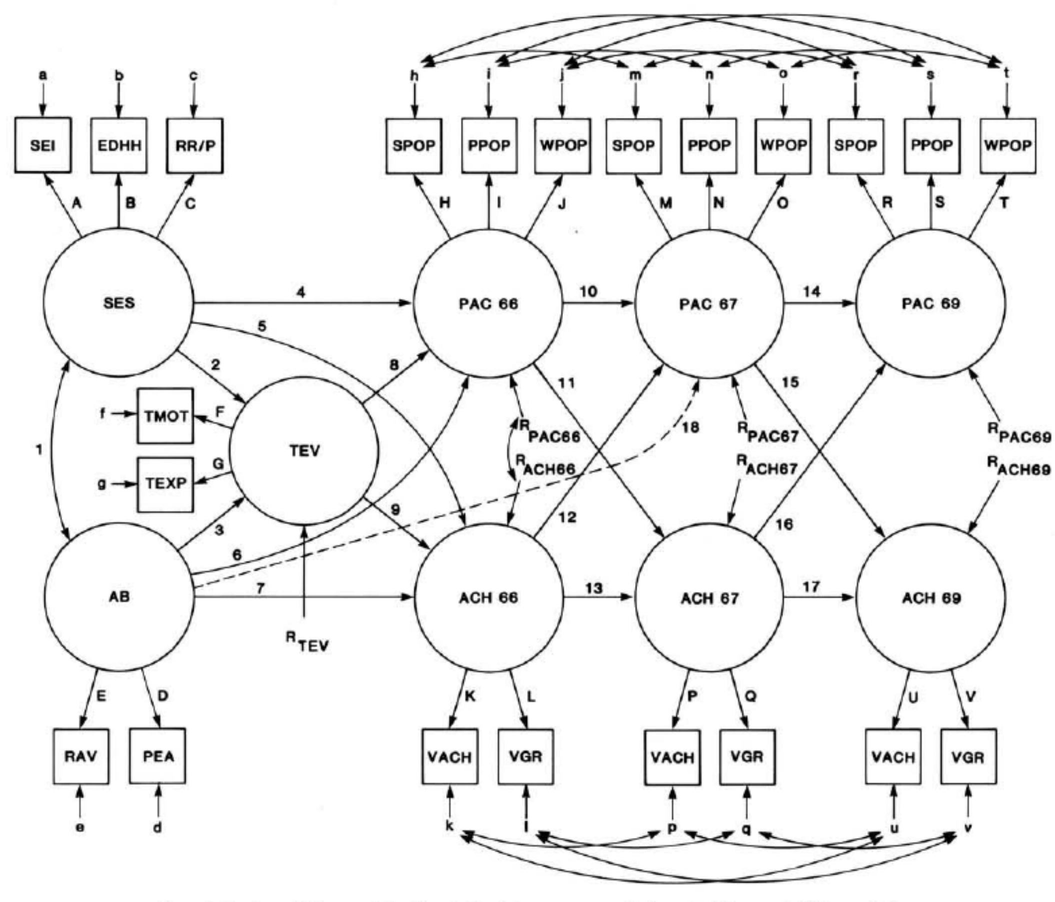


Figure 1. Panel model for examining the relation between peer popularity and achievement. (SES = socioeconomic status, measured by SEI [the Duncan Socioeconomic Index of Occupations], EDHH [educational attainment of the head of the household], and RR/P [the ratio of rooms in the home to people living in the home]; AB = academic ability, measured by RAV [Raven's Progressive Matrices] and PEA [the Peabody Picture Vocabulary Test]; TEV = teacher evaluations of students, measured by TMOT [teachers' ratings of students' motivation] and TEXP [teachers' expectations of their students' eventual educational attainment]; PAC = acceptance by peers, measured by SPOP [seating popularity], PPOP [playground popularity], and WPOP [schoolwork popularity]; and ACH = school achievement, measured by VACH [verbal standardized test performance] and VGR [verbal grades].)