

Longitudinal Data Analysis

Instructor: Natasha Sarkisian

Panel Data Analysis: Mixed Effects Models

So far, when analyzing panel data, we only allowed for the intercepts to vary across units (by having fixed effects or random effects for countries or individuals). A whole other class of models, mixed effects models, also known as multilevel models, hierarchical linear models, or growth curve models, allows for the coefficients themselves to vary across units. That is, we assume that the effects of time-varying variables, and time itself, are not the same across units. We will look at average effect of such variables, the extent to which there is variation around that average, and at level 2 (time-invariant) predictors that may explain that variation (so-called cross-level interactions). But first let's reexamine the equation for random effects model:

$$Y_{ij} = \alpha + X\beta + u_i + e_{ij}$$

We can also rewrite it as:

$$\text{Level 1 model is: } Y_{ij} = \alpha + X\beta + e_{ij}$$

$$\text{Level 2 model is: } \alpha = \pi_0 + u_i$$

Thus, we expressed a random effects model as a two-level model where we can explicitly see that the intercept for each unit equals to grand mean plus unit-specific residual. If our model also contains some time-invariant predictors, we can also write:

$$\text{Level 1 model is: } Y_{ij} = \alpha + X\beta + e_{ij}$$

$$\text{Level 2 model is: } \alpha = \pi_0 + X_i\beta_i + u_i$$

Moving beyond random effects models to mixed models, we can write a similar equation for each of level 1 regression coefficients:

$$\text{Level 1 model is: } Y_{ij} = \alpha + X\beta + e_{ij}$$

$$\text{Level 2 model is: } \alpha = \pi_0 + X_i\beta_i + u_{0i},$$

$$\beta_1 = \pi_1 + X_i\beta_i + u_{1i},$$

...

We will use an example that examines how attitudes toward deviant behavior change over time for teenagers, and what shapes that change. We will use a file called `nys.dta`. 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) and `age` (age in years). Level 2 include person-level variables: `female`, `minority`, and `income`.

```
. use http://www.sarkisian.net/sc706/nys.dta
```

```
. reshape long attit expo, i(id) j(age)
(note: j = 14 15 16 17 18)
```

```
Data                                wide  ->  long
```

```

Number of obs.                241  ->  1205
Number of variables            14  ->    7
j variable (5 values)         ->  age
xij variables:
    attit14 attit15 ... attit18  ->  attit
    expol14 expol15 ... expol18  ->  expo

```

```

-----
. xtset id age, yearly
    panel variable:  id (strongly balanced)
    time variable:  age, 14 to 18
    delta: 1 year

```

Data are considered strongly balanced if all the time points are the same and all cases are observed at all time points. Data are considered balanced if the cases have the same number of time values but these are not exactly the same time points. Data are unbalanced if cases are observed at different numbers of time points.

Focusing just on age, we could estimate a random effects model using both xtreg and xtmixed:

```
. xtreg attit age, re
```

```

Random-effects GLS regression           Number of obs   =   1066
Group variable: id                     Number of groups =    241

R-sq:  within = 0.0674                  Obs per group:  min =    1
        between = 0.0000                  avg   =    4.4
        overall = 0.0207                  max   =    5

Random effects u_i ~ Gaussian           Wald chi2(1)    =   58.31
corr(u_i, X) = 0 (assumed)              Prob > chi2     =   0.0000

```

```

-----
      attit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      age |   .0324074   .0042441     7.64   0.000   .0240892   .0407256
     _cons |  -.0258944   .0692441    -0.37   0.708  -0.1616103  .1098215
-----+-----
     sigma_u |   .21445769
     sigma_e |   .18975623
        rho |   .5608825   (fraction of variance due to u_i)
-----

```

```
. xtmixed attit age || id:
```

```
Performing EM optimization:
```

```
Performing gradient-based optimization:
```

```
Iteration 0:  log restricted-likelihood = 28.838503
Iteration 1:  log restricted-likelihood = 28.838503
```

```
Computing standard errors:
```

```

Mixed-effects REML regression           Number of obs   =   1066
Group variable: id                     Number of groups =    241

Obs per group:  min =    1
                  avg =    4.4
                  max =    5

```

Log restricted-likelihood = 28.838503

Wald chi2(1) = 57.88
 Prob > chi2 = 0.0000

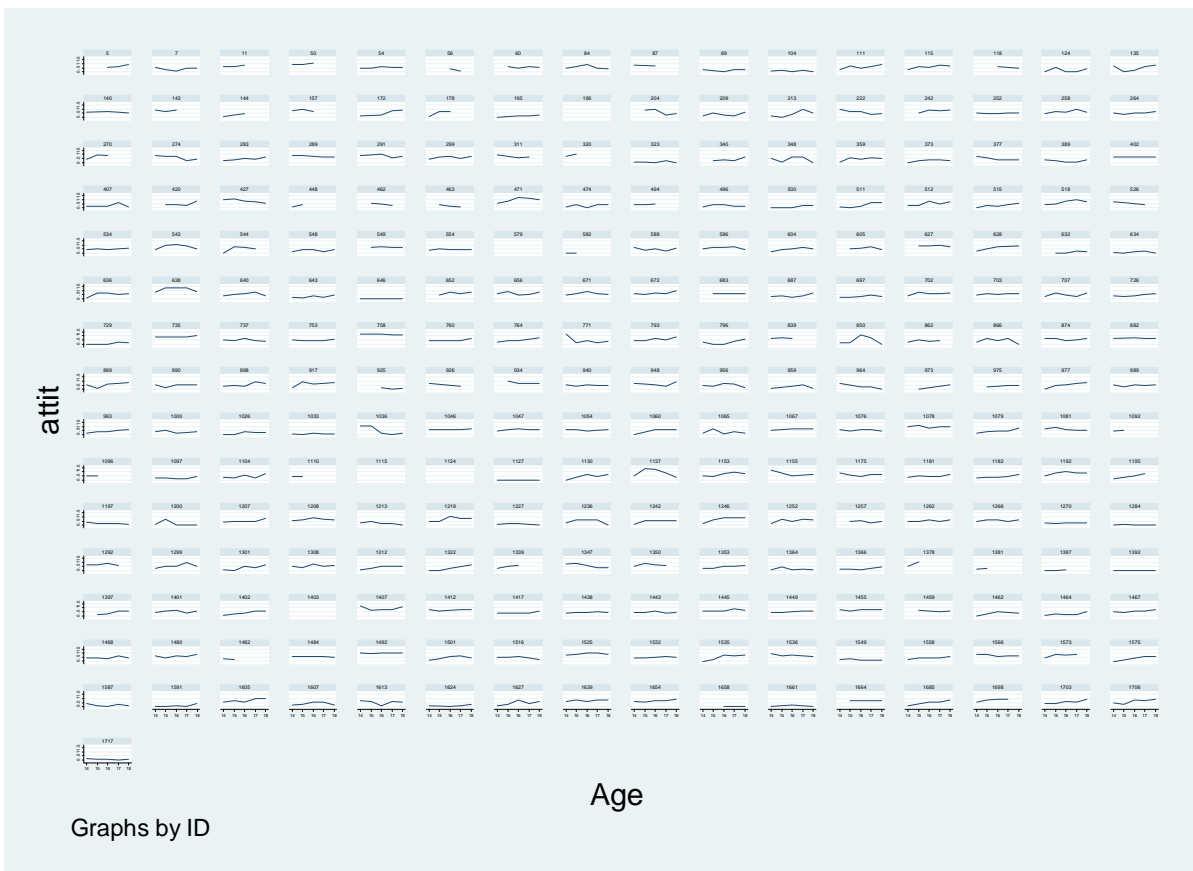
attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0323863	.0042569	7.61	0.000	.0240428	.0407297
_cons	-.0255459	.0694099	-0.37	0.713	-.1615869	.1104951

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity				
sd(_cons)	.2111038	.0116654	.1894347	.2352515
sd(Residual)	.1900667	.0046908	.1810917	.1994865

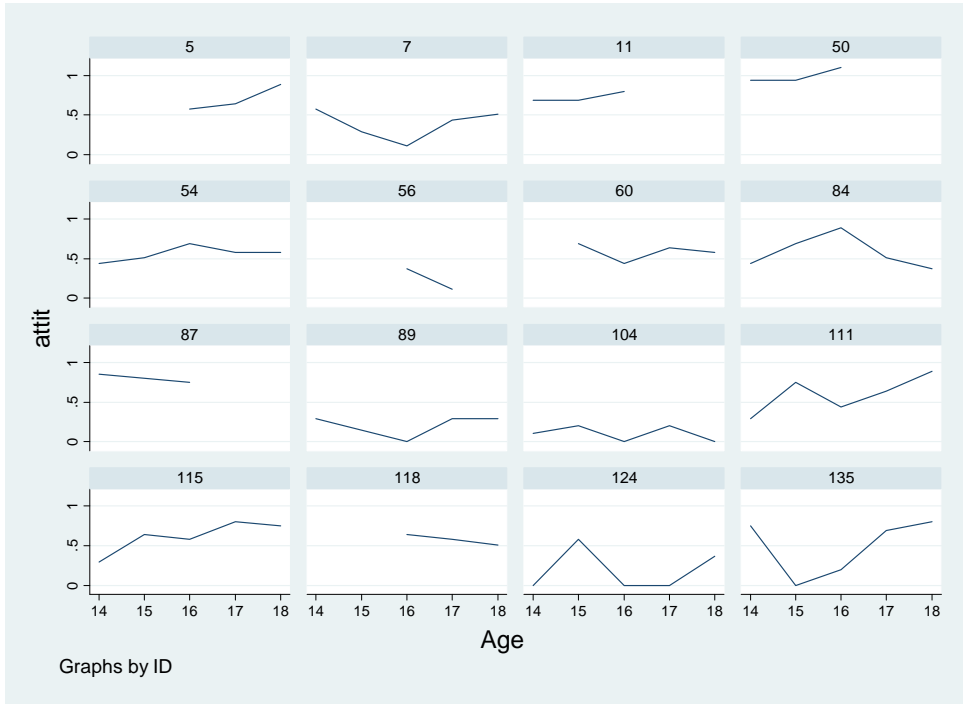
LR test vs. linear regression: chibar2(01) = 397.72 Prob >= chibar2 = 0.0000

Let's examine time trends graphically:

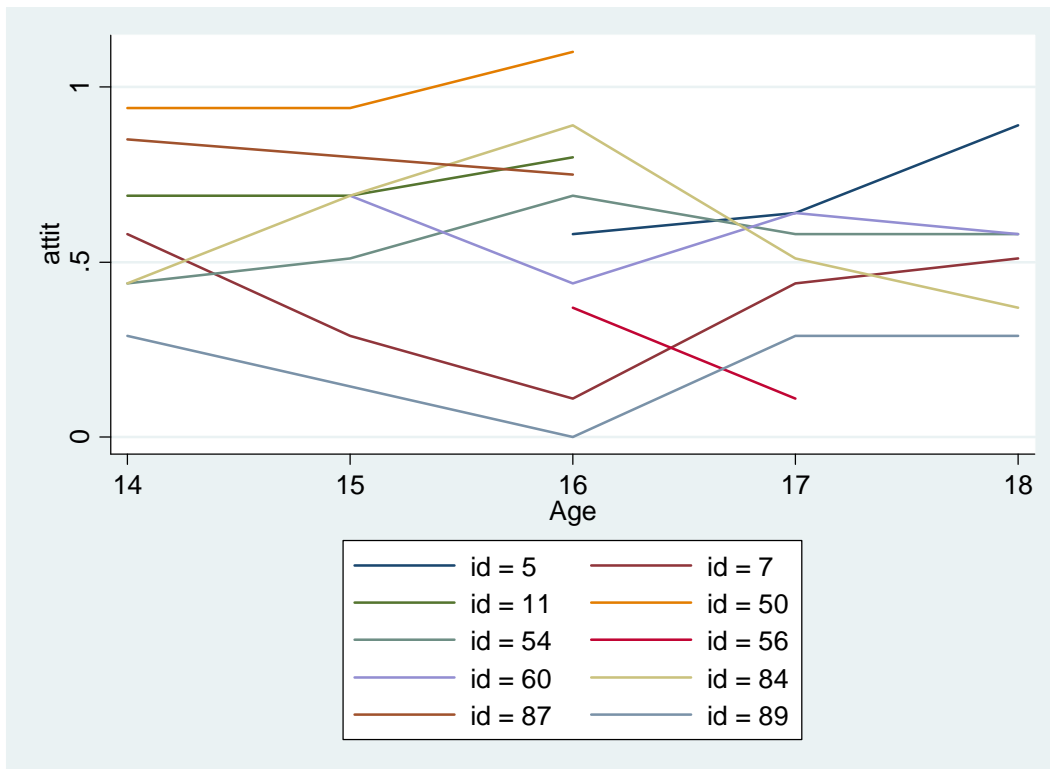
```
. xtline attit
```



```
. xtline attit if id<100
```



```
. xtline attit if id<100, overlay
```



Very often, in this type of analysis, we are interested in understanding why and how the trajectory over time varies across units (that is why these models are also called growth curve models), so we want to explore that variation – that requires estimating a mixed effects model; random effects model cannot assess variation in the slope of age.

```
. xtmixed attit age || id: age, cov(unstructured)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log restricted-likelihood = 30.601124
Iteration 1: log restricted-likelihood = 45.872252
Iteration 2: log restricted-likelihood = 48.997375
Iteration 3: log restricted-likelihood = 49.825208
Iteration 4: log restricted-likelihood = 49.83764
Iteration 5: log restricted-likelihood = 49.838114
Iteration 6: log restricted-likelihood = 49.838114
```

Computing standard errors:

```
Mixed-effects REML regression      Number of obs      =      1066
Group variable: id                 Number of groups   =       241

Obs per group: min =              1
                  avg =             4.4
                  max =              5
```

```
Log restricted-likelihood = 49.838114      Wald chi2(1)      =      36.57
                                          Prob > chi2       =      0.0000
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.0323571	.0053505	6.05	0.000	.0218702 .0428439
_cons	-.024388	.0872417	-0.28	0.780	-.1953785 .1466025

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(age)	.0559502	.0057241	.0457845 .0683731
sd(_cons)	.9364748	.0914994	.7732651 1.134132
corr(age,_cons)	-.9737465	.0059243	-.9831493 -.9592048
sd(Residual)	.1694919	.0048755	.1602004 .1793222

```
LR test vs. linear regression:      chi2(3) = 439.72      Prob > chi2 = 0.0000
```

Note: LR test is conservative and provided only for reference.

Note that we specified covariance option – that is because we want to allow random effects to correlate with each other; if we do not, that would be too restrictive since usually random effects for intercepts and slopes are correlated. So we have two random effects now:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left(0, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right)$$

Our tau matrix now contains the variance in the level-1 intercepts (τ_{00}), the variance in level-1 slopes (τ_{11}), as well as the covariance between level-1 intercepts and slopes ($\tau_{01} = \tau_{10}$). (This covariance is presented as a correlation in our output.) Note that covariance value indicates how much intercepts and slopes covary: in our example, there is a negative correlation between intercepts and slopes. That is, the higher the intercept, the smaller the slope (i.e. if the starting point in terms of deviant attitudes is higher, then the slope is less steep). We can see this as a variance-covariance matrix:

```
. estat recov

Random-effects covariance matrix for level id
-----+-----
          |          age          _cons
-----+-----
    age |   .0031015
   _cons |  -.0505552   .8692899
```

So far we assumed that the time trend is linear but the graph above shows that for many people it is not. Let's estimate a model with a quadratic trend.

```
. tab age

      Age |          Freq.          Percent          Cum.
-----+-----
       14 |             241             20.00             20.00
       15 |             241             20.00             40.00
       16 |             241             20.00             60.00
       17 |             241             20.00             80.00
       18 |             241             20.00            100.00
-----+-----
      Total |           1,205           100.00
```

```
. gen age16=age-16
. gen age16sq=age16^2
```

Note that the intercept will now correspond to value at age 16 rather than at the start of the study.

```
. xtmixed attit age16 age16sq || id: age16 age16sq, cov(unstructured)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:  log restricted-likelihood = 63.779001
Iteration 1:  log restricted-likelihood = 63.88902
Iteration 2:  log restricted-likelihood = 63.889125
Iteration 3:  log restricted-likelihood = 63.889125
```

Computing standard errors:

```
Mixed-effects REML regression
Group variable: id

Number of obs      =      1066
Number of groups   =       241

Obs per group:  min =         1
                avg =         4.4
                max =         5

Wald chi2(2)      =       41.34
```

Log restricted-likelihood = 63.889125 Prob > chi2 = 0.0000

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age16	.0314627	.0053327	5.90	0.000	.0210107 .0419146
age16sq	-.0106962	.0036517	-2.93	0.003	-.0178533 -.003539
_cons	.5140183	.0173066	29.70	0.000	.4800979 .5479387

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(age16)	.060747	.0052145	.0513402 .0718773
sd(age16sq)	.034373	.0044432	.0266801 .0442841
sd(_cons)	.2413507	.0137404	.2158682 .2698413
corr(age16,age16sq)	-.1603304	.1389152	-.4146204 .1171857
corr(age16,_cons)	-.0223911	.0975507	-.2104925 .1673091
corr(age16sq,_cons)	-.5016773	.086103	-.6510172 -.3149473
sd(Residual)	.1513583	.005323	.1412768 .1621592

LR test vs. linear regression: chi2(6) = 471.17 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Let's compare those two models using LR test and BIC. In order to do use LR test, we should reestimate the model using MLE rather than REML. REML is the default estimation method; REML and ML produce similar regression coefficients, but they differ in terms of estimating the variance components -- if the number of level-2 units is small, then ML variance estimates will be smaller than REML and significance tests based on ML will be biased. When we want to use likelihood ratio tests, however, we might have to opt for ML. When fixed effects are the same and one model has fewer random effects than the other, then both REML or ML may be used for LR test (models have to be nested). When one model has fewer fixed effects (and possibly fewer random effects) than the other, then we have to use ML, which is the case here.

```
. xtmixed attit age16 age16sq || id: age16 age16sq, cov(unstructured) mle
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = 76.093606
Iteration 1: log likelihood = 76.206947
Iteration 2: log likel
/ihood = 76.206955
Iteration 3: log likelihood = 76.206955
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1066
Group variable: id              Number of groups   =      241

Obs per group: min =      1
                  avg =      4.4
                  max =      5

Wald chi2(2)                      =      41.54
```

Log likelihood = 76.206955 Prob > chi2 = 0.0000

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age16	.0314681	.0053202	5.91	0.000	.0210407	.0418956
age16sq	-.0106942	.0036435	-2.94	0.003	-.0178353	-.0035532
_cons	.5140137	.0172699	29.76	0.000	.4801654	.547862

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
sd(age16)	.0605119	.0052012	.0511302	.0716151
sd(age16sq)	.0341835	.0044421	.0264974	.0440991
sd(_cons)	.240732	.0136896	.2153421	.2691155
corr(age16,age16sq)	-.1613437	.1392717	-.4161523	.1169595
corr(age16,_cons)	-.0225006	.0975324	-.2105638	.1671686
corr(age16sq,_cons)	-.5017549	.0862317	-.6512845	-.314716
sd(Residual)	.1513555	.0053225	.141275	.1621553

LR test vs. linear regression: chi2(6) = 471.08 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. est store squared

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1066	.	76.20696	10	-132.4139	-82.69722

Note: N=Obs used in calculating BIC; see [R] BIC note

. xtmixed attit age || id: age, cov(unstructured) mle

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = 38.418443 (not concave)
Iteration 1: log likelihood = 39.484174 (not concave)
Iteration 2: log likelihood = 40.06454
Iteration 3: log likelihood = 40.952084 (not concave)
Iteration 4: log likelihood = 42.424539 (not concave)
Iteration 5: log likelihood = 43.623775
Iteration 6: log likelihood = 56.205781
Iteration 7: log likelihood = 57.213181
Iteration 8: log likelihood = 57.441535
Iteration 9: log likelihood = 57.442108
Iteration 10: log likelihood = 57.442108
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1066
Group variable: id              Number of groups   =       241

Obs per group: min =          1
                  avg =         4.4
                  max =          5
```



```

Mixed-effects REML regression
Group variable: id

Number of obs      =      1066
Number of groups   =       241

Obs per group: min =         1
                  avg =         4.4
                  max =         5

Wald chi2(3)      =      269.83
Prob > chi2       =      0.0000

Log restricted-likelihood = 191.56453

```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age16	.0229438	.0048663	4.71	0.000	.0134061 .0324816
age16sq	-.0045771	.0032443	-1.41	0.158	-.0109357 .0017816
expo	.4392177	.0303382	14.48	0.000	.3797559 .4986794
_cons	.2522643	.0215611	11.70	0.000	.2100052 .2945234

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(age16)	.051728	.0051497	.0425584 .0628733
sd(age16sq)	.0261037	.0047828	.0182282 .0373819
sd(expo)	.23522	.0366039	.1733861 .3191056
sd(_cons)	.2071172	.0215611	.1688905 .2539962
corr(age16, age16sq)	-.2236336	.1771147	-.5319718 .1370675
corr(age16, expo)	-.183421	.1671742	-.4812296 .1523468
corr(age16, _cons)	.1768226	.1443242	-.1128169 .4387654
corr(age16sq, expo)	.1805575	.2097605	-.2377791 .5423907
corr(age16sq, _cons)	-.4224947	.1591735	-.6807377 -.0708434
corr(expo, _cons)	-.6382323	.0978611	-.7927606 -.4066177
sd(Residual)	.1410003	.0053562	.1308836 .151899

```

LR test vs. linear regression:      chi2(10) = 290.82   Prob > chi2 = 0.0000

```

Note: LR test is conservative and provided only for reference.

There is significant variation in slopes of all of these three level 1 variables. Next, we add level 2 (time invariant) variables as predictors of attitudes (but not yet of slopes). We have the following level 2 predictors: female, minority, and income.

```

. xtmixed attit age16 age16sq expo female minority income || id: age16 age16sq expo,
cov(unstructured)

```

Performing EM optimization:

Performing gradient-based optimization:

```

Iteration 0:  log restricted-likelihood = 187.55152
Iteration 1:  log restricted-likelihood = 187.96413
Iteration 2:  log restricted-likelihood = 187.96519
Iteration 3:  log restricted-likelihood = 187.96519

```

Computing standard errors:

```

Mixed-effects REML regression
Group variable: id

Number of obs      =      1066
Number of groups   =       241

```

```

Obs per group: min =      1
                avg =      4.4
                max =      5

```

```

Log restricted-likelihood = 187.96519
Wald chi2(6) = 290.41
Prob > chi2 = 0.0000

```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age16	.0228237	.0048686	4.69	0.000	.0132815	.0323659
age16sq	-.0044062	.0032464	-1.36	0.175	-.010769	.0019566
expo	.4416672	.0300694	14.69	0.000	.3827323	.5006021
female	-.0498218	.0220006	-2.26	0.024	-.0929423	-.0067013
minority	.0223901	.0278347	0.80	0.421	-.0321649	.0769451
income	.0141895	.0048562	2.92	0.003	.0046716	.0237074
_cons	.2082864	.0333672	6.24	0.000	.142888	.2736849

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
sd(age16)	.051715	.0051549	.0425372	.062873
sd(age16sq)	.0262153	.0047748	.0183451	.037462
sd(expo)	.2308826	.0363514	.1695796	.3143467
sd(_cons)	.1971736	.0217087	.1589029	.2446615
corr(age16,age16sq)	-.2243084	.1763973	-.5315029	.1350313
corr(age16,expo)	-.1753755	.1690745	-.4770072	.1632151
corr(age16,_cons)	.2036548	.1477114	-.0952044	.468837
corr(age16sq,expo)	.1886991	.2106696	-.2328136	.5505275
corr(age16sq,_cons)	-.4387443	.162634	-.6990387	-.0757856
corr(expo,_cons)	-.6140747	.1078263	-.7836296	-.3593735
sd(Residual)	.1411289	.0053587	.1310074	.1520323

```

LR test vs. linear regression:      chi2(10) = 274.50   Prob > chi2 = 0.0000

```

Note: LR test is conservative and provided only for reference.

Since we are now trying to model variance in the constant (intercept), we should make sure that intercept meaningful by making 0 a meaningful value on all predictors. Dummies are ok as long as they are coded 0/1 but continuous predictors should be mean-centered.

```

. for var expo income: sum X \ gen Xm=X-r(mean)

```

```

-> sum expo

```

Variable	Obs	Mean	Std. Dev.	Min	Max
expo	1066	.5601501	.3106114	0	1.61

```

-> gen expom=expo-r(mean)
(139 missing values generated)

```

```

-> sum income

```

Variable	Obs	Mean	Std. Dev.	Min	Max
income	1205	4.091286	2.346617	1	10

```
-> gen incomem=income-r(mean)

. xtmixed attit agel6 agel6sq expom female minority incomem || id: agel6 agel6sq
expom, cov(unstructured)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log restricted-likelihood = 186.94021
Iteration 1: log restricted-likelihood = 187.96071
Iteration 2: log restricted-likelihood = 187.96519
Iteration 3: log restricted-likelihood = 187.96519
```

Computing standard errors:

```
Mixed-effects REML regression          Number of obs      =      1066
Group variable: id                    Number of groups   =       241

Obs per group: min =          1
                avg =         4.4
                max =          5
```

```
Log restricted-likelihood = 187.96519      Wald chi2(6)       =      290.41
                                           Prob > chi2        =       0.0000
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age16	.0228237	.0048686	4.69	0.000	.0132815 .0323659
age16sq	-.0044062	.0032464	-1.36	0.175	-.010769 .0019566
expom	.4416672	.0300694	14.69	0.000	.3827323 .5006021
female	-.0498218	.0220006	-2.26	0.024	-.0929423 -.0067013
minority	.0223901	.0278347	0.80	0.421	-.0321649 .0769451
incomem	.0141895	.0048562	2.92	0.003	.0046716 .0237074
_cons	.5137396	.0178053	28.85	0.000	.4788418 .5486374

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(age16)	.051715	.0051549	.0425372 .062873
sd(age16sq)	.0262153	.0047748	.0183451 .037462
sd(expom)	.2308826	.0363514	.1695796 .3143468
sd(_cons)	.1558374	.0120352	.1339472 .1813049
corr(age16,age16sq)	-.2243084	.1763973	-.5315029 .1350313
corr(age16,expom)	-.1753755	.1690743	-.4770069 .1632147
corr(age16,_cons)	.1121311	.1200113	-.124952 .3371005
corr(age16sq,expom)	.1886991	.2106692	-.2328129 .550527
corr(age16sq,_cons)	-.3985215	.1260241	-.6141318 -.1275533
corr(expom,_cons)	.0529374	.156538	-.2493168 .3457936
sd(Residual)	.1411289	.0053587	.1310074 .1520323

```
LR test vs. linear regression:      chi2(10) =      274.50      Prob > chi2 = 0.0000
```

Note: LR test is conservative and provided only for reference.

The kind of centering we just applied is called grand-mean centering. The centering issue is important in mixed models.

Centering choices for time-varying (level-1) predictors:

1. Natural metric (X):

You should only use the original metric if the value of 0 for a predictor is a meaningful value. When 0 is not meaningful, the estimate of the intercept will be arbitrary and may be estimated with poor precision. Lack of precision in mixed models can be very problematic. First, because you are estimating within-group intercepts, thus with possibly small N, the estimates may be quite unstable. Second, because you may be trying to model variation in these intercepts, your model will be affected by the unreliability of the estimates.

2. Grand-mean centering (X - grand mean):

This will address the problems with estimation of intercept in original metric. Because the 0 values will fall in the middle of the distribution of the predictors, the intercept estimates will be estimated with much more precision. The intercept is also interpretable. Specifically, it will represent the value for a person with a (grand) average on every predictor. The interpretation of the intercepts is now “adjusted group mean.” The interpretation of slopes does not change. So we can interpret the fixed effect for the intercept as the average attitudes value adjusted for exposure – i.e., the average attitudes level for someone with average exposure to deviant peers.

Note that while it may seem inappropriate at first to center a dummy variable, in mixed models it can actually be quite useful. If uncentered, the intercept in a model with a dummy variable is the average value when the dummy variable is 0. If the dummy variable is centered, the intercept then becomes the mean adjusted for the proportion of time points with the dummy variable=1, so essentially it is the mean for an average case. We would only consider centering dummy variables when we would like to treat them as controls rather than main predictors of interest.

3. Group-mean centering (X – group mean):

Predictors can also be centered around the mean value for a given person (averaged over time). Recall how we used group-mean centered variables to indicate the change component within random effects models along with group means to indicate cross-sectional effects of differences across individuals. The intercept can then be interpreted as the average outcome for each person. This allows interpretation of parameter estimates as effects of change over time within-person. Under grand-mean centering or no centering, the parameter estimates reflect a combination of change over time and differences across individuals. But when we use a group-centered predictor, we only estimate only change effects (within-person component). In order not to discard the effects of differences across individuals, we should include person level variables alongside group-mean centered predictors. This is a common way to separate within and between unit effects in mixed effects model (we did that in random effects model as well):

```
. by id: egen expomean=mean(expo)

. gen expochange=expo-expomean
(139 missing values generated)

. xtmixed attit age16 age16sq expochange expomean female minority incomem || id:
age16 age16sq expochange, cov(unstructured)
```

Performing EM optimization:

Next, we will estimate a model where we will use cross-level interactions to explain variance in slopes across individuals. That is, we will introduce interactions of level 1 predictors with level 2 time-invariant variables and then see what happens to variance of slopes of those level 1 predictors.

```
. for var expomean female minority incomem: gen agel6X=age16*X \ gen
agel6sqX=age16sq*X \ gen expochX=expochange*X

-> gen agel6expomean=age16*expomean
-> gen agel6sqexpomean=age16sq*expomean
-> gen expochexpomean=expochange*expomean
(139 missing values generated)
-> gen agel6female=age16*female
-> gen agel6sqfemale=age16sq*female
-> gen expochfemale=expochange*female
(139 missing values generated)
-> gen agel6minority=age16*minority
-> gen agel6sqminority=age16sq*minority
-> gen expochminority=expochange*minority
(139 missing values generated)
-> gen agel6incomem=age16*incomem
-> gen agel6sqincomem=age16sq*incomem
-> gen expochincomem=expochange*incomem
(139 missing values generated)

. xtmixed attit agel6 agel6expomean agel6female agel6minority agel6incomem agel6sq
agel6sqexpomean agel6sqfemale agel6sqminority agel6sqincomem expochange
expochexpomean expochfemale expochminority expochincomem expomean female minority
incomem || id: agel6 agel6sq expochange , cov(unstructured)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log restricted-likelihood = 173.9332
Iteration 1: log restricted-likelihood = 175.61856
Iteration 2: log restricted-likelihood = 175.63863
Iteration 3: log restricted-likelihood = 175.63863
```

Computing standard errors:

```
Mixed-effects REML regression
Group variable: id
Number of obs      =      1066
Number of groups   =       241

Obs per group: min =         1
                avg =        4.4
                max =         5

Wald chi2(19)      =      426.81
```

Log restricted-likelihood = 175.63863 Prob > chi2 = 0.0000

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age16	.0493045	.0132632	3.72	0.000	.023309 .0752999
age16expom~n	-.0458508	.0191619	-2.39	0.017	-.0834075 -.008294
age16female	.0047919	.0099519	0.48	0.630	-.0147135 .0242972
age16minor~y	-.0073418	.0126937	-0.58	0.563	-.0322211 .0175375
age16incomem	-.0024325	.0021744	-1.12	0.263	-.0066942 .0018292
age16sq	-.0105266	.0088051	-1.20	0.232	-.0277844 .0067311
age16sqexp~n	-.0042772	.0128712	-0.33	0.740	-.0295042 .0209499
age16sqfem~e	.0149019	.0065016	2.29	0.022	.0021589 .0276449
age16sqmin~y	.0047221	.0083582	0.56	0.572	-.0116597 .0211039
age16sqinc~m	.0021889	.0014118	1.55	0.121	-.0005782 .0049561
expochange	.2847031	.1123478	2.53	0.011	.0645054 .5049008
expochexpo~n	.2296443	.1593827	1.44	0.150	-.08274 .5420287
expochfemale	-.0139817	.0758856	-0.18	0.854	-.1627148 .1347513
expochmino~y	-.3022909	.0927468	-3.26	0.001	-.4840713 -.1205104
expochinco~m	-.044656	.0171895	-2.60	0.009	-.0783468 -.0109653
expomean	.6238642	.0495817	12.58	0.000	.5266859 .7210426
female	-.0740025	.0255484	-2.90	0.004	-.1240765 -.0239285
minority	-.0078201	.0322398	-0.24	0.808	-.0710089 .0553687
incomem	.0102604	.0056377	1.82	0.069	-.0007893 .0213102
_cons	.1869384	.034303	5.45	0.000	.1197058 .2541711

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(age16)	.0519192	.0052175	.0426372 .063222
sd(age16sq)	.0245313	.0048939	.0165924 .0362687
sd(expoc~ge)	.2452769	.0440905	.1724433 .3488727
sd(_cons)	.1596776	.0108551	.1397585 .1824356
corr(age16, age16sq)	-.2635438	.1876893	-.5818363 .1247552
corr(age16, expoc~ge)	-.2494407	.1852378	-.5662355 .1315645
corr(age16, _cons)	.1554511	.1139626	-.072047 .3675742
corr(age16sq, expoc~ge)	.2379141	.2394154	-.249461 .6291221
corr(age16sq, _cons)	-.4062904	.1228207	-.6165855 -.1418793
corr(expoc~ge, _cons)	.0437171	.1729702	-.287574 .3656647
sd(Residual)	.1383646	.0052703	.1284112 .1490895

LR test vs. linear regression: chi2(10) = 278.69 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Let's simplify the model by omitting non-significant cross-level interactions; we will use LR test and BIC to make sure we do not omit anything important:

```
. xtmixed attit age16 age16expomean age16female age16sq age16sqexpomean
age16sqfemale expochange expochminority expochincomem expomean female minority
incomem || id: age16 age16sq expochange , cov(unstructured) mle
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = 239.02435
Iteration 1: log likelihood = 240.79649
Iteration 2: log likelihood = 240.81527
```

Iteration 3: log likelihood = 240.81527

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1066
Group variable: id              Number of groups   =       241

                                Obs per group: min =        1
                                avg   =       4.4
                                max   =        5

                                Wald chi2(13)    =      430.09
                                Prob > chi2     =       0.0000

Log likelihood = 240.81527
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age16	.0448121	.0125932	3.56	0.000	.0201299	.0694943
age16expom~n	-.0408274	.0187166	-2.18	0.029	-.0775112	-.0041436
age16female	.0039725	.0095473	0.42	0.677	-.0147399	.0226849
age16sq	-.0071441	.0083618	-0.85	0.393	-.0235329	.0092447
age16sqexp~n	-.0091057	.0124199	-0.73	0.463	-.0334482	.0152368
age16sqfem~e	.0151555	.0062652	2.42	0.016	.0028759	.0274351
expochange	.4191731	.0422462	9.92	0.000	.3363721	.5019741
expochmino~y	-.3121521	.0882582	-3.54	0.000	-.4851349	-.1391693
expochinco~m	-.0550901	.016154	-3.41	0.001	-.0867514	-.0234288
expomean	.6318958	.049011	12.89	0.000	.535836	.7279556
female	-.0738328	.0252351	-2.93	0.003	-.1232926	-.024373
minority	.0038862	.0267865	0.15	0.885	-.0486143	.0563868
incomem	.0152012	.0047064	3.23	0.001	.0059768	.0244256
_cons	.1799113	.0336683	5.34	0.000	.1139226	.2458999

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
sd(age16)	.0511533	.0051439	.0420028	.0622974
sd(age16sq)	.0239896	.0048743	.0161092	.0357251
sd(expoc~ge)	.2395901	.0432503	.1681945	.3412918
sd(_cons)	.1580128	.0107077	.13836	.1804571
corr(age16,age16sq)	-.301639	.1891484	-.6163842	.0962109
corr(age16,expoc~ge)	-.2493234	.1860113	-.5672316	.1332521
corr(age16,_cons)	.1621463	.1134869	-.0647553	.3731057
corr(age16sq,expoc~ge)	.2665695	.2434294	-.2359136	.6565793
corr(age16sq,_cons)	-.4161707	.1226707	-.6254132	-.1510905
corr(expoc~ge,_cons)	.04502	.174465	-.2891048	.3693672
sd(Residual)	.1383104	.005244	.128405	.14898

LR test vs. linear regression: chi2(10) = 278.16 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

```
. est store reduced
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1066	.	240.8153	25	-431.6305	-307.3388

Note: N=Obs used in calculating BIC; see [R] BIC note

```
. xtmixed attit age16 age16expomean age16female age16minority age16incomem age16sq
age16sqexpomean age16sqfemale age16sqminority age16sqincomem expochange
expochexpomean expochfemale expochminority expochincomem expomean female minority
incomem || id: age16 age16sq expochange , cov(unstructured) mle
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log likelihood = 241.71886
Iteration 1: log likelihood = 243.56957
Iteration 2: log likelihood = 243.59055
Iteration 3: log likelihood = 243.59055
```

Computing standard errors:

```
Mixed-effects ML regression      Number of obs      =      1066
Group variable: id              Number of groups   =       241

Obs per group: min =           1
                  avg =           4.4
                  max =           5
```

```
Wald chi2(19) = 439.44
Prob > chi2   = 0.0000
Log likelihood = 243.59055
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age16	.0492662	.013066	3.77	0.000	.0236573	.0748752
age16expom~n	-.045851	.0188732	-2.43	0.015	-.0828419	-.0088602
age16female	.0048259	.0097991	0.49	0.622	-.0143799	.0240318
age16minor~y	-.0073877	.0124986	-0.59	0.554	-.0318845	.017109
age16incomem	-.0024181	.0021411	-1.13	0.259	-.0066145	.0017784
age16sq	-.0105712	.0086834	-1.22	0.223	-.0275903	.006448
age16sqexp~n	-.0041202	.0126952	-0.32	0.746	-.0290023	.0207619
age16sqfem~e	.0148405	.0064114	2.31	0.021	.0022744	.0274066
age16sqmin~y	.0047483	.0082399	0.58	0.564	-.0114016	.0208982
age16sqinc~m	.0022079	.0013923	1.59	0.113	-.0005208	.0049367
expochange	.2869866	.110384	2.60	0.009	.0706379	.5033353
expochexpo~n	.2282904	.156432	1.46	0.144	-.0783107	.5348915
expochfemale	-.0151889	.0744531	-0.20	0.838	-.1611144	.1307366
expochmino~y	-.3040498	.0908928	-3.35	0.001	-.4821965	-.1259031
expochinco~m	-.0445934	.0168649	-2.64	0.008	-.077648	-.0115387
expomean	.6235837	.04901	12.72	0.000	.5275259	.7196416
female	-.0738721	.0252524	-2.93	0.003	-.123366	-.0243783
minority	-.0079024	.0318651	-0.25	0.804	-.0703567	.054552
incomem	.0102222	.0055722	1.83	0.067	-.000699	.0211435
_cons	.187023	.0339053	5.52	0.000	.12057	.2534761

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
sd(age16)	.0506134	.0051428	.041474	.0617668
sd(age16sq)	.0234736	.0049506	.015526	.0354895
sd(expoc~ge)	.2334688	.0436144	.1618883	.3366992
sd(_cons)	.1571853	.0106749	.1375955	.1795641
corr(age16,age16sq)	-.2765475	.1934959	-.6009231	.1260373
corr(age16,expoc~ge)	-.2525177	.1898236	-.5753521	.1383928
corr(age16,_cons)	.1594468	.1142585	-.0688563	.3718804

```

corr(agel6sq,expoc~ge) | .2475761 .2498039 -.2624586 .6494836
corr(agel6sq,_cons) | -.4012455 .126131 -.6167747 -.1297468
corr(expoc~ge,_cons) | .0476106 .1768763 -.2911408 .3757532
-----+-----
sd(Residual) | .1383681 .0052563 .1284401 .1490635
-----+-----
LR test vs. linear regression:      chi2(10) = 278.01 Prob > chi2 = 0.0000

```

Note: LR test is conservative and provided only for reference.

```
. lrtest . reduced
```

```

Likelihood-ratio test                LR chi2(6) = 5.55
(Assumption: reduced nested in .)    Prob > chi2 = 0.4754

```

```
. estat ic
```

```

-----+-----
Model | Obs ll(null) ll(model) df AIC BIC
-----+-----
. | 1066 . 243.5905 31 -425.1811 -271.0594
-----+-----

```

Note: N=Obs used in calculating BIC; see [R] BIC note

No significant difference in model fit indicated by LR test, and BIC is substantially smaller in the reduced model; therefore, we can use the reduced model.

To summarize model building in mixed effects models, we have a number of options:

- The effects of level 1 predictors can be estimated as either fixed effects or random effects
- Level 2 predictors can be used to predict the intercept (i.e., as direct predictors of DV)
- Level 2 predictors can explain the variation in slopes of level 1 predictors (i.e., as cross-level interactions)

Because so many components are involved, it is best to proceed incrementally.

1. Start by fitting a model with only the time variable. Evaluate level 2 variance in intercepts and time slopes to see if a mixed effects model is necessary.
2. Estimate a model with random intercept and slopes using only level 1 variables (all slopes should be random effects). Evaluate slope variance and decide whether some slopes should be fixed (i.e., no random component included for it).
3. Estimate a model with both level 1 variables and level 2 variables used as predictors of intercepts.
4. For slopes with significant variance, use level 2 predictors to explain that variance (i.e., estimate a model with cross-level interactions).
5. If the slope variance remaining after entering level 2 predictors is not statistically significant, estimate that slope as non-randomly varying (i.e., keep cross-level interactions but do not include a random component for that slope).
6. When making decisions what variables to include and whether to estimate random or fixed effects, use LR tests and BIC values to select a model with best fit and parsimony.

Next, let's explore how much variance in slopes of each of level 1 predictors our final model explain. For that, let's compare it to a model without any level 2 predictors:

```
. xtmixed attit age16 age16sq expochange || id: age16 age16sq expochange,
cov(unstructured)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0: log restricted-likelihood = 127.87781
Iteration 1: log restricted-likelihood = 129.02377
Iteration 2: log restricted-likelihood = 129.0304
Iteration 3: log restricted-likelihood = 129.0304
```

Computing standard errors:

```
Mixed-effects REML regression      Number of obs      =      1066
Group variable: id                 Number of groups   =       241

Obs per group: min =           1
                  avg =          4.4
                  max =           5
```

```
Log restricted-likelihood = 129.0304      Wald chi2(3)      = 135.39
                                          Prob > chi2       = 0.0000
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age16	.0244345	.0049234	4.96	0.000	.0147849	.0340841
age16sq	-.00557	.003245	-1.72	0.086	-.01193	.0007901
expochange	.3406091	.0375955	9.06	0.000	.2669232	.414295
_cons	.5031748	.0164454	30.60	0.000	.4709425	.5354071

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Unstructured				
sd(age16)	.0522794	.0051892	.043037	.0635068
sd(age16sq)	.0263832	.0047014	.0186057	.0374118
sd(expoc~ge)	.2714305	.0445711	.1967363	.3744837
sd(_cons)	.2302087	.0130801	.2059481	.2573272
corr(age16,age16sq)	-.2029267	.1761232	-.5122707	.1530259
corr(age16,expoc~ge)	-.207126	.1738767	-.5125884	.1448746
corr(age16,_cons)	-.0792529	.1034457	-.2760959	.1239708
corr(age16sq,expoc~ge)	.0267284	.2158694	-.377091	.4220088
corr(age16sq,_cons)	-.3444214	.1171253	-.5508209	-.0983232
corr(expoc~ge,_cons)	.2299758	.1514832	-.0791513	.4987532
sd(Residual)	.1379758	.0052833	.1279997	.1487294

```
LR test vs. linear regression:      chi2(10) = 546.30   Prob > chi2 = 0.0000
```

Note: LR test is conservative and provided only for reference.

```
. estat recov
```

Random-effects covariance matrix for level id

	age16	age16sq	expocha-e	_cons
age16	.0027331			
age16sq	-.0002799	.0006961		

```

expochange | -.0029392   .0001914   .0736745
      _cons | -.0009538  -.0020919   .0143702   .052996

```

```

. xtmixed attit agel6 agel6expomean agel6female agel6sq agel6sqexpomean
agel6sqfemale expochange expochminority expochincomem expomean female minority
incomem || id: agel6 agel6sq expochange , cov(unstructured)

```

Performing EM optimization:

Performing gradient-based optimization:

```

Iteration 0: log restricted-likelihood = 192.14796
Iteration 1: log restricted-likelihood = 193.81989
Iteration 2: log restricted-likelihood = 193.83833
Iteration 3: log restricted-likelihood = 193.83833

```

Computing standard errors:

```

Mixed-effects REML regression
Group variable: id

Number of obs      =      1066
Number of groups   =       241

Obs per group: min =         1
                avg =        4.4
                max =         5

Wald chi2(13)      =      420.27
Prob > chi2        =      0.0000

Log restricted-likelihood = 193.83833

```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
agel6	.044842	.0127063	3.53	0.000	.0199381 .069746
agel6expom~n	-.0408385	.0188866	-2.16	0.031	-.0778556 -.0038215
agel6female	.0039596	.0096366	0.41	0.681	-.0149277 .022847
agel6sq	-.0071184	.0084317	-0.84	0.399	-.0236442 .0094073
agel6sqexp~n	-.0091831	.0125238	-0.73	0.463	-.0337293 .0153631
agel6sqfem~e	.0151793	.0063176	2.40	0.016	.002797 .0275616
expochange	.4184038	.0427378	9.79	0.000	.3346392 .5021684
expochmino~y	-.3111393	.089373	-3.48	0.000	-.4863071 -.1359714
expochinco~m	-.0551251	.0163477	-3.37	0.001	-.087166 -.0230842
expomean	.6320613	.0494977	12.77	0.000	.5350476 .7290749
female	-.0738951	.0254879	-2.90	0.004	-.1238505 -.0239397
minority	.0038667	.0270767	0.14	0.886	-.0492027 .0569361
incomem	.0151963	.0047581	3.19	0.001	.0058707 .024522
_cons	.1798652	.0340044	5.29	0.000	.1132177 .2465127

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
id: Unstructured			
sd(agel6)	.051896	.0051872	.0426632 .0631269
sd(agel6sq)	.0245891	.0048454	.0167112 .0361807
sd(expoc~ge)	.2465608	.0435606	.1743974 .3485845
sd(_cons)	.1601864	.0108543	.1402646 .1829378
corr(agel6,agel6sq)	-.2924035	.1859938	-.6042509 .0971238
corr(agel6,expoc~ge)	-.2427525	.1829635	-.5571936 .1325751
corr(agel6,_cons)	.1587813	.1132003	-.0673683 .3694128
corr(agel6sq,expoc~ge)	.256062	.2371067	-.2311838 .6406166
corr(agel6sq,_cons)	-.4168253	.1210471	-.6236643 -.1554373
corr(expoc~ge,_cons)	.044171	.171939	-.2853101 .3643155

```

          sd(Residual) |      .1383125      .0052534      .1283899      .149002
-----
LR test vs. linear regression:      chi2(10) =      280.32      Prob > chi2 = 0.0000

```

Note: LR test is conservative and provided only for reference.

```
. estat recov
```

Random-effects covariance matrix for level id

```

          |      age16      age16sq  expocha~e      _cons
-----+-----
age16    |      .0026932
age16sq  |     -.0003731      .0006046
expochange |     -.0031061      .0015524      .0607922
_cons    |      .00132     -.0016418      .0017446      .0256597

```

To see both matrices back to back, I copied the one from the model above:

```

          |      age16      age16sq  expocha~e      _cons
-----+-----
age16    |      .0027331
age16sq  |     -.0002799      .0006961
expochange |     -.0029392      .0001914      .0736745
_cons    |     -.0009538     -.0020919      .0143702      .052996

```

So the variance in age16 slope went from .0027331 to .0026932, variance in age16sq slope from .0006961 to .0006046, variance in expochange slope from .0736745 to .0607922, and variance in intercepts from .052996 to .0256597. We can calculate all of these as percentages:

```

. di (.0027331-.0026932)/.0027331
.01459881

. di (.0006961-.0006046)/.0006961
.13144663

. di (.0736745-.0607922)/.0736745
.17485426

. di (.052996-.0256597)/.052996
.51581817

```

We might also want to know overall how much of within-person and between-person variance we are explaining here. Typically, in growth curve models, we compare our full model to a model with only time variables (in our case, age and age squared) that does not include any random slopes, only random intercept. Here's our null model:

```
. xtmixed attit age16 age16sq || id:, variance
```

Performing EM optimization:

Performing gradient-based optimization:

```

Iteration 0:  log restricted-likelihood = 28.467641
Iteration 1:  log restricted-likelihood = 28.467641

```

Computing standard errors:

```
Mixed-effects REML regression      Number of obs      =      1066
```


expochinco~m	-.0478129	.0141427	-3.38	0.001	-.075532	-.0200938
expomean	.6297096	.0486829	12.93	0.000	.5342929	.7251263
female	-.0724012	.0250864	-2.89	0.004	-.1215697	-.0232327
minority	.0022848	.0272989	0.08	0.933	-.05122	.0557897
incomem	.014601	.0047976	3.04	0.002	.005198	.0240041
_cons	.1798456	.0334747	5.37	0.000	.1142364	.2454548

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity				
var(_cons)	.0199164	.0025003	.0155723	.0254724
var(Residual)	.0293565	.0014523	.0266438	.0323455

LR test vs. linear regression: chibar2(01) = 216.93 Prob >= chibar2 = 0.0000

So for level 1, we have var(Residual) .0293565 vs .0358148 and for level 2, var(_cons) is .0199164 vs .0445029. Now we can calculate R squared:

```
* Level 1:
. di (.0358148-.0293565)/.0358148
.18032489
* Level 2:
. di (.0445029-.0199164) /.0445029
.55246961

* Total R-squared:
. di (.0358148+.0445029-.0293565-.0199164)/(.0358148+.0445029)
.38652501
```

We can also run xtreg to obtain approximate R-squared values (since we do not need random slopes for that) but those R squared values will compare our model to a null model without any variables (so without age variables as well), so they will be somewhat different:

```
. xtreg attit agel6 agel6expomean agel6female agel6sq agel6sqexpomean agel6sqfemale
expochange expochminority expochincomem expomean female minority incomem
```

```
Random-effects GLS regression                Number of obs    =    1066
Group variable: id                          Number of groups  =     241

R-sq:  within = 0.2477                      Obs per group:  min =     1
        between = 0.5058                      avg   =     4.4
        overall = 0.4028                      max   =     5

Random effects u_i ~ Gaussian                Wald chi2(13)    =    504.52
corr(u_i, X) = 0 (assumed)                  Prob > chi2      =     0.0000
```

attit	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
agel6	.0443167	.0103721	4.27	0.000	.0239877	.0646457
agel6expom~n	-.0426375	.0153122	-2.78	0.005	-.0726489	-.0126261
agel6female	.0039313	.0077485	0.51	0.612	-.0112554	.0191181
agel6sq	-.0077005	.0085027	-0.91	0.365	-.0243655	.0089645
agel6sqexp~n	-.0071247	.0126492	-0.56	0.573	-.0319167	.0176674
agel6sqfem~e	.0148111	.0063908	2.32	0.020	.0022854	.0273368
expochange	.4450367	.0360808	12.33	0.000	.3743196	.5157538
expochmino~y	-.3099835	.0741428	-4.18	0.000	-.4553007	-.1646663
expochinco~m	-.0478129	.0141375	-3.38	0.001	-.075522	-.0201038

expomean		.6297261	.0487233	12.92	0.000	.5342302	.7252221
female		-.0724031	.0251082	-2.88	0.004	-.1216142	-.023192
minority		.0022791	.0273332	0.08	0.934	-.0512929	.0558512
incomem		.0146007	.0048037	3.04	0.002	.0051855	.0240159
_cons		.1798357	.033503	5.37	0.000	.114171	.2455004

sigma_u		.14136818					
sigma_e		.17125674					
rho		.40526077	(fraction of variance due to u_i)				

Diagnostics for mixed effects models

To conduct diagnostics for mixed effects models, we can obtain predicted values and various residuals; here is what we can obtain using predict command:

xb	xb, linear predictor for the fixed portion of the model
stdp	standard error of the fixed-portion linear prediction xb
fitted	fitted values, linear predictor of the fixed portion plus contributions based on predicted random effects
residuals	residuals, response minus fitted values
rstandard	standardized residuals
reffects	best linear unbiased predictions (BLUPs) of the random effects. By default, BLUPs for all random effects in the model are calculated. You must specify q new variables, where q is the number of random-effects terms in the model.
reses	standard errors of the best linear unbiased predictions (BLUPs) of the random effects. By default, standard errors for all BLUPs in the model are calculated. You must specify q new variables.

Thus, residuals will give you level 1 residuals, and reffects will give you level 2 residuals for each level 2 random component. You should examine both types of residuals to assess normality. To assess linearity, you can plot residuals (level 1 and level 2) against each predictor using lowess (there should be no relationship between them). You can also examine both level 1 and level 2 residuals for potential outliers (i.e., look for observations with large standardized residuals).

There is no robust option available with xtmixed, but to obtain standard errors and significance tests that are less dependent on assumptions, you can use bootstrapping, although it does take time to calculate. E.g., compare to the model on pp.9-10:

```
. bootstrap: xtmixed attit age16 age16sq expo || id: age16 age16sq expo,
cov(unstructured)
(running xtmixed on estimation sample)

Bootstrap replications (50)
-----+----- 1 -----+----- 2 -----+----- 3 -----+----- 4 -----+----- 5
..... 50

Mixed-effects REML regression      Number of obs      =      1066
Group variable: id                 Number of groups   =       241

Obs per group: min =                1
                        avg =         4.4
                        max =                5
```

