656 Lab: Growth Curve Modeling (from pages 78-87 and 91-94 of the textbook)

Data: Weight gain in Asian children in Britian.

Variables
- id: child identifier
- weight: weight in Kg
- age: age in years
- gender: gender (1: male, 2: female)

Goal: Compare `xtmixed` and `gllamm` for modeling quadratic growth curve trajectories.

```
. use http://www.stata-press.com/data/mlmus/asian, clear
. label def g 1 "boy" 2 "girl"
. label values gender g
```

Exploratory Data Analysis
What does the data look like? First, we will find out how many children we have in the study and how often they had their weight measured. Note that we have to generate a time variable because in order to use the `xtdes` command, STATA needs the time variable to be an integer and age is reported in (non-integer) years.

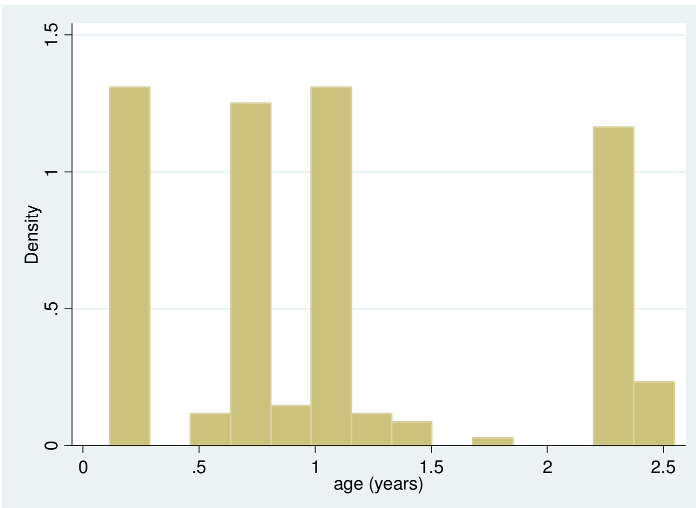
```
. by id: gen time=_n
. tsset id time
. xtdes

      id:  45, 258, ..., 4975                                n =         68
    time:  1, 2, ..., 5                                      T =          5
           Delta(time) = 1; (5-1)+1 = 5
           (id*time uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       2         3         4       4       5

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
       27     39.71   39.71 |  111..
       19     27.94   67.65 |  11...
       15     22.06   89.71 |  1111.
        4      5.88   95.59 |  1....
        3      4.41  100.00 |  11111
 ---------------------------+---------
       68    100.00         |  XXXXX
```

We have 68 children, with a maximum of 5 observations per child (3 children) and minimum of 1 observation per child (4 children). The most common number of observations per child (the mode) is 3, since 27 children have 3 observations. Note that missing observations always occur as the child ages and we have no 'gaps' in our observations on weight.

```
. sum age
```

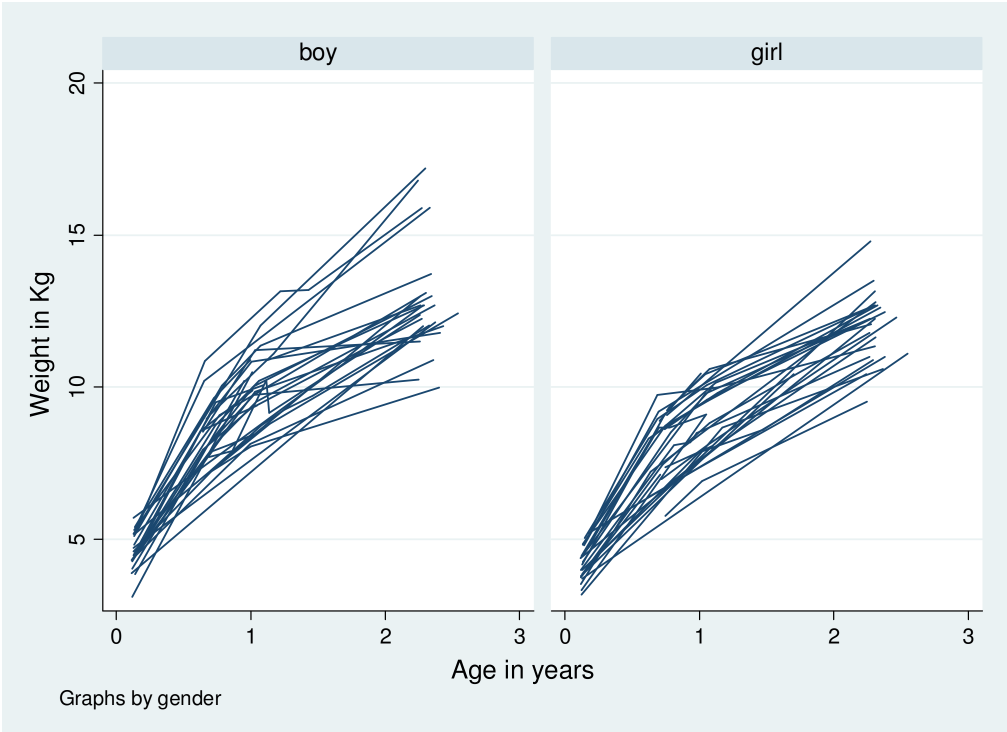
```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |       198    1.080552     .787069   .1149897   2.546201
```

```
. hist age, xtitle(age (years))
```



Weights are generally measured on children at ages 6 weeks, and at 8, 12 and 27 months. Now let's take a look how weight changes over time for each child.

```
. sort id age

. graph twoway (line weight age, connect(ascending)), by(gender)
xtitle(Age in years) ytitle(Weight in Kg)
```



The childrens' growth appears to be non-linear in relation to time. Since the relationship between weight and age is non-linear, we will include a quadratic term in our model. Note also that at the first weight measurement, it appears that each child has his or her own starting weight and that we could consider these starting weights to be an approximately normally distributed random variable. We will build a random intercept into our initial model.

xtmixed

Quadratic growth with random intercept model where U_{1j} is the random intercept for child *j*:

$$weight_{ij} = \beta_1 + \beta_2 age_{ij} + \beta_3 age_{ij}^2 + U_{1j} + \varepsilon_{ij}$$

```
. ** quadratic growth with random intercept **
. gen age2 = age^2

. xtmixed weight age age2 || id:, mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -276.83266
Iteration 1:   log likelihood = -276.83266

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       198
Group variable: id                              Number of groups   =        68
```

```
                                                Obs per group: min =         1
                                                               avg =       2.9
                                                               max =         5


                                                Wald chi2(2)       =   2623.63
Log likelihood = -276.83266                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.817918   .2896529    26.99   0.000     7.250209    8.385627
        age2 |  -1.705599   .1085984   -15.71   0.000    -1.918448    -1.49275
       _cons |   3.432859   .1810702    18.96   0.000     3.077968     3.78775
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   sd(_cons) |   .9182256   .0973788      .7458965    1.130369
-----------------------------+------------------------------------------------
                sd(Residual) |   .7347063   .0452564      .6511507    .8289837
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    78.07 Prob >= chibar2 = 0.0000
```

The estimated standard deviation of the random intercept is 0.918.

Quadratic growth with random intercept U_{1j} and random slope U_{2j} for child *j*:

$$weight_{ij} = \beta_1 + \beta_2 age_{ij} + \beta_3 age_{ij}^2 + U_{1j} + U_{2j} age_{ij} + \varepsilon_{ij}$$

By including a random slope, we allow children to have different rates of growth.

```
. ** quadratic growth with random intercept and random slope **
.
. xtmixed weight age age2 || id: age, cov(unstr) mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -258.13527
Iteration 1:   log likelihood =  -258.0782
Iteration 2:   log likelihood = -258.07784
Iteration 3:   log likelihood = -258.07784

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       198
Group variable: id                              Number of groups   =        68

                                                Obs per group: min =         1
                                                               avg =       2.9
                                                               max =         5


                                                Wald chi2(2)       =   1978.20
Log likelihood = -258.07784                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.703998   .2394082    32.18   0.000     7.234767    8.173229
        age2 |  -1.660465   .0885229   -18.76   0.000    -1.833967   -1.486963
```

```
       _cons |   3.494512   .1372636    25.46   0.000      3.22548    3.763544
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                     sd(age) |   .5040802   .0879337       .358107    .7095558
                   sd(_cons) |   .6359558   .1293523      .4268684    .9474578
             corr(age,_cons) |   .2747814   .3309063     -.3965135    .7546038
-----------------------------+------------------------------------------------
                sd(Residual) |   .5757751   .0505985      .4846745    .6839993
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   115.58   Prob > chi2 = 0.0000

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

Quadratic growth with random intercept U_{1j} and random slope U_{2j} for child j that includes a child-level covariate, an indicator of gender:

$$weight_{ij} = \beta_1 + \beta_2 age_{ij} + \beta_3 age_{ij}^2 + \beta_4 girl_j + U_{1j} + U_{2j} age_{ij} + \varepsilon_{ij}$$

```
. ** including a child-level covariate **
.
. gen girl = gender - 1

.
. xtmixed weight age age2 girl || id: age , cov(unstr) mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -253.91218
Iteration 1:   log likelihood = -253.86704
Iteration 2:   log likelihood = -253.86692
Iteration 3:   log likelihood = -253.86692

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       198
Group variable: id                              Number of groups   =        68

                                                Obs per group: min =         1
                                                               avg =       2.9
                                                               max =         5


                                                Wald chi2(3)       =   1975.44
Log likelihood = -253.86692                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.697967   .2382121    32.32   0.000      7.23108    8.164855
        age2 |  -1.657843   .0880529   -18.83   0.000    -1.830423   -1.485262
        girl |  -.5960093   .1963689    -3.04   0.002    -.9808853   -.2111332
       _cons |   3.794769   .1655053    22.93   0.000     3.470385    4.119153
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                     sd(age) |   .5097089   .0871791      .3645317    .7127039
                   sd(_cons) |    .594731   .1289891      .3887823    .9097762
```

```
             corr(age,_cons) |   .1571086   .3240801     -.4564674    .6694143
-----------------------------+------------------------------------------------
                sd(Residual) |   .5723301   .0496274      .4828786    .6783521
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   104.17   Prob > chi2 = 0.0000

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

gllamm

When modeling random effects beyond a random intercept in gllamm, we need to use the `eq` command to specify the equation for the variable multiplying each random effect and include the name of each equation in the `eqs` option of gllamm.

```
. ** quadratic growth with random intercept **

. gen cons = 1

.
. eq inter: cons

.
. gllamm weight age age2, i(id) eqs(inter) adapt

Running adaptive quadrature
Iteration 0:    log likelihood = -303.31828
Iteration 1:    log likelihood = -279.21855
Iteration 2:    log likelihood = -276.88181
Iteration 3:    log likelihood = -276.83266
Iteration 4:    log likelihood = -276.83266


Adaptive quadrature has converged, running Newton-Raphson
Iteration 0:   log likelihood = -276.83266
Iteration 1:   log likelihood = -276.83266

number of level 1 units = 198
number of level 2 units = 68

Condition Number = 14.785391

gllamm model

log likelihood = -276.83266

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.817871   .2899873    26.96   0.000     7.249507    8.386236
        age2 |  -1.705589   .1086957   -15.69   0.000    -1.918629    -1.49255
       _cons |   3.432893   .1811779    18.95   0.000      3.07779    3.787995
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------

  .53966034 (.06647545)

Variances and covariances of random effects
------------------------------------------------------------------------------


***level 2 (id)

    var(1): .84334423 (.17887769)
```

```
------------------------------------------------------------------------------

. ** quadratic growth with random intercept and random slope **
```

By defining this matrix, we are storing the parameter estimates from the previous model, which we will use as starting values for the parameter estimates in the next model.

```
. matrix a = e(b)

.
. eq slope: age
```

The option `nrf(2)` specifies that we now have two random effects (intercept and slope). The `ip(m) nip(15)` specifies that we are using a spherical integration rule of degree 15 (don't need to worry about this – just know that it speeds up the estimation).

```
. gllamm weight age age2, i(id) nrf(2) eqs(inter slope) ip(m) nip(15) from(a)
adapt

Running adaptive quadrature
Iteration 0:    log likelihood = -276.83266
Iteration 1:    log likelihood = -264.70282
Iteration 2:    log likelihood = -258.46797
Iteration 3:    log likelihood = -258.40577
Iteration 4:    log likelihood = -258.08334
Iteration 5:    log likelihood = -258.07834
Iteration 6:    log likelihood = -258.07802
Iteration 7:    log likelihood =   -258.078


Adaptive quadrature has converged, running Newton-Raphson
Iteration 0:   log likelihood =   -258.078
Iteration 1:   log likelihood =   -258.078  (backed up)
Iteration 2:   log likelihood = -258.07784
Iteration 3:   log likelihood = -258.07784

number of level 1 units = 198
number of level 2 units = 68

Condition Number = 8.938685

gllamm model

log likelihood = -258.07784

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.703998      .24026    32.07   0.000     7.233097    8.174899
        age2 |  -1.660465    .0890109   -18.65   0.000    -1.834923   -1.486007
       _cons |   3.494512    .1376254    25.39   0.000     3.224771    3.764253
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------

  .33151691 (.05826674)

Variances and covariances of random effects
------------------------------------------------------------------------------
```

```
***level 2 (id)

    var(1): .4044401 (.16452478)
    cov(2,1): .08808734 (.08802551) cor(2,1): .27478094

    var(2): .25409703 (.08865128)
------------------------------------------------------------------------------
```

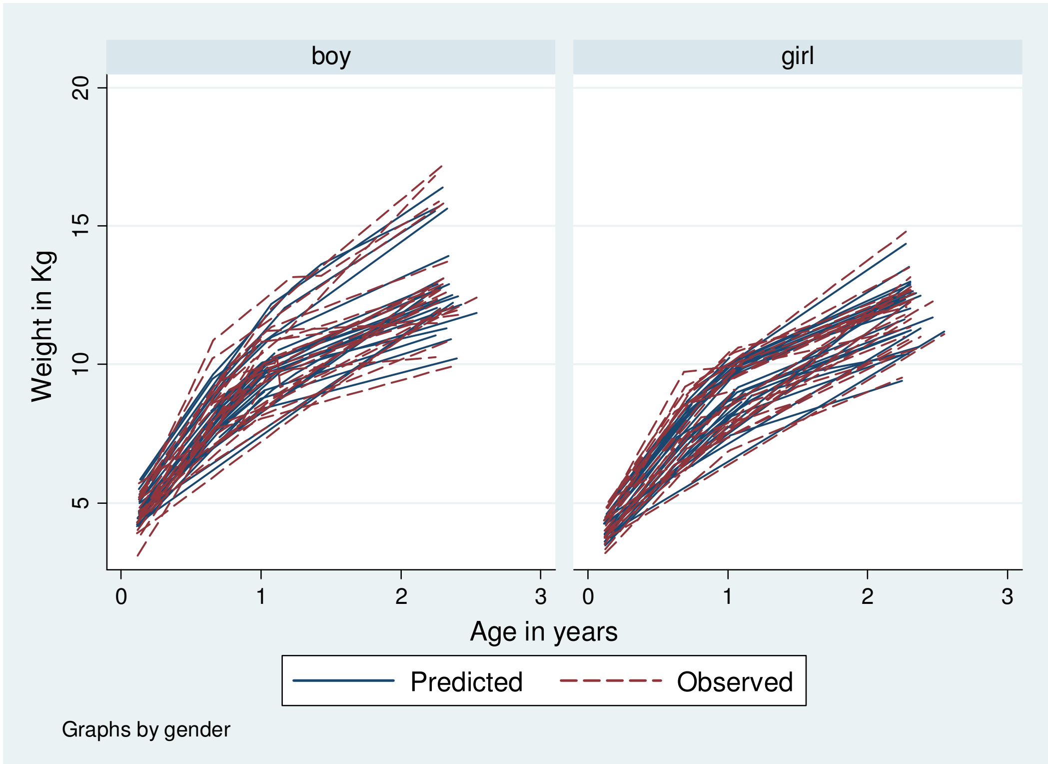
Look at Table 3.2 in the textbook (page 82) that compares the results from xtmixed to those from gllamm. The results from xtmixed and gllamm are identical for the coefficient estimates and standard errors of the betas (the fixed part of the model) however, the estimates of the random parts of the models vary according to the stata procedure.

Predicting trajectories for each child

- **`xtmixed`**

Get the empirical Bayes estimates of the random intercepts and random slopes

```
. * re-run the xtmixed including the child-level covariate
. xtmixed weight age age2 girl || id: age , cov(unstr) mle

. predict traj, fitted

. sort id age

. graph twoway (line traj age, connect(ascending)) (line weight age,
connect(ascending) clpatt(dash)), by(gender) xtitle(Age in years)
ytitle(Weight in Kg) legend(order(1 "Predicted" 2 "Observed"))
```



The model appears to fit the data adequately based on a comparison of the fitted trajectories to the observed trajectories.

- **`gllamm`**

Get the empirical Bayes estimates of the random intercepts and random slopes

```
. * re-run the gllamm including the child-level covariate
. gllamm weight age age2, i(id) nrf(2) eqs(inter slope) ip(m) nip(15)
from(a) adapt

. gllapred traj, linpred

. graph twoway (line traj age, connect(ascending)) (line weight age,
connect(ascending) clpatt(dash)), by(gender) xtitle(Age in years)
ytitle(Weight in Kg) legend(order(1 "Predicted" 2 "Observed"))
```

This will produce the same graph as before.