Multilevel Mixed (hierarchical) models

Christopher F Baum

EC 823: Applied Econometrics

Boston College, Spring 2013

Introduction to mixed models

Stata supports the estimation of several types of *multilevel mixed models*, also known as hierarchical models, random-coefficient models, and in the context of panel data, repeated-measures or growth-curve models. These models share the notion that individual observations are grouped in some way by the design of the data.

Mixed models are characterized as containing both fixed and random effects. The fixed effects are analogous to standard regression coefficients and are estimated directly. The random effects are not directly estimated but are summarized in terms of their estimated variances and covariances. Random effects may take the form of random intercepts or random coefficients.

For instance, in hierarchical models, individual students may be associated with schools, and schools with school districts. There may be coefficients or random effects at each level of the hierarchy. Unlike traditional panel data, these data may not have a time dimension.

In repeated-measures or growth-curve models, we consider multiple observations associated with the same subject: for instance, repeated blood-pressure readings for the same patient. This may also involve a hierarchical component, as patients may be grouped by their primary care physician (PCP), and physicians may be grouped by hospitals or practices.

Linear mixed models

The simplest sort of model of this type is the *linear mixed model*, a regression model with one or more random effects. A special case of this model is the one-way random effects panel data model implemented by `xtreg, re`. If the only random coefficient is a random intercept, that command should be used to estimate the model.

For more complex models, the command `xtmixed` may be used to estimate a multilevel mixed-effects regression. Consider a dataset in which students are grouped within schools (from Rabe-Hesketh and Skrondal, *Multilevel and Longitudinal Modeling Using Stata, 3rd Edition*, 2012). We are interested in evaluating the relationship between a student's age-16 score on the GCSE exam and their age-11 score on the LRT instrument.

As the authors illustrate, we could estimate a separate regression equation for each school in which there are at least five students in the dataset:

```
. use gcse, clear

. egen nstu = count(gcse + lrt), by(school)

. statsby alpha = _b[_cons] beta = _b[lrt], by(school) ///
> saving(indivols, replace) nodots: regress gcse lrt if nstu > 4

      command:  regress gcse lrt if nstu > 4
        alpha:  _b[_cons]
         beta:  _b[lrt]
           by:  school
```

That approach gives us a set of 64 αs (intercepts) and βs (slopes) for the relationship between `gcse` and `lrt`. We can consider these estimates as data and compute their covariance matrix:

```
. use indivols, clear
(statsby: regress)

. summarize alpha beta

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       alpha |        64   -.1805974    3.291357  -8.519253   6.838716
        beta |        64    .5390514    .1766135   .0380965   1.076979

. correlate alpha beta, covariance
(obs=64)

             |    alpha     beta
-------------+------------------
       alpha |   10.833
        beta |  .208622  .031192
```

To estimate a single model, we could consider a fixed-effects approach (`xtreg, fe`), but the introduction of random intercepts and slopes for each school would lead to a regression with 130 coefficients.

Furthermore, if we consider the schools as a random sample of schools, we are not interested in the individual coefficients for each school's regression line, but rather in the mean intercept, mean slope, and the covariation in the intercepts and slopes in the population of schools.

A more sensible approach is to specify a model with a school-specific random intercept and school-specific random slope for the i^{th} student in the j^{th} school:

$$y_{i,j} = (\beta_1 + \delta_{1,j}) + (\beta_2 + \delta_{2,j})x_{i,j} + \epsilon_{i,j}$$

We assume that the covariate x and the idiosyncratic error ϵ are both independent of δ_1, δ_2.

The random intercept and random slope are assumed to follow a bivariate Normal distribution with covariance matrix:

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

Implying that the correlation between the random intercept and slope is

$$\rho_{12} = \frac{\psi_{21}}{\sqrt{\psi_{11}\psi_{22}}}$$

We could estimate a special case of this model, in which only the intercept contains a random component, with either `xtreg, re` or `xtmixed, mle`.

The syntax of Stata's `xtmixed` command is

`xtmixed` *depvar fe_eqn* [|| *re_eqn*] [|| *re_eqn*] [, *options*]

The `fe_eqn` specifies the fixed-effects part of the model, while the `re_eqn` components optionally specify the random-effects part(s), separated by the double vertical bars (`||`). If a `re_eqn` includes only the level of a variable, it is listed followed by a colon (`:`). It may also specify a linear model including an additional *varlist*.

```
. xtmixed gcse lrt || school: if nstu > 4, mle nolog
Mixed-effects ML regression                     Number of obs      =      4057
Group variable: school                          Number of groups   =        64

                                                Obs per group: min =         8
                                                               avg =      63.4
                                                               max =       198

                                                Wald chi2(1)       =   2041.42
Log likelihood = -14018.571                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5633325   .0124681    45.18   0.000     .5388955    .5877695
       _cons |   .0315991   .4018891     0.08   0.937    -.7560891    .8192873
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Identity             |
                   sd(_cons) |   3.042017   .3068659      2.496296     3.70704
-----------------------------+------------------------------------------------
                sd(Residual) |    7.52272   .0842097       7.35947    7.689592
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   403.32 Prob >= chibar2 = 0.0000
```

By specifying `gcse lrt || school:`, we indicate that the fixed-effects part of the model should include a constant term and slope coefficient for `lrt`. The only random effect is that for `school`, which is specified as a random intercept term which varies the school's intercept around the estimated (mean) constant term.

These results display the `sd(_cons)` as the standard deviation of the random intercept term. The likelihood ratio (LR) test shown at the foot of the output indicates that the linear regression model in which a single intercept is estimated is strongly rejected by the data.

This specification restricts the school-specific regression lines to be parallel in `lrt-gcse` space. To relax that assumption, and allow each school's regression line to have its own slope, we add `lrt` to the random-effects specification. We also add the `cov(unstructured)` option, as the default is to set the covariance (ψ_{21}) and the corresponding correlation to zero.

```
. xtmixed gcse lrt || school: lrt if nstu > 4, mle nolog ///
> covariance(unstructured)
Mixed-effects ML regression                     Number of obs      =      4057
Group variable: school                          Number of groups   =        64
                                                Obs per group: min =         8
                                                               avg =      63.4
                                                               max =       198
                                                Wald chi2(1)       =    779.93
Log likelihood = -13998.423                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5567955   .0199374    27.93   0.000     .5177189    .5958721
       _cons |  -.1078456   .3993155    -0.27   0.787    -.8904895    .6747984
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                    sd(lrt)  |   .1205424   .0189867      .0885252    .1641394
                  sd(_cons)  |   3.013474    .305867      2.469851    3.676752
            corr(lrt,_cons)  |    .497302   .1490473      .1563124    .7323728
-----------------------------+------------------------------------------------
               sd(Residual)  |   7.442053   .0839829      7.279257    7.608491
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   443.62   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

These estimates present the standard deviation of the random slope parameters (`sd(lrt)`) as well as the estimated correlation between the two random parameters (`corr(lrt,_cons)`). We can obtain the corresponding covariance matrix with `estat`:

```
. estat recovariance
Random-effects covariance matrix for level school

             |       lrt      _cons
-------------+----------------------
         lrt |   .0145305
       _cons |   .1806457   9.081027
```

These estimates may be compared with those generated by school-specific regressions. As before, the likelihood ratio (LR) test of the model against the linear regression in which these three parameters are set to zero soundly rejects the linear regression model.

The dataset also contains a school-level variable, `schgend`, which is equal to 1 for schools of mixed gender, 2 for boys-only schools, and 3 for girls-only schools. We interact this qualitative factor with the continuous `lrt` model to allow both intercept and slope to differ by the type of school:

```
. xtmixed gcse c.lrt##i.schgend || school: lrt if nstu > 4, mle nolog ///
> covariance(unstructured)
Mixed-effects ML regression                     Number of obs      =      4057
Group variable: school                          Number of groups   =        64
                                                Obs per group: min =         8
                                                               avg =      63.4
                                                               max =       198
                                                Wald chi2(5)       =    804.34
Log likelihood = -13992.533                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lrt |   .5712245   .0271235    21.06   0.000     .5180634    .6243855
             |
     schgend |
           2 |   .8546836    1.08629     0.79   0.431    -1.274405    2.983772
           3 |    2.47453   .8473229     2.92   0.003     .8138071    4.135252
             |
    schgend#|
       c.lrt |
           2 |  -.0230016    .057385    -0.40   0.689    -.1354742    .0894709
           3 |  -.0289542   .0447088    -0.65   0.517    -.1165818    .0586734
             |
       _cons |  -.9975795   .5074132    -1.97   0.049    -1.992091   -.0030679
------------------------------------------------------------------------------
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school: Unstructured				
sd(lrt)	.1198846	.0189169	.0879934	.163334
sd(_cons)	2.801682	.2895906	2.287895	3.43085
corr(lrt,_cons)	.5966466	.1383159	.2608112	.8036622
sd(Residual)	7.442949	.0839984	7.280122	7.609417

```
LR test vs. linear regression:       chi2(3) =   381.44   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

The coefficients on `schgend` levels 2 and 3 indicate that girls-only schools have a significantly higher intercept than the other school types. However, the slopes for all three school types are statistically indistinguishable. Allowing for this variation in the intercept term has reduced the estimated variability of the random intercept (`sd(_cons)`).

Just as `xtmixed` can estimate multilevel mixed-effects linear regression models, `xtmelogit` can be used to estimate logistic regression models incorporating mixed effects, and `xtmepoisson` can be used for Poisson regression (count data) models with mixed effects.

More complex models, such as ordinal logit models with mixed effects, can be estimated with the user-written software `gllamm` by Rabe-Hesketh and Skrondal (see their earlier-cited book, or `ssc describe gllamm` for details).