Chapter 4 Random Coefficient Models

Paul E. Johnson

February 13, 2016

1 Glossary

New in this chapter: Separate random effects applied to intercept and slope.

Notation continuing from previous

ξ_{ij}: "xi" total (or combined) individual row-level error term for group i row j,

ζ_j: "zeta" a group level random error, j indexes a grouping variable, $Var(\zeta_j) = \psi$

ϵ_{ij}: "epsilon" $Var(\epsilon_{ij}) = \theta$. Individual row-level error uncorrelated with ζ_j

$Cov(y_{ij}, y_{i'j}|\zeta_j) = 0$. Apart from ζ_j, the observed scores are "conditionally uncorrelated"

The combined error term has two variance components

$$\xi_{ij} = \zeta_j + \epsilon_{ij} \tag{1}$$

2 Highlights

1. Notation is tricky because in matrix algebra the first subscript is supposed to be the row variable and the second subscript is the columns. In this book, however, they write the intercept random error as ζ_{1j} and the slope error ζ_{2j}, but if you put those in a matrix, you see the subscripts are clearly backwards:

	intercept ζ_{1j}	slopeζ_{2j}
group 1	ζ_{11}	ζ_{21}
group 2	ζ_{12}	ζ_{22}
$\vdots$		
group J	ζ_{1J}	ζ_{2J}

   whereas, in proper matrix algebra, the first subscript would be j and the second would be the column number.

   If you find other notes I've written in the past, you will notice I hate these backward subscripts and so I'd rather refer to the intercept and slope errors as, for example, $\zeta 1_j$ and $\zeta 2_j$

2. p. 182. Within school model. Take each school as isolated subset. Or could estimate unique intercept and slope for each in one big model (using factor notation).

3. p. 188. Specification. Assert that the intercept is $(\beta_1 + \zeta_{1j})$ while the sloe is $(\beta_2 + \zeta_{2j})$. That means the regression is

$$y_{ji} = \{\beta_1 + \zeta_{1j}\} + \{\beta_2 + \zeta_{2j}\}x_{ji} + \epsilon_{ji} \tag{2}$$

$$\begin{aligned} y_{ji} &= \beta_1 + \beta_2 x_{ji} + \{\zeta_{1j} + \zeta_{2j}x_{ji} + \epsilon_{ji}\} \\ &= \{fixed\, predictors\} + \{random\, effects\} \end{aligned}$$

   Assumptions

   a) error terms have expected values 0. Given the value of the group level variables, $E[\zeta_j|X] = E[\zeta_{2j}|X] = 0$

   b) variance/covariance of errors ζ_{1j} and ζ_{2j}. This is more difficult for me to tease out of RHS p. 190.

      i. TODO: return and figure out way to explain this to students.

4. Skip to p. 210: Two Stage Model Formulation

   Multi-level models are all estimated at one level: the individual level. Group variables and random effects enter the analysis because they are "copied" among all the rows in a group. In a sense, there is really no such thing as a multi-level model because, when the rubber hits the road, everything is written out at the individual level.

   Nevertheless, there are benefits to thinking through the problem with several levels. The most famous way of doing this is popularized in the famous book Hierarchical Linear Models by Raudenbush and Bryk. The main idea is that we formulate a models for the intercept and slope coefficients at the group level ("level 2") and then we insert them into the individual regression model.

   I think the notation in the RHS book is somewhat baffling, so I'll write it in a way I understand and we can talk over the differences. I'll still use their 1, 2 subscripts for groups, I just don't want to introduce γ as a separate layer of unneeded notation.

   a) Level 1 (individual level) regression. Start with the idea that there is a regression model that has different intercept β^*_{0j} and slope β^*_{1j} for each group j

   $$y_{ji} = \beta^*_{1j} + \beta^*_{2j}x_{ji} + \varepsilon_{ji} \tag{3}$$

   The modeling procedure is

      i. step 1. create a separate model for the level 2 variables, the ones subscripted by j

      ii. step 2. algebraically insert them into the level 1 model (3)

      iii. step 3. rearrange to re-group the fixed coefficients and random effect coefficients.

b) Random intercept model.

Level 2 model: the intercept does vary, but the slope is the same for all. How to write that out?

Level 2 Slope model: Simple! (same for all)

$$\beta^*_{2j} = \beta_2 \tag{4}$$

Level 2 Intercept model: Suppose there is normal random noise added to a "mean of the intercepts" (expected value, population value) β_1

$$\beta^*_{1j} = \beta_1 + \zeta_j \tag{5}$$

Insert those into the regression model in (3)

$$y_{ji} = \beta_1 + \zeta_j + \beta_2 x_{ji} + \varepsilon_{ji} \tag{6}$$

Rearrange to see that you get back the model we've been discussing

$$y_{ji} = \beta_1 + \beta_2 x_{ji} + \{\varepsilon_{ji} + \zeta_j\} \tag{7}$$

The combined error term is thus heteroskedastic,

$$\xi_{ji} = \zeta_j + \varepsilon_{ji}. \tag{8}$$

You could estimate that with OLS regression, but you are violating some assumptions, which means 1) the OLS estimates of the β's have higher variance than the alternative GLS estimator and 2) the standard errors from OLS are wrong.

c) Random intercept and random slope

Level 2 model: We need two random effects, one for the intercept (ζ_{1j}) and one for the slope (ζ_{2j}).

Level 2 Intercept model: Lets say that is the same as previous, except we need ζ_{1j} rather than ζ_j:

$$\beta^*_{1j} = \beta_1 + \zeta_{1j} \tag{9}$$

Level 2 Slope model: let the group j slope be a common element β_2 plus a random shock

$$\beta^*_{2j} = \beta_2 + \zeta_{2j} \tag{10}$$

$$y_{ji} = \beta_1 + \zeta_{1j} + (\beta_2 + \zeta_{2j})x_{ji} + \varepsilon_{ji} \tag{11}$$

$$y_{ji} = \beta_1 + \beta_2 x_{ji} + \{\varepsilon_{ji} + \zeta_{1j} + \zeta_{2j} x_{ji}\} \tag{12}$$

The combined error term is thus heteroskedastic (and a little horrible, actually):

$$\xi_{ji} = \zeta_{1j} + \zeta_{2j} x_{ji} + \varepsilon_{ji}. \tag{13}$$

Again, you could estimate that with OLS, but the GLS model would have lower variance of parameter estimates.

5. How go get a fancier looking model with predictors at level 2. This is where the two stage model notation can help us.

   a) Fancier Level 2 intercept model. Let a group level predictor $x3_j$ exist. The value of this is the same for all observations in group j. This thing is predictive of the intercept as

   $$\beta^*_{1j} = \beta_1 + \beta_3 x3_j + \zeta_{1j} \tag{14}$$

   b) We have to cheat on notation now. This variable $x3_j$ does not vary within group, but I need to have an i-subscripted version of it. So let $x3_{ji} = x3_j$. For all i in j, the value is the same thing. We need $x3_{ji}$ when we do step 3 in the analysis procedure

   c) Insert that group intercept model into (3):

   $$y_{ji} = \beta_1 + \beta_3 x3_{ji} + \zeta_{1j} + \beta^*_{2j} x_{ji} + \varepsilon_{ji} \tag{15}$$

   $$y_{ji} = \beta_1 + \beta_3 x3_{ji} + \beta^*_{2j} x_{ji} + \zeta_{1j} + \varepsilon_{ji} \tag{16}$$

   The end result of inserting a level-2 predictor in the intercept model () is that a new predictor appears in the final regression model.

   d) Now watch what (bad thing) happens when we say there is a level 2 predictor in the slope model:

   $$\beta^*_{2j} = \beta_2 + \beta_4 x4_j + \zeta_{2j} \tag{17}$$

   Insert that in (15):

   $$y_{ji} = \beta_1 + \beta_3 x3_{ji} + \zeta_{1j} + \{\beta_2 + \beta_4 x4_j + \zeta_{2j}\} x_{ji} + \varepsilon_{ji} \tag{18}$$

   $$= \beta_1 + \beta_3 x3_{ji} + \beta_2 x_{ji} + \beta_4 x4_{ji} x_{ji} + \{\zeta_{1j} + \zeta_{2j} x_{ji} + \varepsilon_{ji}\} \tag{19}$$

   OUT POPS an interaction! $\beta_4 x4_{ji} x_{ji}$.

6. Now back to RHS p. 191. We want to show that the variance of the intercept, and the covariance of the intercept with the slope random errors, are ill defined in some sense. They are both dependent on the scaling of the x predictor, which is arbitrary. Hence, any changes in estimates that flow by adding or substracting from x are purely superficial changes, they don't solve anything, but they may make the estimates easier to understand.

   a) The combined error is
   $$\xi_{ji} = \zeta_{1j} + \zeta_{2j}x_{ji} + \varepsilon_{ji}. \tag{20}$$

   b) Variance $Var(\xi_{ji}) = \psi_{11} + 2\psi_{21}x_{ij} + \psi_{22}x_{ij}^2 + \theta$.

      Recall

      $\theta = Var(\varepsilon_{ji})$

      $\psi_{11} = Var(\zeta_{1j})$

      $\psi_{21} = Cov(\zeta_{1j}, \zeta_{2j})$

      Background.

      i. Remember fundamental law
      $$Var(k_1X + k_2Y) = k_1^2Var(X) + k_2^2Var(Y) + 2k_1k_2Cov(X,Y) \tag{21}$$
      For constants k_1, k_2 and random variables X and Y.

      ii. Remember in (), that x_{ji} is a constant, it is a pre-determined value. The randoms are the ζs and ε.

7. VERY VITAL graph on p. 193.

   a) The variance of the slope exaggerates the variance of the intercept,

   b) Even if intercept variance is 0, the slope variance causes apparent variation in the intercept.

   c) Depending on where you draw the y axis, the intercept's variance appears larger or smaller.

   d) IMPLICATIONS

      i. Centering (subtract a constant from x_{ji}) causes an apparent "shift" in y axis and a change in calculated intercept. CLEARLY, this is not a substantively important change, it should not alter our conclusions at all. But it alters the calculations in an entirely superficial/frustrating way that confuses many researchers.

      ii. But it does alter the observed diversity among intercepts of the true regression lines for the subgroups.

iii. p. 192 RHS emphasize "lack of invariance" in variance estimates. These estimates from random intercept variance is changed by centering x, and the estimated covariance between ζ_1 and ζ_s is altered as well. This is the same old "mean-centering does not matter" debate combing back. THESE ESTIMATE CHANGES ARE ENTIRELY SUPERFICIAL and NOT WORTH EMPHASIZING, they don't solve any statistical problems and they don't make one model more precise than the other.

e) Other implications

i. IF you fit a model that includes a random slope, it is NECESSARY to include a random intercept as well.

ii. Why? Draw a graph of a random slope when the intercept is the same for all groups.

f) p. 194. "To make ψ_{11} and ψ_{21}interpretable, it makes sense to translate x_{ij} so that the value $x_{ij} = 0$ is a useful reference point in some way. Typical choices are either mean-centering (as for lrt) or, if x_{ij} is time, as in growth-curve models, defining 0 to be the initial time in some sense. Because the magnitude and interpretation of ψ_{21}depend on the location (or translation) of x_{ij}, which is often arbitrary, it generally does not make sense to set $\psi_{12}$to 0 by specifying uncorrelated intercepts and slopes."

8. Test slope variance

a) Use an LR test of the model with random slope compared with one that omits this.

b) MUST not use REML estimates for this test. Only MLE will do.

c) The χ^2 value and p values are not exactly correct, but we believe they are "conservative" in the sense that the p value is larger than it ought to be. Hence, if we reject the null, we are probably correct.

d) However, suppose p=0.06. Temptation is to say we ought to still reject, since p estimate is larger than it ought to be.

9. p. 199. Interpreting variance estimates.

a) Calculate the variance of the error given X .

b) Graph on p. 200 perhaps not so informative, should make an hourglass sort of plot. Will draw on board.

10. Estimating random intercepts and slopes
    a) RHS says "likelihood" because of interpretation difficulty."In general, we therefore donot recommend using this method and suggest using empirical Bayes prediction instead." So lets don't bother with this.
    b) Rather use Empriical Bayes estimates, they are conditional modes. See Stata codes below.

3 Stata code

1. p. 182: plot predicted value from a single regression. TODO: See why the sort is needed in here? Let z be a grouping variable

   ```
   regress y x if z == 1
   predict yhat, xb
   twoway (scatter y x)(line yhat x, sort) if z == 1
   ```

   That's for group z == 1 only

2. Trellis Graphic. Stata's implementation uses twoway, but with a by statement so that many separate models are plotted. I understand how the scatters are produced, but don't fully understand the "lfit" keyword in the code

   ```
   twoway (scatter y x)(lfit y x, sort lpatt(solild)), ///
   by(z, compact legend(off) cols(5)) ///
   xtitle(whatever) ytitle(somethingelse) ysize(3) xsize(2)
   ```

   TODO. find out
    a) if twoway with lfit can do this, why did they do the other way with just one school?
    b) what is the effect of ysize and xsize?
    c) What are possible other settings for lpatt?
    d) Test variations with legend off, on.

3. p. 185 Fit an OLS regression within each grouping
    a) count number within each sub-group

       ```
       egen num = count(y), by(z)
       ```

    b) output a new data set ols.dta that has slopes and coefficients
    c) ```
       statsby inter=_b[_cons] slope=_b[lrt], by(z) saving(ols):
           ///
       regress y x if num>4
       ```

       TODO: find out about inter and slope arguments of statsby
       ```

d) I suggest we inspect ols.dta by itself. But Stata still has tragic “one data set at a time” policy so we have close what we are doing to inspect ols.dta. I still do that, but RHS goes other way

e) Merge the OLS estimate from the groups back onto the individual level data set

```
sort z
merge M:1 z using ols
```

That’s a many-to-one merge, matching observations on the grouping variable z.

i. NOTE WELL: that leaves behind a variable “_merge” that will be in the way if we try more merges in future. Stata will refuse until we drop “_merge”

f) p. 186 RHS create a scatter of slope on intercept

```
twoway scatter slope inter, xtitle(Intercept) ytitle(Slope
    )
```

PJ Notes: that plots individual points piled on top of one another.

Cleaner approach would open ols.dta and plot that.

g) p. 186 This method is used to find one respondent within each group.

```
egen pickone = tag(z)
```

i. Here is another relic of Stata’s stubborn insistence on one data set at a time. We have individual data, but are going to use one row per group.

ii. I suggest instead analyzing ols.dta directly

4. p. 187 Spaghetti plot: This one plots predicted values for the various groups.

```
generate pred = inter + slope*x
sort z x
twoway (line pred x, connect(ascending)), xtitle(LRT) ytitle(
    Fitted)
```

TODO:

a) Remember the connect(ascending) argument. See what happens without it

b) In p. 187, why don’t all of the lines have same left and right end points?

c) Same question over again. What if we open ols.dta and draw lines with indicated slope and intercept?

5. p. 194-6 Using xtmixed

a) Random intercept. Create label “ri” for saved fit

```
xtmixed y x || z:, mle
estimates store ri
```

b) Random slope

```
xtmixed y x || z:x, covariance(unstructured) mle
estimates store rc
```

c) Default display from xtmixed includes estimated standard deviation and correlation of random effects.

d) Want Variance and Covariance instead? Can "replay" the model (p. 196)

```
xtmixed, variance
```

or use the postestimation command

```
estat recovariance
```

TODO: Find how to make Stata tell me the full list of things that estat can do for a fitted model of a given type.

e) NOTE: Necessary to run "estimates store rc" after either of these changes.

f) What about `covariance(unstructured)`?

   i. Means that the covariance of ζ_1 and ζ_2 is "freely estimated".

g) Alternative variance structures are discussed later in book.

6. LR Test

```
lrtest rc ri
```

7. p. 201 Random effect estimates

   a) "likelihood" p. 201 "In general, we therefore donot recommend using this method and suggest using empirical Bayes prediction instead." So lets don't bother with this.

   b) Empirical Bayes

   ```
   estimates restore rc
   predict ebs ebi, reffects
   ```

   The predict method knows there are 2 columns of random effects, it expects 2 names for them. (p. 202 explains why intercept is last).

   PJ: This notation will become terrible when there are more random effects to track.

8. Visualization of estimated lines:Spaghetti plots of predicted lines

```
predict yhatre, fitted
sort z x
twoway (line yhatre, connect(ascending), xtitle(x) ytitle(y)
```

9. Regression Diagnostics. Suppose the random effect estimates are epi and ebs, saved in data frame. We can't just plot them as they are, since there are "repeats" in the data frame. Need only one individual per group. Hence use pickone.

```
histogram ebi if pickone==1, normal xtitle(whatever)
histogram ebs if pickone==1, normal xtitle(whatever)
```

10. p. 206 has a very fancy graph that puts histograms on sides of a scatterplot. The code (p. 205) uses "saving(name_me, replace)" as function argument, so that the graph is not drawn separately, but later is added in with the "graph combine" function. Study this sequence

    ```
    scatter ebs ebi if pickone==1, saving(yx, replace) xtitle(x)
       ytitle(y) ylabel(, nogrid)
    histogram ebs if pickone==1, freq horizontal saving(hy,
       replace) normal yscale(alt) ytitle(" ") fxsize(35) ylabel(,
       nogrid)
    histogram ebi if pickone==1, freq saving(hx, replace) normal
       xscale(alt) xtitle(" ") fysize(35) ylabel(,nogrid)
    graph combine hx.gph yx.gph hy.gph, hole(2) imargin(0 0 0 0)
    ```

11. p. 206. Histogram of residuals

    ```
    predict res1, residuals
    histogram res1, normal xtitle(blah)
    ```

12. p. 207-209. Caterpillar plot for group level estimates.

    a) The standard errors of the random slope and intercept can be produced with the "reses" argument to predict

       ```
       estimates restore rc
       predict slope_se inter_se, reses
       ```

       Note User duty to name the things in the right order.

    b) p. 208: Don't know what gsort + in this code does. TODO: Find out. Perhaps is typographical error?

       ```
       gsort + ebi − pickone
       generate rank = sum(pickone)
       ```

    c) twoway CATERPILLAR plot

       ```
       serrbar ebi inter_se rank if pickone == 1, addplot(scatter
            labpos rank, mlabel(school) msymbol(none) mlabpos(0))
          scale(1.96) xtitle(Rank) ytitle(pred) legend(off)
       ```

    d) p. 209. Previous requires magic serrbar function. Here's one I actually understand, I could produce on my own:.

       "An alternative method for producing a caterpillar plot is to first generate the confidence limits lower and upper and then use the rcap plot type to produce the intervals"

```
generate lower = ebi - 1.96*inter_se
generate upper = epi + 1.96*inter_se
towway(rcap lower upper rank, blpatt(solid) lcol(black))
 ///
  (scatter ebi rank) ///
  (scatter labpos rank, mlabel(school) msymbol(none)
    mlabpos(0) ///
  mlabcol(black) mlabsize(medium)), ///
  xtitle(Rank) ytitle(Pred) legend (off) ///
  xscale(range(1 65)) xlabel(1/65) ysize(1)
```

e) p. 209. WARNING Estimates of standard errors of predictions from Stata assume the benchmark value of x^0. RHS claim that the gllamm package can provide se estimates for various values of x^0 but Stata does not.