

# Multilevel Models

## 3. Random Coefficients

Germán Rodríguez

Princeton University

April 2, 2018

# Group-specific regressions

We return to our analysis of language scores by verbal IQ in 131 schools in the Netherlands.

A simple approach to the analysis of the data would fit separate regressions in each school using the model

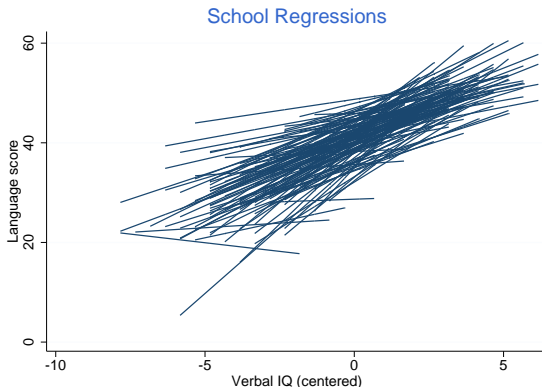
$$Y_{ij} = \alpha_i + \beta_i x_{ij} + e_{ij}$$

where  $e_{ij} \sim N(0, \sigma_e^2)$ , fitted just to school  $i$  (so in fact the error variance could vary across schools).

This is easy to do as shown in the computing logs, where we use the `statsby` command in Stata and `dplyr`'s `group_by()` in R, to run a simple linear regression in each school and gather the intercepts and slopes.

# School regression lines

We can then plot the 131 regression lines



Some of these lines are based on relatively small schools and are thus rather poorly estimated.

# Random-coefficient models

An alternative approach is to view the intercept and slope of the regression lines as random:

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)x_{ij} + e_{ij}$$

where  $\alpha$  is the average intercept and  $\beta$  the average slope,  $a_i$  is a school effect on the intercept and  $b_i$  is a school effect on the slope, and  $e_{ij}$  is the usual error term, with  $e_{ij} \sim N(0, \sigma_e^2)$ .

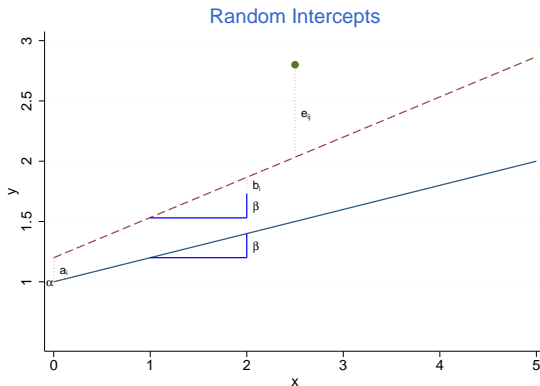
The distribution of the school effects on the intercept and slope is *bivariate* normal with mean zero and a general variance-covariance

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}\right),$$

that allows for correlation between the intercept and slope effects.

# Random Slopes

The figure below shows the model in graphical form



Interpretation of the intercept (and its variance) depends on the predictor and is more natural if the predictor is centered.

# Moments and variances

In this model the expected outcome is a linear function of  $x_{ij}$  reflecting the average or pooled regression line

$$E(Y_{ij}) = \alpha + \beta x_{ij}$$

The variance of the outcome turns out to be

$$\text{var}(Y_{ij}) = \sigma_a^2 + 2x_{ij}\sigma_{ab} + x_{ij}^2\sigma_b^2 + \sigma_e^2$$

and depends on  $x_{ij}$ , so the model is *heteroscedastic*.

The covariance between two outcomes in the same group, is

$$\text{cov}(Y_{ij}, Y_{ik}) = \sigma_a^2 + (x_{ij} + x_{ik})\sigma_{ab} + x_{ij}x_{ik}\sigma_b^2, j \neq k$$

and depends on the two values of the predictor,  $x_{ij}$  and  $x_{ik}$ .

# Intraclass correlation

As a result, the intraclass correlation is now a function of the covariates! With one predictor

$$\rho_{ijk} = \frac{\sigma_a^2 + (x_{ij} + x_{ik})\sigma_{ab} + x_{ij}x_{ik}\sigma_b^2}{\sqrt{\sigma_a^2 + 2x_{ij}\sigma_{ab} + x_{ij}^2\sigma_b^2 + \sigma_e^2} \sqrt{\sigma_a^2 + 2x_{ik}\sigma_{ab} + x_{ik}^2\sigma_b^2 + \sigma_e^2}}$$

To obtain a single-number summary we can compute the intraclass correlation for 'average' individuals

$$\bar{\rho} = \frac{\sigma_a^2 + 2\bar{x}\sigma_{ab} + \bar{x}^2\sigma_b^2}{\sigma_a^2 + 2\bar{x}\sigma_{ab} + \bar{x}^2\sigma_b^2 + \sigma_e^2}$$

where  $\bar{x}$  is the overall mean of the predictor.

If the predictor is centered we obtain a more familiar result

$$\bar{\rho} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

# Language scores: model

The computing logs show how to fit a random-slope model to the data on language scores by verbal IQ in Stata and R. In Stata the command is

```
mixed langpost iqvc || schoolnr: iqvc, ///  
      mle covariance(unstructured)
```

The two vertical bars separate the fixed and random parts of the model. We treat the constant and verbal IQ as random at the school level. It is important to request an *unstructured* covariance matrix, otherwise Stata assumes independence.

In R we call the `lmer()` function with a two-sided formula

```
lmer(langpost ~ iqvc + (iqvc | schoolnr),  
      data = snijders, REML = FALSE)
```

Here it is important to request ML as the default in R is REML.



# Language scores: estimates

The average relationship across schools is

$$E(Y_{ij}) = 40.71 + 2.53(x - \bar{x})$$

but we find substantial variation across schools in both the intercept and slope, with

$$\hat{\sigma}_a^2 = 3.056^2 \quad \text{and} \quad \hat{\sigma}_b^2 = 0.458^2$$

There's also a negative correlation between the intercept and slope (not to be confused with the intraclass correlation)

$$r_{ab} = -0.817$$

The large negative correlation is not unusual, which is why it is important to use an unstructured covariance matrix.

Finally, the error variance is estimated as

$$\hat{\sigma}_e^2 = 6.44^2$$

# Language scores: intraclass correlation

The correlation between the language scores of two students with average verbal IQ in the same school is

$$\hat{\rho}(\bar{x}) = \frac{3.06^2}{3.06^2 + 6.44^2} = 0.184$$

which also means that 18.4% of the variance in language scores at average IQ occurs between schools.

*Exercise:* calculate the intraclass correlation for different values of verbal IQ and plot it.

You should find that the correlation in language scores for children in the same school who have verbal IQs one standard deviation below the overall mean is 0.279. What happens above the mean?

# Testing Hypotheses

Tests of hypotheses proceed along the same general lines as in random-intercept models. Here are the highlights:

- We can test the significance of the average effect of verbal IQ using a Wald test. We get  $z = 31.0$ , equivalent to  $\chi^2_1 = 962.0$ .
- We can test the significance of the variance of the slope across schools using a likelihood ratio test. Removing verbal IQ from the random part saves two parameters. Equivalently,  $\sigma_a^2 = 0$  implies  $\sigma_{ab} = 0$ , so we test that both are zero. We get  $\chi^2 = 21$  using ML. The test can also be done using REML.
- Because the above test is on a boundary of the parameter space it does not have the usual  $\chi^2_2$  distribution. Stata treats it as conservative, whereas MALMUS recommend using a 50:50 mixture of  $\chi^2_1$  and  $\chi^2_2$ . Either way it is highly significant.
- Removing verbal IQ from the fixed and random parts of the model saves 3 parameters. The LR test is conservatively  $\chi^2_3$ .

# Estimating intercepts and slopes

The ML estimates of  $a_i$  and  $b_i$  can be obtained by calculating the residuals from the fixed part of the model

$$y_{ij} - (\hat{\alpha} + \hat{\beta}x_{ij})$$

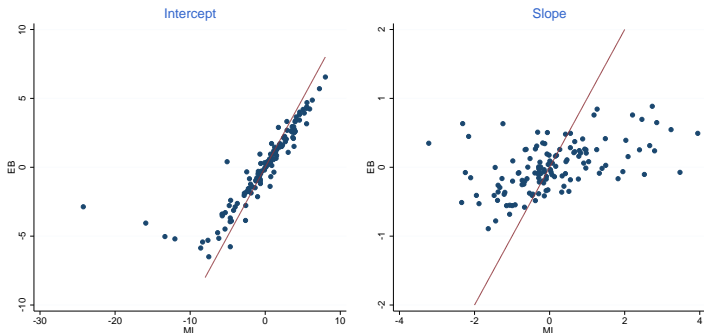
and then fitting school-specific regressions of those residuals on  $x_{ij}$  by OLS.

The EB estimates can be obtained in Stata by using `predict bi ai, reffects` after `mixed` (warning: Stata outputs slopes *before* intercepts) and in R by calling `ranef()`.

Both methods treat the fixed effects as known and substitute estimates. Typically the EB estimates show shrinkage towards zero compared to the ML estimates.

# ML and EB estimates

The following figure compares ML and EB estimates of the school effects on the intercept and slope for the language score data

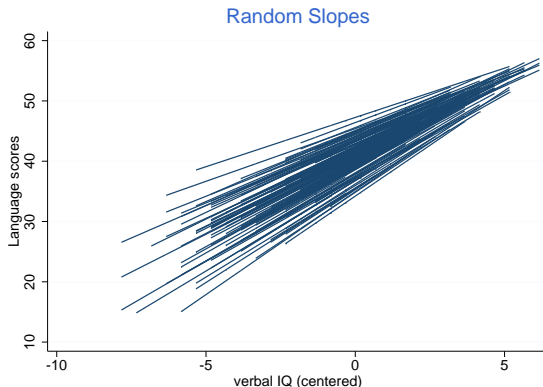


We see substantial shrinkage of the intercept effects for a few small schools, and of the slope effects all around.

Note: The online log has very similar plots comparing empirical Bayes with school-specific regressions

# Predicted regression lines

We can also plot the predicted regression lines for all schools



The figure shows that school differences in language scores are more pronounced at low verbal IQs than at the high end.

## Level-2 predictors

We now consider introducing a level-2 predictor  $z_i$ , illustrated by school SES. As usual I center the predictor on the overall mean, but will leave that implicit to simplify the notation.

We can introduce school SES as a main effect

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)x_{ij} + \gamma z_i + e_{ij}$$

We can also interact school SES with student verbal IQ

$$Y_{ij} = (\alpha + a_i) + (\beta + b_i)x_{ij} + \gamma z_i + \delta x_{ij}z_i + e_{ij}$$

The additional term is called a *cross-level* interaction.

Estimation and testing proceeds as usual, with main effects and interactions in the fixed part interpreted in the standard way.

# Level-specific models

There's another way to think about random-coefficient models. We start with a level-1 model

$$Y_{ij} = A_i + B_i x_{ij} + e_{ij}$$

where we predict language scores as a function of verbal IQ with coefficients that vary by school plus an error term.

We then add a level-2 model for the coefficients

$$\begin{aligned} A_i &= \alpha + a_i, \quad \text{and} \\ B_i &= \beta + b_i \end{aligned}$$

In this case the intercept and slope are viewed as just a constant plus a residual; in other words we have *null* models at level 2.

Substituting the second set of equations on the first yields the random-slope model in slide 4.



## Level-2 models

We now introduce school level SES as a predictor in the level-2 equations, writing

$$\begin{aligned}A_i &= \alpha_0 + \alpha_1 z_i + a_i, \quad \text{and} \\B_i &= \beta_0 + \beta_1 z_i + b_i\end{aligned}$$

where the level-2 residuals  $a_i$  and  $b_i$  have a bivariate normal distribution with mean zero and unstructured covariance matrix.

In this model both the intercept and slope are linear functions of school SES, but we could let just the intercept depend on SES.

Substituting these equations in the original model gives

$$Y_{ij} = (\alpha_0 + \alpha_1 z_i + a_i) + (\beta_0 + \beta_1 z_i + b_i)x_{ij} + e_{ij}$$

and rearranging terms leads back to the model with a cross-level interaction in slide 15.

# Level-specific and reduced forms

The package HLM uses the level-specific formulation of the model, whereas Stata and R use the reduced form and requires the user to specify the cross-level interaction terms, but the models are equivalent and the estimates are exactly the same.

MALMUS notes that users of HLM tend to include more cross-level interactions than users of Stata because they are built-in.

In the computing logs we fit a model using centered verbal IQ, centered school SES, and the interaction term. In Stata

```
mixed langpost iqvc sesc iqvcXsesc ///  
|| schoolnr: iqvc, mle covariance(unstructured)
```

The specification of the random part is exactly the same as before we introduced SES, but the fixed part now has the main effect of SES and its cross-level interaction with verbal IQ.

Here are the ML estimates of the fixed effects

Log likelihood = -7607.8383      Wald chi2(3) = 1059.58  
Prob > chi2 = 0.0000

langpost	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iqvc	2.515011	.0789096	31.87	0.000	2.360351	2.669671
sesc	.2410466	.0658997	3.66	0.000	.1118856	.3702077
iqvcXsesc	-.0470625	.0174818	-2.69	0.007	-.0813263	-.0127988
_cons	40.71326	.2910762	139.87	0.000	40.14276	41.28376

These can be written as the following estimated equation

$$E(Y_{ij}|a_i, b_i) = (40.713 + 0.241z_i + a_i) + (2.515 - 0.047z_i + b_i)x_{ij}$$

The expected score for an average student in the average school is 40.71, it is 1.07 higher if school SES is one sd more, 5.20 higher when verbal IQ is one sd more, and 5.84 higher if both conditions obtain. Clearly, verbal IQ makes more of a difference in schools with low SES.

Note: Verbal IQ has sd=2.07 and school SES has sd=4.43

Here are the parameters for the random part

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
-----+-----				
schoolnr: Unstructured				
sd(iqvc)	.3859035	.1208045	.2089373	.7127571
sd(_cons)	2.86771	.2384242	2.436495	3.375242
corr(iqvc,_cons)	-.8008473	.2294731	-.9821521	.1519006
-----+-----				
sd(Residual)	6.443356	.1005435	6.249277	6.643462
-----+-----				
LR test vs. linear regression:	chi2(3) =	215.75	Prob > chi2 = 0.0000	

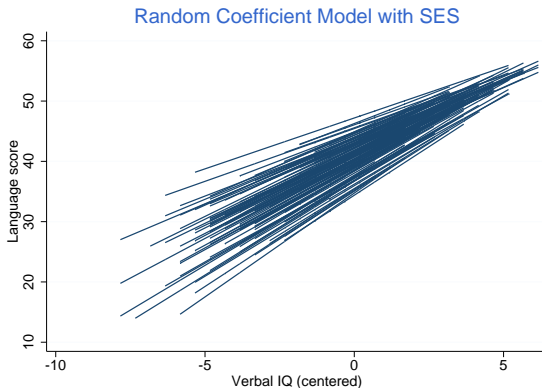
Note: LR test is conservative and provided only for reference.

The estimates are similar to the previous model, with substantial effects of unobserved school characteristics after accounting for school SES. These effects exceed those of SES.

The intraclass correlation for average students in the average school is 0.165, and now represents variation in the scores of average students across schools with the same SES.

# Predicted school lines

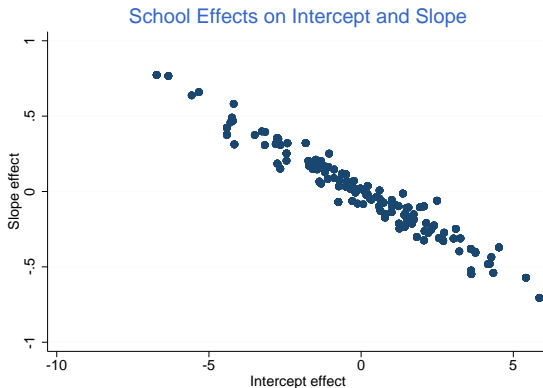
We can calculate EB estimates of the random effects as usual and plot the predicted school-level regressions



The figure looks very similar to the previous analysis, with smaller school differences at higher verbal IQs, all at observed SES.

# Empirical Bayes estimates

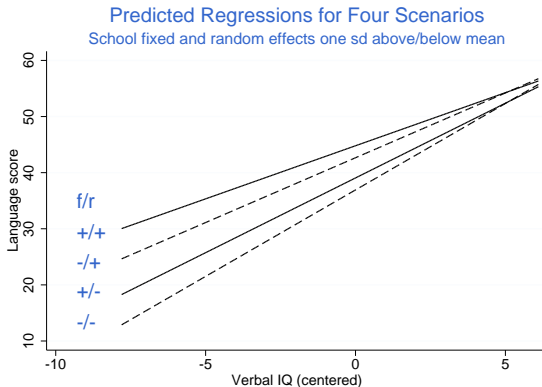
We can also plot the empirical Bayes estimates of the school effects on the intercept and slope



The prior correlation is  $-0.801$  and the posterior correlation is  $-0.971$ . Schools with higher language scores at average verbal IQ show smaller differences by verbal IQ.

# Observed and unobserved effects

To compare observed and unobserved school effects we look at four school scenarios, setting SES one sd above/below the mean and the correlated random effects one sd above/below the mean.



Clearly there are large unobserved school effects on language scores which persist at high verbal IQs.