HW3

Neal Marquez

January 30th, 2019

Homework 2

This homework uses data from the 1975 U.S. Sustaining Effects Study of elementary education, available as the egsingle dataset in the `mlmRev` R package. This gives data on 1,721 students in 60 schools. We will take `math` (Mathematics achievement score) as the outcome variable, and `schoolid` (the code for the school the student attends) as the grouping variable. For this homework, we will use the data for year 0.5 only.

Question 1

We now consider the micro-level predictor variable `retained`, and the macro-level predictor variable `lowinc`. Consider all the possible multilevel models resulting from including or not each of the `school`, `retained` and `lowinc`. This yields 8 different models. Estimate each of them and compare them. What conclusions do you draw from this analysis?

With the addition of the two covariates `retained` and `lowinc` we now have four versions of covariate inclusion to consider as well as whether or not to include a set of random effects accounting for the school level effect. For clarity below we show the functional forms for each model.

$$\text{Model 1: Intercept}$$
$$y_i = \beta_0 + \epsilon_i$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$
$$\text{Model 2: Low Income}$$
$$y_i = \beta_0 + \beta_1 \times \text{LowIncome}_{j[i]} + \epsilon_i$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$
$$\text{Model 3: Intercept}$$
$$y_i = \beta_0 + \beta_1 \times \text{Retained}_j + \epsilon_i$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$
$$\text{Model 4: Intercept}$$
$$y_i = \beta_0 + \beta_1 \times \text{LowIncome}_{j[i]} + \beta_1 \times \text{Retained}_j + \epsilon_i$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$

Models 5-8 match up with models 1:4 respectively however include the additional term for the school level effect. Below we show the structure for model 5.

$$\text{Model 5: HM Intercept}$$
$$y_i = \beta_0 + \alpha_{j[i]} + \epsilon_i$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$
$$\alpha_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2) \; j = 1, \dots, J$$

Note that prior to using the covariates in each model we adjust `lowinc` so that it is zero centered and we re-code the "dummy" variable `retained` such that a true is represented by 1 and a false is represented by
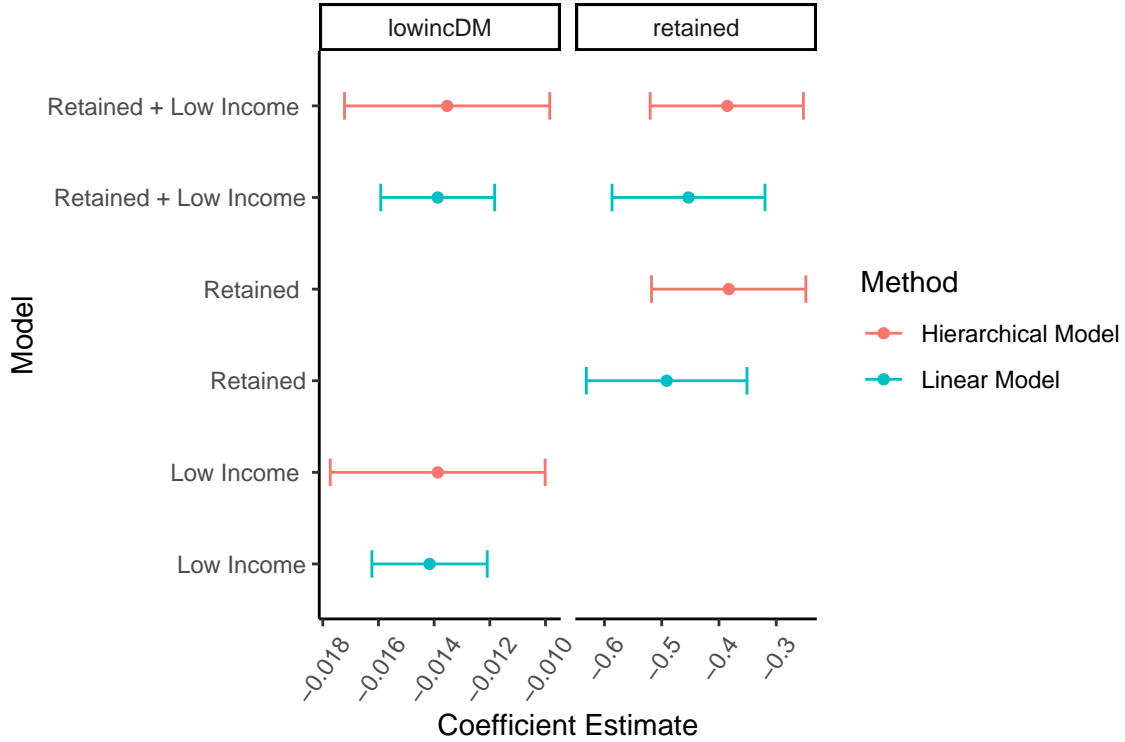
Figure 1: Coefficient Estimates Across Models

-1, as suggested in Gelman and Hill 2007. The coefficient for `retained` is then seen as half the amount of marginal change we observe when we move from `retained` being false to `retained` being true. After we fit the models we may compare the results using BIC. Table 1 shows the BIC calaculate for the 4 linear and the 4 Hierarchical models listed above. We see that in each case the hierarchical model provides a better fit to the data by at least 30 which indicates a "very strong" preference for the hierarchical model as suggested in Raftery 1995. In addition we also find that among the hierarchical models, model 8, which includes both covariates is the model most preferred although the evidence is not as strong.

Table 1: BIC for 8 Evaluated models

Model Covariates	Linear Model	Heirarchical Model
Intercept	5378.605	5186.481
Low Income	5214.945	5168.217
Retained	5339.424	5166.663
Retained + Low Income	5178.562	5147.601

Figure 1 shows the effect of how different models estimate coefficients differently. Each panel represents a different covariate being estimate and the y axis show which model was used to estimate that particular instance. notice that for certain panel-model combination we see no estimate of a coefficient because it was not included in that model. What we can see is that the effect estimate does not greatly change among the different model specifications, however, we do see increased confidence intervals for coefficient estimate in the hierarchical models. This is because the models account for the correlated nature of the observations, thus, decreasing the amount of information that a single data point gives us if it is in the same school.
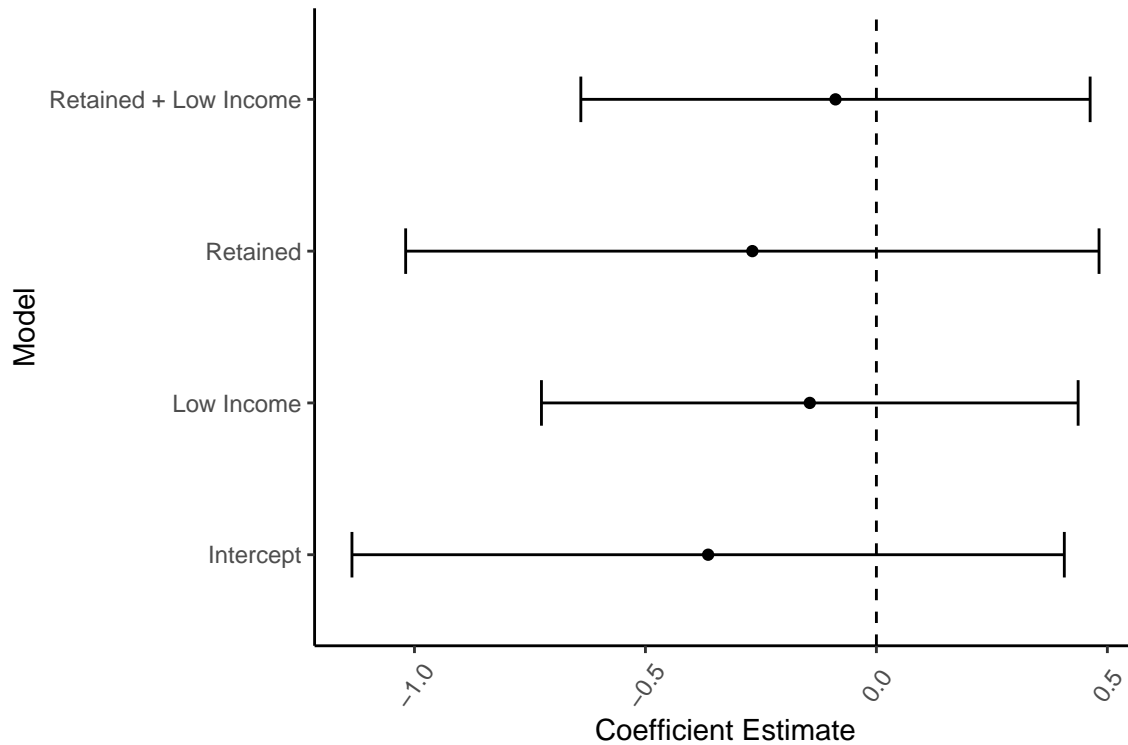
Figure 2: School ID 3221 Estaimate Across HM Models

Question 2

Consider the school for which the difference between the estimated school mean under the random effects model was most different from that under the classical model in Question 2 of Homework 2. Calculate and plot the estimated school effect and its confidence interval under each of the 8 models considered here. What do you conclude about this school from this analysis?

Figure 2 shows the estimate for school 3221 across the four hierarchical models. We do not include the four linear models because they do not contain a specific school effects as shown in the equations above. For each of the models the confidence interval overlaps with a zero value. The addition of different covariates into the model pulls the mean estimate of this school towards zero. This is not surprising as we have only three observations from this school and as we add covariates to the model more of the variance between schools can be explained by the covariates. This is especially noticeable when we include just the school level covariate `lowinc`, and can be evidenced by examining Figure 3 where we see a reduction in the estimate of σ_α with the inclusion of the school level predictor.
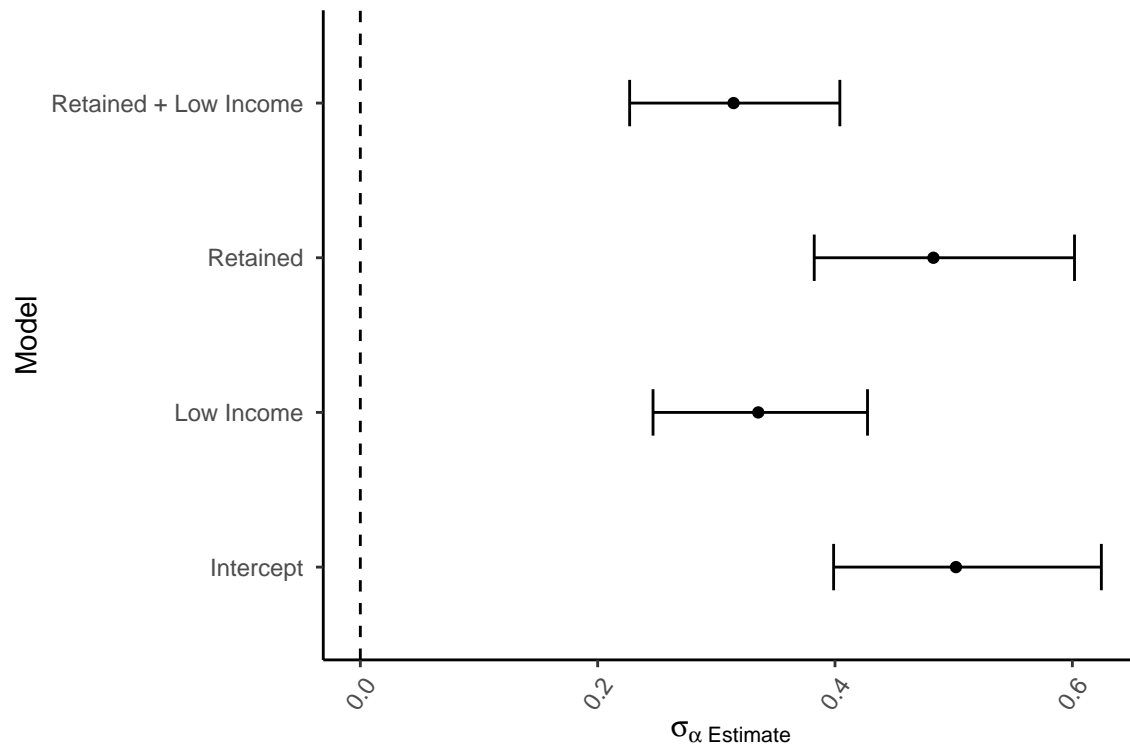
Figure 3: Estimate of Sigma Alpha with 95% Confidence Intervals

References

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research". Sociological Methodology. Vol 25, 111-163.
Gelman, Andrew and Jennifer Hill 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.