Multilevel Model for Physics

Brendi Ang

17/10/2021

Contents

Physics **2**

Exploring the dataset with basic linear model . . . 2

Getting the data ready for modelling . . . 3

Removing zero enrolments . . . 3

Linearise response variable using log transformation . . . 3

Unconditional means model . . . 4

Intraclass correlation (ICC) . . . 4

Unconditional growth model . . . 5

Testing fixed effects . . . 6

Parametric bootstrap to test random effects . . . 6

Confidence inetrval . . . 7

Interpreting final model . . . 8

Composite model . . . 8

Fixed effects . . . 9

Random effects . . . 11

Predictions . . . 13

Physics

Exploring the dataset with basic linear model

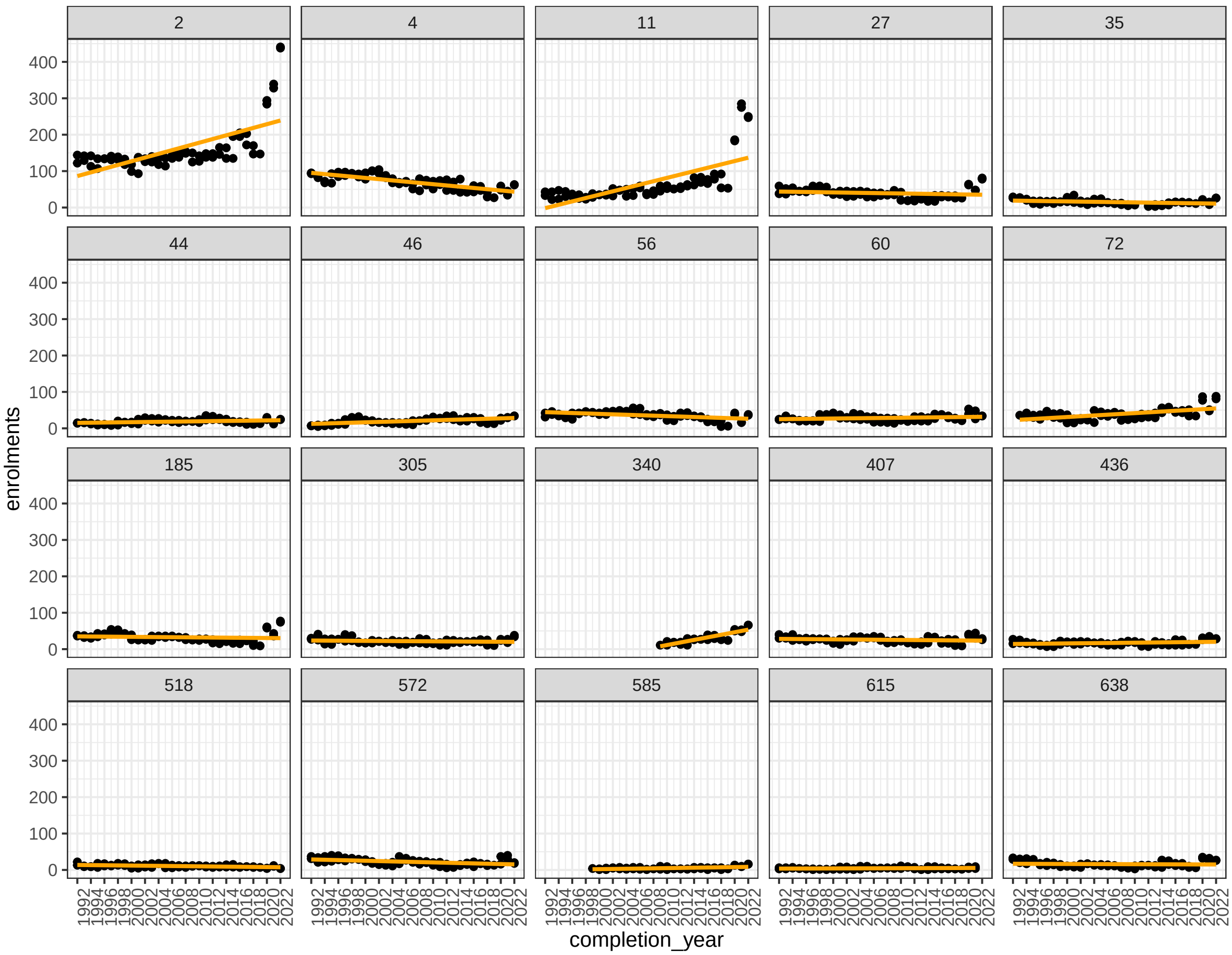

Figure 1: Basic linear model for 20 randomly selected schools to provide an at-a-glance visualisation of enrolment trends within schools for Physics

Figure 1 fits a linear model for enrolments for a random sample of 20 schools. School cohorts vary greatly in sizes; For instance, most schools in these random draws have approximately less than 50 enrolments per cohort except for school 2 and school 11. As each panel have the same y and x-axis, the fitted line are computed relative to one another. It is apparent that school 2 and school 11 have significantly larger increase in enorlments than all the other randomly selected schools.

Getting the data ready for modelling

Removing zero enrolments

All zero enrolments in a given year will be removed for modelling. As aforementioned, most of the zero enrolments in year 11 (refer to Figure **??**) were attributed to the 2007 prep year cohort while zero enrolments in year 12 relates to the first year in which a school introduces the subject. Other zero enrolments mostly relates to smaller schools with little to no enrolments in the subject for a given year. These zero enrolments will be removed for modelling purposes.

Linearise response variable using log transformation

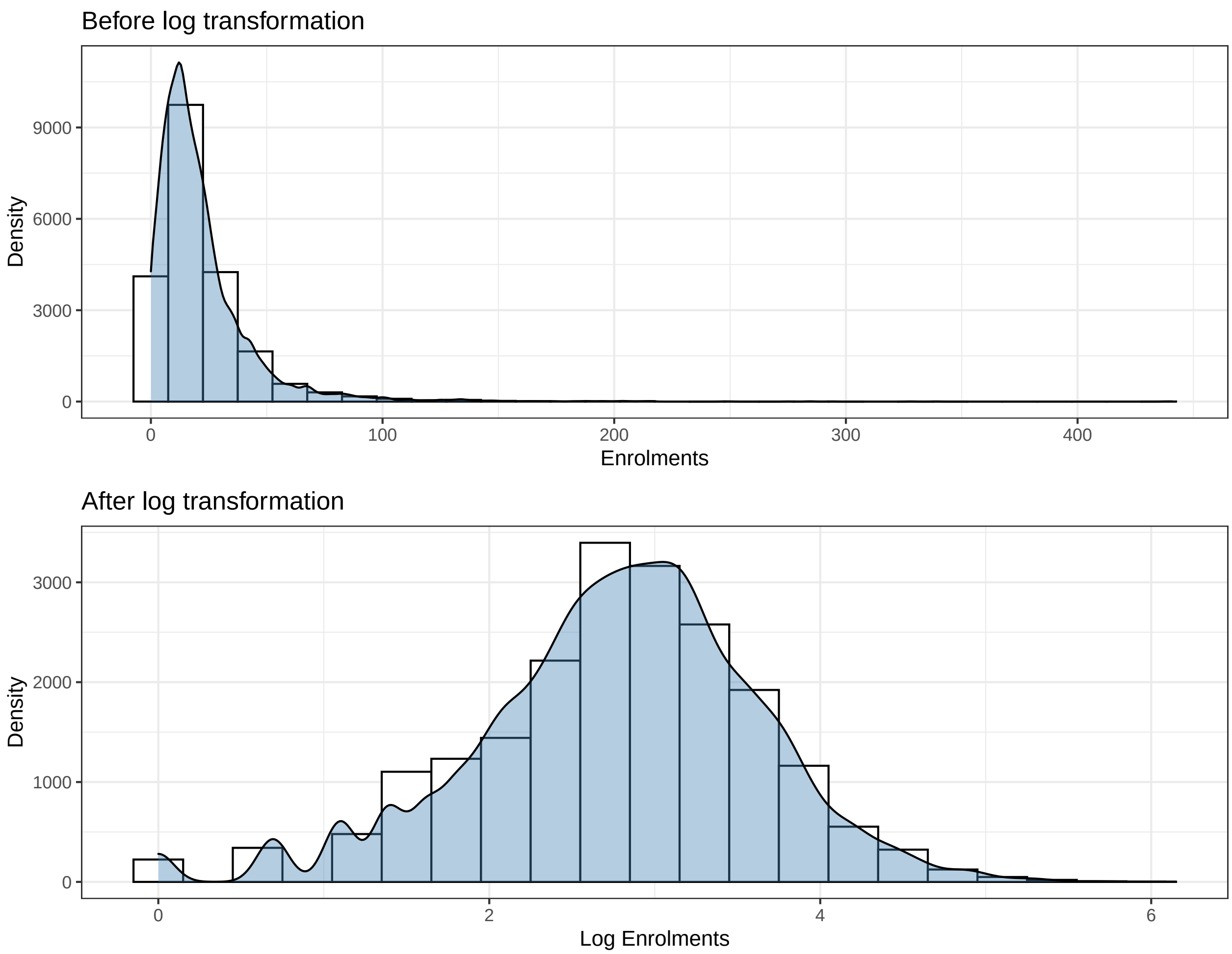

Figure 2: Effects of log transformation for response variable (enrolments) in Specialist Mathematics subject

As multilevel model assumes normality in the error terms, a log transformation is utilised to allow models to be estimated by the linear mixed models. The log transformation allows enrolment numbers to be approximately normally distributed (Figure 2.

Unconditional means model

Table 1: AIC values for all candidate models for Physics

	df	AIC
Model0.2: Schools nested within districts	4	28367.49
Model0.1: Schools nested within postcodes	4	28413.95
Model0.0: Within schools	3	28422.00

As per Step 3, the three potential models are fitted, with the AIC shown in Table **??**. Based on the AIC, `model0.2`, corresponding the schools nested within districts is the best model and will be used in the subsequent analysis.

Intraclass correlation (ICC)

```
## Random effects:

##  Groups                       Name        Variance Std.Dev.
##  qcaa_school_id:qcaa_district (Intercept) 0.52937  0.72758
##  qcaa_district                (Intercept) 0.11779  0.34321
##  Residual                                 0.21883  0.46780

##
##  Fixed effects:

##             Estimate Std. Error  t value
## (Intercept) 2.639354  0.1015033 26.00263

##
##  Number of schools (level-two group) = 445
##  Number of district (level-three group) = 13
```

This model takes into account 445 schools nested in 13 districts. In a three-level multilevel model, two intraclass correlations can be obtained using the model summary output above:

The **level-two ICC** is the correlation between a school i from a certain district j in time t and in time t^*:

$$\text{level-two ICC} = \frac{\tau_{00}^2}{\tau_{00}^2 + \phi_{00}^2 + \sigma^2} = \frac{0.5294}{(0.5294 + 0.1178 + 0.2188)} = 0.6113$$

The **level-three ICC** refers to the correlation between different schools i and i^* from a specific district j.

$$\text{level-three ICC} = \frac{\phi_{00}^2}{\tau_{00}^2 + \phi_{00}^2 + \sigma^2} = \frac{0.1178}{(0.5294 + 0.1178 + 0.2188)} = 0.1360$$

Unconditional growth model

```
summary(model)
```

```
##  Groups                       Name        Variance   Std.Dev.  Corr
##  qcaa_district:qcaa_school_id (Intercept) 1.3036e+00 1.1417664
##                               year92      1.5154e-03 0.0389283 -0.728
##  qcaa_district                (Intercept) 5.4646e-02 0.2337643
##                               year92      6.4797e-05 0.0080496 0.654
##  Residual                                 1.6862e-01 0.4106371
```

```
##               Estimate Std. Error   t value
## (Intercept) 2.34184885 0.08659973 27.042219
## year92      0.01183235 0.00300599  3.936256
```

```
##  Number of Level Two groups =  445
##  Number of Level Three groups =  13
```

- $\pi_{0ij} = 2.3418$: Initial enrolments for school i in district j (*i.e.* expected log enrolments when time = 0)
- $\pi_{1ij} = 0.01183$: Growth rate for student i in school j
- $\epsilon_{tij} = 0.1686$: Residual associated with student score at a specific time point

When the subject was first introduced in 1992, school were expected to have 10.40 ($e^{2.3418}$), on average. Enrolments are expected to increase by 1.19% (($e^{0.01183} - 1) \times 100$) per year. The estimated within-schools variance decreased by 22.9342% (0.2188 to 0.16862), implying that 22.9342% of within-school variability can be explained by the linear growth over time.

Testing fixed effects

Table 2: AIC for all possible models with different combinations of fixed effects

model	AIC
model4.7	24501.28
model4.4	24501.43
model4.1	24503.01
model4.5	24503.07
model4.0	24506.68
model4.9	24542.26
model4.2	24542.26
model4.8	24542.26
model4.10	24542.27
model4.6	24544.00
model4.3	24544.08

As highlighted in step 6, `sector` and `unit` will be added as predictors to the model. The largest possible model will be fitted, before removing fixed effects one by one while recording the AIC for each model. In this case, `model4.0` corresponds to the largest possible model while `model4.10` is the smallest possible model. The model with the optimal (lowest) AIC is `model4.7` (Table **??**). The next section will test the selected model's random effects to build the final model.

Parametric bootstrap to test random effects

Table 3: Parametric Bootstrap to compare larger and smaller, nested model

npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr_boot(>Chisq)
14	24504.51	24615.13	-12238.25	24476.51	NA	NA	NA
16	24501.28	24627.71	-12234.64	24469.28	7.2276	2	0.003

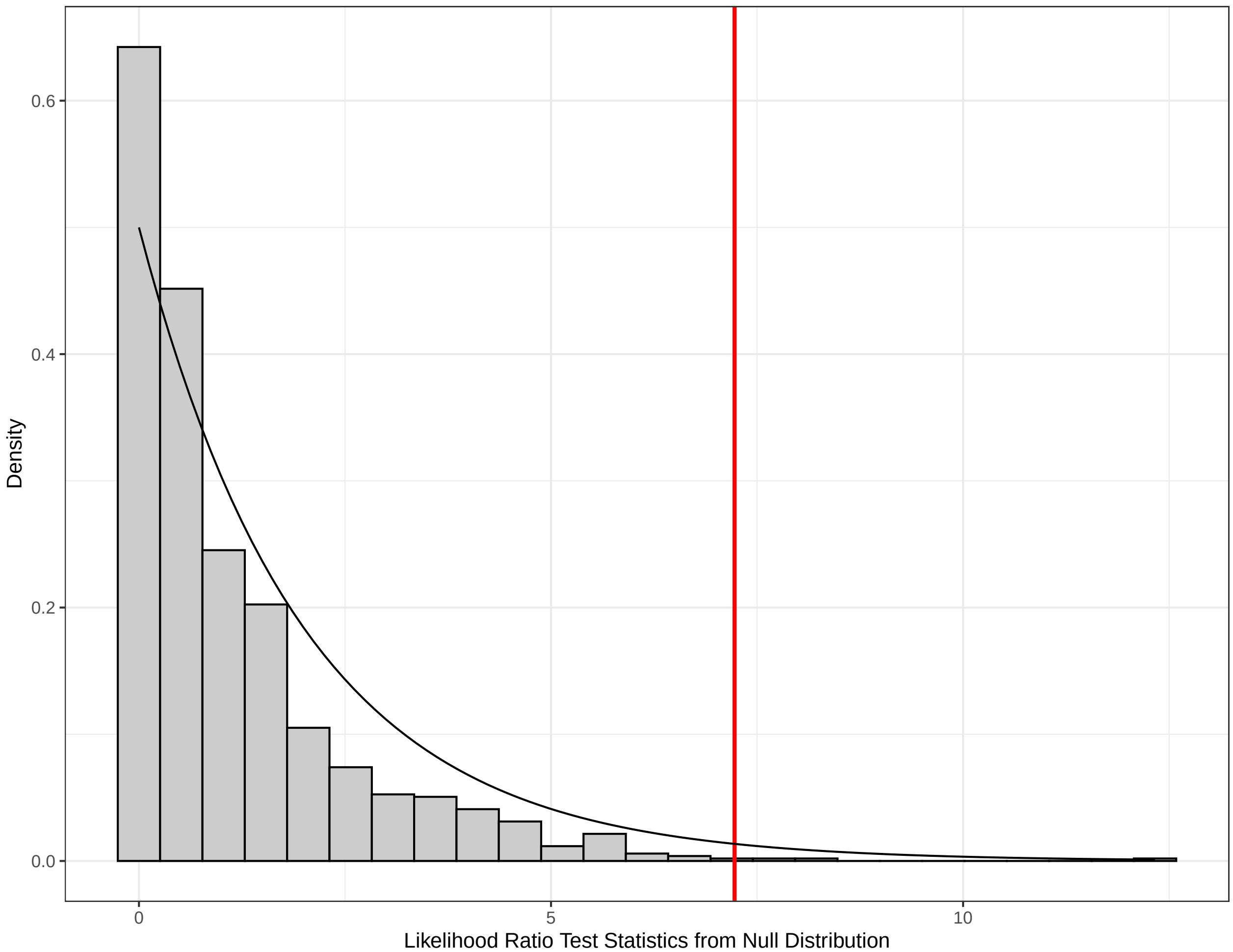

Figure 3: Histogram of likelihood ratio test statistic, with a red vertical line indicating the likelihood ratio test statistic for the actual model

The parametric bootstrap is used to approximate the likelihood ratio test statistic to produce a more accurate p-value by simulating data under the null hypothesis (detailed explanation can be found in step 7. Figure **??** displays the likelihood ratio test statistic from the null distribution, with the red line representing the likelihood ratio test statistic using the actual data. The bootstrap likelihood ratio test is significant at the 1% level ($\chi^2 = 7.2276$, p-value $= 0.006$ – see Table), suggesting that the larger model (with random slope at level three) is the better alternative.

Confidence inetrval

Table 4: 95% confidence intervals for fixed and random effects in the final model

var	2.5 %	97.5 %
sd_(Intercept)|qcaa_district:qcaa_school_id	1.0233122	1.1868086
cor_year92.(Intercept)|qcaa_district:qcaa_school_id	-0.7505522	-0.6442933
sd_year92|qcaa_district:qcaa_school_id	0.0336116	0.0390317
sd_(Intercept)|qcaa_district	0.0651241	0.4278079
cor_year92.(Intercept)|qcaa_district	-0.3047833	1.0000000
sd_year92|qcaa_district	0.0018526	0.0116635
sigma	0.4057217	0.4140799
(Intercept)	2.1566730	2.7659652
year92	0.0040382	0.0230794
sectorGovernment	-0.1569509	0.4788706
sectorIndependent	-1.0238517	-0.3247649
unityear_12_enrolments	-0.0564014	-0.0051342
year92:sectorGovernment	-0.0231053	-0.0018089
year92:sectorIndependent	0.0068328	0.0307081
sectorGovernment:unityear_12_enrolments	-0.0617423	-0.0005128
sectorIndependent:unityear_12_enrolments	-0.0555152	0.0122242

The parametric bootstrap is utilised to construct confidence intervals (as detailed in step 8). If the confidence intervals for the random effects does not include 0, it provides statistical evidence that the p-value is less than 0.5. In other words, it suggests that the random effects and the correlation between the random effects are significant at the 5% level. The 95% confidence interval is shown above (Table 4), and the random effects all exclude 0, further reiterating that they are statistically significant at the 5% level.

Interpreting final model

Composite model

- Level one (measurement variable)

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} year92_{tij} + \epsilon_{tij}$$

- Level two (schools within districts) will contain new predictor(`sector`)

$$\pi_{0ij} = \beta_{00j} + \beta_{01j} sector_{ij} + \beta_{02j} unit_{ij} + \beta_{03j} sector_{ij} unit_{ij} + u_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} sector_{ij} + u_{1ij}$$

- Level three (districts)

$$\beta_{00j} = \gamma_{000} + r_{00j}$$
$$\beta_{01j} = \gamma_{010} + r_{01j}$$
$$\beta_{02j} = \gamma_{020} + r_{02j}$$
$$\beta_{03j} = \gamma_{030} + r_{03j}$$
$$\beta_{10j} = \gamma_{100} + r_{10j}$$
$$\beta_{11j} = \gamma_{110} + r_{11j}$$

Therefore, the composite model can be written as

$$\begin{aligned}
Y_{tij} &= \pi_{0ij} + \pi_{1ij} year92_{tij} + \epsilon_{tij} \\
&= (\beta_{00j} + \beta_{01j} sector_{ij} + \beta_{02j} unit_{ij} + \beta_{03j} sector_{ij} unit_{ij} + u_{0ij}) + (\beta_{10j} + \beta_{11j} sector_{ij} + u_{1ij}) year92_{tij} + \epsilon_{tij} \\
&= \; [(\gamma_{000} + r_{00j}) + (\gamma_{010} + r_{01j}) sector_{ij} + (\gamma_{020} + r_{02j}) unit_{ij} + (\gamma_{030} + r_{03j}) sector_{ij} unit_{ij} + u_{0ij}] + \\
&\quad\; [(\gamma_{100} + r_{10j}) + (\gamma_{110} + r_{11j}) sector_{ij} + u_{1ij}] \, year92_{tij} + \epsilon_{tij} \\
&= \; [\gamma_{000} + \gamma_{010} sector_{ij} + \gamma_{020} unit_{ij} + \gamma_{020} sector_{ij} unit_{ij} + \gamma_{100} year92_{tij} + \gamma_{110} year92_{tij} sector_{ij}] + \\
&\quad\; [r_{00j} + r_{01j} sector_{ij} + r_{02j} unit_{ij} + r_{03j} sector_{ij} unit_{ij} + u_{0ij} + r_{10j} year92_{tij} + r_{11j} year92_{tij} sector_{ij} + u_{1ij} \epsilon_{tij}]
\end{aligned}$$

Fixed effects

```
##  Groups                        Name        Variance   Std.Dev. Corr
##  qcaa_district:qcaa_school_id (Intercept) 1.2224e+00 1.105612
##                                year92      1.3296e-03 0.036463 -0.702
##  qcaa_district                 (Intercept) 7.1922e-02 0.268183
##                                year92      4.5643e-05 0.006756 0.620
##  Residual                                  1.6788e-01 0.409733
```

```
##                                            Estimate  Std. Error   t value
## (Intercept)                              2.46349218 0.149836105 16.441245
## year92                                   0.01310726 0.004790919  2.735856
## sectorGovernment                         0.15333884 0.150567068  1.018409
## sectorIndependent                       -0.66726007 0.168419418 -3.961895
## unityear_12_enrolments                  -0.02817815 0.013299922 -2.118670
## year92:sectorGovernment                 -0.01202594 0.005109688 -2.353557
## year92:sectorIndependent                 0.01877807 0.005711365  3.287842
## sectorGovernment:unityear_12_enrolments  -0.03146220 0.015438491 -2.037906
## sectorIndependent:unityear_12_enrolments -0.02406566 0.017479537 -1.376791
```

```
##  Number of Level Two groups =  445
##  Number of Level Three groups =  13
```

Using the model output above (see step 9 for detailed explanation on fixed effects), the estimated increase in mean enrolments for government schools is 0.1082% ($(e^{0.0131073-0.0120259}-1)\times 100$), which is 1.1954% ($(e^{0.01626900}-1)\times 100$) less than that of catholic schools. On the other hand, the mean enrolments for independent schools are estimated to increase by 3.2399% ($(e^{0.0131073+0.0187781}-1)\times 100$) each year, which is 1.8956% ($(e^{0.0187781}-1)\times 100$) more than catholic schools.

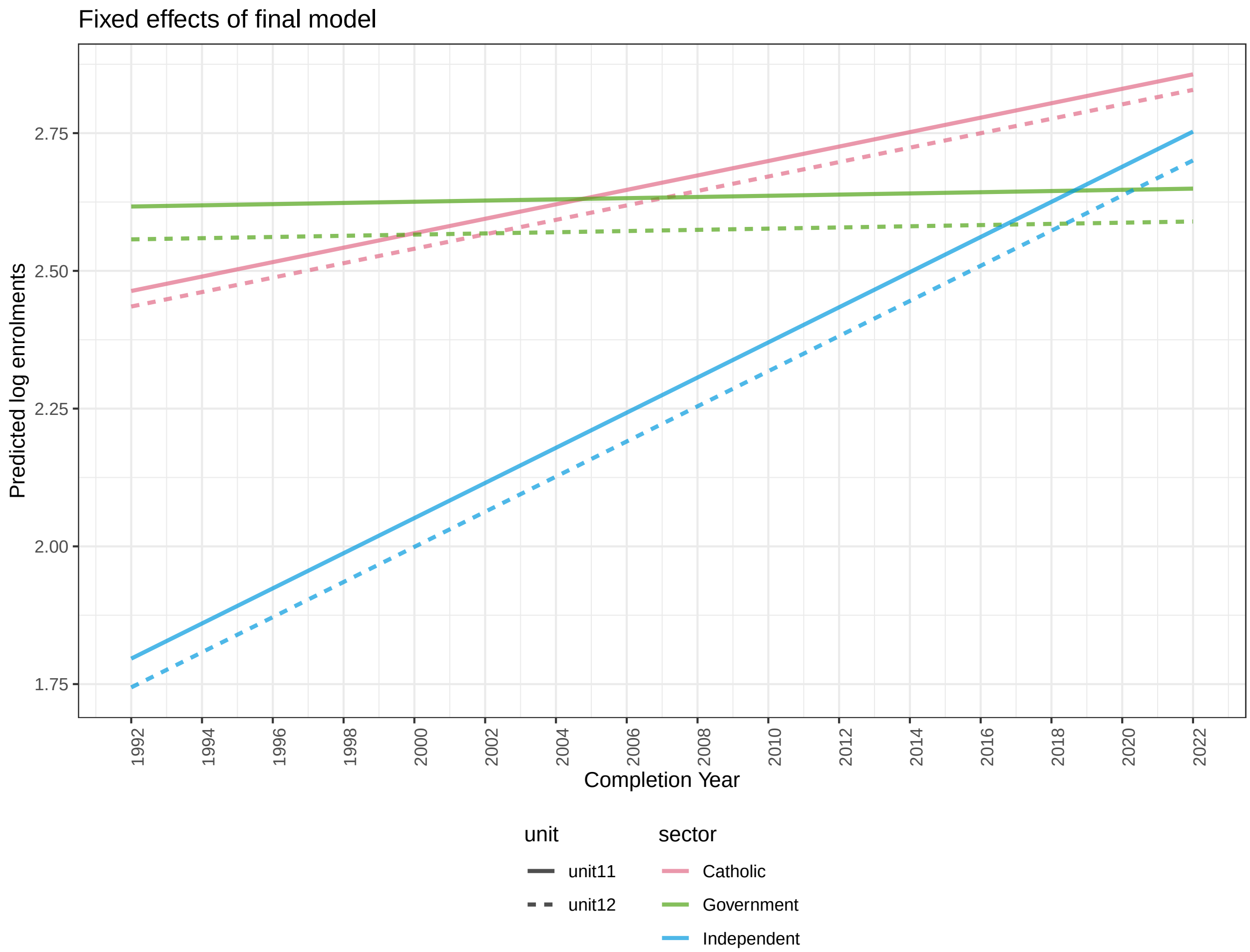

Figure 4: Fixed effects of the final model for Physics

The fixed effects can be visualised in Figure 4. It is manifest that independent schools shows the largest increase in enrolments over the years, while government schools shows a slight decrease in enrolments across the years, on average. Each sector demonstrated that the average year 11 enrolments was less than year 12 enrolments, why may imply that students are not continuing their year 12 syllabus after completing year 11.

Random effects

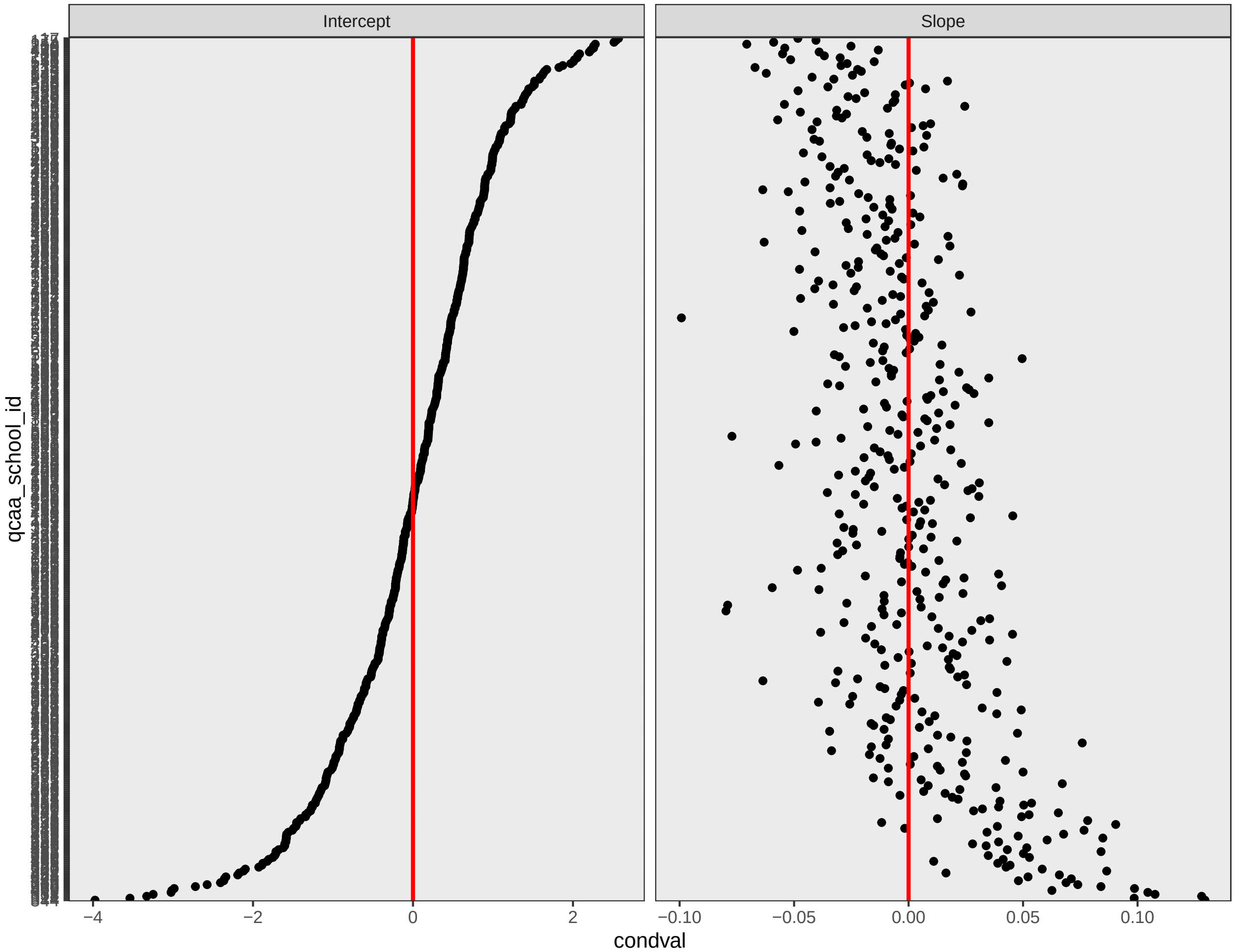

Figure 5: Random effects for schools

Figure 5 displays the random effects for a given school. It is apparent that the random intercepts and slopes are negatively correlated, where a large intercept is associated with a smaller random slope (correlation = -0.70, as shown in the model output). This indicates that a larger school is associated with a smaller increase (decrease) in enrolments over the years while smaller schools are predicted to have larger increase in enrolments over the years.

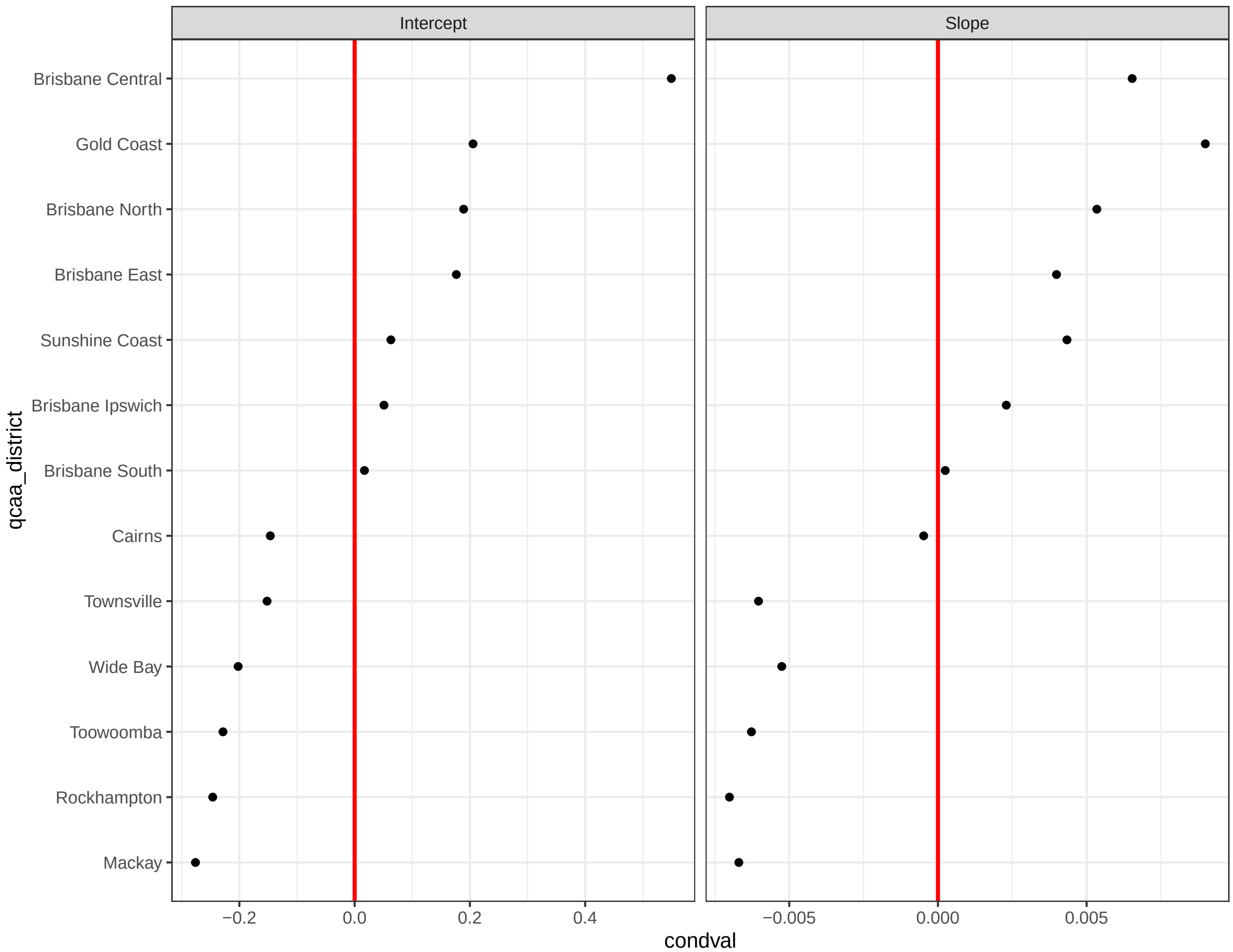

Figure 6: Random intercept for districts

The model includes a random slope at level three, and the effects of the random intercept and slope in level three is shown in Figure 6. As shown in the model output, correlation between the random intercept and slope is 0.62. Loosely speaking, this suggests that districts with large enrolments are going to be larger, while smaller districts are going to increase (decrease) at a slower rate. Based on Figure 6, Gold Coast are estimated to have the highest slope, indicating that the rate of change in enrolment is the greatest compared to the other districts.

Predictions

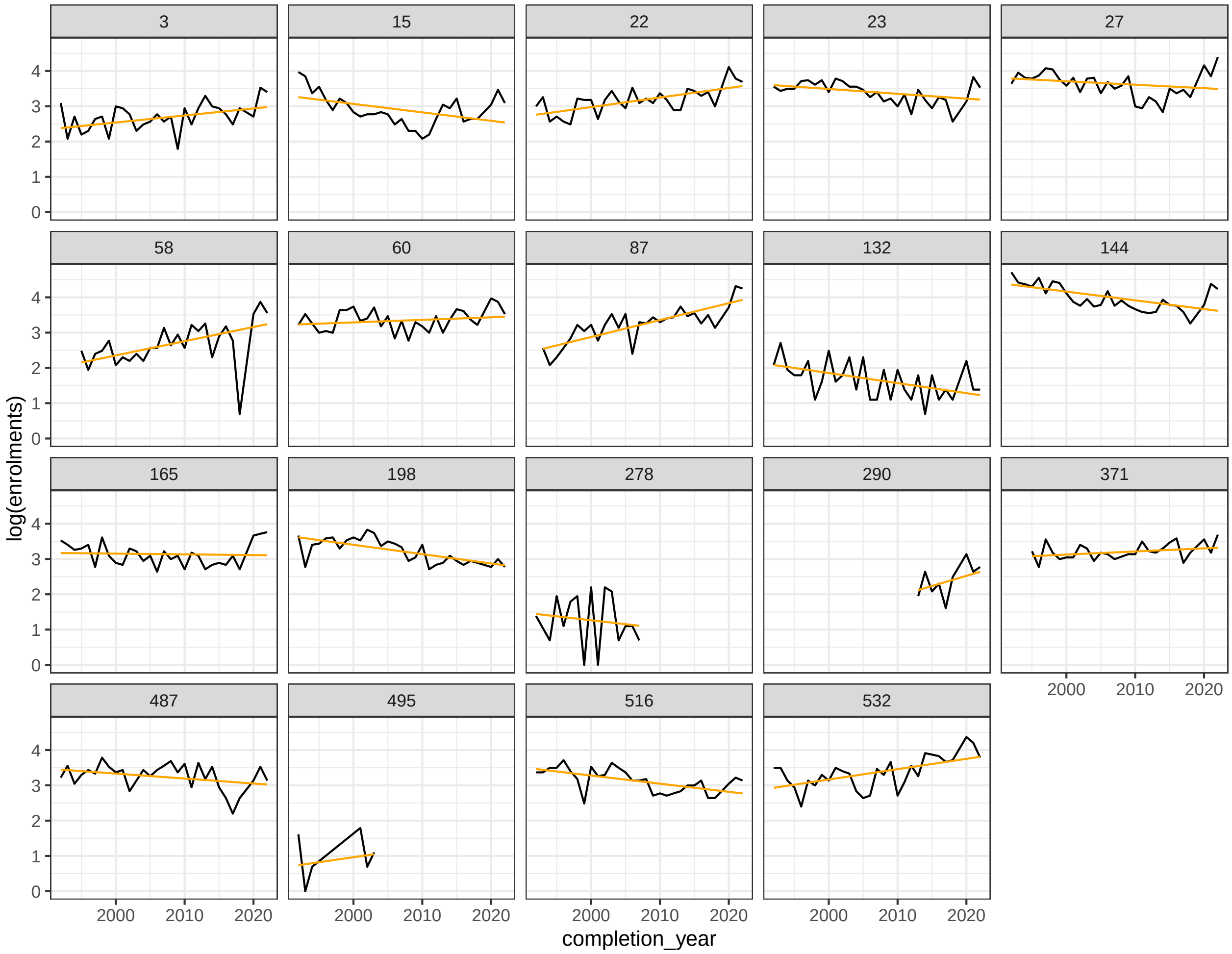

Figure 7: Model predictions for 20 randomly selected schools

Figure 7 above shows the predictions for 20 randomly selected schools.