Multilevel Model for Agricultural Practices

Brendi Ang

17/10/2021

Contents

Agricultural Practices

Exploring the dataset with basic linear model for each school

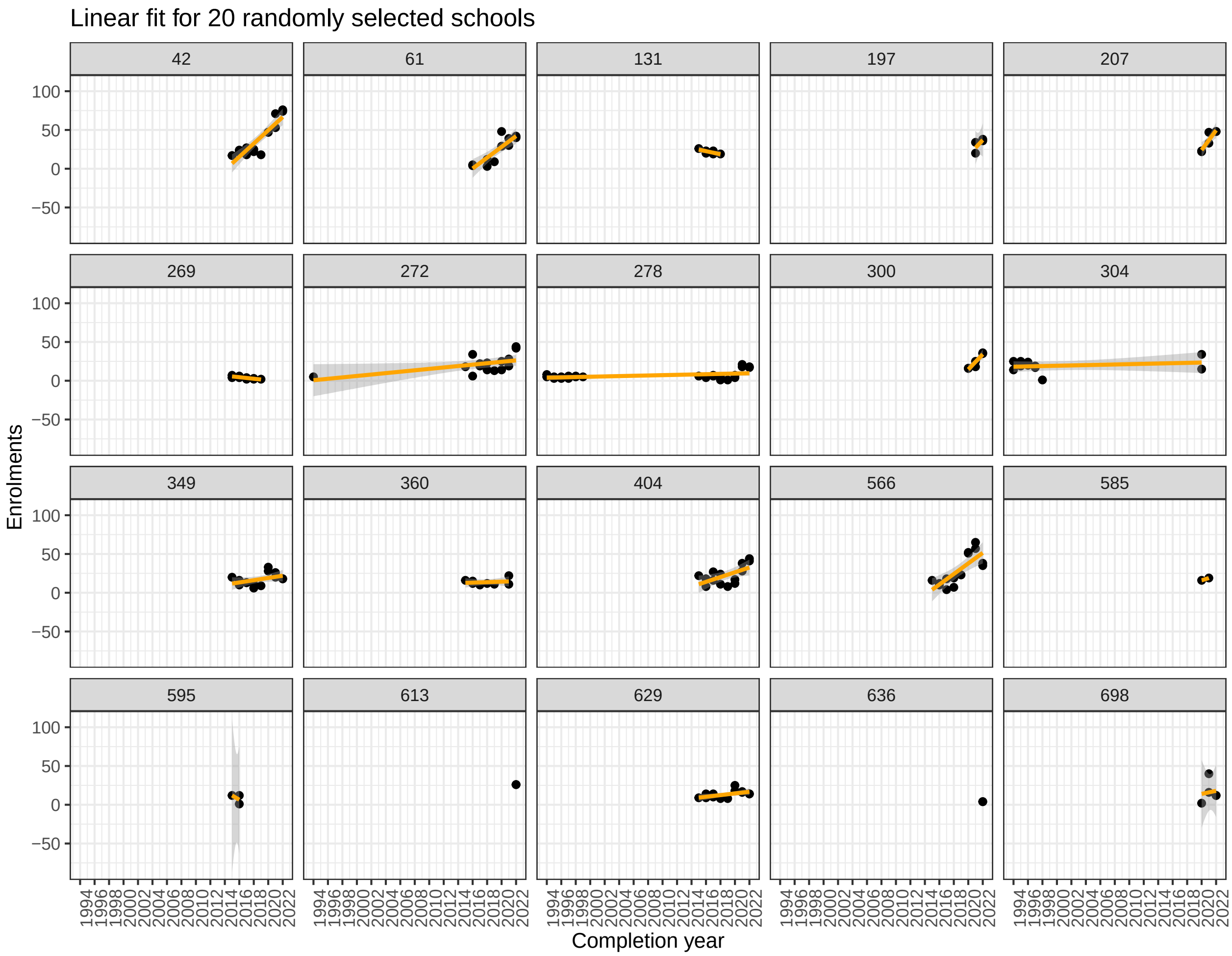

Figure 1: Basic linear model for 20 randomly selected schools to provide an at-a-glance visualisation of enrolment trends within schools for Agricultural Practices

As described in step 1, a basic linear model was plotted for each school to provide insights of the enrolment trends for each school. Figure 1 demonstrates a distinct feature, in which there was a halt in the subject from 2000 to 2014. Schools 272 and 278 and 304 appers to re-introduce subject again after the halt, while some schools introduced the subject in 2015. Some enrolment patterns are rather erratic, such as having only one cohort taking the subject (school 613 and 636) or having a start increase in enrolment in one particular year (school 698).

Getting the data ready for modelling

Removing enrolments before year 2000 and graduating cohort 2019

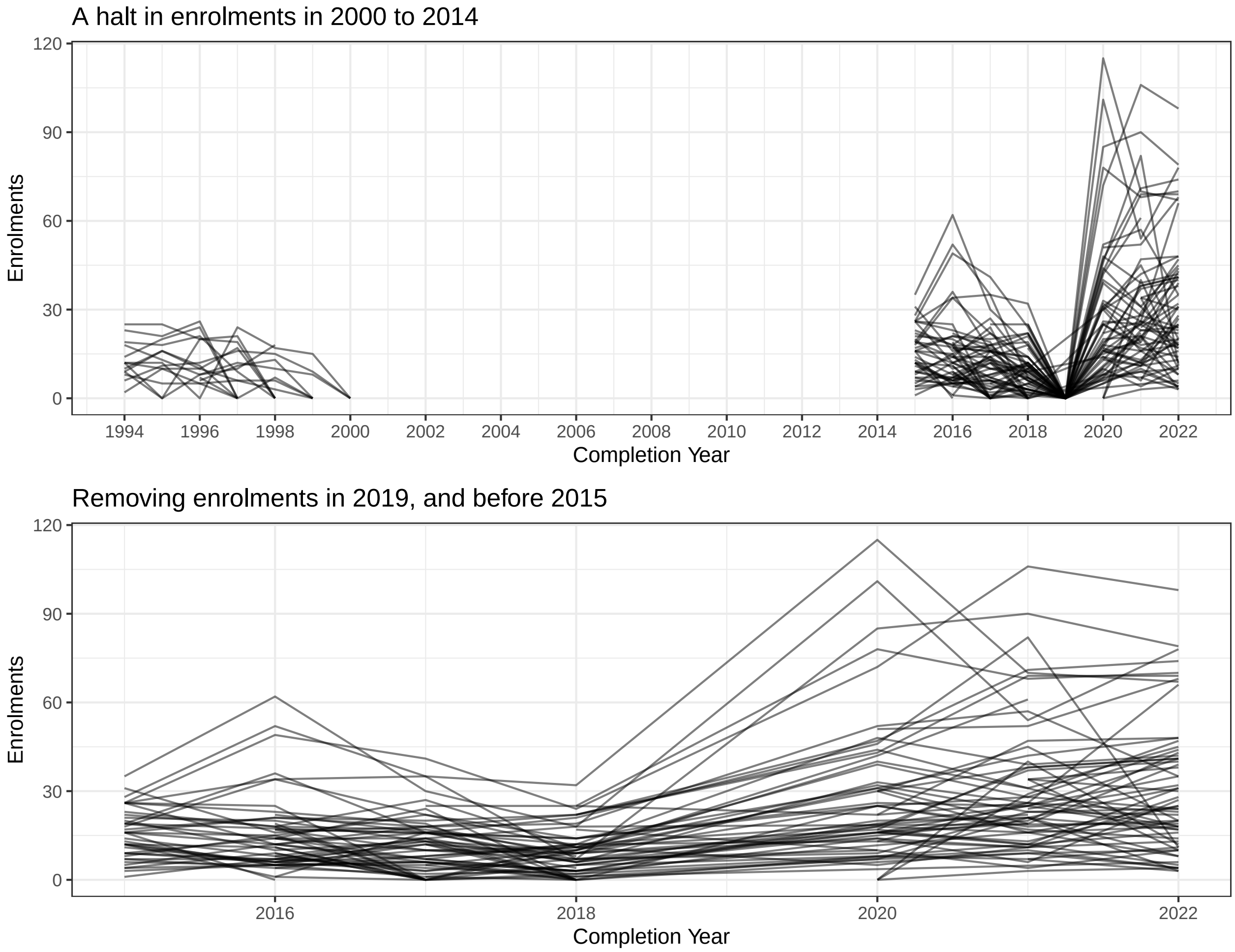

Figure 2: Time plot of enrolments, demonstrating a halt in the subject from 2000 to 2014

A halt in Agricultural practices subject in 2000, before being re-introduced again in 2015 (Figure 2. Enrolments before year 2000 were rather small, as compared to enrolments from 2015 onwards. Therefore, enrolments before year 2000 will be removed.

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. For these reasons, all completion years with zero enrolments will also be removed for modelling.

Linearise response variable using log transformation

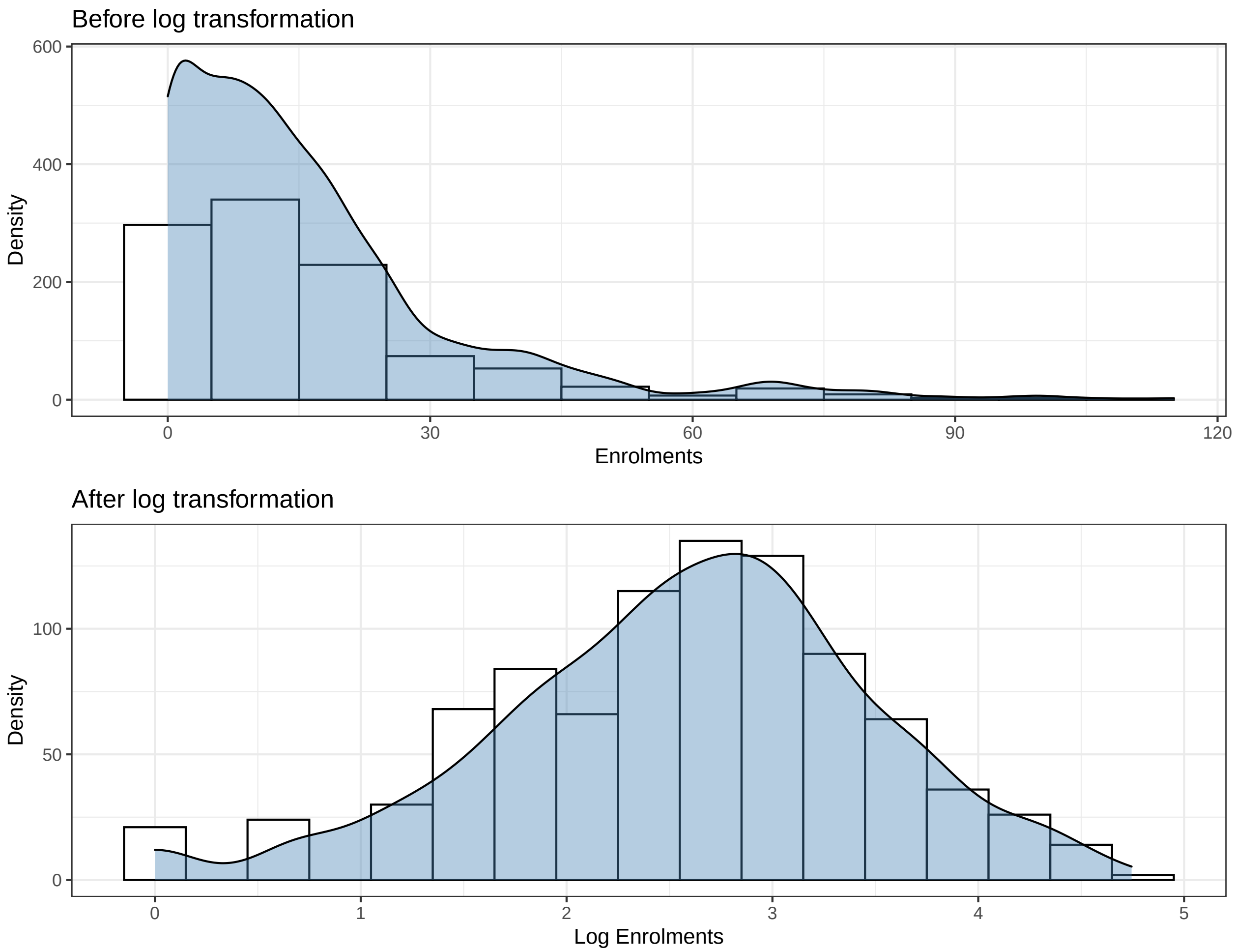

Figure 3: Effects of log transformation for response variable (enrolments) in Agricultural Practices subject

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 3).

Unconditional means model

Table 1: AIC values for all candidate models for Agricultural Practices

	df	AIC
Model0.0: Within schools	3	1573.659
Model0.2: Schools nested within districts	4	1574.922
Model0.1: Schools nested within postcodes	4	1575.659

As underlined in step 3, the three candidate models are fitted and their AIC is shown in Table 1. Based on the AIC, the two-level model (`model0.0`) is the best model and will be used in the subsequent analysis.

Intraclass correlation (ICC)

```
summary(model0.0)
```

```
## Random effects:

##  Groups        Name        Variance Std.Dev.
##  qcaa_school_id (Intercept) 0.48659  0.69756
##  Residual                  0.39251  0.62651

##
##  Fixed effects:

##             Estimate Std. Error  t value
## (Intercept) 2.647161 0.07777087 34.03796

##
##  Number of schools (level-two group) = 93
##  Number of district (level-three group) = NA
```

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

$$\text{Level-two ICC} = \frac{\tau_{00}^2}{\tau_{00}^2 + \phi_{00}^2 + \sigma^2} = \frac{0.4866}{(0.4866 + 0.3925)} = 0.5535$$

This can be conceptualised as the correlation between the enrolments of a selected school at two randomly drawn year (*i.e.* two randomly selected cohort from the same school). In other words, 55.35% of the total variability is attributable to the differences in enrolments within schools at different time periods.

Unconditional growth model

```
summary(model1.0)
```

```
##  Groups        Name        Variance Std.Dev. Corr
##  qcaa_school_id (Intercept) 0.550453 0.741925
##                year15      0.006227 0.078912 -0.402
##  Residual                  0.257708 0.507650
```

```
##              Estimate Std. Error  t value
## (Intercept) 2.0127475 0.09392102 21.43021
## year15      0.1514085 0.01398980 10.82278

##  Number of Level Two groups =  93
##  Number of Level Three groups =  NA
```

The next step involves incorporating the linear growth of time into the model. The model output is shown above.

- $\pi_{0ij} = 2.0128$: Initial status for school i in district j (*i.e.* expected log enrolments when time = 0)
- $\pi_{1ij} = 0.1514$: Growth rate for school i in district j
- $\epsilon_{tij} = 0.2577$: Variance in within-school residuals after accounting for linear growth overtime

When the subject was re-introduced into the QCE system in 2015, schools were expected to have 7.4842 ($e^{2.0128}$) enrolment. On average, enrolments were expected to increase by a staggering 16.35% ($(e^{0.151408} - 1) \times 100$) per year. The estimated within-schools variance decreased by 34.34% (0.3925 to 0.2577), implying that 34.34% 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	npar	AIC	BIC	logLik
model4.5	12	1327.094	1381.961	-651.5468
model4.4	11	1328.231	1378.526	-653.1157
model4.3	10	1329.521	1375.244	-654.7606
model4.1	14	1330.494	1394.506	-651.2470
model4.7	13	1331.572	1391.012	-652.7862
model4.6	12	1333.059	1387.926	-654.5293
model4.0	16	1333.719	1406.875	-650.8593
model4.9	11	1334.117	1384.412	-656.0585
model4.2	11	1334.117	1384.412	-656.0585
model4.8	11	1334.117	1384.412	-656.0585
model4.10	11	1334.117	1384.412	-656.0585

As outlined in step 6, `sector` and `unit` will be added as predictors to the model. The largest possible model (`model4.0`) will then be fitted, before removing the fixed effects one at a time (with `model4.10` being the smallest of all 10 candidate models), while recording the AIC for each model. `model4.5` appears to have the optimal (smallest) AIC (Table 2), and will be used in the next section 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)
10	1330.958	1376.681	-655.4792	1310.958	NA	NA	NA
12	1327.094	1381.961	-651.5468	1303.094	7.864871	2	0.009

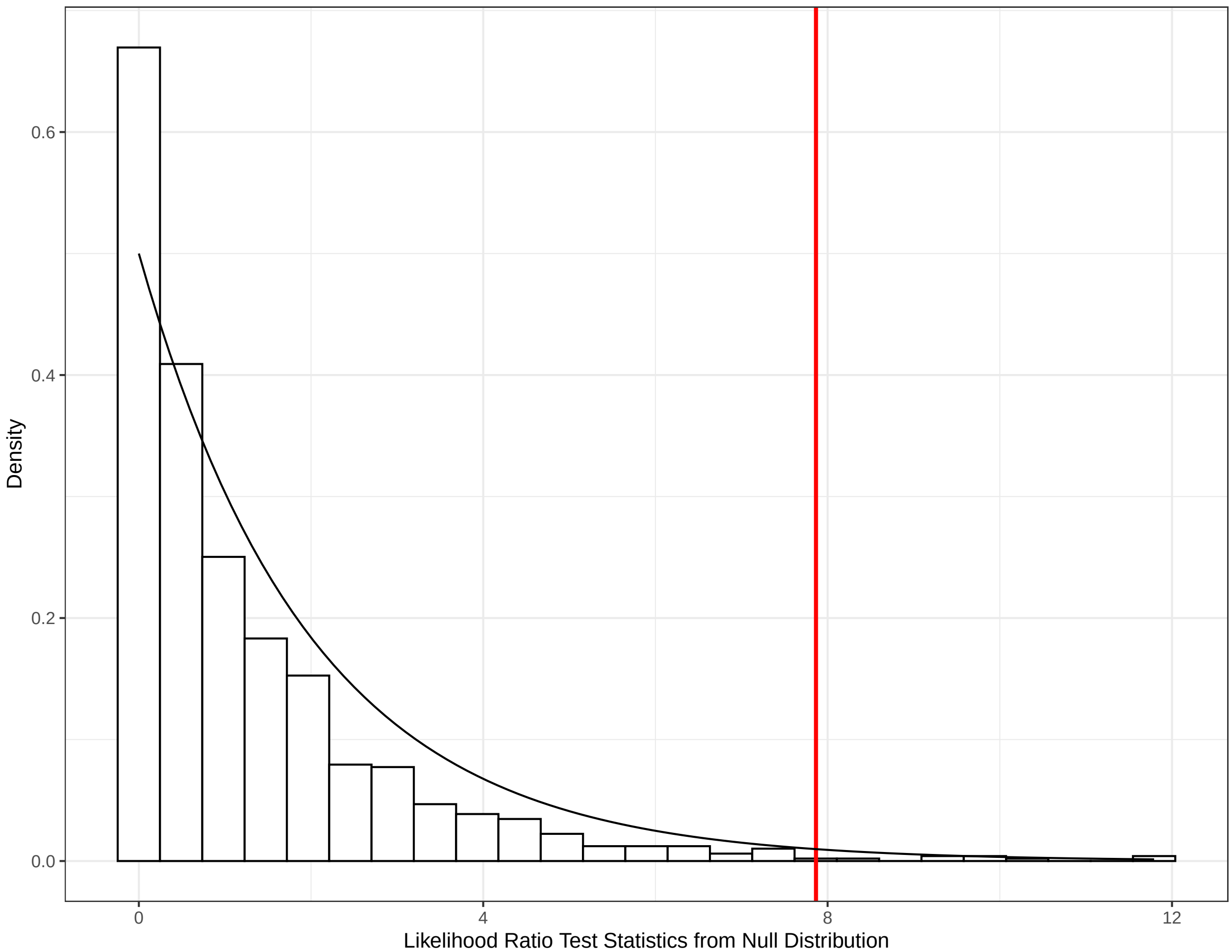

Figure 4: Histogram of likelihood ratio test statistic, with a red vertical line indicating the likelihood ratio test statistic for the actual model

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 4 displays the likelihood ratio test statistic from the null distribution, with the red line indicates the likelihood ratio test statistic using the actual data.

The p-value indicates the proportion of times in which the bootstrap test statistic is greater than the observed test statistic. The low estimated p-value is 0.009 < 0.05 (Table 3) rejects the null hypothesis at the 5% level, indicating that the larger model (including random slope at level two) is preferred.

Confidence interval

Table 4: 95% confidence interval for the fixed and random effects

var	2.5 %	97.5 %
sd_(Intercept)|qcaa_school_id	0.5410851	0.8275173
cor_year15.(Intercept)|qcaa_school_id	-0.6286851	0.1177694
sd_year15|qcaa_school_id	0.0360913	0.0971561
sigma	0.4719375	0.5290075
(Intercept)	1.5118458	2.6445580
year15	-0.0251663	0.1676654
sectorGovernment	-0.3715192	0.7707218
sectorIndependent	-1.3320193	0.2141525
unityear_12_enrolments	-0.4283373	-0.1388208
year15:sectorGovernment	-0.0312332	0.1723089
year15:sectorIndependent	0.0272586	0.2881371
year15:unityear_12_enrolments	-0.0002587	0.0607488

The parametric bootstrap is utilised to construct confidence intervals (detailed explanation in step 8) for the random effects. If the confidence intervals between 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 confidence interval for the random effects all exclude 0 (Table 4), indicating that they're different from 0 in the population (*i.e.* statistically significant).

Interpreting the final model

Composite model

- Level one (measurement variable)

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} year15_{tij} + \epsilon_{tij}$$

- Level two (schools within postcodes)

$$\pi_{0ij} = \beta_{00j} + \beta_{01} sector_{ij} + \beta_{02} unit_{ij} + u_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} sector_{ij} + \beta_{12j} unit_{ij} + u_{1ij}$$

Therefore, the composite model can be written as

$$
\begin{aligned}
Y_{tij} &= \pi_{0ij} + \pi_{1ij} year15_{tij} + \epsilon_{tij} \\
&= (\beta_{00j} + \beta_{01} sector_{ij} + \beta_{02} unit_{ij} + u_{0ij}) + (\beta_{10j} + \beta_{11j} sector_{ij} + \beta_{12j} unit_{ij} + u_{1ij}) year15_{tij} + \epsilon_{tij} \\
&= \left[\beta_{00j} + \beta_{01} sector_{ij} + \beta_{02} unit_{ij} + \beta_{10j} year15_{tij} + \beta_{11j} sector_{ij} year15_{tij} + \beta_{12j} unit_{ij} year15_{tij}\right] \left[u_{0ij} + u_{1ij} + \epsilon_{tij}\right]
\end{aligned}
$$

Fixed effects

```
summary(model_f)
```

```
##  Groups         Name        Variance  Std.Dev. Corr
##  qcaa_school_id (Intercept) 0.4842622 0.695890
##                 year15      0.0048578 0.069698 -0.337
##  Residual                   0.2509081 0.500907
```

```
##                                 Estimate Std. Error   t value
## (Intercept)                   2.03638008 0.29670877  6.863228
## year15                        0.06877268 0.04919228  1.398038
## sectorGovernment              0.22111834 0.31239875  0.707808
## sectorIndependent            -0.53317803 0.37319648 -1.428679
## unityear_12_enrolments       -0.27868693 0.07555397 -3.688581
## year15:sectorGovernment       0.06990061 0.05093108  1.372455
## year15:sectorIndependent      0.14926780 0.06131800  2.434323
## year15:unityear_12_enrolments  0.02866851 0.01629460  1.759388
```

```
##  Number of Level Two groups =  93
##  Number of Level Three groups =  NA
```

Based on the model output, the estimated mean enrolments for government schools are estimated to be 24.75% ($e^{0.2211} - 1 \times 100$) more than that of catholic schools when the subject was re-introduced in 2015. Government schools are also estimated to have an estimated increase in mean enrolments of 14.87% ($(e^{0.0687727+0.0699006} - 1) \times 100$) per year, which is 7.24% more than that of catholic schools.

On the other hand, independent schools are estimated to have mean enrolments of 41.32% ($(e^{-0.5331780} - 1) \times 100$) less than that of catholic schools in 2015. This low initial status is matched with a relatively large increase of 24.36% in mean enrolments per year, which is a staggering 16.10% more than that of catholic schools, on average.

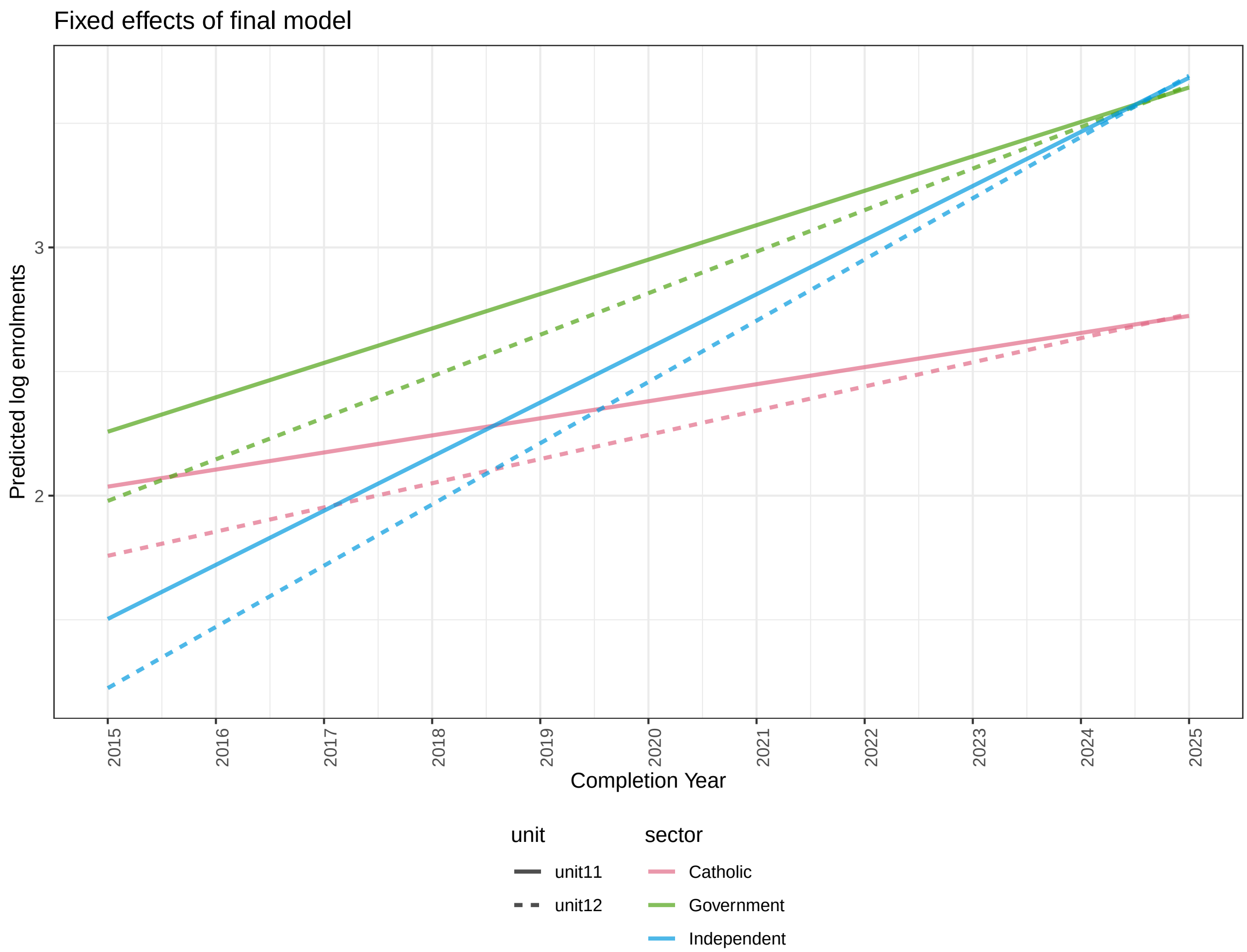

Figure 5: Fixed effects of the final model for Agricultural Practices subject

Based on the fixed effects (see step 9 for detailed explanation), independent schools show the greatest rate of change (as indicated by the slope in Figure 5), indicating that enrolment numbers are increasing at the highest rate as compare to all other sectors. For all sectors, unit 12 appears to have a larger slope than that of unit 11, indicating that there are increasingly more unit 12 enrolments than unit 11. It appears that after 2025, unit 12 enrolments are generally more than that of unit 11. Catholic showed a rather slow increase in enrolments relatively to the other sectors.

Random effects

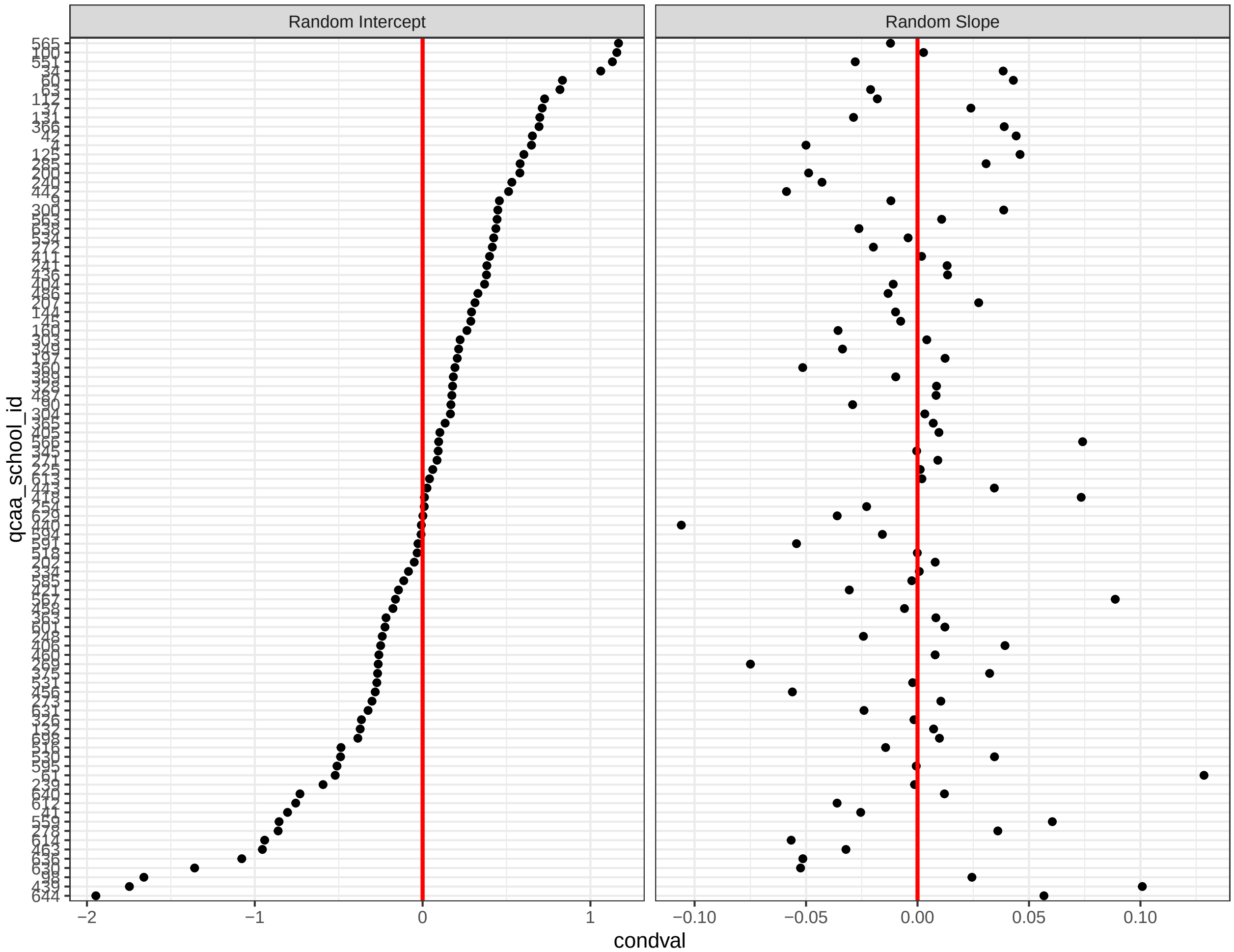

Figure 6: Random effects for all schools

Figure 6 represents the random intercept and slope of the random effects for all 93 schools who offered Agricultural Practices subject. There are no distinct correlations between the random intercept and slope.

Predictions

Figure 7: Model predictions for year 11 enrolments for 20 randomly selected schools

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