

MLM Assignment 2

Christine Hedde-von Westernhagen

Emilia Löscher

Kyuri Park

Utrecht University, 07 March, 2022

Data Description

In this assignment, we analyze the `curran_wide.csv` data set, which contains the information about the age, antisocial behavior, reading skills, emotional support, cognitive stimulation, and mother's age of 221 sampled children. Antisocial behavior and reading skills are measured over 4 occasions. In this analysis, we do not use the variables for child's gender and emotional support.

The data specifics are as follows:

- `id`: child id
- `time`: measurement occasion
- `anti`: antisocial behavior (time-variant)
- `read`: reading recognition skills (time-variant)
- `momage`: mother's age measured at the first occasion (time-invariant)
- `homecog`: cognitive stimulation measured at the first occasion (time-invariant)

1. Convert the wide data file into a long format. Check the data and recode if necessary.

As seen below, the original data (`curran_wide.csv`) is converted into a long format.

`time` is re-coded from 1 - 4 to 0 - 3 such that the first measurement is set to 0.

`read`, `momage`, and `homecog` are centered at the grand mean.

The reasons we center them at the grand mean are as follows:

- 1) it can give a meaningful interpretation for the overall intercept such that the intercept represents the expected value of antisocial behavior for an average child (with average reading skills, average mom's age and average cognitive stimulation) when these predictors are included in the model.
- 2) in case of including an interaction effect, grand-mean centering can make the main effect also more interpretable, as 0 now represents the average value of the explanatory variables.
- 3) it can also make the intercept variance more meaningful (e.g., the expected difference between children at the average level of a predictor variable).

```
## # A tibble: 6 x 6
##   id time anti  read momage homecog
##   <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1   34     0     3 -2.25   2.40 -0.0995
## 2   34     1     6 -1.45   2.40 -0.0995
## 3   34     2     4  0.155  2.40 -0.0995
## 4   34     3     5  0.155  2.40 -0.0995
## 5   58     0     0 -2.05   2.40 -0.0995
## 6   58     1     2  0.155  2.40 -0.0995
```

Table 1

Descriptive statistics

	n	mean	sd	median	min	max	skew	se
id	884	3679	2495	3410	34	8870	0.39	83.92
time	884	1.5	1.12	1.5	0	3	0	0.04
anti	884	1.82	1.82	1	0	10	1.12	0.06
read	884	0	1.62	0.05	-3.65	4.05	0.11	0.05
momage	884	0	1.87	0.4	-4.6	3.4	-0.14	0.06
homecog	884	0	2.45	-0.1	-6.1	4.9	-0.37	0.08

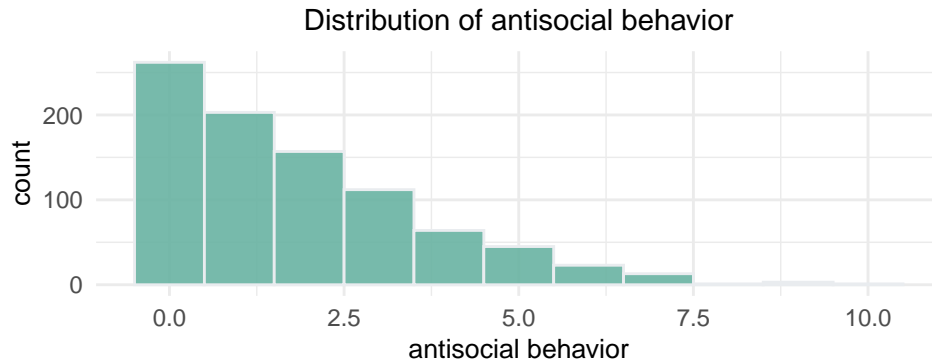


Figure 1: Right-skewed Antisocial behavior

Table 1 shows the descriptive statistics of each variable in the re-coded data. There are total 884 observations for 221 children. The mean values of **read**, **momage**, and **homecog** are 0, as expected (grand-mean centered), and they seem to be not overly skewed given that they have small skewness (i.e., $<|1|$). The skewness value of **anti** is higher than 1, which indicates that the distribution is skewed. Hence, we checked the distribution of **anti** using a histogram and as shown in Figure 1, the distribution is right-skewed where most values are clustered around the left tail. This is in alignment with the (top-left) boxplot shown in Figure 2.

1a. Check the linearity assumption, report and include plots.

Based on Figure 3, it is concluded that the linearity assumption is met in both level 1 and level 2, even though the relationships between the variables seem to be rather weak (i.e., slopes are quite flat), except for (d) *Cognitive stimulation - Average Antisocial behavior*.

1b. Check for outliers.

Figure 2 shows that *antisocial behavior* has several univariate outliers and *cognitive stimulation* also has one outlier in the left tail. In Figure 3, we can check the bivariate outliers at each level. At level 1, although there are few points spotted in the upper-side of plots (a) and (b), they do not seem to be influential. Also, at level 2, no such influential outlier is observed.

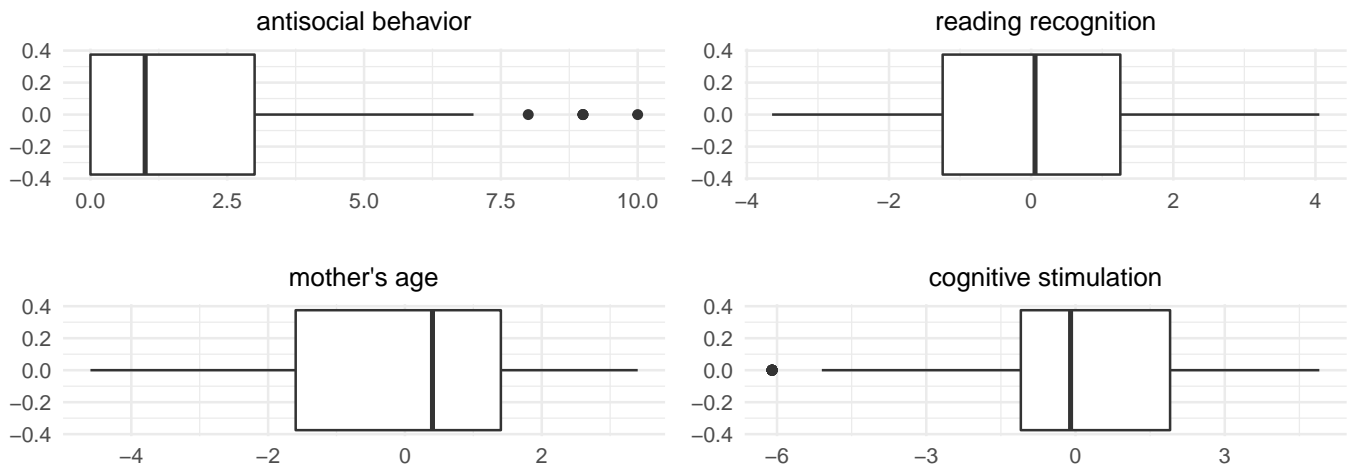


Figure 2: Overall distribution of each variables

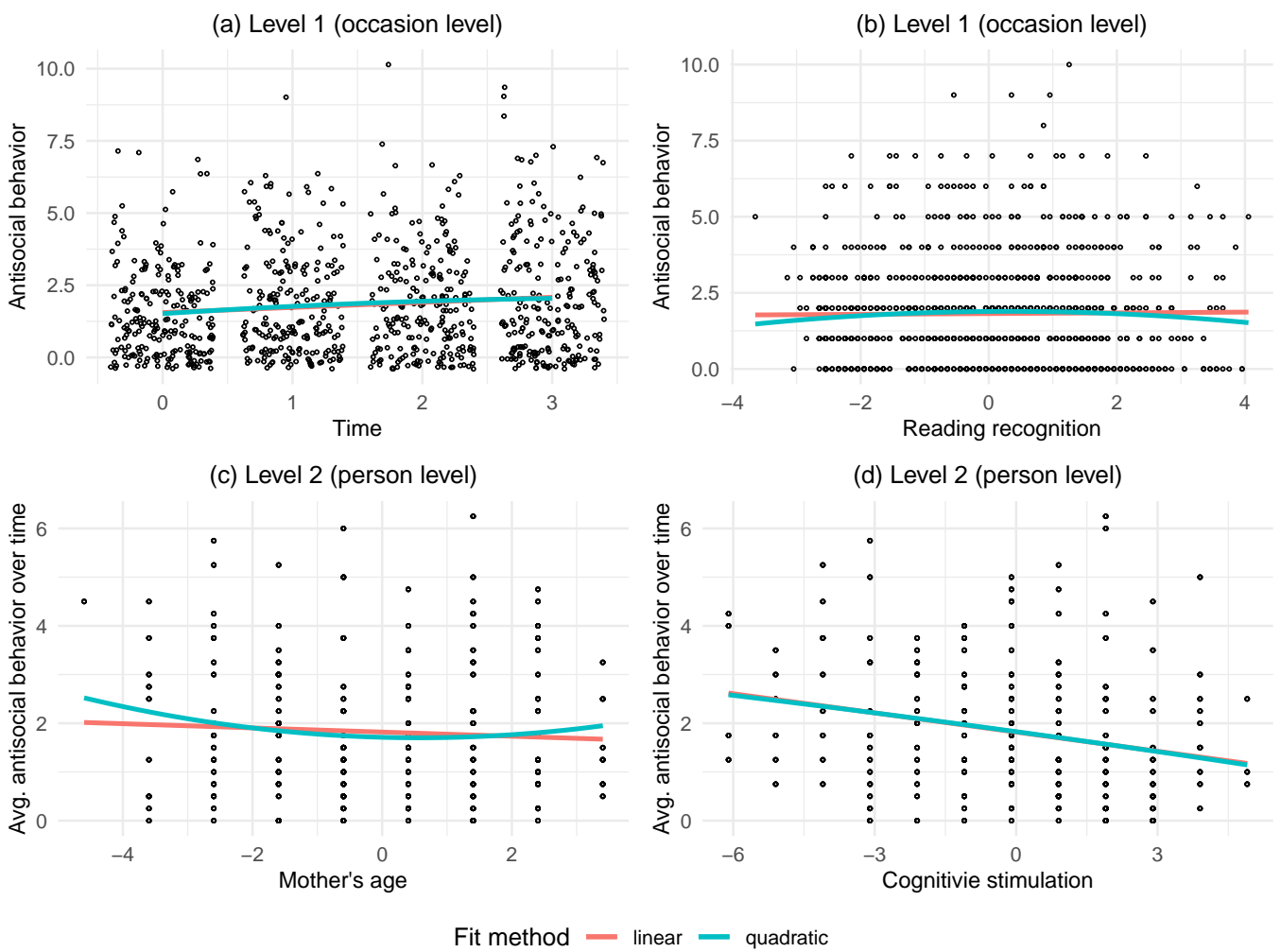


Figure 3: Scatterplots to inspect linearity and outliers

2. Answer the question: should you perform a multilevel analysis?

2a. What is the mixed model equation?

– Mixed Model Equation

$$anti_{ti} = \beta_{00} + u_{0i} + e_{ti}$$

- $anti_{ti}$ refers to antisocial behavior of child i at time t .
- β_{00} refers to the overall intercept, which is the average antisocial behavior over all children.
- u_{0i} refers to the random residual error at the person level (level 2), which represents the deviation from the overall intercept (β_{00}) of child i .
- e_{ti} refers to the residual error at the occasion level (level 1) of child i at time t .

2b. Provide and interpret the relevant results.

As shown in *Table 2*, the intercept term is identical as 1.82 in the single-level model M_0 and the random intercept model M_1 . This overall intercept represents the average antisocial behavior across all children. M_1 decomposes the variance term in a variance at level 1 (Var_{occ}) and level 2 (Var_{sub}).

The deviance for the random intercept model turns out to be significantly smaller than that of the single-level model, $\chi^2(1) = 231.97$, $p < .001$, indicating that Var_{sub} is significantly greater than 0. The AIC value is also lower for the random intercept model ($AIC_{M_1} = 3343.5$) compared to the single-level model ($AIC_{M_0} = 3573.5$), which is in accordance with the deviance difference test result. This means that initial antisocial behavior level (at the first occasion since time is recoded to 0 - 3) differs across the children.

```
# model0: single-level regression model for comparison
model0 <- lm(anti ~ 1, data = curran_long)
summary(model0)

# model1: random intercept model ((benchmark model to compute ICC))
model1 <- lmer(anti ~ 1 + (1|id), REML = FALSE, data = curran_long)
summary(model1)

# check the significance of random intercept
anova(model1, model0)
```

2c. What is the intraclass correlation?

The intraclass correlation (ρ) is calculated as follows:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

```
ICC <- 1.579/(1.579+1.741)
```

The ICC in this model is equal to $\rho = 0.48$.

Table 2

Single-level model and Intercept-only model

Model	M_0 : single-level model	M_1 : random intercept
<i>Fixed part</i>	Coefficient(SE)	Coefficient(SE)
Intercept	1.82(.06)	1.82(.10)
<i>Random part</i>		
Var_{occ}	3.32	1.74
Var_{sub}		1.58
Deviance	3569.5	3357.5
AIC	3573.5	3343.5
Deviance difference^a		231.97***

Note

*** $p < .001$, ** $p < .01$ ^a p-value for χ^2 test is one-sided p-value.

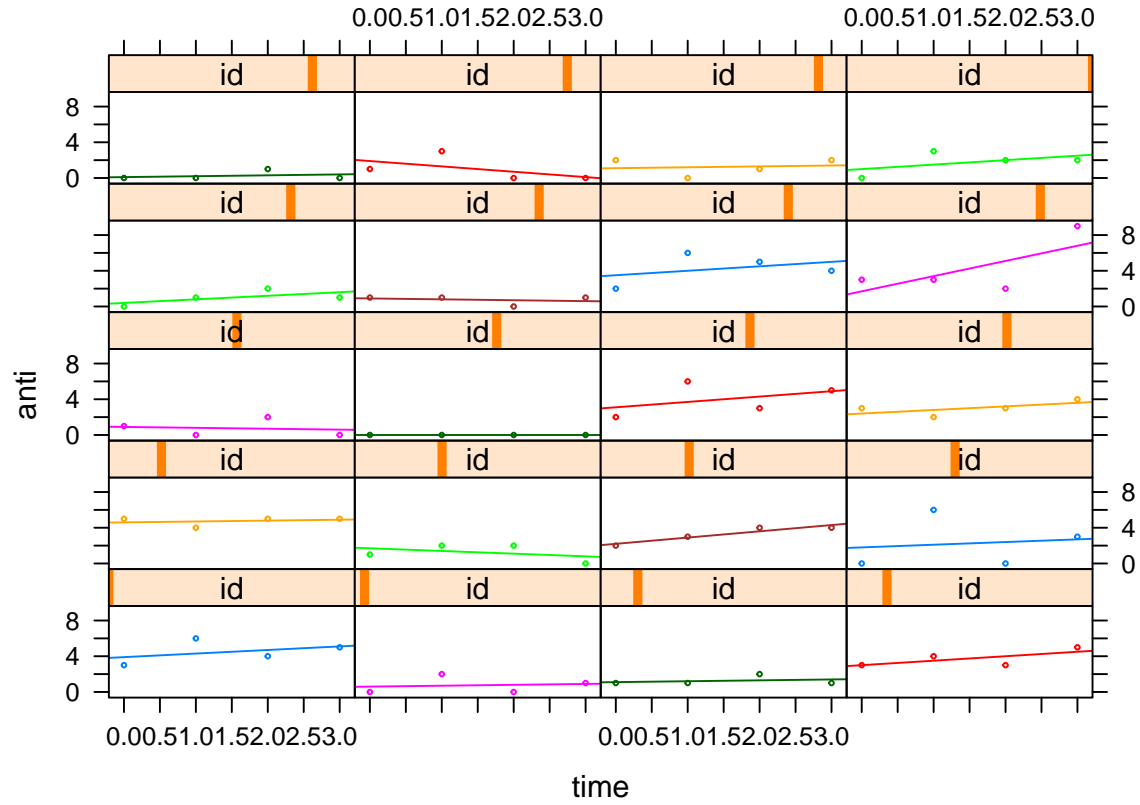
2d. What is your conclusion regarding the overall question regarding the necessity of performing a multilevel analysis?

Yes, we should perform the multilevel analysis in this case, because not only is the data structure nested (i.e., multiple measurements within each individual), but also the difference between individuals accounts for about 48% of the total variance. In other words, the intraclass correlation – ICC: the proportion of the total variance explained by the between-individual differences – is 0.476, which is high enough that the multilevel analysis is warranted.

Also, as shown above, the chi-square difference test comparing models M_0 and M_1 results in the significant outcome in favor of the random intercept multilevel model (M_1), $\chi^2(1) = 231.97$, $p < .001$, which indicates that the M_1 fits the data significantly better than the single-level model (M_0).

In addition, *Figure 4* shows the relationships between *antisocial behavior - time* and *antisocial behavior - reading skills* in each children (first 20). Here we see that the slopes and intercepts vary across the children, which again suggests the necessity of multilevel analysis that can properly address this between-person variabilities.

(a) Antisocial behavior – Time



(b) Antisocial behavior – Reading skills

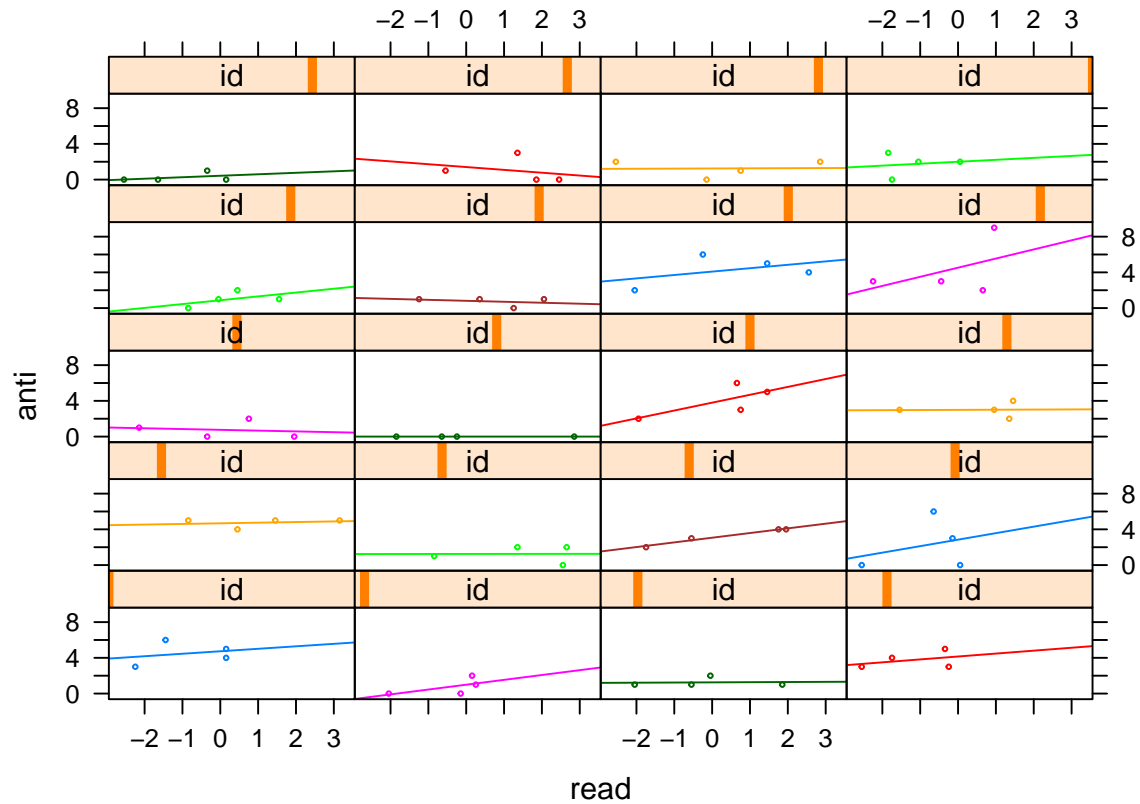


Figure 4: Scatterplots of children to inspect the variabilities in intercepts/slopes

3. Add the time-varying predictor(s).

Provide and interpret the relevant results and provide your overall conclusion.

As shown in Table 3, the intercept term is now 1.55 in M_2 with `time` predictor, which refers to the average antisocial behavior at the first occasion ($\text{time} = 0$) over all children. `time` turns out to be a statistically significant predictor of `anti`, $b_{\text{time}} = 0.18, t(663) = 4.513, p < .001$, meaning that antisocial behavior is expected to increase by 0.18 on each succeeding occasion. By adding `time`, the occasion-level residual error variance (Var_{occ}) goes down to 1.69, but the subject-level variance (Var_{sub}) goes up to 1.60. This is the case because in the longitudinal data with fixed occasion measurements, there is no variation in time-points between subjects. Subsequently, Var_{occ} is overestimated and Var_{sub} is underestimated in the intercept-only model, M_1 . In order to correct this, a measurement occasion variable needs to be included in the model. Hence, M_2 with the predictor `time` correctly decomposes the variance terms and correspondingly can be used as a benchmark model for computing the R^2 .

The significance of adding `time` to the random intercept model(M_1):

The deviance of M_2 is significantly smaller than the deviance of M_1 , $\chi^2(1) = 20.06, p < .001$ (see Table 2 and Table 3). It indicates that M_2 has a significantly better model fit than the intercept-only model, M_1 . The AIC values also reflect this, as $\text{AIC}_{M_2} = 3325.5$ is lower than $\text{AIC}_{M_1} = 3343.5$.

In M_3 with `time` and `read` together, `time` is still a significant predictor $b_{\text{time}} = 0.21, t(882) = 2.73, p < .01$, but `read` is not a significant predictor of `anti`, $b_{\text{read}} = -0.03, t(831) = -0.54, p = .588$. The deviance of M_3 is 3317.2, which is almost the same as M_2 , and the difference in the deviance is not significant as expected, $\chi^2(1) = 0.29, p = .297$. Also, the AIC of M_3 is slightly higher than the AIC of M_2 (see Table 3), which again suggests that adding `read` does not provide a better model fit. Therefore, we drop `read` and proceed with M_2 including only the predictor `time`.

```
# model2: add a time-varying predictor, time ((benchmark model for computing R2))
model2 <- lmer(anti ~ 1 + time + (1|id), REML = FALSE, data = curran_long)
summary(model2)
anova(model2, model1)

# model3: add a time-varying predictor, read
model3 <- lmer(anti ~ 1 + time + read + (1|id), REML = FALSE, data = curran_long)
summary(model3)
anova(model3, model2)
```

Table 3

Adding time-varying predictors

Model	M_2 : add time	M_3 : add read
Fixed part	Coefficient(SE)	Coefficient(SE)
Intercept	1.55(.11)	1.50(.15)
time	0.18(.04)***	0.21(.07)**
read		-0.03(.06)
Random part		
Var_{occ}	1.69	1.69
Var_{sub}	1.60	1.58
Deviance	3317.5	3317.2
AIC	3325.5	3327.2
Deviance difference^a	20.06***	0.29

Note

*** $p < .001$, ** $p < .01$

^a p-value for χ^2 test is one-sided p-value.

4. On which level or levels can you expect explained variance?

Calculate and interpret the explained variances.

In theory, we can expect explained variances (R^2) at both occasion (level 1) and subject level (level 2), as the level 1 predictor in general can explain the variances at both levels. However, in this case where the additional time-varying predictor **read** is not significant and correspondingly does not reduce the variance at the occasion level (but rather increase it slightly), $R^2_{occasion}$ becomes negative and therefore, is not defined.

```
m2var.lv1 <- 1.689 # level 1 variance in model 2 (benchmark model)
m2var.lv2 <- 1.592 # level 2 variance in model 2 (benchmark model)
m3var.lv1 <- 1.693 # level 1 variance in model 3
m3var.lv2 <- 1.576 # level 2 variance in model 3

# explained variance at level 1 (occasion level)
R2.lv1 <- (m2var.lv1 - m3var.lv1) / m2var.lv1
R2.lv1

# explained variance at level 2 (subject level)
R2.lv2 <- (m2var.lv2 - m3var.lv2) / m2var.lv2
R2.lv2
```

The computed R^2 values for each level using M_2 as our benchmark:

- $R^2_{occasion}$ = not defined
- $R^2_{subject}$ = 0.0101

This means that reading recognition skill (**read**) further explains about 1% of the variance between children, which is not a lot. Hence, it is concluded that **read** has a low explanatory power.

5. Add the time invariant predictor(s) to the model.

Provide and interpret the relevant results and provide your overall conclusion.

In M_{4a} we add both time invariant predictors, mother's age (**momage**) and cognitive stimulation (**homecog**). As shown in Table 4, **momage** turns out to be not significant, $b_{momage} = -0.001, t(221) = -0.02, p = .985$, and only **homecog** is significant, $b_{homecog} = -0.13, t(221) = -3.35, p < .001$. This means that for each one point increase in **homecog** (the more cognitive stimulation a child receives), the average **anti** is expected to decrease by 0.13 (the average antisocial behavior of a child goes down) when the other predictors are held constant. The intercept term value remains almost the same as in the previous model (i.e., 1.55), but the interpretation of it now with the added predictors is the expected antisocial behavior for a child with the average cognitive stimulation and average mom's age at the first occasion (as **homecog** and **momage** are centered at the grand mean). By adding these level 2 predictors, we see that the variance at level 2 (subject level) is decreased and this correspondingly produces $R^2_{sub} = .065$ (see Q6 for the derivation of R^2_{sub}).

The significance of adding **momage** and **homecog** to M_2 ¹:

The deviance of M_{4a} is significantly smaller than the deviance of M_2 , $\chi^2(2) = 11.642, p < .01$. It indicates that M_{4a} fits significantly better than M_2 with only **time**. The lower $AIC_{M_{4a}} = 3317.8$ also suggests that M_{4a} is preferable.

However, given that **momage** is not a significant predictor, we drop **momage** and only include **homecog** in the model, M_{4b} . The fixed part of model remains nearly the same; both **time**, $b_{time} = 0.18, t(663) = 4.513, p < .001$, and **homecog**

¹We choose M_2 for the comparison because M_3 includes the non-significant predictor **read** which is not included in M_{4a} and M_{4b}

are significant, $b_{homecog} = -0.13, t(221) = -3.35, p < .001$. It means that antisocial behavior is expected to increase by 0.18 on each succeeding time point, and the average antisocial behavior is expected to decrease by 0.13 with each point higher on cognitive stimulation. The intercept term here would represent the average antisocial behavior for a child with the average cognitive stimulation at the first occasion. By keeping only `homecog` in the model, we see that the variance at level 2 (subject level) is decreased by same amount as in M_{4a} and accordingly results in the same $R_{sub}^2 = .065$.

The significance of adding only `homecog` to M_2 :

The deviance of M_{4b} is significantly smaller than the deviance of M_2 , $\chi^2(1) = 11.641, p < .001$. It indicates that M_{4b} fits significantly better than M_2 with only `time`. In addition, $AIC_{M_{4b}} = 3315.8$ is not only lower than the AIC for M_2 but is also slightly lower than the AIC for M_{4a} . Given these overall results, we decided to proceed with M_{4b} with predictors `time` and `homecog`.

```
## proceed with a model without the non-significant predictor, 'read'
# model4a: add time-invariant predictors, momage & homecog
model4a <- lmer(anti ~ 1 + time + momage + homecog + (1|id), REML = FALSE, data= curran_long)
summary(model4a)
anova(model4a, model2)

# model4b: remove the non-significant predictor, 'momage'
model4b <- lmer(anti ~ 1 + time + homecog + (1|id), REML = FALSE, data= curran_long)
summary(model4b)
anova(model4b, model2)
```

Table 4

Adding time-invariant predictors

Model	M_{4a} : add momage + homecog	M_{4b} : add only homecog
Fixed part	Coefficient(SE)	Coefficient(SE)
Intercept	1.55(.11)	1.55(.11)
time	0.18(.04)***	0.18(.04)***
momage	0.00(.05)	
homecog	-0.13(.04)***	-0.13(.01)***
Random part		
Var_{occ}	1.70	1.69
Var_{sub}	1.50	1.49
Deviance	3305.8	3305.8
AIC	3317.8	3315.8
Deviance difference^{ab}	11.64**	11.64***

Note

*** $p < .001$, ** $p < .01$

^a p-value for χ^2 test is one-sided p-value.

^b Here the deviance is compared with M_2 .

6. On which level or levels can you expect explained variance?

Calculate and interpret the explained variances.

We can expect the explained variances (R^2) at the subject level (level 2), as the level 2 predictor can only explain the variance at level 2. (As seen below, the level 1 variance does not change).

```

m2var.lv1 <- 1.689 # level 1 variance in model 2 (benchmark model)
m2var.lv2 <- 1.592 # level 2 variance in model 2 (benchmark model)
m4var.lv1 <- 1.689 # level 1 variance in model 4b
m4var.lv2 <- 1.488 # level 2 variance in model 4b

# explained variance at level 2 (subject level)
R2.lv2 <- (m2var.lv2 - m4var.lv2) / m2var.lv2
R2.lv2

```

The computed R^2 value for the subject level using M_2 as our benchmark:

- $R^2_{subject} = 0.0653$

This means that cognitive stimulation provided at home (`homecog`) explains about 6.5% of the variance between children.

7. For the time-varying predictor(s), check if the slope is fixed or random.

7a. What are the null- and alternative hypotheses?

The null- and alternative hypotheses for checking whether the slope for the time-varying predictor `time` is fixed or random are:

- $H_0: \sigma_{u1}^2 = 0$; The slope for the time variable is equal across the children.
- $H_1: \sigma_{u1}^2 > 0$; The slope for the time variable varies across the children.

The null- and alternative hypotheses for checking whether the slope for the time-varying predictor `read` is fixed or random are:

- $H_0: \sigma_{u2}^2 = 0$; The slope for the read variable is equal across the children. - $H_1: \sigma_{u2}^2 > 0$; The slope for the read variable varies across the children. ** either we add model 5c situation here **

OR

*** figure out this line ** - H_0 : The slope for the time and/or read variable is equal across the children.

- H_1 *: The slope for the time and/or read variable varies across the children.

We considered both of the time-varying predictors `time` as `read`, as it is possible for a predictor to have no significant average regression slope but to have a significant variance component for the slope.

7b. Provide and interpret the relevant results.

First, we fit the models including a random slope for each of the time-varying predictors separately: `time` and `read` one-by-one.

The deviance difference between the model including random slopes in `time` (M_{5a}) and model with fixed slope (M_{4b}) is significant, $\chi^2(2) = 26.561$, $p < .001$. $AIC_{M_{5a}} = 3293.3$ is lower than $AIC_{M_{4b}} = 3315.8$.

The deviance difference between the model including random slopes in `read` (M_{5b}) and model with fixed slope (M_{4b}) is also significant, $\chi^2(3) = 18.872$, $p < .001$. $AIC_{M_{5a}} = 3303.0$ is again lower than $AIC_{M_{4b}} = 3315.8$.

However, when we attempt to let both predictors have random slopes (M_{5c}), the model fails to converge.

Given that M_{5a} has the lower deviance as well as lower AIC value, we decided to proceed with M_{5a} including random slopes in `time` and interpret the results further.

```

# model5a: let 'time' have random slopes
model5a <- lmer(anti ~ 1 + time + homecog + (1 + time|id), REML = FALSE, data = curran_long)
summary(model5a)
anova(model5a, model4b)

# model5b: let 'read' have random slopes
model5b <- lmer(anti ~ 1 + time + homecog + read + (1 + read|id),
               REML = FALSE, data= curran_long)
summary(model5b)
anova(model5b, model4b)

# model5c: let both have random slopes --> model fails to converge
model5c <- lmer(anti ~ 1 + time + homecog + (1 + time + read|id),
               REML = FALSE, data = curran_long)

```

Table 5

Adding random slopes

Model	M_{5a} : random slopes in time
Fixed part	Coefficient(SE)
Intercept	1.55 (.10)
time	0.18(.04)***
homecog	-0.10(.04)**
Random part	
Var_{occ}	0.95
Var_{sub}	1.53
Var_{time}	0.10
$Cor_{sub*time}$	0.41
Deviance	3279.3
AIC	3293.3
Deviance difference^{ab}	26.56***

Note

*** $p < .001$, ** $p < .01$

^a p-value for χ^2 test is one-sided p-value.

^b Here the deviance is compared with M_{4b} .

As seen in Table 5, the fixed part of model does not change much in M_{5a} where the slope of **time** is allowed to vary across children. The intercept term is the same as 1.55, **time** is still significant with the same regression coefficient, $b_{time} = 0.18, t(221) = 4.137, p < .001$, and **homecog** is also significant but with a bit lower coefficient, $b_{homecog} = -0.10, t(221) = -2.793, p < .01$, compared to the previous model (M_{4b}). Now that we have random slopes for time, the regression coefficient of **time** represents the average slope across all children. The interpretation of regression coefficient of **homecog** as well as the intercept term remains the same as previously described (see Q5).

Having introduced the random slopes for time, now we also have the correlation between the intercept and slope. As shown in Table 5, $Cor_{sub*time}$ is estimated to be 0.41, which can be interpreted as positive intercepts (expected **anti** at the first occasion) tend to result in positive slopes (rate of change) and vice versa, in a magnitude of 0.41. This means that if a child starts with a high initial level of antisocial behavior, then he/she will show a faster increase in antisocial behavior in the following time.

Adding random slope for **time** (M_{5a}) turns out to fit significantly better than M_{4b} with fixed slope, $\chi^2(2) = 26.561, p < .001$. In addition, $AIC_{M_{5a}} = 3293.3$ is lower than $AIC_{M_{4b}} = 3315.8$, which again suggests the better fit of M_{5a} . This means that the size of the effect of **time** on antisocial behavior significantly varies across different children.

7c. Provide an overall conclusion.

Given the overall results above, we decide that M_{5a} with the random slope in **time** has a significantly better model fit than M_{4b} , and correspondingly conclude that the children significantly vary in their rate of change in antisocial behavior over time.

8. If there is a random slope, set up a model that predicts the slope variation.

We have found that there is a random slope for the variable **time**. In the following, we are fitting two models to check whether the variable **momage** or the variable **homecog** can explain the slope variation in **time**.

Provide and interpret the relevant results and provide your overall conclusion.

The estimate for the interaction term is $b_{time*homecog} = -0.05$. This means that for children who have more cognitive stimulation (higher values for **homecog**), their antisocial behavior increases less strongly over time or even decreases in some cases. This is visualized in the *Figure 5*.

I am not sure if the they will decrease...?

From M_{6a} , it can be seen that **homecog** can predict the slope variation of **time**, as the interaction of **homecog** and **time** is significant, $b_{time*homecog} = -0.05, t(221) = -2.643, p < .01$. This subsequently means that the effect of **time** on **anti** decreases by 0.05 with every point increase in **homecog**. In other words, the rate of change in antisocial behavior over time increases in a slower pace when cognitive stimulation is higher. This is visualized in *Figure 5* and *Figure 6*.

Note, however, that the fixed effect for **homecog** is not significant anymore after including the interaction term, $b_{homecog} = -0.06, t(221) = -1.625, p = .106$. Nevertheless, the fixed effect of **homecog** is not removed from the model, as main effects of an interaction should always remain in the model. **time** is still a significant predictor, $b_{time} = 0.18, t(221) = 4.202, p < .001$, and the interpretation of it now is that with each time point increase, antisocial behavior of a child is expected to increase by 0.18, when the received cognitive stimulation (**homecog**) is average, as it is grand-mean centered (**homecog** = 0 refers to the mean of **homecog**). The interpretation of overall intercept stays the same: the average antisocial behavior for a child with the average cognitive stimulation at the first occasion. The estimate of correlation between intercept and slope ($Cor_{sub*time}$) is estimated to be 0.47, which is a bit higher than the previous model, and the interpretation remains the same: if a child starts with a high initial level of antisocial behavior, then he/she will show a faster increase in antisocial behavior in the following time.

Comparison of the M_{6a} and M_{5a} reveals that the variance component of time (Var_{time}) decreases a bit (see *Table 5* and *Table 6*). It indicates that **homecog** explains a part of the slope variation in **time**. Using M_{5a} as a benchmark, the explained variance is computed: the interaction term **time*homecog** explains about 13% of the variation in the slope for **time**; $R^2_{randomslope} = 0.13$. The deviance difference between M_{6a} and M_{5a} turns out to be significant, $\chi^2(1) = 6.8778, p < .01$, meaning that M_{6a} fits significantly better than M_{5a} . Additionally, $AIC_{M_{6a}} = 3288.4$ is lower than $AIC_{M_{5a}} = 3293.3$, which again tells us that M_{6a} is preferred to M_{5a} .

In M_{6b} , the interaction of **time** and **momage** is added to the model. As seen in *Table 6*, it can be seen that **momage** cannot predict the slope variation in **time**, as the interaction of **momage** and **time** is not significant, $b_{time*momage} = -0.003, t(221) = -0.142, p = .888$. Not surprisingly, the fixed effect of **momage** is also not significant, $b_{momage} = -0.006, t(222) = -0.119, p = .905$. This was already the case in M_{4a} , which led us to drop **momage** from the model. Furthermore, the deviance difference test suggests that M_{6b} does not fit significantly better than M_{5a} , $\chi^2(2) = 0.0535, p = .974$. AIC value of M_{6b} is also higher not only than $AIC_{M_{6a}}$ but also slightly higher than $AIC_{M_{5a}}$, which indicates that M_{6a} is a better model than M_{6b} .

Given these overall results, we conclude that M_{6a} with `homecog` can predict the slope variation in `time`. In addition, M_{6a} provides a better fit compared to M_{5a} and M_{6b} , given its significant deviance difference and lower AIC value. Hence, we decide to stick with M_{6a} as our final model.

```
## add a cross-level interaction
# model6a: check if homecog can explain the slope variation in time
model6a <- lmer(anti ~ 1 + time + homecog + homecog*time + (1+time|id),
               REML = FALSE, data = curran_long)
summary(model6a)
anova(model6a, model5a)

# model6b: check if momage can explain the slope variation in time
model6b <- lmer(anti ~ 1 + time + momage + homecog + momage*time + (1+time|id),
               REML = FALSE, data = curran_long)
summary(model6b)
anova(model6b, model5a)
```

Table 6

Adding a cross-level interaction

Model	M_{6a} : homecog*time	M_{6b} : momage*time
<i>Fixed part</i>	Coefficient(SE)	Coefficient(SE)
Intercept	1.55(.10)	1.55(.10)
time	0.18(.04)***	0.18(.04)***
momage		-0.01(.05)
homecog	-0.06(.04)	-0.10(.04)**
time*homecog	-0.05(.02)**	
time*momage		0.00(.02)
<i>Random part</i>		
Var_{occ}	1.53	1.53
Var_{sub}	0.94	0.95
Var_{time}	0.08	0.10
$Cor_{sub*time}$	0.47	0.41
Deviance	3272.4	3279.2
AIC	3288.4	3297.2
Deviance difference^{ab}	6.88**	0.05

Note

*** $p < .001$, ** $p < .01$

^a p-value for χ^2 test is one-sided p-value.

^b Here the deviance is compared with M_5 .

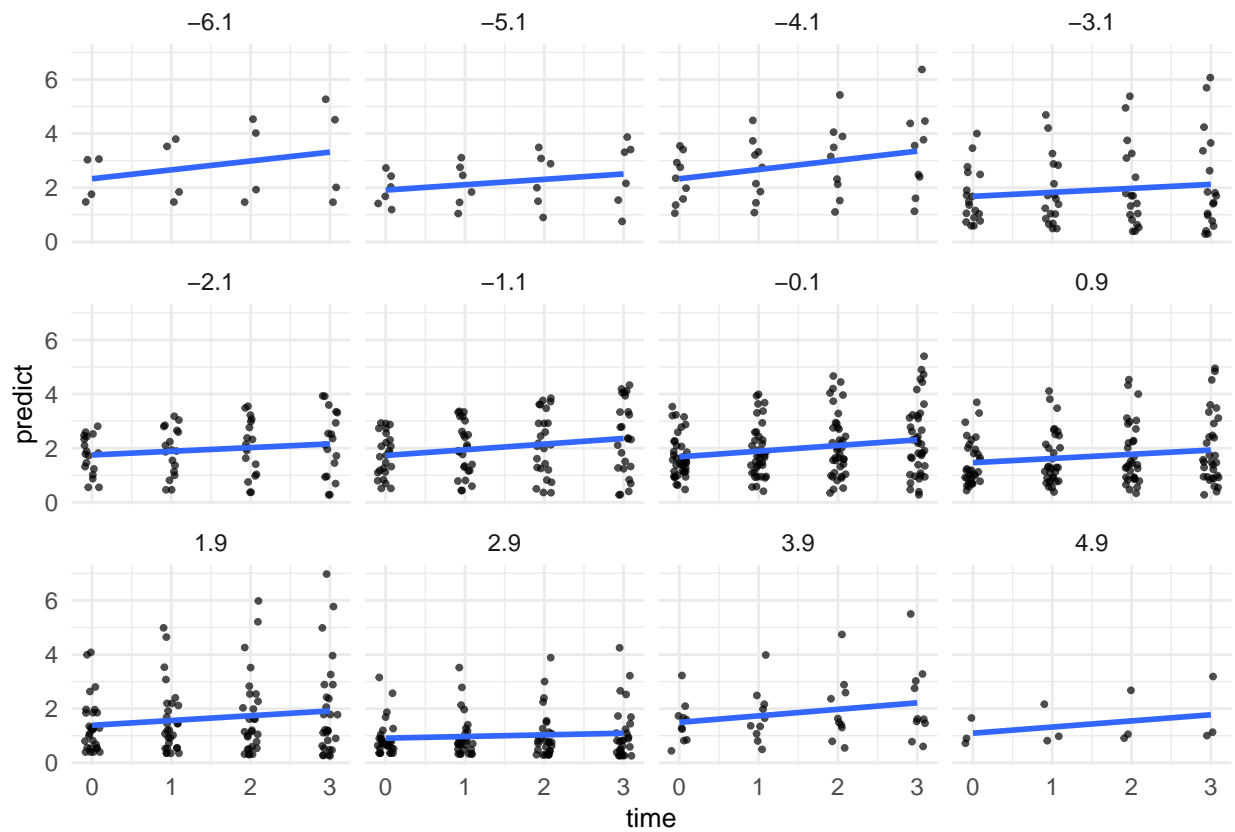


Figure 5: Predicted values of antisocial behavior over time by levels of 'homecog'

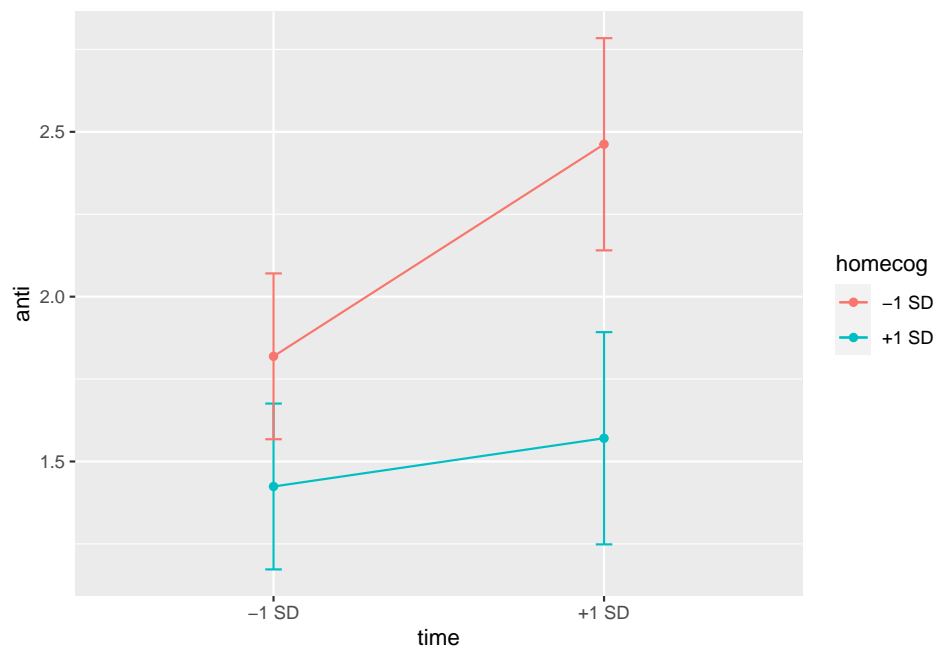


Figure 6: Interaction between time and homecog

9. Decide on a final model.

We choose M_{6a} to be our final model as it has the best model fit among all the tested models. The level 1 and 2 model equations, as well as the mixed model equation of our final model are given below.

9a. Provide the separate level 1 and 2 model equations, as well as the mixed model equation.

– Level 1 Model Equation

$$anti_{ti} = \pi_{0i} + \pi_{1i}time_{ti} + e_{ti}$$

– Level 2 Model Equations

$$\pi_{0i} = \beta_{00} + \beta_{01}homecog_i + u_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}homecog_i + u_{1i}$$

– Mixed Model Equation

$$anti_{ti} = \beta_{00} + \beta_{10}time_{ti} + \beta_{01}homecog_i + \beta_{11}homecog_i \times time_{ti} + u_{0i} + u_{1i}time_{ti} + e_{ti}$$

9b. Check the normality assumption for both the level-1 and level-2 errors, report.

We checked the normality assumption on level 1 and on level 2. In all three cases, the normality assumption is (somewhat) violated. This can be seen from the respective Q-Q plots in *Figure 7* as the points are not on the straight diagonal line.

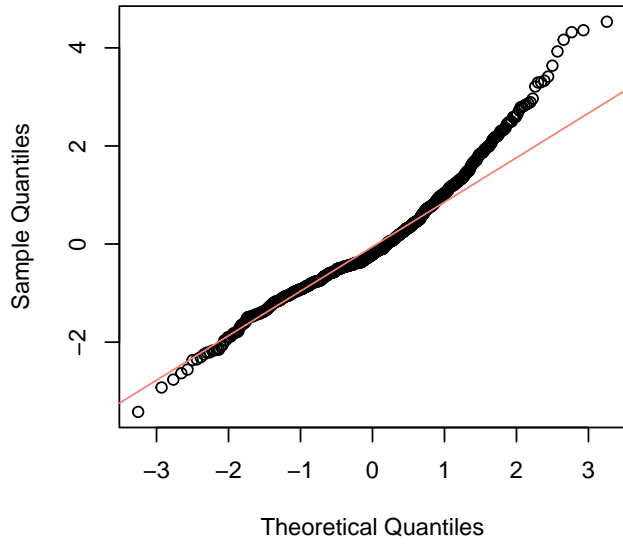
- In *Figure 7(a)*, the upper end of the Q-Q plot deviates from the diagonal line forming a s-shaped curve. Therefore, it is concluded that the normality assumption is slightly violated in level 1.
- *Figure 7(b)* and *Figure 7(c)* show a similar pattern – slightly u-shaped curve – where both tails deviate upward from the straight diagonal line. Hence, we conclude that normality assumption is again slightly violated in level 2, with both intercept and slope residuals.
- A possible explanation & corresponding solution :
In *Figure 1*, we observe that the distribution of the outcome variable antisocial behavior (**anti**) is very right skewed. This might have something to do with the deviation of the residual points in the Q-Q plots below. Thus, transforming the outcome variable could make the normality assumption more justifiable. For example, a log-transformation of the variable **anti** could be considered.²

Contribution

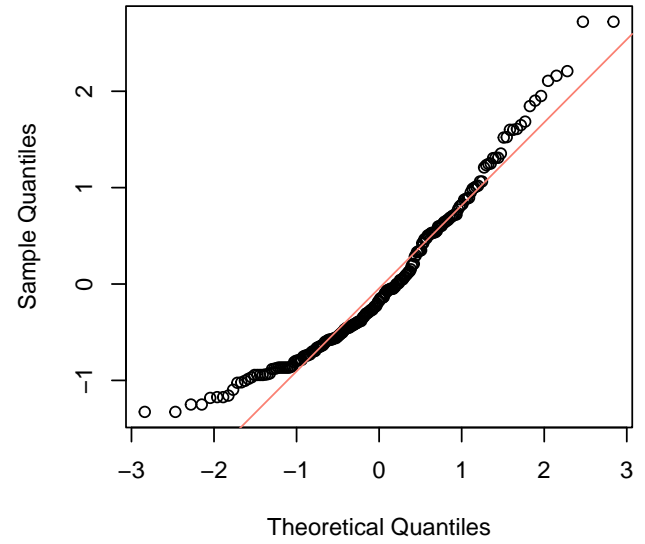
- Christine: Coding, interpretation, writing
- Emilia: Coding, interpretation, writing, Excel
- Kyuri: Coding, interpretation, writing

²When we tested this alternative specification of the variable **anti**, we saw that for the log-transformed outcome variable **anti**, the distribution of the residuals resembled a normal distribution closely. Hence, it is advisable to re-analyze the data with a log-transformed version of the outcome variable **anti**.

(a) Normal QQ plot for Level 1 Residual Var.



(b) Normal QQ plot for Level 2 Intercept Var.



(c) Normal QQ plot for Level 2 Slope Var.

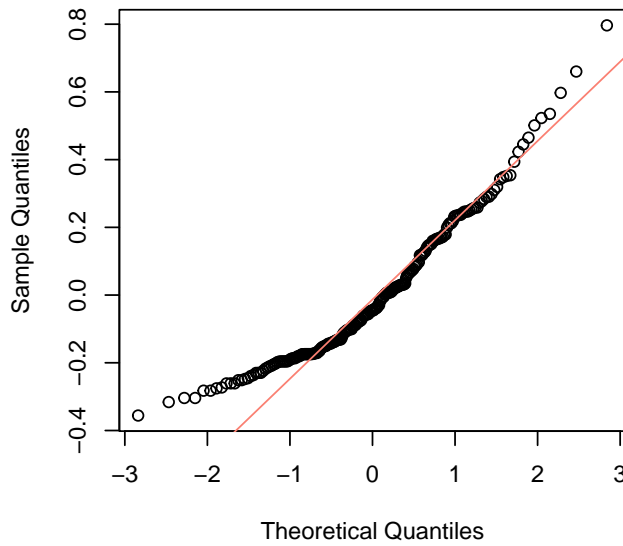


Figure 7: Q-Q plots for level 1 and level 2 residuals