Section 2 pre Final

Colin Linke

2024-06-03

```
full.model <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWriting + Fa

full.model.reduced.interactions <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVer

small.model <- lme(GPA ~ Cluster_current * Semester, random = ~ 1 | BARCS_ID, data = data.file.long, na

pseudo.trajectory <- lme(GPA ~ 1 + Fager4_binary + FH_binary + Sex + Cluster_SEM1 + Semester + Age1stro

fm.time.slope <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWriting +
```

```
pbkrtest::KRmodcomp(full.model.lmer, full.model.lmer.reduced.interactions) ## only allows lmer -.-
```

```
## large : GPA ~ Cluster_current + Sex + Age1stround + SATMath + SATVerbal +
##     SATWriting + Fager4_binary + FH_binary + STAI_SELF_Total +
##     BDI_SELF_Total + Parental_SES + Semester + (1 | BARCS_ID) +
##     Cluster_current:Sex + Cluster_current:Semester + Sex:Semester
## small : GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal +
##     SATWriting + Fager4_binary + FH_binary + STAI_SELF_Total +
##     BDI_SELF_Total + Parental_SES + Semester * Cluster_current +
##     Semester + (1 | BARCS_ID)
##            stat       ndf       ddf F.scaling p.value
## Ftest    0.5628    5.0000 2945.1811   0.99993  0.7286
```

```
anova(full.model.reduced.interactions, full.model)
```

```
##                                  Model df      AIC      BIC    logLik   Test
## full.model.reduced.interactions     1 24 5470.785 5617.665 -2711.393
## full.model                          2 29 5477.951 5655.431 -2709.976 1 vs 2
##                                   L.Ratio p-value
## full.model.reduced.interactions
## full.model                        2.833905  0.7256
```

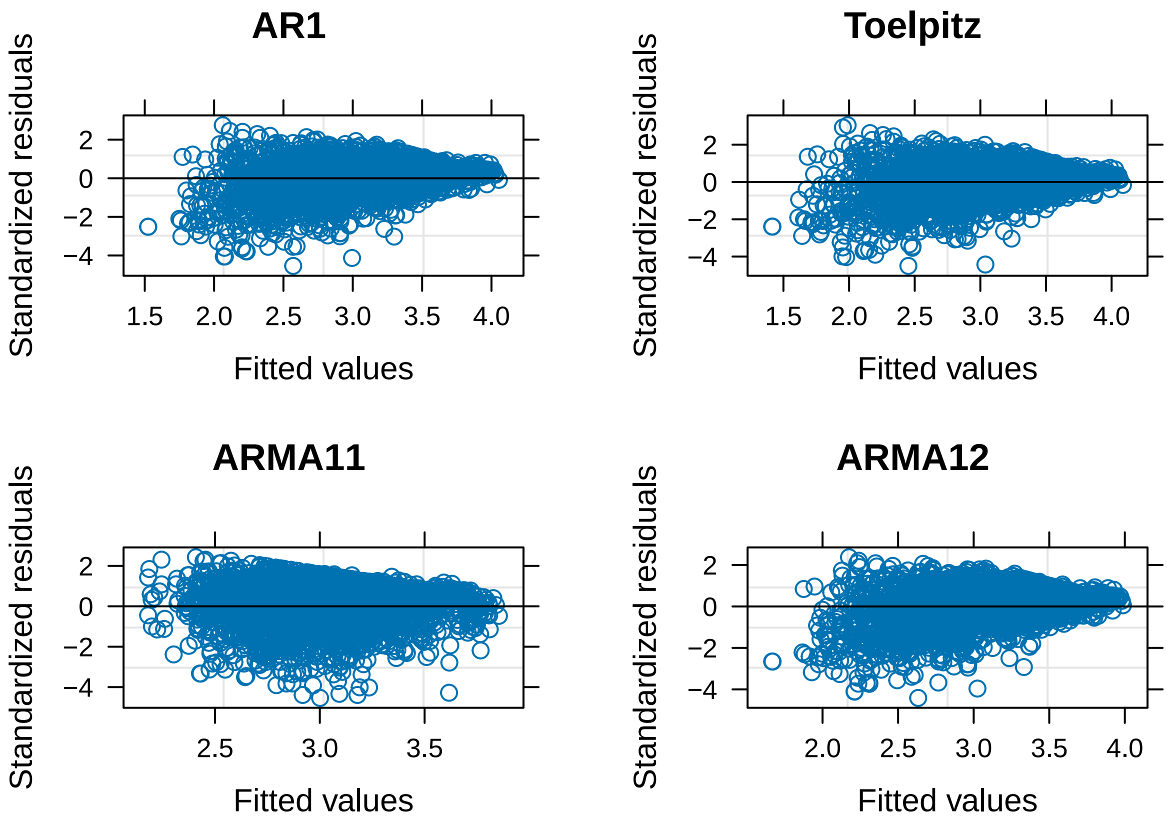
```
full.model.AR1 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWriting
                     random = ~ 1 | BARCS_ID, correlation = corAR1(form = ~ Time| BARCS_ID),
                     data = data.file.long, na.action = na.exclude, method = "ML" )  ##N = 3361

full.model.Unstructured <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SA
                       random = ~ 1 | BARCS_ID, correlation = corSymm(form = ~ Time| BARCS_ID),
                     data = data.file.long, na.action = na.exclude, method = "ML"  ) ## N = 3361

full.model.CompSymm <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWri
                       random = ~ 1 | BARCS_ID, correlation = corCompSymm(form = ~ Time| BARCS_ID),
                     data = data.file.long, na.action = na.exclude, method = "ML"  ) ## N = 3361
```

```r
full.model.Toelpitz <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWri
                          random = ~ 1 | BARCS_ID, correlation = corARMA(p = 0, q = 3, form = ~ Time| BARCS
                        data = data.file.long, na.action = na.exclude, method = "ML"  ) ## N = 3361

full.model.Exponential <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATW
                          random = ~ 1 | BARCS_ID, correlation = corExp(form = ~ Time| BARCS_ID),
                        data = data.file.long, na.action = na.exclude, method = "ML"  ) ## N = 3361

full.model.gaussian <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWri
                          random = ~ 1 | BARCS_ID, correlation = corGaus(form = ~ Time| BARCS_ID),
                        data = data.file.long, na.action = na.exclude, method = "ML"  ) ## N = 3361

full.model.MA1 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWriting +
                          random = ~ 1 | BARCS_ID, correlation = corARMA(q = 1, form = ~ Time| BARCS_ID),
                        data = data.file.long, na.action = na.exclude, method = "ML" ) ## N = 3361

full.model.ARMA11 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWritin
                          random = ~ 1 | BARCS_ID, correlation = corARMA(p = 1, q = 1, form = ~ Time| BARCS
                        data = data.file.long, na.action = na.exclude , method = "ML"  ) ## N = 3361

full.model.ARMA12 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWritin
                          random = ~ 1 | BARCS_ID, correlation = corARMA(p = 1, q = 2, form = ~ Time| BARCS
                        data = data.file.long, na.action = na.exclude, method = "ML"   ) ## N = 3361

full.model.ARMA21 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWritin
                          random = ~ 1 | BARCS_ID, correlation = corARMA(p = 2, q = 1, form = ~ Time| BARCS
                        data = data.file.long, na.action = na.exclude , method = "ML" ) ## N = 3361

full.model.ARMA22 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWritin
                          random = ~ 1 | BARCS_ID, correlation = corARMA(p = 2, q = 2, form = ~ Time| BARCS
                        data = data.file.long, na.action = na.exclude , method = "ML") ## N = 3361

AIC_values <- AIC(full.model.AR1, full.model.Unstructured, full.model.CompSymm, full.model.Toelpitz, ful
BIC_values <- BIC(full.model.AR1, full.model.Unstructured, full.model.CompSymm, full.model.Toelpitz, ful
df_scores_Cor <- data.frame(AIC_values, BIC_values)
df_scores_Cor[,-3]
```

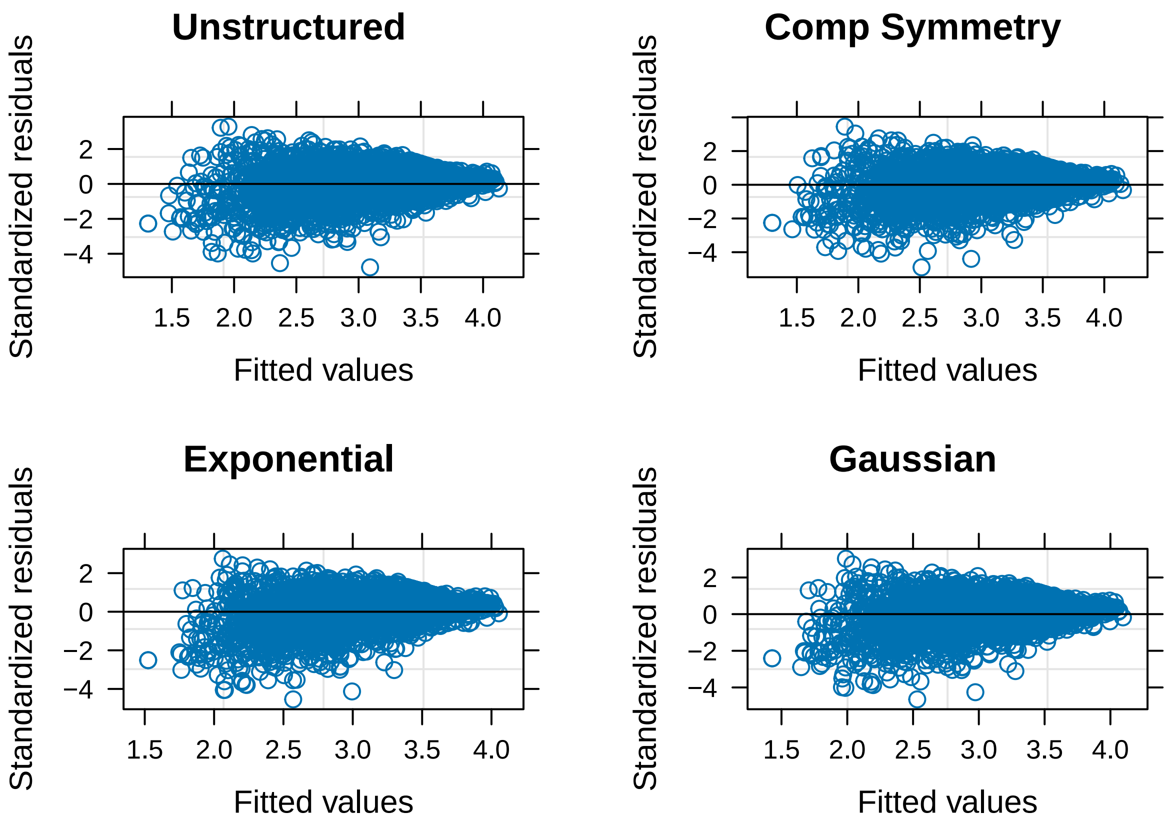
```
##                         df      AIC      BIC
## full.model.AR1          25 5417.072 5570.072
## full.model.Unstructured 30 5411.590 5595.190
## full.model.CompSymm     25 5472.785 5625.785
## full.model.Toelpitz     27 5411.958 5577.198
## full.model.Exponential  25 5417.072 5570.072
## full.model.gaussian     25 5424.321 5577.321
## full.model.MA1          25 5424.587 5577.586
## full.model.ARMA11       26 5411.011 5570.130
## full.model.ARMA12       27 5411.958 5577.198
## full.model.ARMA21       27 5411.958 5577.198
## full.model.ARMA22       28 5413.958 5585.318
```

```r
residuals.AR1 <- plot(full.model.AR1, main = "AR1")
residuals.Unstructured <- plot(full.model.Unstructured, main = "Unstructured")
residuals.CompSymm <- plot(full.model.CompSymm, main = "Comp Symmetry")
residuals.Toelpitz <- plot(full.model.Toelpitz, main = "Toelpitz")
```

```r
residuals.Exponential <- plot(full.model.Exponential, main = "Exponential")
residuals.Gaussian <- plot(full.model.gaussian, main = "Gaussian")
residuals.MA1 <- plot(full.model.MA1, main = "MA1")
residuals.ARMA11 <- plot(full.model.ARMA11, main = "ARMA11")
residuals.ARMA12 <- plot(full.model.ARMA12, main = "ARMA12")
residuals.ARMA21 <- plot(full.model.ARMA21, main = "ARMA21")
residuals.ARMA22 <- plot(full.model.ARMA22, main = "ARMA22")

cowplot::plot_grid(residuals.AR1, residuals.Toelpitz, residuals.ARMA11, residuals.ARMA12, nrow = 2)
```




```r
cowplot::plot_grid(residuals.Unstructured, residuals.CompSymm, residuals.Exponential, residuals.Gaussia
```



There seems to be a flat optimization plane and similar correlation structures would tend to estimate similarly. AIC –> ARMA (1,1), BIC –> AR(1).

choosing the ARMA (1,1) structure, due to the paper deciding on the AIC.

```
small.model.ARMA11 <- lme(GPA ~ Cluster_current * Semester, random = ~ 1 | BARCS_ID, data = data.file.lo

pseudo.trajectory.ARMA11 <- lme(GPA ~ 1 + Fager4_binary + FH_binary + Sex + Cluster_SEM1 + Semester + Ag

time.slope.ARMA11 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWritin

small.model.AR1 <- lme(GPA ~ Cluster_current * Semester, random = ~ 1 | BARCS_ID, data = data.file.long

pseudo.trajectory.AR1 <- lme(GPA ~ 1 + Fager4_binary + FH_binary + Sex + Cluster_SEM1 + Semester + Age1s

time.slope.AR1 <- lme(GPA ~ 1 + Cluster_current + Sex + Age1stround + SATMath + SATVerbal + SATWriting +
```

This is only to compare the three (plus the alternative Time as numeric variable) models to themselves with the ARMA11 and AR1 structures. the ARMA11 structure would be chosen for all models based on the AIC, however if it were based on the BIC, then the AR1 would be prefered, with the exception of the Small Model.

the first one with manually computed significance with the SIdak correction, the second regular Anova output. The extra commands dont seem to do anything really unfortunately

```
print(as.data.frame(anova.pt))
```

```
##                              numDF denDF     Fvalue       pvalue corrected.pvalue
## (Intercept)                      1  2695 33932.3480 0.000000e+00     0.000000e+00
## Fager4_binary                    1   985    25.0115 6.744727e-07     4.215456e-08
## FH_binary                        1   985     6.4890 1.100532e-02     6.914064e-04
## Sex                              1   985    24.7649 7.641016e-07     4.775637e-08
```

```
## Cluster_SEM1                     2   985    19.4252 5.319448e-09     3.324655e-10
## Semester                         3  2695     4.6134 3.179569e-03     1.990199e-04
## Age1stround                      1   985     4.4715 3.471572e-02     2.205853e-03
## SATMath                          1   985   178.9563 0.000000e+00     0.000000e+00
## SATVerbal                        1   985    24.8150 7.449722e-07     4.656078e-08
## SATWriting                       1   985    19.4518 1.145224e-05     7.157688e-07
## STAI_SELF_Total                  1   985     0.3131 5.759329e-01     5.220446e-02
## BDI_SELF_Total                   1   985     4.3129 3.808356e-02     2.423789e-03
## Parental_SES                     1   985     4.3523 3.721633e-02     2.367601e-03
## Group_transition1                2   985     4.0941 1.695495e-02     1.068199e-03
## Cluster_SEM1:Semester            6  2695     1.1905 3.081967e-01     2.276523e-02
## Semester:Group_transition1       6  2695     0.7936 5.748391e-01     5.205186e-02
##                              significant
## (Intercept)                          ***
## Fager4_binary                        ***
## FH_binary
## Sex                                  ***
## Cluster_SEM1                         ***
## Semester                               *
## Age1stround
## SATMath                              ***
## SATVerbal                            ***
## SATWriting                           ***
## STAI_SELF_Total
## BDI_SELF_Total
## Parental_SES
## Group_transition1
## Cluster_SEM1:Semester
## Semester:Group_transition1
```

```r
Anova(pseudo.trajectory.ARMA11)#, test.statistic = "F",  robust= "hc3", correction = "Sidak")
```

```
## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted

## Analysis of Deviance Table (Type II tests)
##
## Response: GPA
##                                Chisq Df Pr(>Chisq)
## Fager4_binary                 7.1226  1   0.007612 **
## FH_binary                     1.4046  1   0.235954
## Sex                          21.3295  1  3.867e-06 ***
## Cluster_SEM1                 47.1070  2  5.900e-11 ***
## Semester                     14.0422  3   0.002848 **
## Age1stround                   4.7450  1   0.029383 *
## SATMath                      28.9591  1  7.392e-08 ***
```

```
## SATVerbal                   1.6565  1   0.198082
## SATWriting                 20.2297  1  6.868e-06 ***
## STAI_SELF_Total             0.8879  1   0.346035
## BDI_SELF_Total              4.2848  1   0.038455 *
## Parental_SES                3.9228  1   0.047635 *
## Group_transition1           8.2109  2   0.016483 *
## Cluster_SEM1:Semester       4.7318  6   0.578640
## Semester:Group_transition1  4.8003  6   0.569672
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted
```

```
## Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)
##
## Response: GPA
##                                  F Df  Df.res    Pr(>F)
## Fager4_binary               7.1522  1  993.14  0.007610 **
## FH_binary                   1.2379  1  990.71  0.266149
## Sex                        20.4954  1  980.59 6.711e-06 ***
## Cluster_SEM1               22.9339  2  985.96 1.839e-10 ***
## Semester                    4.7585  3 2733.37  0.002595 **
## Age1stround                 4.5078  1 1025.82  0.033980 *
## SATMath                    29.3119  1  973.12 7.763e-08 ***
## SATVerbal                   1.4649  1  975.50  0.226442
## SATWriting                 20.5234  1  974.99 6.620e-06 ***
## STAI_SELF_Total             1.0099  1  980.92  0.315182
## BDI_SELF_Total              3.8749  1  986.39  0.049294 *
## Parental_SES                3.7321  1  976.97  0.053664 .
## Group_transition1           3.9939  2  981.76  0.018727 *
## Cluster_SEM1:Semester       0.9069  6 2739.90  0.488726
## Semester:Group_transition1  0.9043  6 2734.68  0.490655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

replicates the table with significances on page 9

```r
AIC(full.model.ARMA11, full.model.ARMA11.slope)
```

```
##                         df      AIC
## full.model.ARMA11       26 5411.011
## full.model.ARMA11.slope 28 5414.360
```

```r
BIC(full.model.ARMA11, full.model.ARMA11.slope)
```

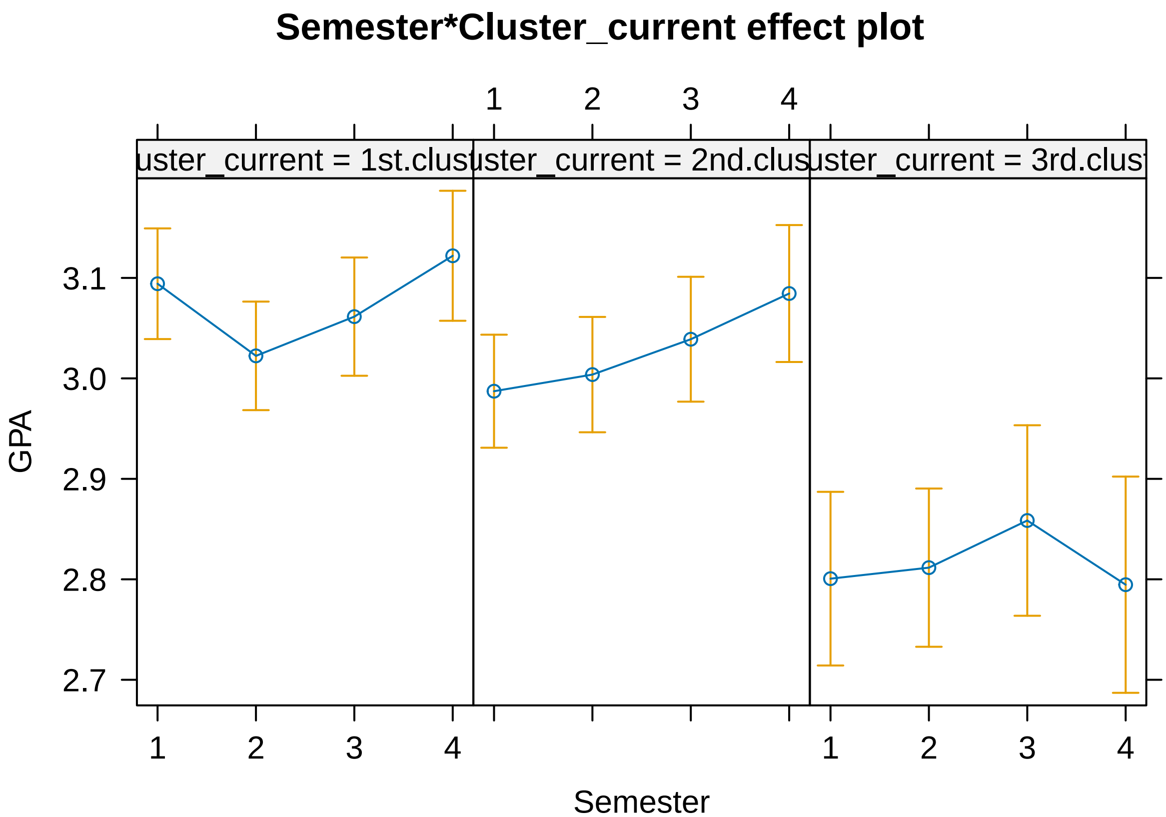
```
##                         df      BIC
## full.model.ARMA11       26 5570.130
```

```
## full.model.ARMA11.slope 28 5585.719
```

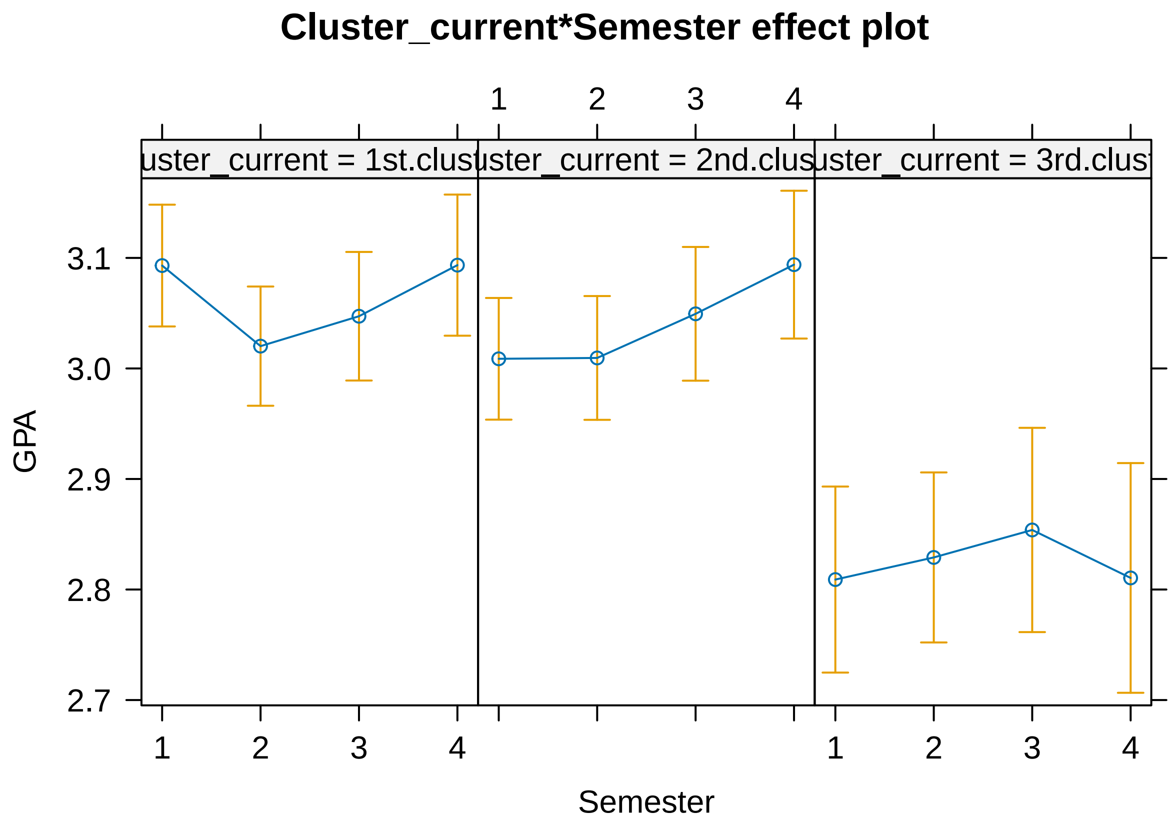
Based on the AIC and the BIC criterium, the inclusion of the Time variable into the random variable instead of the correlation structure is not beneficial. A model in which Time is included in both the correlation structure in addition to being a random effect isn't feasible, because in such a case the coeffecient matrix isn't invertible anymore.

page 7

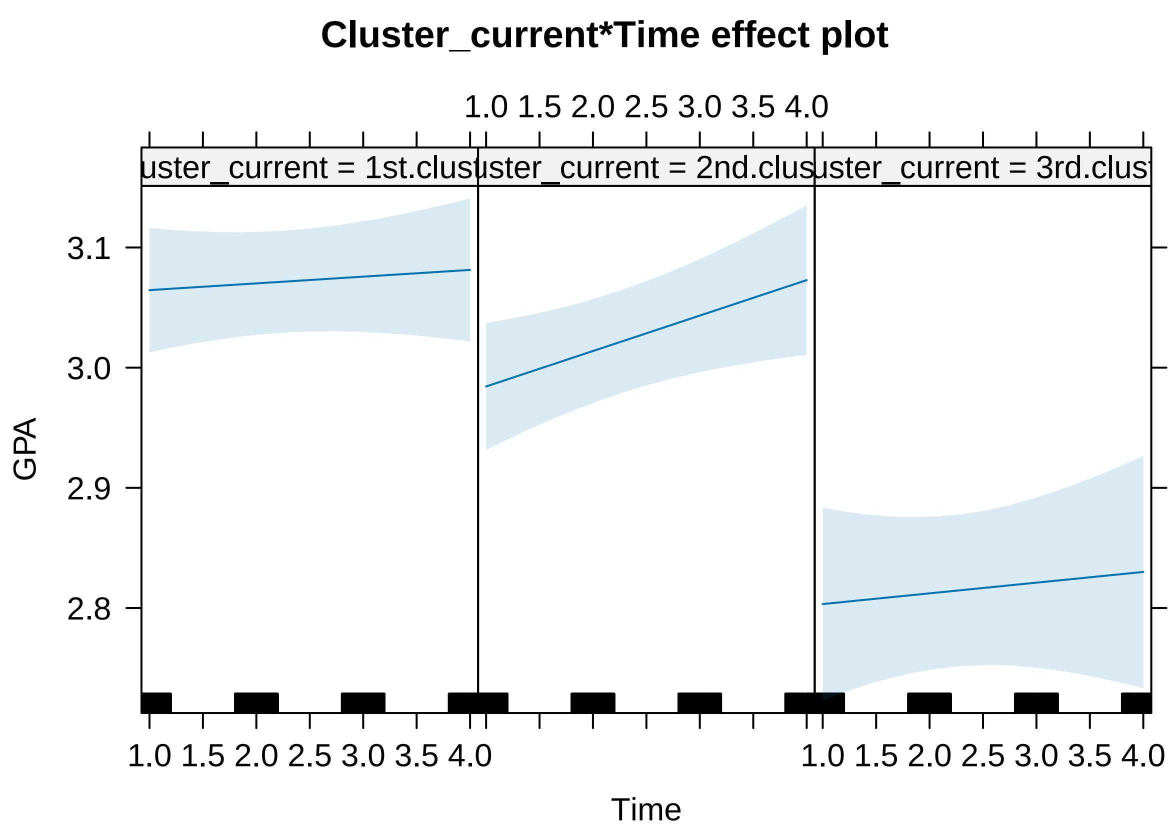
```
plot(effect("Semester*Cluster_current", full.model.ARMA11, robust= "hc3", correction = "Sidak"))
```




```
plot(effect("Cluster_current*Semester", small.model.ARMA11, robust= "hc3", correction = "Sidak"))
```

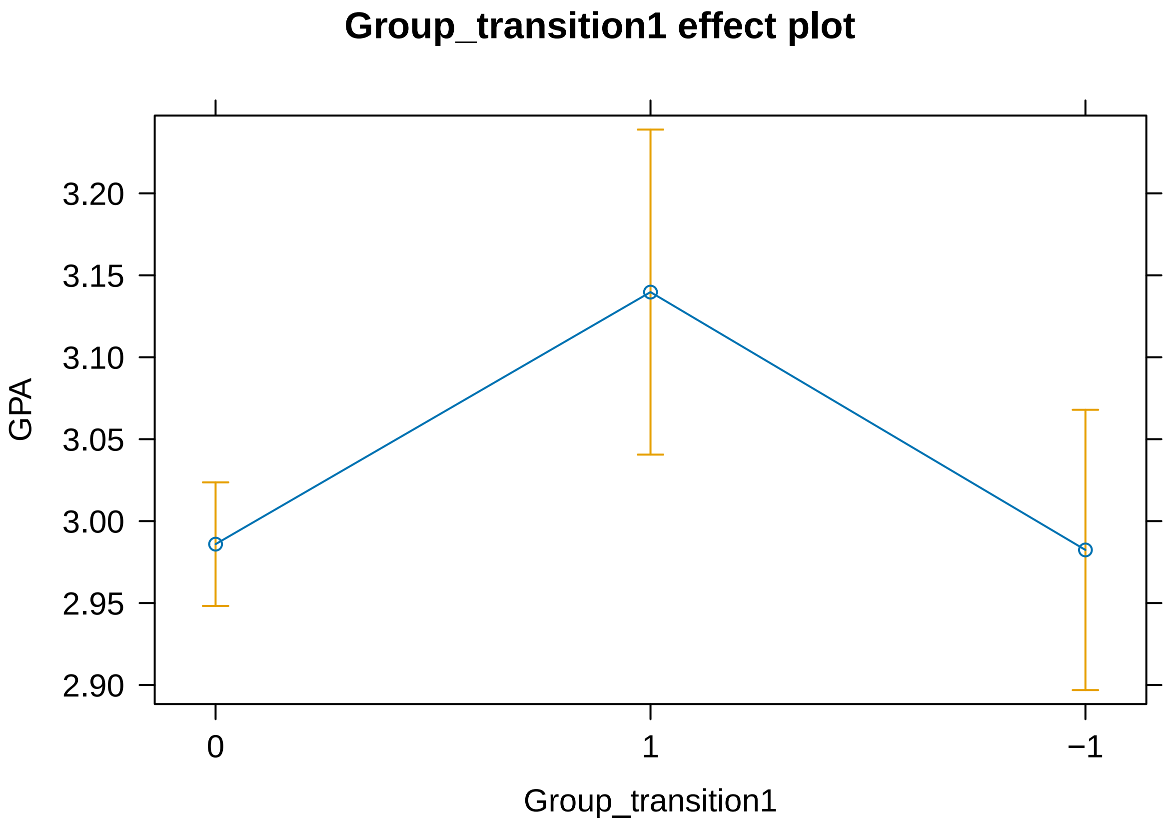



```
plot(effect("Cluster_current*Time", time.slope.ARMA11, robust= "hc3", correction = "Sidak"))
```

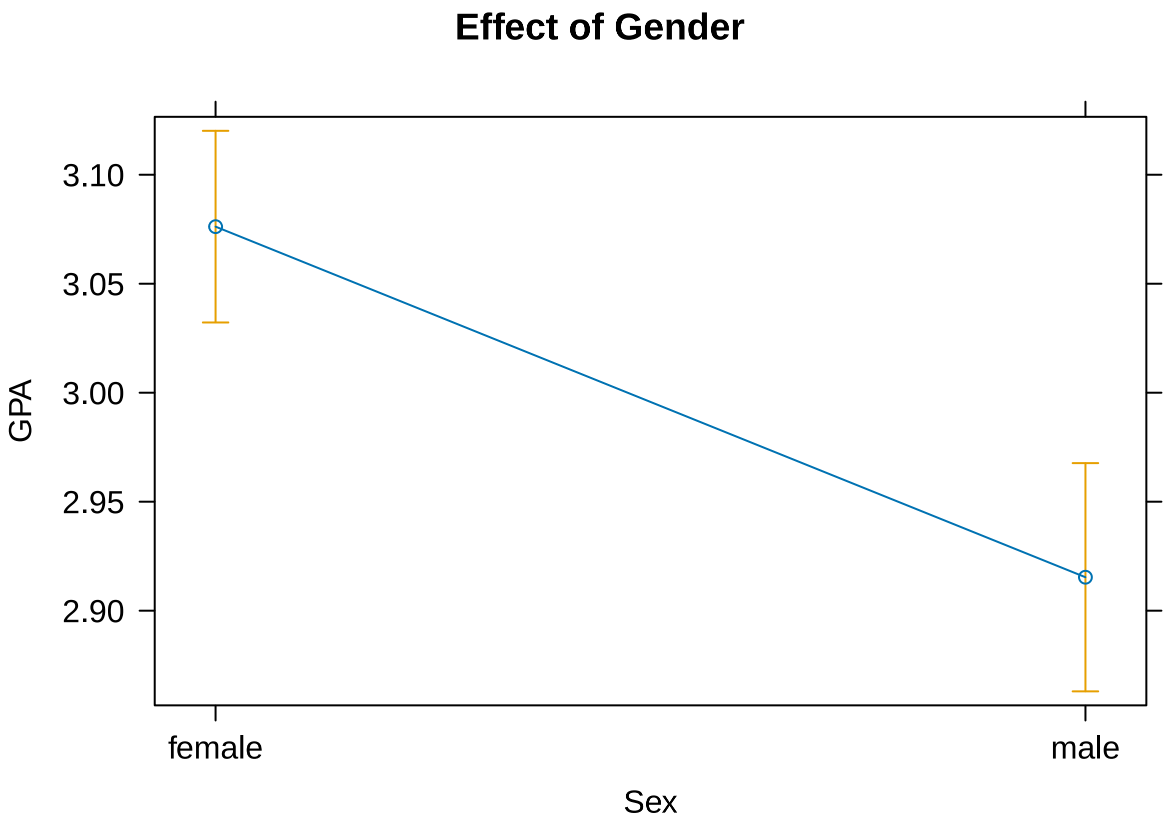


```r
plot(effect("Group_transition1", pseudo.trajectory.ARMA11, robust= "hc3", correction = "Sidak"))
```

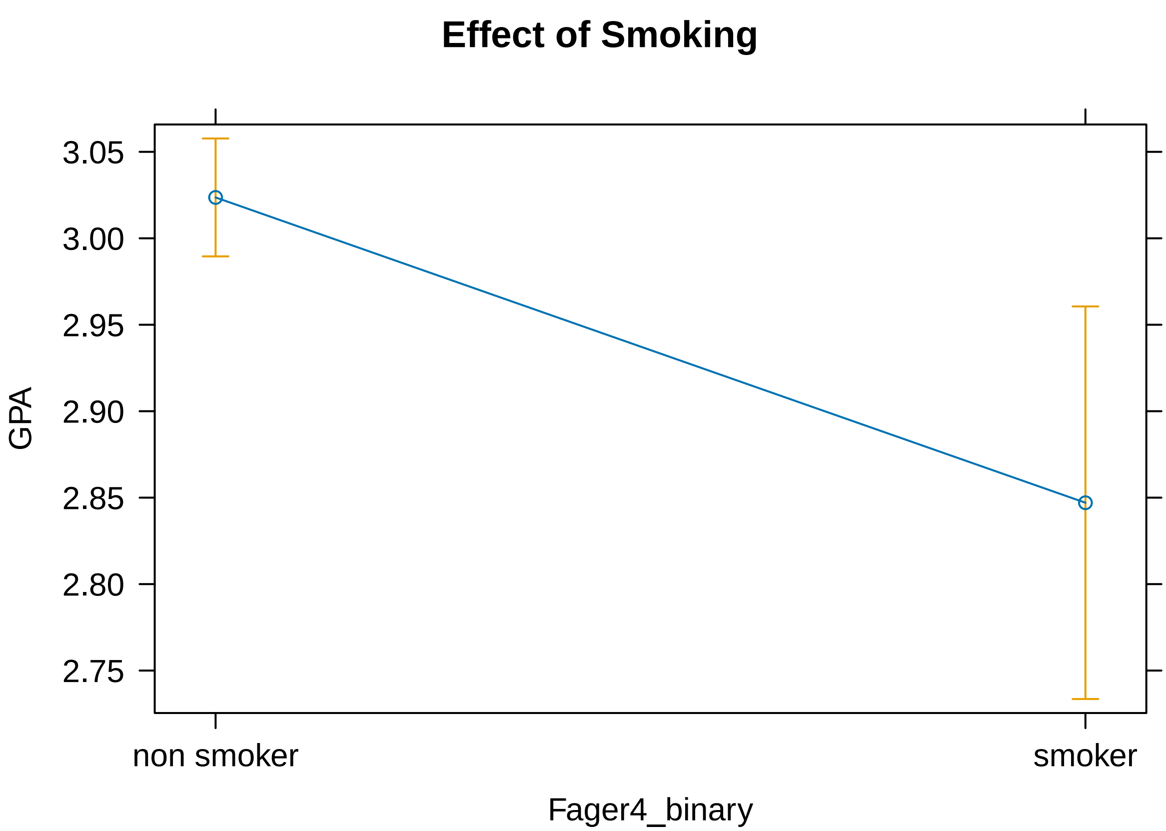
```
## NOTE: Group_transition1 is not a high-order term in the model
```



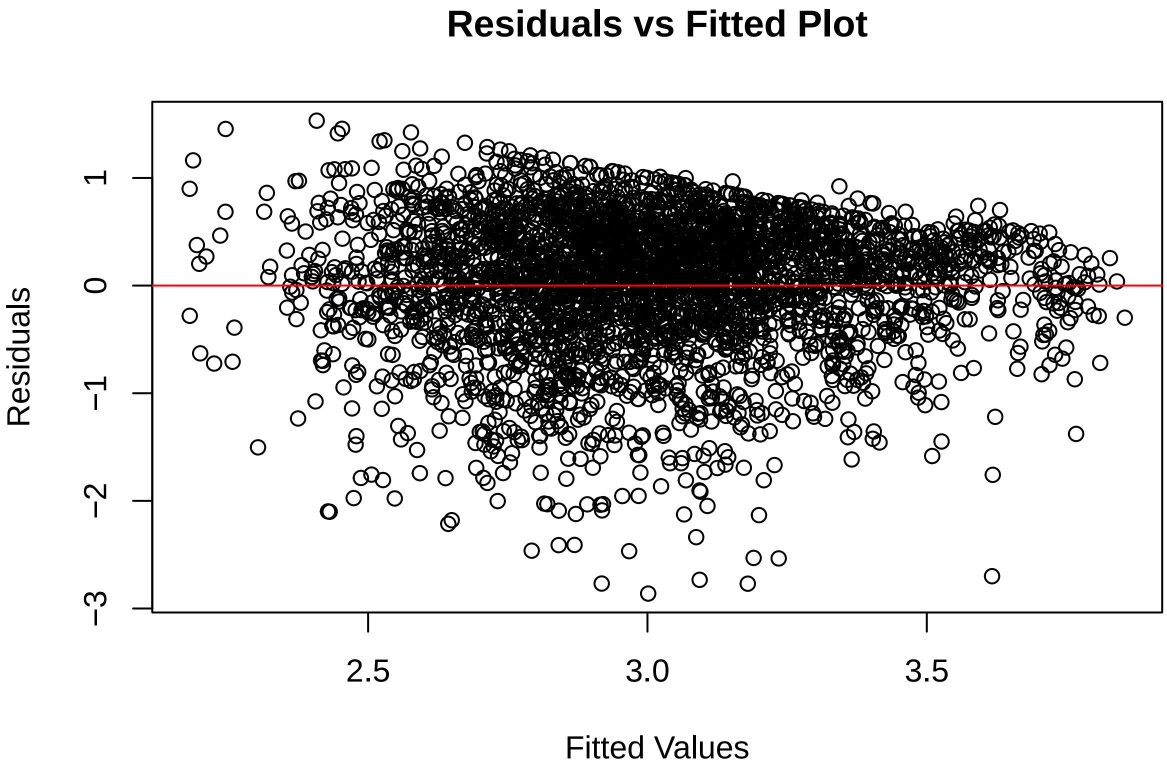

```r
plot(effect("Sex", full.model.ARMA11, robust= "hc3", correction = "Sidak"), main = "Effect of Gender")
```



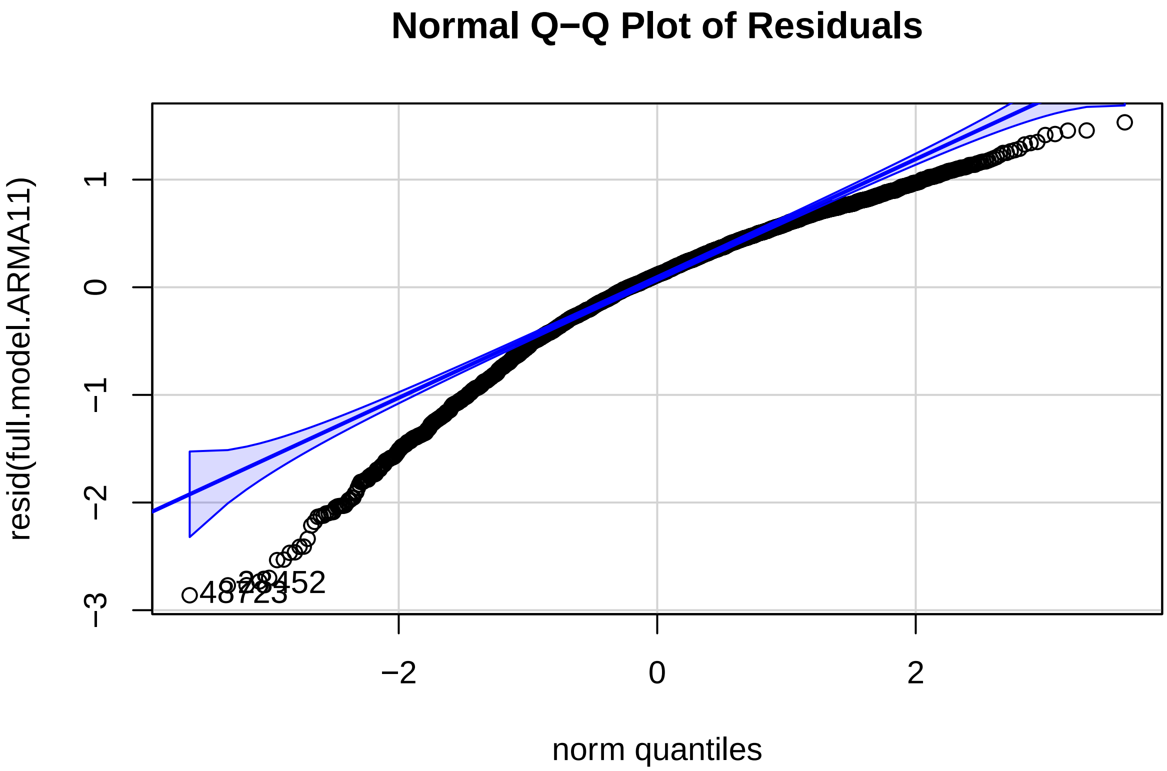

```
plot(effect("Fager4_binary", full.model.ARMA11, robust= "hc3", correction = "Sidak"), main = "Effect of
```



```r
plot(fitted(full.model.ARMA11), resid(full.model.ARMA11),
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Plot")
abline(h = 0, col = "red")
```

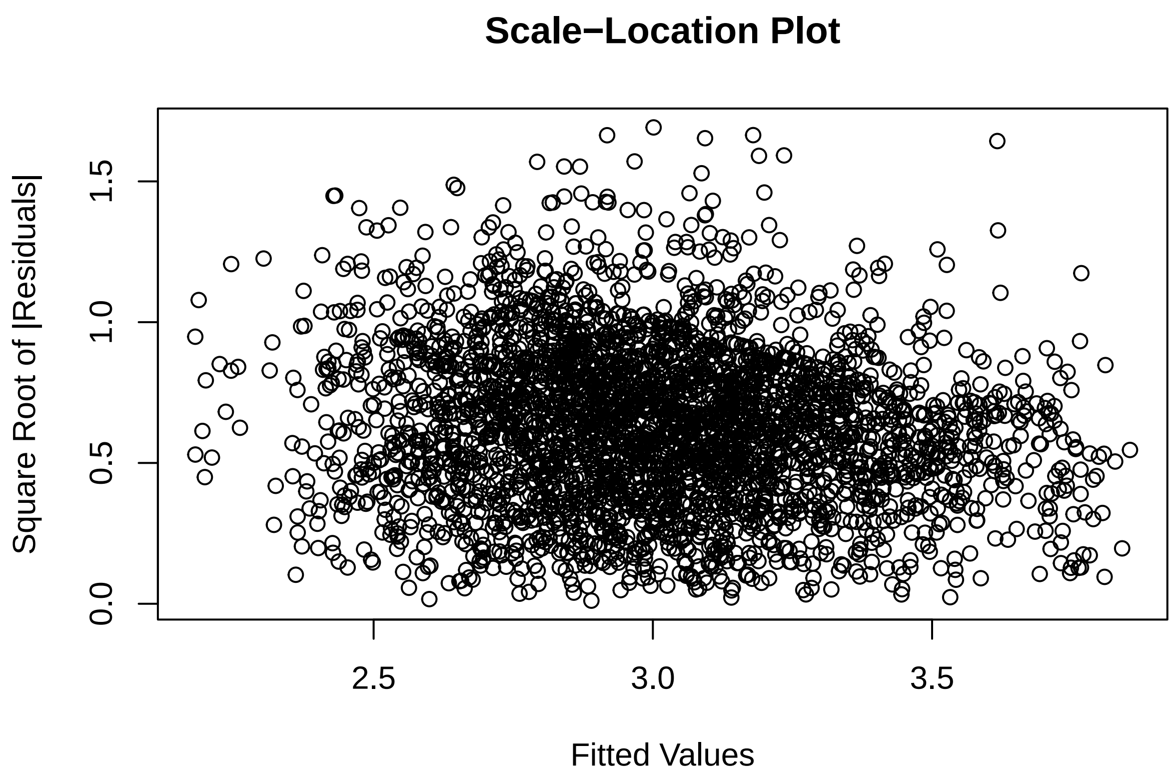



```r
qqPlot(resid(full.model.ARMA11),
       main = "Normal Q-Q Plot of Residuals")
```

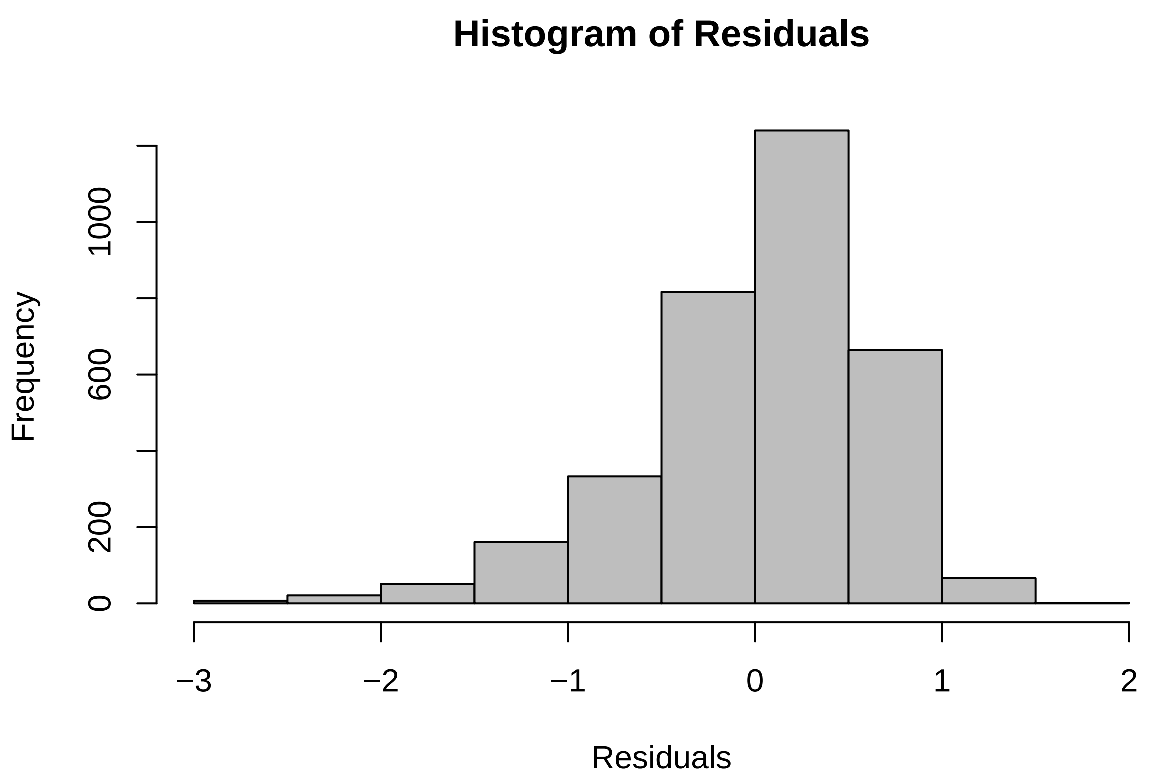



```
## 48723 28452
##  1415   391
```

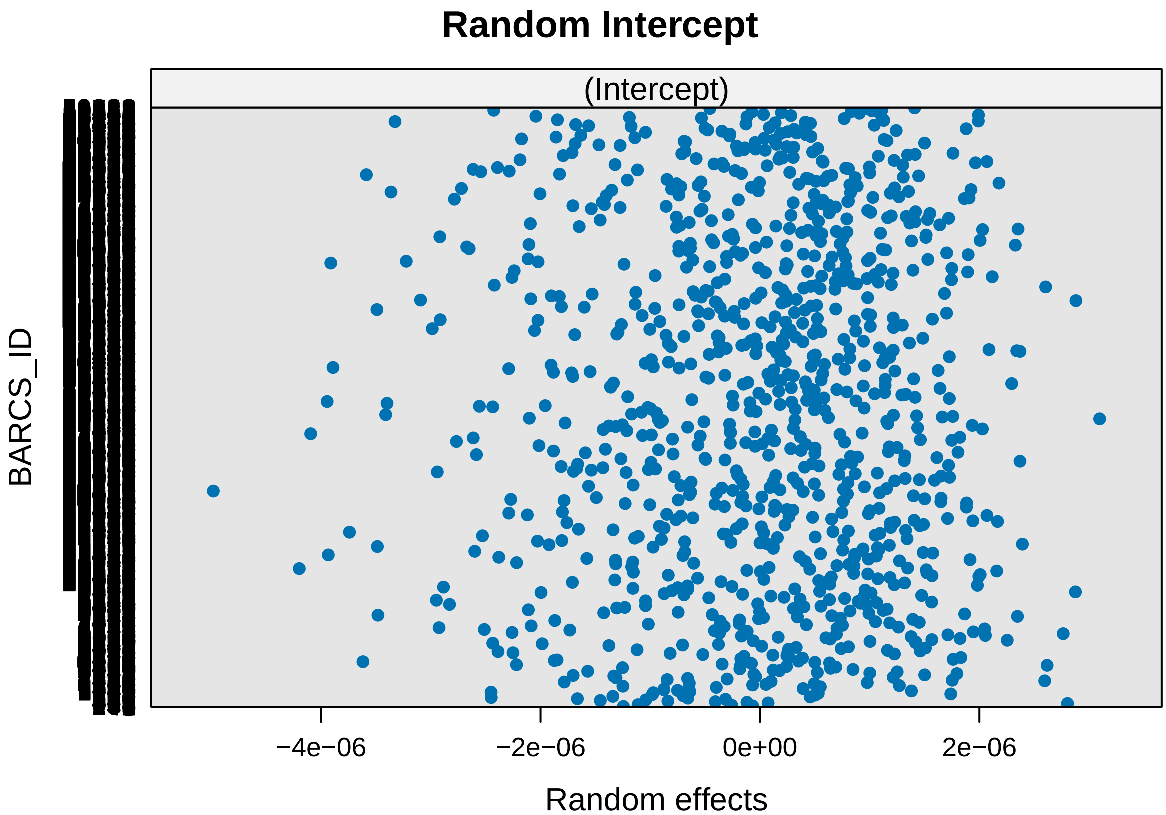
```
sqrt_abs_resid <- sqrt(abs(resid(full.model.ARMA11)))
plot(fitted(full.model.ARMA11), sqrt_abs_resid,
     xlab = "Fitted Values",
     ylab = "Square Root of |Residuals|",
     main = "Scale-Location Plot")
```




```r
hist(resid(full.model.ARMA11),
     main = "Histogram of Residuals",
     xlab = "Residuals",
     breaks = 10,
     col = "gray")
```



```r
plot(ranef(full.model.ARMA11),
       main = "Random Intercept")
```




```r
dim(ranef(full.model.ARMA11))
```

```
## [1] 1004    1
```

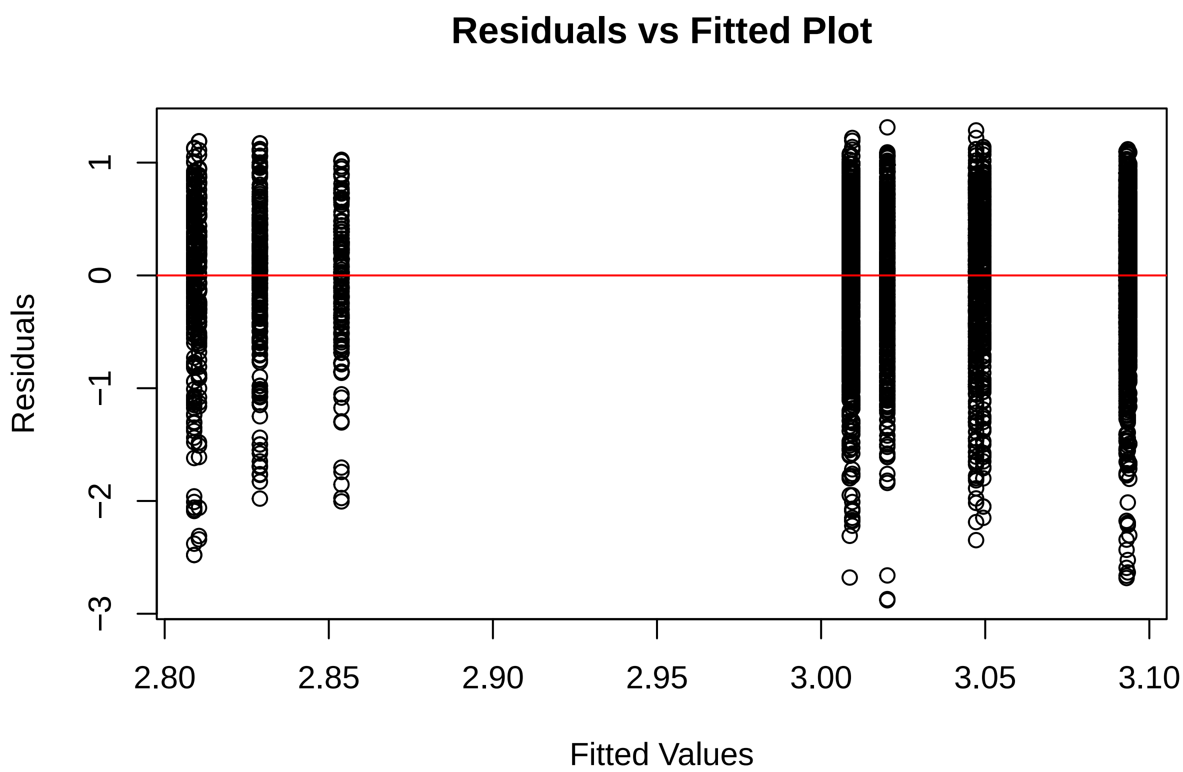
```r
print(range(ranef(full.model.ARMA11)))
```

```
## [1] -4.982805e-06  3.096283e-06
```

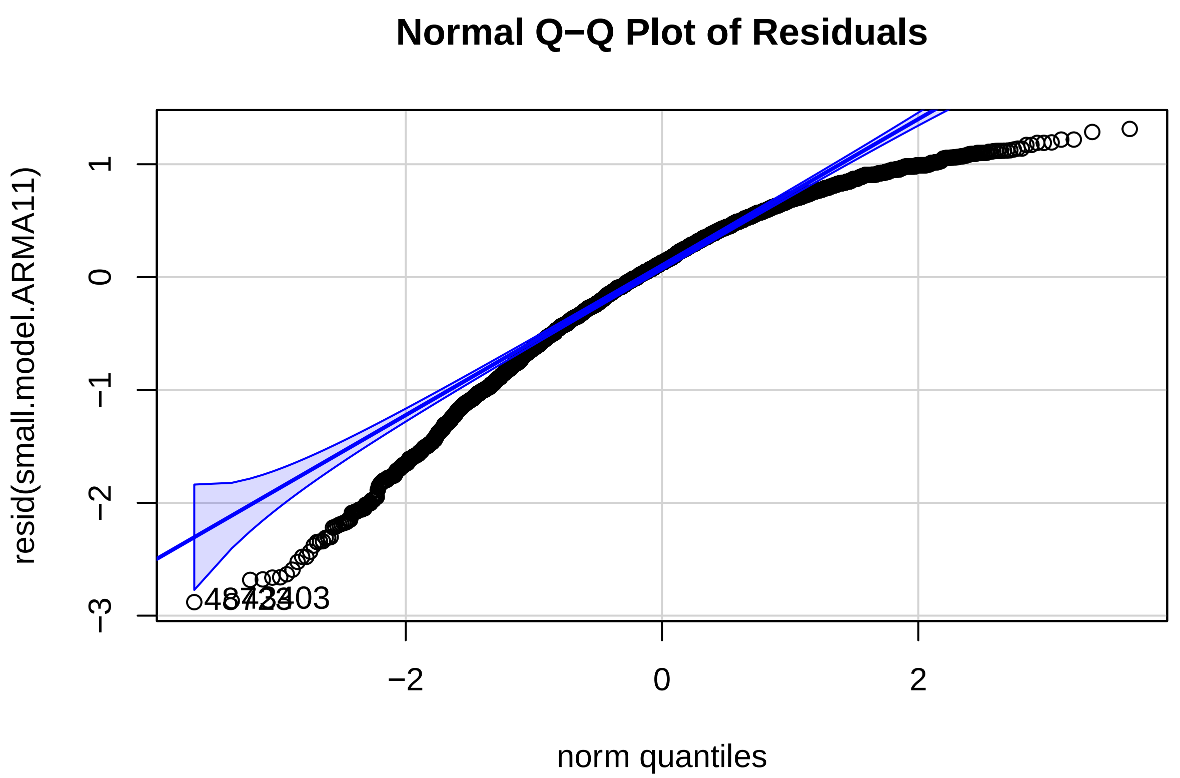
```r
print(r.squaredGLMM(full.model.ARMA11)) ####TERRRIBLE!!!!!!!!!!!!
```

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.

##            R2m       R2c
## [1,] 0.1778163 0.1778177
```

```r
plot(fitted(small.model.ARMA11), resid(small.model.ARMA11),
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Plot")
abline(h = 0, col = "red")
```

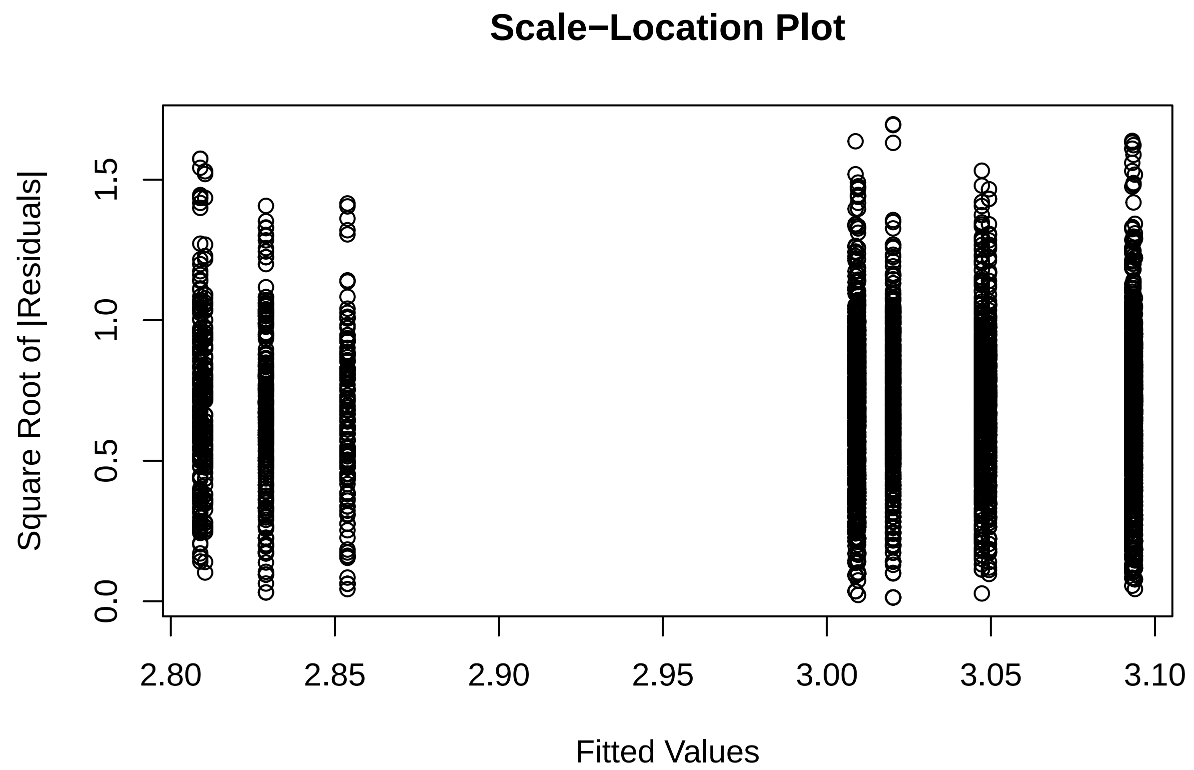



```r
qqPlot(resid(small.model.ARMA11),
       main = "Normal Q-Q Plot of Residuals")
```

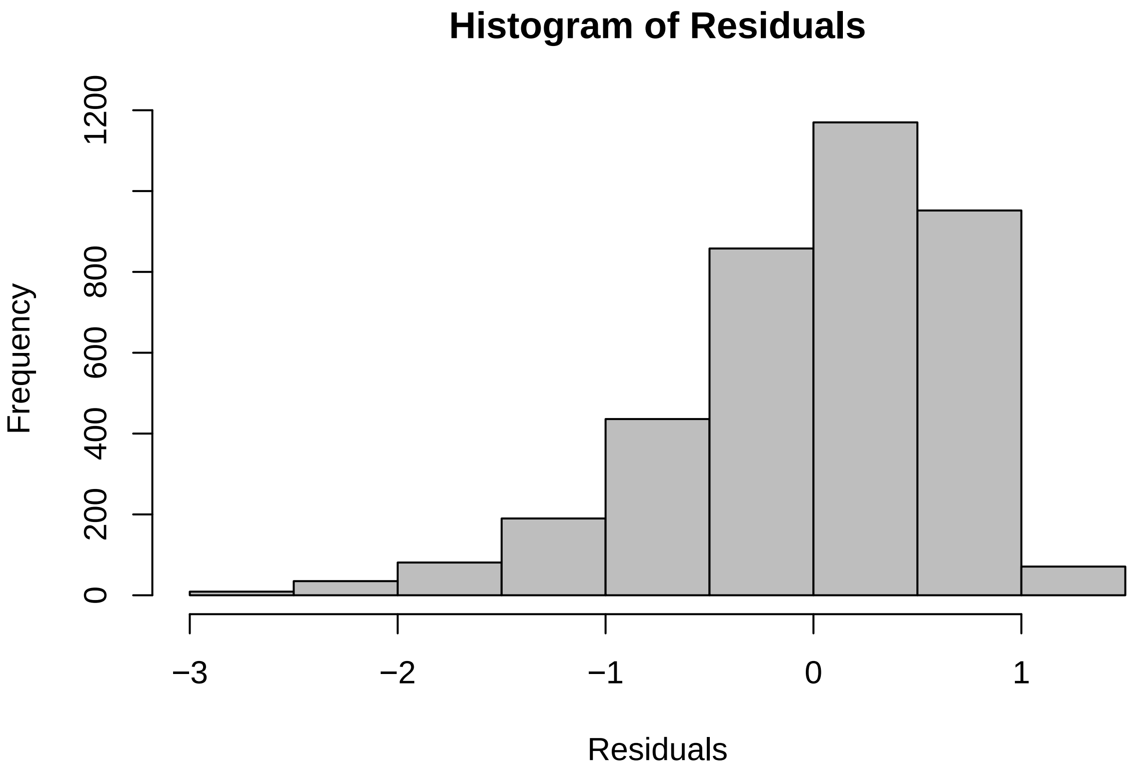



```
## 48723 43403
##  1415  1409
```

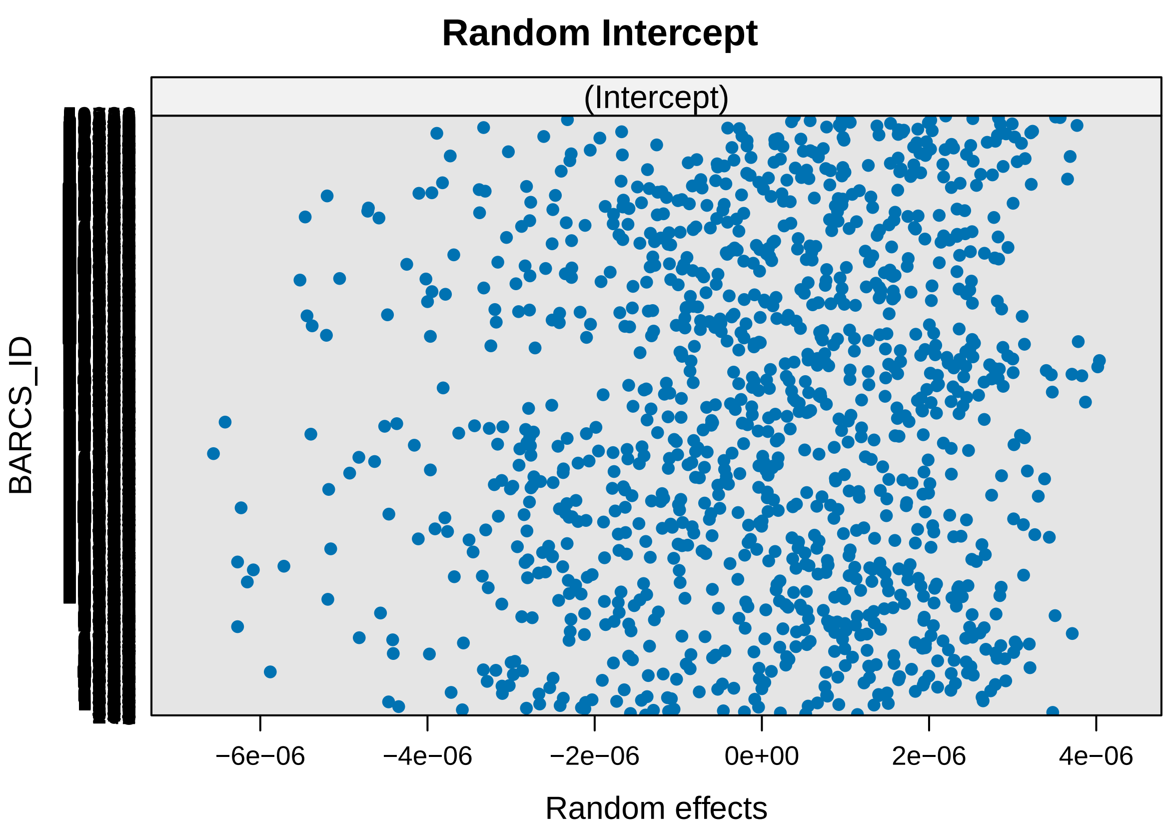
```r
sqrt_abs_resid <- sqrt(abs(resid(small.model.ARMA11)))
plot(fitted(small.model.ARMA11), sqrt_abs_resid,
     xlab = "Fitted Values",
     ylab = "Square Root of |Residuals|",
     main = "Scale-Location Plot")
```




```r
hist(resid(small.model.ARMA11),
     main = "Histogram of Residuals",
     xlab = "Residuals",
     breaks = 10,
     col = "gray")
```




```r
plot(ranef(small.model.ARMA11),
       main = "Random Intercept")
```



```r
print(dim(ranef(small.model.ARMA11)))
```

```
## [1] 1140    1
```

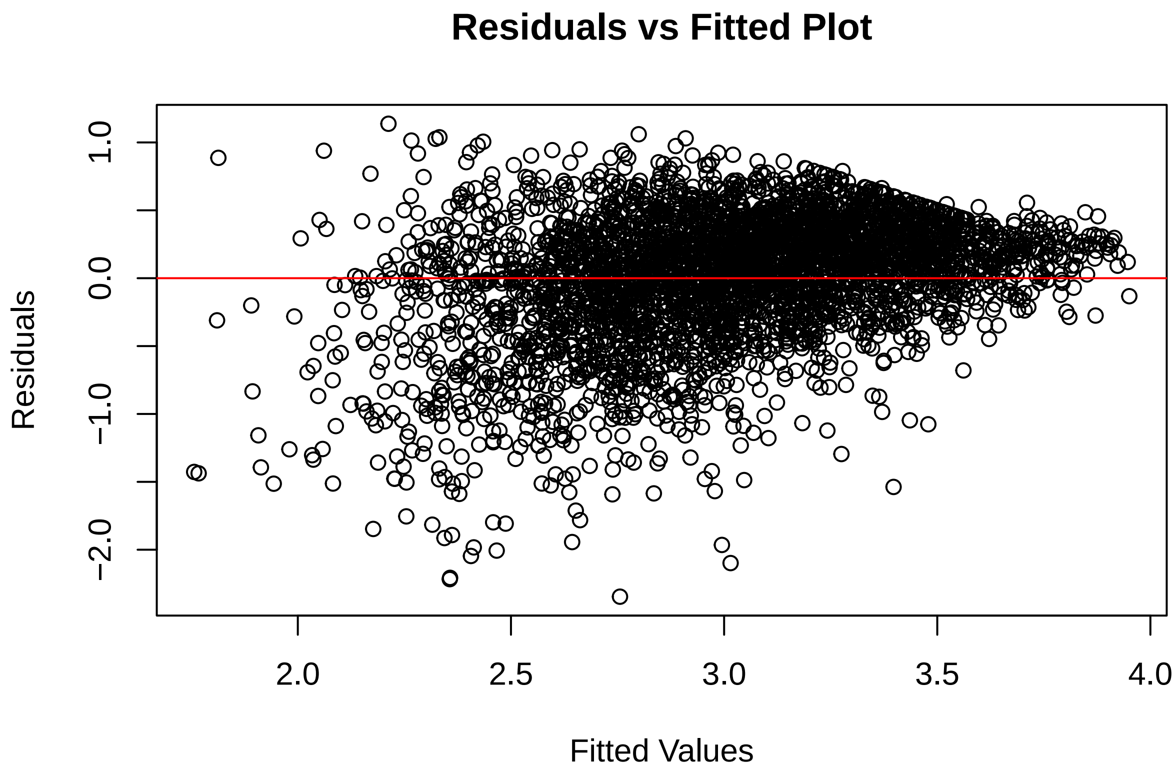
```r
print(range(ranef(small.model.ARMA11)))
```

```
## [1] -6.560243e-06  4.037665e-06
```

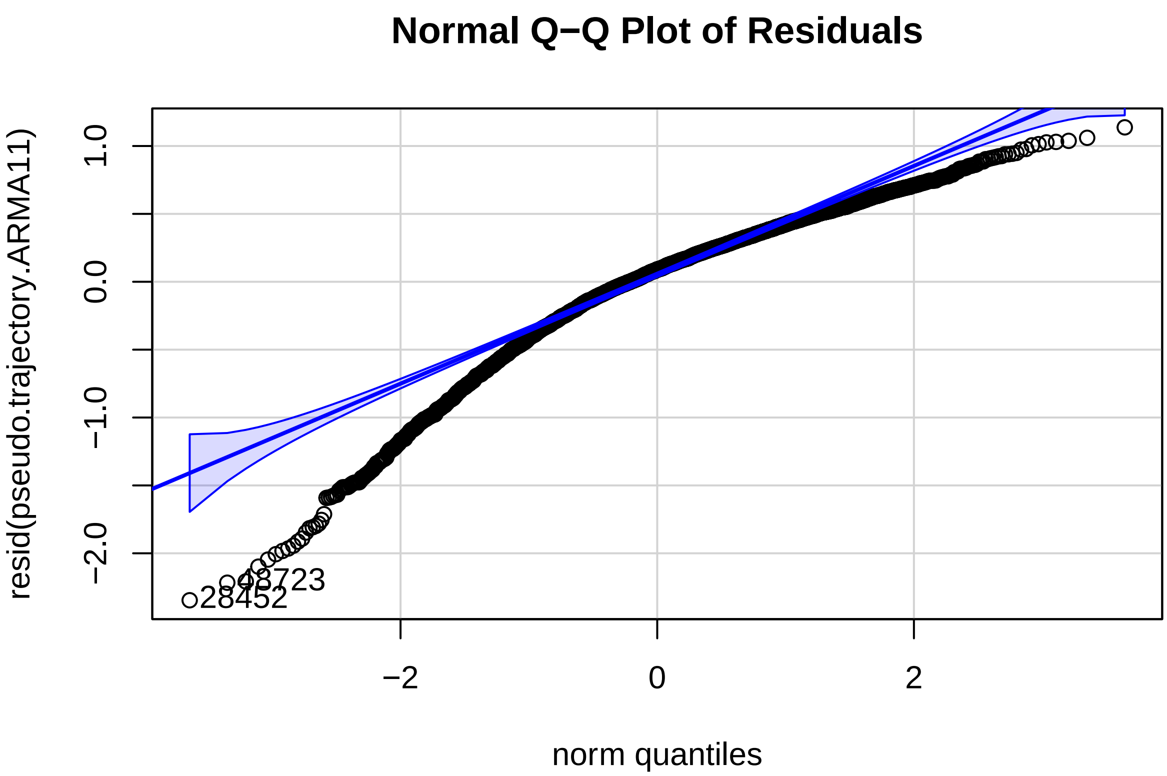
```r
print(r.squaredGLMM(small.model.ARMA11)) ####TERRRIBLE!!!!!!!!!!!!
```

```
##             R2m        R2c
## [1,] 0.01678364 0.01678612
```

```r
plot(fitted(pseudo.trajectory.ARMA11), resid(pseudo.trajectory.ARMA11),
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Plot")
abline(h=0, col = "red")
```

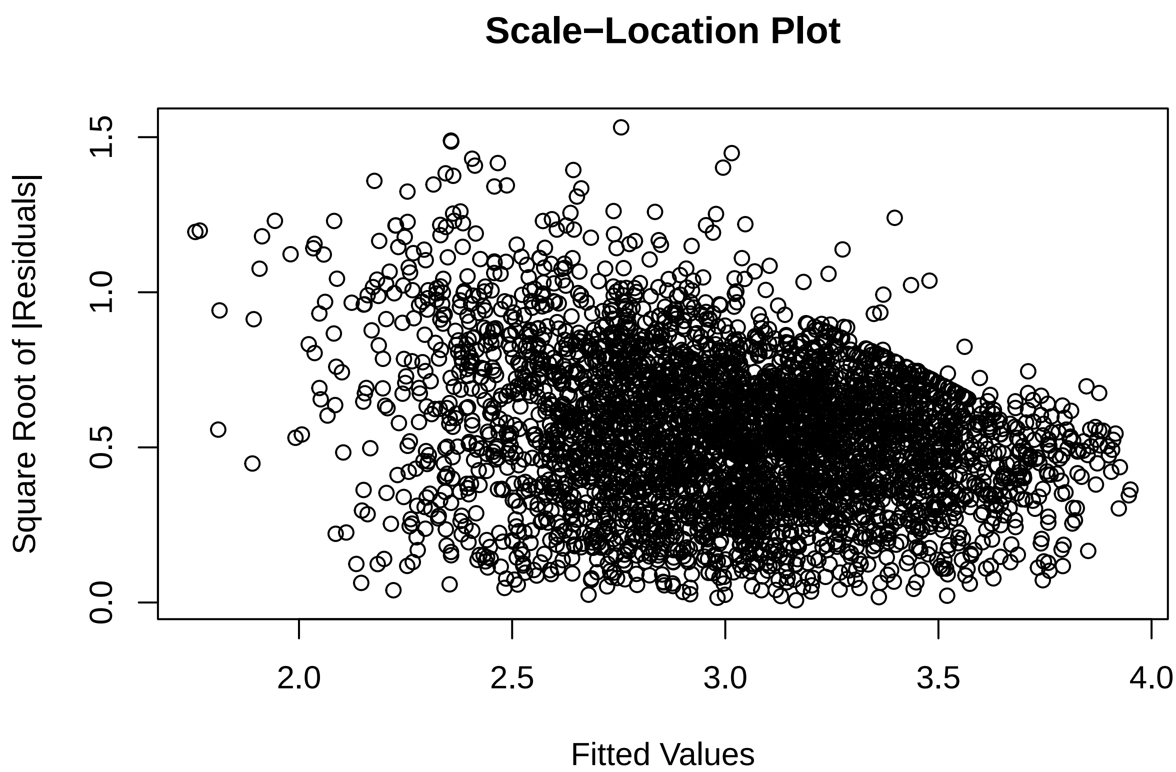



```r
qqPlot(resid(pseudo.trajectory.ARMA11),
       main = "Normal Q-Q Plot of Residuals")
```

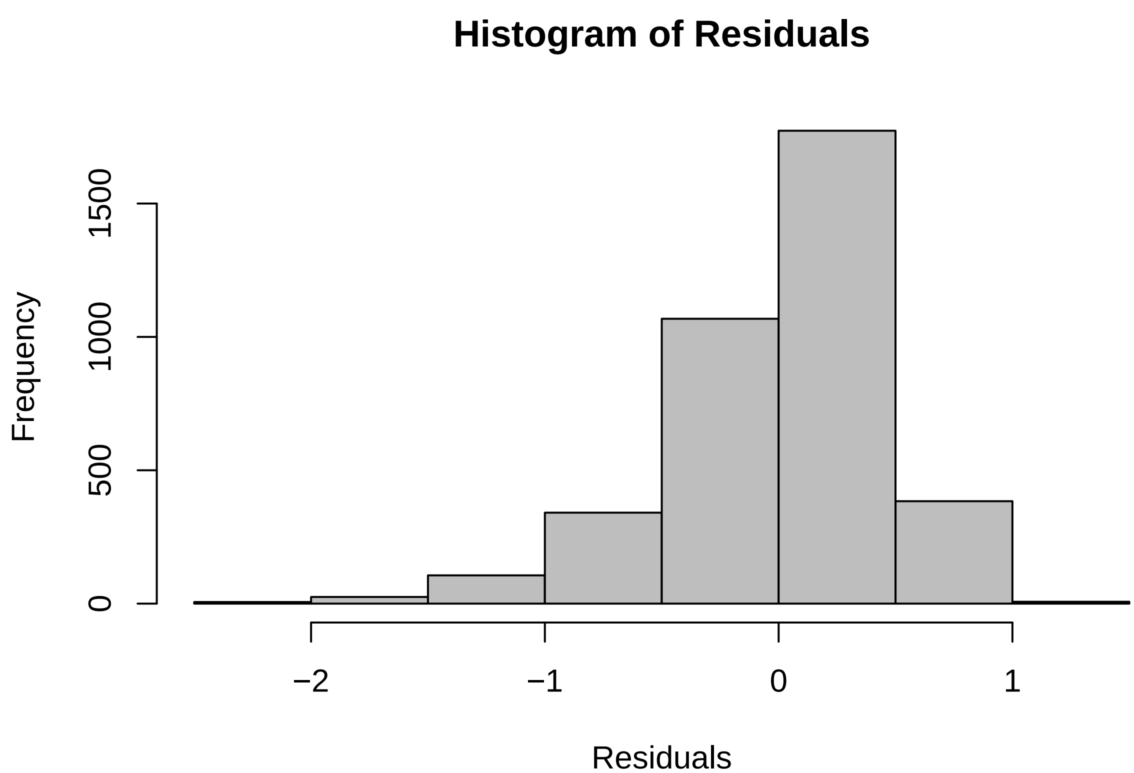



```
## 28452 48723
##   391  1415
```

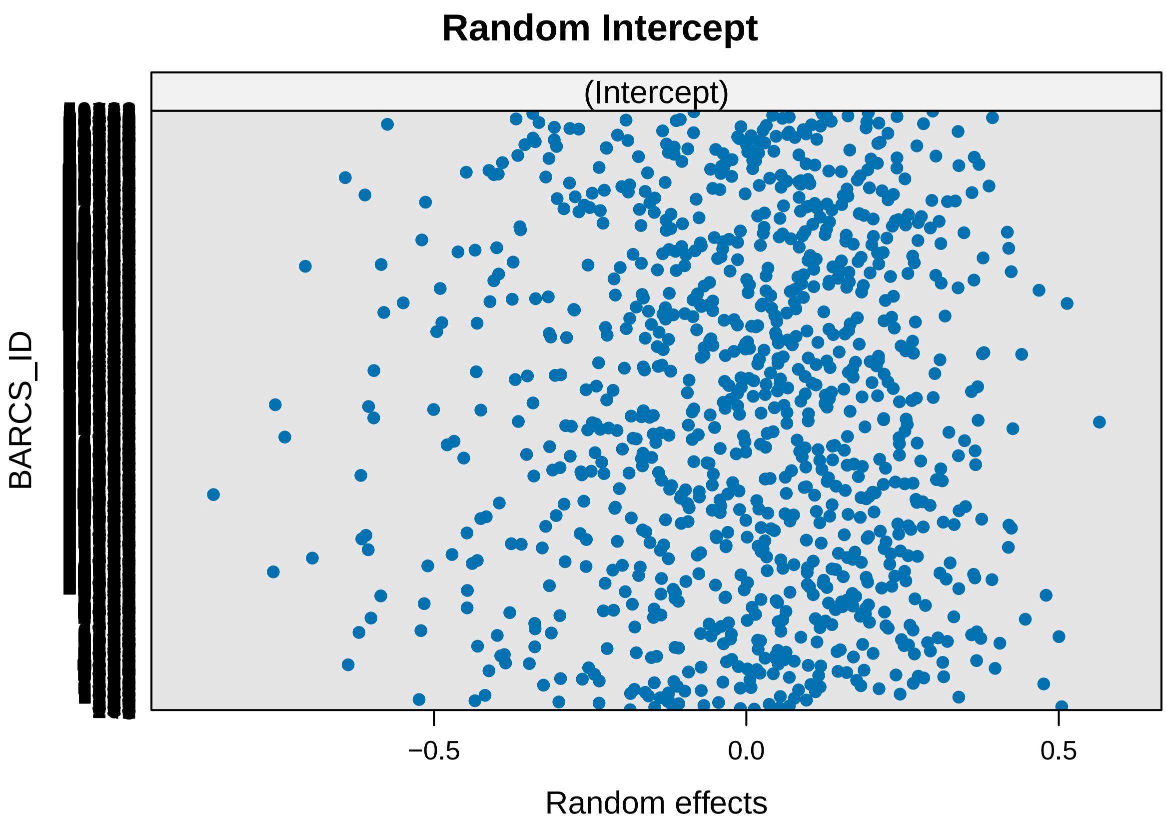
```
sqrt_abs_resid <- sqrt(abs(resid(pseudo.trajectory.ARMA11)))
plot(fitted(pseudo.trajectory.ARMA11), sqrt_abs_resid,
     xlab = "Fitted Values",
     ylab = "Square Root of |Residuals|",
     main = "Scale-Location Plot")
```




```r
hist(resid(pseudo.trajectory.ARMA11),
     main = "Histogram of Residuals",
     xlab = "Residuals",
     breaks = 10,
     col = "gray")
```



```r
plot(ranef(pseudo.trajectory.ARMA11),
       main = "Random Intercept")
```




```r
dim(ranef(pseudo.trajectory.ARMA11))
```

```
## [1] 1000    1
```

```r
print(range(ranef(pseudo.trajectory.ARMA11)))
```

```
## [1] -0.8528631  0.5650472
```

```r
print(r.squaredGLMM(pseudo.trajectory.ARMA11)) ### somewhat better
```

```
##            R2m       R2c
## [1,] 0.1835358 0.4254371
```

the pseudo trajectory is doing a lot better here.

```r
full.model.gls <- gls(GPA ~ 1 + Sex + Age1stround + SATMath + SATVerbal + SATWriting + Fager4_binary + I
anova(full.model.gls, full.model.ARMA11)
```

```
##                    Model df      AIC      BIC    logLik   Test      L.Ratio
## full.model.gls         1 25 5562.076 5714.911 -2756.038
## full.model.ARMA11      2 26 5564.076 5723.025 -2756.038 1 vs 2 3.089681e-06
##                    p-value
## full.model.gls
## full.model.ARMA11   0.9986
```

```r
small.model.gls <- gls(GPA ~ Semester * Cluster_current, data = data.file.long, na.action = na.exclude,
anova(small.model.gls, small.model.ARMA11)
```

```
## Warning in nlme::anova.lme(object = small.model.gls, small.model.ARMA11):
## fitted objects with different fixed effects. REML comparisons are not
## meaningful.
```

```
##                     Model df      AIC      BIC    logLik   Test      L.Ratio
## small.model.gls         1 15 6445.394 6538.996 -3207.697
## small.model.ARMA11      2 16 6447.394 6547.236 -3207.697 1 vs 2 4.999507e-06
##                     p-value
## small.model.gls
## small.model.ARMA11  0.9982
```

```
pseudo.trajectory.gls <- gls(GPA ~ 1 + Fager4_binary + FH_binary + Sex + Cluster_SEM1 + Semester + Age1s
anova(pseudo.trajectory.ARMA11, pseudo.trajectory.gls)
```

```
##                           Model df      AIC      BIC    logLik   Test   L.Ratio
## pseudo.trajectory.ARMA11      1 34 6147.618 6358.781 -3039.809
## pseudo.trajectory.gls         2 33 6145.947 6350.899 -3039.974 1 vs 2 0.3287907
##                           p-value
## pseudo.trajectory.ARMA11
## pseudo.trajectory.gls      0.5664
```

RIP lol. PT isn't as clear probably due to the fact that the Random effects actually do something

```
vif(full.model.gls)
```

```
##                                 GVIF Df GVIF^(1/(2*Df))
## Sex                         1.145947  1        1.070489
## Age1stround                 1.036041  1        1.017861
## SATMath                     1.910025  1        1.382037
## SATVerbal                   2.729602  1        1.652151
## SATWriting                  3.252019  1        1.803336
## Fager4_binary               1.048463  1        1.023945
## FH_binary                   1.039356  1        1.019488
## STAI_SELF_Total             1.947916  1        1.395678
## BDI_SELF_Total              1.993410  1        1.411882
## Parental_SES                1.081658  1        1.040028
## Semester                   11.867058  3        1.510279
## Cluster_current             4.657988  2        1.469094
## Semester:Cluster_current   37.907462  6        1.353818
```

```
vif(pseudo.trajectory.gls)
```

```
##                                   GVIF Df GVIF^(1/(2*Df))
## Fager4_binary                 1.078203  1        1.038366
## FH_binary                     1.041129  1        1.020357
## Sex                           1.155155  1        1.074782
## Cluster_SEM1                  2.846460  2        1.298902
## Semester                     18.208709  3        1.621984
## Age1stround                   1.041380  1        1.020480
## SATMath                       1.908960  1        1.381651
## SATVerbal                     2.749887  1        1.658278
## SATWriting                    3.246544  1        1.801817
## STAI_SELF_Total               1.954151  1        1.397909
## BDI_SELF_Total                2.003921  1        1.415599
## Parental_SES                  1.103742  1        1.050591
## Group_transition1             2.702576  2        1.282167
## Cluster_SEM1:Semester        33.305212  6        1.339294
```

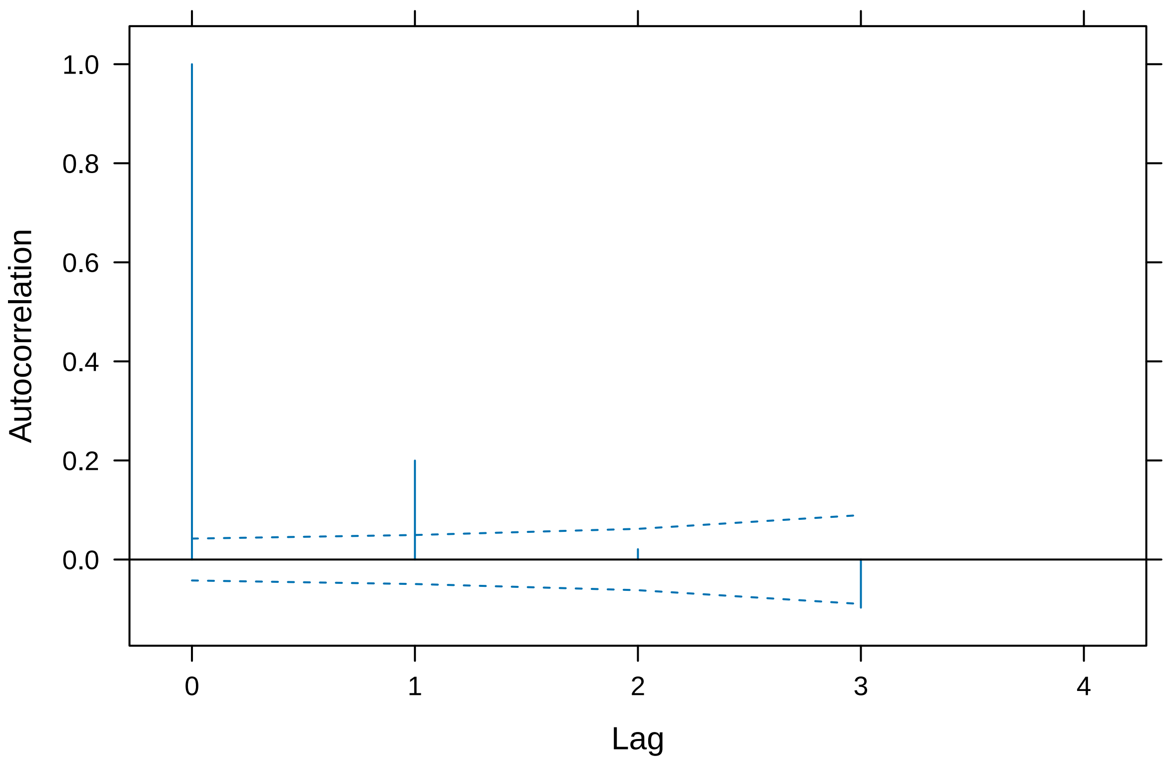
```
## Semester:Group_transition1  8.317589  6        1.193071
```

Semester + Interaction have high / very high multicollinearity, the other fixed effects seem okay. For the pt, the Semester and Cluster categorization in Semester are highly multicollinear, as well as the group transition variable having a high collinear value. This makes of course sense as they are related.

```
(ACF.pt <- ACF(pseudo.trajectory.ARMA11, maxLag=4))
```

```
##   lag         ACF
## 1   0  1.00000000
## 2   1  0.19978412
## 3   2  0.02089272
## 4   3 -0.09718593
## 5   4         NaN
```

```
plot(ACF.pt, alpha=0.01)
```




```
#stat.ethz.ch/R-manual/R-devel/library/nlme/html/ACF.lme.html
```

expected. There is an ARMA(1,1) Cov structure in the model assumption, a lag of 1 should be showcassed here and it is.

```
shapiro.test(resid(full.model.ARMA11))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(full.model.ARMA11)
## W = 0.95119, p-value < 2.2e-16
```

```
shapiro.test(resid(small.model.ARMA11))
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  resid(small.model.ARMA11)
## W = 0.948, p-value < 2.2e-16
```

```r
shapiro.test(resid(pseudo.trajectory.ARMA11))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(pseudo.trajectory.ARMA11)
## W = 0.94165, p-value < 2.2e-16
```

```r
shapiro.test(resid(time.slope.ARMA11)) ## additional tests are somewhat trivial
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(time.slope.ARMA11)
## W = 0.94257, p-value < 2.2e-16
```

strong evidence that the data is not normally distributed