Model Specification for Autism study

Longitudinal Data: Autism study

General Model Specification

- M-1: model M-1 includes the fixed effects of age, age-squared, SICD group, and SICD group by age interaction and the SICD by age-squared interaction
- Model M-1 also includes three random effects associated with each child: a random intercept, a random age effect, and a random age-squared effect

Let y_{ij} be an individual response, $VSAE_{ij}$ on child i at the j-th visit (j = 1,2,3,4,5) corresponding to ages $(2, 3, 5, 9, 13)$. A general random coefficient model is written as follows:

$$M_1 : y_{ij} = \beta_0+\beta_1 x_{1ij}+\beta_2 x_{2ij}+\beta_3 x_{3i}+\beta_4 x_{4i}+\beta_5 x_{1ij}\times x_{3i}+\beta_6 x_{1ij}\times x_{4i}+\beta_7 x_{2ij}\times x_{3i}+\beta_8 x_{2ij}\times x_{4i}+u_{0i}+u_{1i}\times x_{1ij}+u_{2i}\times x_{2ij}+\epsilon_{ij}$$

- x_{1ij} represents the value of Age minus 2 so that VSAE scores at other ages are comparable to that at age 2. x_{2ij} represents the squared of x_{1ij}
- $$x_{3i} = \begin{cases} 1 & \text{SICDEGP is 1;} \\ 0 & \text{otherwise.} \end{cases}$$
-
- $$x_{4i} = \begin{cases} 1 & \text{SICDEGP is 2;} \\ 0 & \text{otherwise.} \end{cases}$$
- $\beta_0 - \beta_8$ represent the fixed effects associated with the intercept, the covariates and the interaction terms in the model
- β_0 represents the mean predicted VSAE score for children at 2 years of age for the SICD group 3
- β_1 and β_2 represent the fixed effects of age and age-squared for the SICD group 3
- β_3 and β_4 represent the fixed effects associated with the SICD group 1 and 2
- In particular, β_3 represents the difference in the intercept for the SICD group 1 and the reference group SICD at 3
- Similar interpretation for β_4
- β_5 and β_6 represent the differences in the linear effect of age between SICD group 1 and 2 and the linear effect of age in the SICD group 3 respectively
- β_7 and β_8 represent the differences in the quadratic effect of age between SICD group 1 and 2 and the quadratic effect of age in the SICD group 3 respectively
- u_{0i}, u_{1i}, u_{2i} represent the random effects associated with the intercept, linear effect of age and the quadratic effect of age for child i.
- Assume the distribution of the random vector of three random effects $\mathbf{u}_i$ associated with child i as multivariate normal. That is,

$$\mathbf{u}_i = \begin{pmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{pmatrix} \sim N_3(\mathbf{0}, \mathbf{D}).$$

Here the variance-covariance matrix of the three random effects is defined as

$$\mathbf{D} = \begin{pmatrix} \sigma^2_{int} & \sigma_{int,age} & \sigma_{int,age^2} \\ \sigma_{int,age} & \sigma^2_{age} & \sigma_{age,age^2} \\ \sigma_{int,age^2} & \sigma_{age,age^2} & \sigma^2_{age^2} \end{pmatrix},$$

where $\sigma^2_{int}, \sigma^2_{age}, \sigma^2_{age^2}$ are the variances of the random intercept, the random effect of age and the random effect of the age-sqaured. The off-diagonals are the covariances of the pair-wise random effects.

- Finally, ϵ_{ij} is the residual associated with the observation at time j on child i. The residuals are assumed to independent and identically distirbuted conditional on the random effects

$$\epsilon_{ij} \sim N(0, \sigma^2).$$

- Residuals are also assumed to be independent of the random effects

Hierarchical Model Specification

- Level 1 Model (Time)

$$y_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_{2i}x_{2ij} + \epsilon_{ij}$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$.

- Level 1 model has the child specific quadratic regressions of VSAE of age and age-squared
- The random intercept β_{0i}, the random linear and quadratic effects of age β_{1i}, β_{2i} vary between children
- $\beta_{0i}, \beta_{1i}, \beta_{2i}$ in the level 1 model depend on fixed effects associated with level 2 covariates and random effetcs
- Level 2 model (child)

$$\begin{aligned} \beta_{0i} &= \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i} + u_{0i} \\ \beta_{1i} &= \beta_1 + \beta_5 x_{3i} + \beta_6 x_{4i} + u_{1i} \\ \beta_{2i} &= \beta_2 + \beta_7 x_{3i} + \beta_8 x_{4i} + u_{2i} \end{aligned}$$

- where,

$$\mathbf{u}_i = \begin{pmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{pmatrix} \sim N_3(\mathbf{0}, \mathbf{D}).$$

- The intercept β_{0i} for child i depends on the fixed overall intercept β_0, fixed effects (β_3, β_4) of the child level covariates, indicators of SICD group 1 (x_{3i}) and SICD group 2 (x_{4i}) and a random effect u_{0i} associated with child i
- The child specific linear effect of age β_{1i} depends on the overall fixed effect of age β_1, fixed effects (β_5, β_6) of indicators of SICD group 1 and SICD group 2, and a random effect u_{1i} associated with child i
- The child specific quadratic effect of age β_{2i} depends on the overall fixed effect of age-squared β_2, fixed effects (β_7, β_8) of indicators of SICD group 1 and SICD group 2, and a random effect u_{2i} associated with child i

Hypothesis Tests:

1. M-2: Fit a model without the random child-specific intercepts. The rationale for this model is that there was little variability in the VSAE scores at age 2 which can be attributed to random error, instead of between-subject variability.
2. M-2 implies child-specific predicted trajectories within a given level of SICD group have a common VSAE score at age 2
3. M-2.A: The random effects associated with the quadratic effect of age can be ommitted from the model M-1. This is equivalent to testing the following hypotheses:
4. $$H_0 : \mathbf{D} = \begin{pmatrix} \sigma^2_{age} & 0 \\ 0 & 0 \end{pmatrix}$$

   $$H_A : \mathbf{D} = \begin{pmatrix} \sigma^2_{age} & \sigma_{age,age^2} \\ \sigma_{age,age^2} & \sigma^2_{age^2} \end{pmatrix}.$$
5. REML based likelihood ratio test is used for testing this hypothesis. For large sample this test statistic follows a mixture of χ^2_1 and χ^2_2 distributions with equal weights 0.5.
6. The fixed effects associated with the age-squared by SICD group interaction can be excluded from the model

   $$H_0 : \beta_7 = \beta_8 = 0$$

   $$H_A : \beta_7 \neq 0 \ \text{ Or } \ \beta_8 \neq 0$$
7. ML based likelihood ratio test
8. The fixed effects associated with the age by SICD group interaction can be excluded from the model

   $$H_0 : \beta_5 = \beta_6 = 0$$

   $$H_A : \beta_5 \neq 0 \ \text{ Or } \ \beta_6 \neq 0$$
9. ML based likelihood ratio test

Model Fitting with R

Prepare data for fitting random coefficient model

```
# Read in Rat Brain data in long format.
autism <- read.csv("/Users/munnibegum/Library/CloudStorage/OneDrive-BallStateUniversity/Mydocs_21/Teachi
autism$age_2 = autism$age - 2
autism$age_2sq = (autism$age_2)^2

# Define SICD factors

autism$sicdF[autism$sicdegp ==3]<- 0
autism$sicdF[autism$sicdegp ==2]<- 2
autism$sicdF[autism$sicdegp ==1]<- 1
autism$sicdF <- factor(autism$sicdF)
```

```
#head(autism)

autism2 <- autism[complete.cases(autism), ]

#summary(autism2)
```

Fit a random coefficient model M-1

```
library(nlme)

# Model M-1: Random intercept, random age and age-squared effects model
#M.1_fit <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF, random = ~
#age_2+age_2sq | childid, method = "REML", data = autism2)

# You will get an error when execute the above code. Because the REML estimation algorithm did
#not converge to a solution for the parameter estimates
#
#summary(M.1_fit) # No summary is available

# Model M-2: Model without random intercept, random age and age-squared effects model
M.2_fit <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF, random = ~
age_2+age_2sq-1 | childid, method = "REML", data = autism2)

summary(M.2_fit)
```

```
## Linear mixed-effects model fit by REML
##   Data: autism2
##        AIC      BIC    logLik
##   4641.276 4698.457 -2307.638
##
## Random effects:
##  Formula: ~age_2 + age_2sq - 1 | childid
##  Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev    Corr
## age_2    3.8298024 age_2
## age_2sq  0.3625824 -0.317
## Residual 6.2047262
##
## Fixed effects:  vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF
##                    Value Std.Error  DF   t-value p-value
## (Intercept)    13.769637 0.8093779 446 17.012618  0.0000
## age_2           5.603481 0.7938042 446  7.059021  0.0000
## age_2sq         0.203999 0.0816167 446  2.499473  0.0128
## sicdF1         -5.415784 1.0934935 155 -4.952735  0.0000
## sicdF2         -4.037378 1.0293605 155 -3.922220  0.0001
## age_2:sicdF1   -3.296437 1.0915657 446 -3.019916  0.0027
## age_2:sicdF2   -2.746559 1.0270067 446 -2.674334  0.0078
## age_2sq:sicdF1 -0.134615 0.1133821 446 -1.187271  0.2358
## age_2sq:sicdF2 -0.129699 0.1055228 446 -1.229104  0.2197
##  Correlation:
##                (Intr) age_2  ag_2sq sicdF1 sicdF2 ag_2:sF1 ag_2:sF2 ag_2s:F1
## age_2          -0.418
## age_2sq         0.318 -0.586
```

```
## sicdF1         -0.740  0.310 -0.235
## sicdF2         -0.786  0.329 -0.250  0.582
## age_2:sicdF1    0.304 -0.727  0.426 -0.426 -0.239
## age_2:sicdF2    0.323 -0.773  0.453 -0.239 -0.425  0.562
## age_2sq:sicdF1 -0.229  0.422 -0.720  0.321  0.180 -0.592   -0.326
## age_2sq:sicdF2 -0.246  0.453 -0.773  0.182  0.320 -0.329   -0.590    0.557
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.22352306 -0.37936451 -0.05010264  0.28909578  6.88643142
##
## Number of Observations: 610
## Number of Groups: 158
```

```
# Model M-2A: Random quadratic age effect can be ommitted
M.2A_fit <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF, random = ~
age_2-1 | childid, method = "REML", data = autism2)

summary(M.2A_fit)
```

```
## Linear mixed-effects model fit by REML
##   Data: autism2
##        AIC      BIC    logLik
##   4721.203 4769.587 -2349.601
##
## Random effects:
##  Formula: ~age_2 - 1 | childid
##            age_2 Residual
## StdDev: 3.935294 7.728103
##
## Fixed effects:  vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF
##                    Value Std.Error  DF   t-value p-value
## (Intercept)    13.545646 1.0009262 446 13.533111  0.0000
## age_2           6.267595 0.8587043 446  7.298898  0.0000
## age_2sq         0.108030 0.0575702 446  1.876493  0.0612
## sicdF1         -5.229967 1.3501195 155 -3.873707  0.0002
## sicdF2         -3.792041 1.2719984 155 -2.981168  0.0033
## age_2:sicdF1   -3.919624 1.1811068 446 -3.318602  0.0010
## age_2:sicdF2   -3.611116 1.1103421 446 -3.252256  0.0012
## age_2sq:sicdF1 -0.042638 0.0806569 446 -0.528640  0.5973
## age_2sq:sicdF2 -0.001807 0.0747335 446 -0.024185  0.9807
##  Correlation:
##                (Intr) age_2  ag_2sq sicdF1 sicdF2 ag_2:sF1 ag_2:sF2 ag_2s:F1
## age_2          -0.457
## age_2sq         0.495 -0.643
## sicdF1         -0.741  0.338 -0.367
## sicdF2         -0.787  0.359 -0.389  0.583
## age_2:sicdF1    0.332 -0.727  0.468 -0.461 -0.261
## age_2:sicdF2    0.353 -0.773  0.498 -0.262 -0.461  0.562
## age_2sq:sicdF1 -0.353  0.459 -0.714  0.488  0.278 -0.648   -0.355
## age_2sq:sicdF2 -0.381  0.496 -0.770  0.283  0.493 -0.360   -0.649    0.550
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.68744221 -0.36518799 -0.05471469  0.31700669  5.52389989
```

```
##
## Number of Observations: 610
## Number of Groups: 158
```

```
anova(M.2A_fit,M.2_fit)
```

```
##          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## M.2A_fit     1 11 4721.203 4769.587 -2349.601
## M.2_fit      2 13 4641.276 4698.457 -2307.638 1 vs 2 83.92683  <.0001
```

REML log-likelihood Ratio Test Statistic: $-2[REMLlog-likelihoo(M.2A)-REMLlog-likelihoo(M.2)] = -2[-2349.601+2307.638] = 83.9$.

The test statistic follows a mixture of χ^2 distributions with equal weights 0.5 and 1 and 2 degrees of freedom. The p-value for this test is calculated as follows:

```
0.5*(1-pchisq(83.9,1)+0.5*(1-pchisq(83.9,2)))
```

```
## [1] 0
```

Since the p-value is close to 0, the likelihood ratio test indicates that the random quadratic effetcs os age should be retained in the model.

Test on Fixed Effects

Fixed effects associated with the age-sqaured by SICD group interaction can be ommitted

```
# Model M-2 with ML fit
M.2_ML <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF + age_2sq:sicdF, random = ~
age_2+age_2sq-1 | childid, method = "ML", data = autism2)

# Model M-3 with ML fit
M.3_ML <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF, random = ~ age_2+ age_2sq-1 |
childid, method = "ML", data = autism2)
anova(M.3_ML,M.2_ML)
```

```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## M.3_ML     1 11 4634.314 4682.862 -2306.157
## M.2_ML     2 13 4636.444 4693.819 -2305.222 1 vs 2 1.869704  0.3926
```

Based on the more conservative ML based likelihood ratio test, we fail to reject the null hypothesis that the fixed effects of the age-sqaured by SICD group interaction is zero. Thus we drop this interaction from the model.

Fixed effects associated with the age by SICD group interaction can be ommitted

```
# Model M-4 with ML fit
M.4_ML <- lme(vsae ~ age_2 + age_2sq + sicdF, random = ~ age_2+ age_2sq-1 | childid, method =
"ML", data = autism2)
anova(M.4_ML,M.3_ML)
```

```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## M.4_ML     1  9 4653.696 4693.417 -2317.848
## M.3_ML     2 11 4634.314 4682.862 -2306.157 1 vs 2 23.38232  <.0001
```

Based on the ML based likelihood ratio test, we reject the null hypothesis that the fixed effects of the age by SICD group interaction is zero. Thus we keep this interaction in the model. Therefore, model M-3 can be considered as the final model

```r
library(nlme)
M.3_final <- lme(vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF, random = ~ age_2+ age_2sq-1 |
childid, method = "REML", data = autism2)
summary(M.3_final)
```

```
## Linear mixed-effects model fit by REML
##   Data: autism2
##       AIC      BIC    logLik
##   4633.57 4681.991 -2305.785
##
## Random effects:
##  Formula: ~age_2 + age_2sq - 1 | childid
##  Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev    Corr
## age_2    3.8110274 age_2
## age_2sq  0.3556805 -0.306
## Residual 6.2281389
##
## Fixed effects:  vsae ~ age_2 + age_2sq + sicdF + age_2:sicdF
##                  Value Std.Error  DF   t-value p-value
## (Intercept)  13.463533 0.7815177 448 17.227419  0.0000
## age_2         6.148750 0.6882639 448  8.933711  0.0000
## age_2sq       0.109008 0.0427795 448  2.548125  0.0112
## sicdF1       -4.987639 1.0379064 155 -4.805480  0.0000
## sicdF2       -3.622820 0.9774516 155 -3.706394  0.0003
## age_2:sicdF1 -4.068041 0.8797676 448 -4.623995  0.0000
## age_2:sicdF2 -3.495530 0.8289509 448 -4.216812  0.0000
##  Correlation:
##              (Intr) age_2  ag_2sq sicdF1 sicdF2 a_2:F1
## age_2        -0.341
## age_2sq       0.177 -0.356
## sicdF1       -0.730  0.211 -0.005
## sicdF2       -0.775  0.224 -0.003  0.583
## age_2:sicdF1  0.218 -0.683  0.000 -0.309 -0.174
## age_2:sicdF2  0.230 -0.723 -0.006 -0.174 -0.309  0.567
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.26516725 -0.39753349 -0.05448662  0.29145900  6.84667726
##
## Number of Observations: 610
## Number of Groups: 158
```

```r
anova(M.3_final)
```

```
##             numDF denDF  F-value p-value
## (Intercept)     1   448 947.3622  <.0001
## age_2           1   448 145.2867  <.0001
## age_2sq         1   448   6.2235   0.013
## sicdF           2   155  23.0142  <.0001
## age_2:sicdF     2   448  12.5627  <.0001
```

Final Model: Interpretation of the parameter estimates

Fixed-Effect Parameter Estimates

1. The intercept $\hat{\beta}_0 = 13.46$ represented the estimated mean VSAE score for the children at 2 years of age in the reference category of SICD group 3
2. Children in SICD group 3 had the highest initial language scores
3. The parameter estimates $\hat{\beta}_1 = 6.15$ and $\hat{\beta}_2 = 0.11$ are positive and significant
4. And suggests an increasing trend in VSAE scores as a function of age for children in SICD group 3
5. $\hat{\beta}_3 = -4.99$, suggesting that the mean initial VSAE score for children in SCID group 1 is 4.99 units lower than that of children in the reference group (SICD group 3)
6. $\hat{\beta}_4 = -3.62$, suggesting that the mean initial VSAE score for children in SCID group 2 is 3.62 units lower than that of children in the reference group (SICD group 3)
7. Recall the interaction effect β_5 represents the difference in the linear effect of age in SICD group 1 versus the SICD group 3
8. $\hat{\beta}_5 = -4.07$ meaning that the linear age effect for children in SICD group 1 is 4.07 units less than that for children in SICD group 3
9. Note that estimated linear age effect for children in SICD group 1 is $6.15 - 4.07 = 2.08$ which is still positive
10. $\hat{\beta}_6 = -3.50$ meaning that the linear age effect for children in SICD group 2 is 3.5 units less than that for children in SICD group 3
11. That means the estimated linear age effect for children in SICD group 2 is $6.15 - 3.5 = 2.65$ and is positive and similar to that for children in SICD group 1

Variance-covariance matrix for ith child

The marginal variance-covariance matrix for the i-th child implied by the final model can be written as follows:

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{R}_i$$

For example, for a child with id 1, $\mathbf{V}_1$ is estimated as below:

```
#Variance covariance for child 1
getVarCov(M.3_final, individual="1",type="marginal")
```

```
## childid 1
## Marginal variance covariance matrix
##       1       2       3        4       5
## 1 38.79   0.000   0.000    0.000    0.00
## 2  0.00  52.610  39.728   84.617  120.27
## 3  0.00  39.728 157.330  273.610  425.25
## 4  0.00  84.617 273.610  769.400 1293.00
## 5  0.00 120.270 425.250 1293.000 2543.20
##   Standard Deviations: 6.2281 7.2533 12.543 27.738 50.43
```

```
#Variance covariance for child 6
getVarCov(M.3_final, individual="6",type="marginal")
```

```
## childid 6
## Marginal variance covariance matrix
##       1       2        3       4
## 1 38.79   0.000    0.000    0.00
## 2  0.00  52.610   84.617  120.27
## 3  0.00  84.617  769.400 1293.00
## 4  0.00 120.270 1293.000 2543.20
##   Standard Deviations: 6.2281 7.2533 27.738 50.43
```

Variance-covariance for child with id 1

1. Note, first that the estimated marginal varinces of the VSAE scores increases dramatically with age (as noted in the EDA)
2. The mariginal covariances associated with age 2 is 0 [in the first row and first column]
3. This is due to exclusion of the random intercept from the model, and to using Age-2 as a covariate instead of age
4. Note the values of the first row of the $\mathbf{Z}_i$ matrix correspond to the values of Age - 2 (x_{1ij}) and $(Age-2)^2 = x_{2ij}$
5. The Z_i matrix can be written as follows:

$$\mathbf{Z}_i = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 3 & 9 \\ 7 & 49 \\ 11 & 121 \end{pmatrix}$$

6. Also note that since there is no intercept in the final model, variance in VSAE score at age 2 is due to residual variance, contributed by $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$