BIOS6643. L10 Model Building

Exercise involving G and R matrices

Consider a basic science experiment conducted where cell counts are measured at 4 time points for samples taken from individual subjects or animals. A linear mixed model will be fit for the data (perhaps after log transformation), and fixed effects will be included for time, and treatment group as well as their interaction. (To answer this question we do not need to know the specific form of $\boldsymbol{X\beta}$.)

Determine the structure for $\boldsymbol{V}_i$ if a random intercept for subjects will be included, plus an AR(1) structure for the error covariance matrix ($\boldsymbol{R}_i$). What does the combination of non-simple $\boldsymbol{R}$ and $\boldsymbol{G}$ allow you to do in modeling covariance that using only one cannot do? Discuss in a few sentences.

See solution in file "fig-sol-ex1-lab10.png" under *Files > ComputingLabs.*

FEV Study

Description

The data set gives characteristics of children patients with a diagnosis of Cystic Fibrosis (CF) who are patients at the Colorado Children's Hospital. Data pull was conducted in Summer 2020. This dataset may only be used for BIOS6643.

VARIABLE DESCRIPTIONS:
id: patient ID
race: race
Genotypes1: mutation of copy one of the gene that codes for the CFTR protein Genotypes2: mutation of copy two of the gene that codes for the CFTR protein. age: Age in years
gender : gender fev: % of predicted forced expiratory volume in 1 second

- **Objectives** of the study included:
  - Determine if fev values over time are larger on average for males than for females
  - Determine if the rate of change of fev over time is different for males and females.

Selecting the mean structure

```
#  Read in the data

##  Read in the data
dat <- read.csv("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/BIOS664
                header=TRUE)
length(unique(dat$id)) ## number of individuals
```

```
## [1] 723
```

```
head(dat,3)
```

```
##   id      Race Genotypes1 Genotypes2   age gender      fev
## 1  1 Caucasian       F508      R553X 10.32      m 55.45775
## 2  1 Caucasian       F508      R553X 10.47      m 53.48924
## 3  1 Caucasian       F508      R553X 10.72      m 49.62416
```

```
summary(dat$age) ## age in years
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.01    9.83   13.50   13.73   17.29   25.00
```

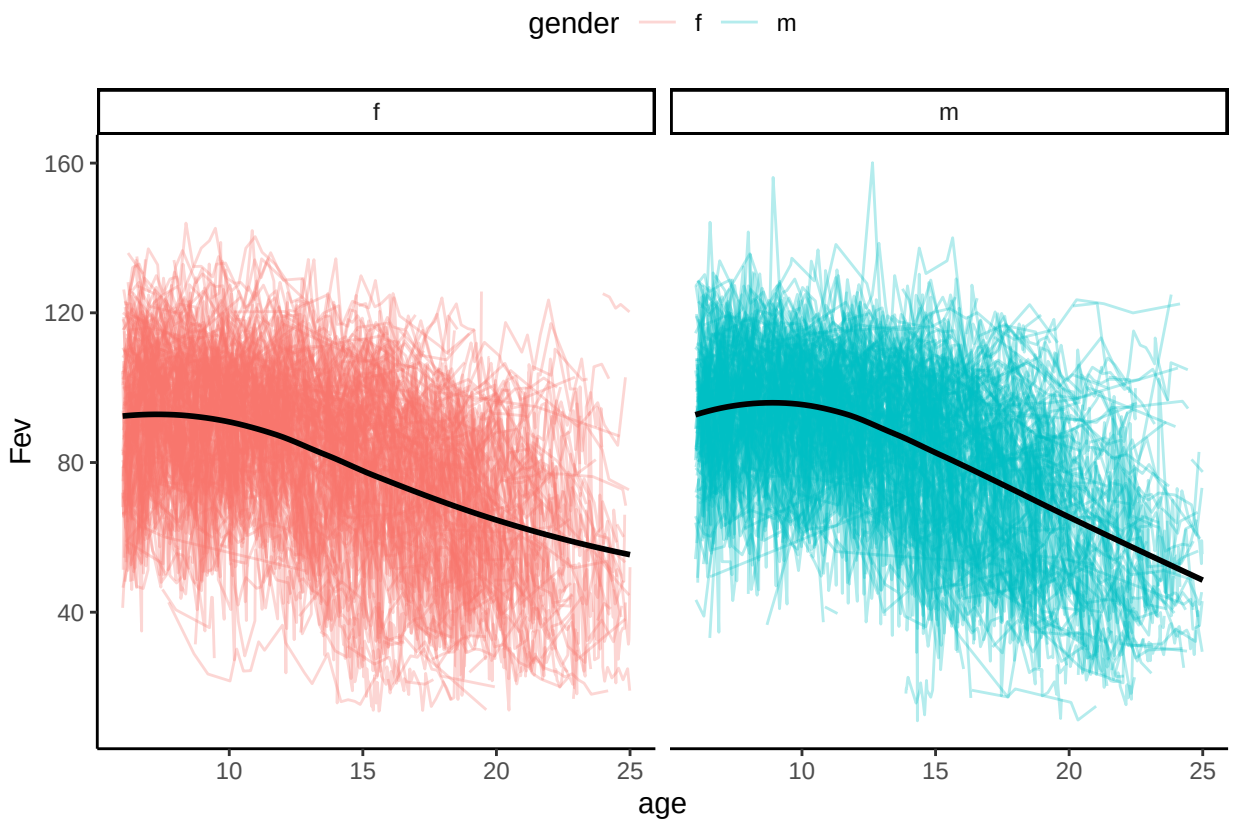
```
summary(dat$fev)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.99   66.72   85.53   81.90   99.31  160.10
```

```
######
#############################################################################################################
### selecting the mean form
## ------------------------------------------------------------------------------------------------

ggplot(dat, aes(x = age, y = fev, group = id)) +
    facet_wrap(~gender) +
    geom_line(aes(color = gender), alpha = 0.3) +
    geom_smooth(aes(group = 1), method="loess", color="black", se=FALSE) +
    theme_classic() +
    theme(legend.position = "top") +
    ylab("Fev")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```r
## age as linear
fit.lme.0 <- lme(fev ~ -1 + gender + age:gender, data=dat,
                 random = ~ 1 | id,method="ML")
summary(fit.lme.0)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   145661.2 145708.1 -72824.58
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept) Residual
## StdDev:    17.76981 11.57677
##
## Fixed effects:  fev ~ -1 + gender + age:gender
##                 Value Std.Error    DF   t-value p-value
## genderf     106.70242 1.0952620   721  97.42182       0
## genderm     107.72595 1.0864343   721  99.15550       0
## genderf:age  -1.76779 0.0370579 17770 -47.70342       0
## genderm:age  -1.79357 0.0409169 17770 -43.83454       0
##  Correlation:
##             gendrf gendrm gndrf:
## genderm      0.000
## genderf:age -0.439  0.000
## genderm:age  0.000 -0.499  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -7.07828572 -0.53190983  0.05024462  0.60213684  5.64679392
##
## Number of Observations: 18494
## Number of Groups: 723
```

```r
## adding quadratic terms of age
fit.lme.1 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender,
                 data=dat,
                 random = ~ 1 | id,method="ML")
summary(fit.lme.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   145213.8 145276.4 -72598.92
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept) Residual
## StdDev:    17.67878 11.43327
##
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender
##                         Value Std.Error    DF   t-value p-value
## genderf              93.62488 1.5128318   721  61.88717  0.0000
## genderm              87.90149 1.5695259   721  56.00512  0.0000
```

```
## genderf:age           0.41730 0.1794696 17768   2.32517  0.0201
## genderm:age           1.49555 0.1934694 17768   7.73017  0.0000
## genderf:I(age * age) -0.07836 0.0063016 17768 -12.43528  0.0000
## genderm:I(age * age) -0.11718 0.0067410 17768 -17.38287  0.0000
##  Correlation:
##                      gendrf gendrm gndrf: gndrm: gndrf:I(*a)
## genderm               0.000
## genderf:age          -0.744  0.000
## genderm:age           0.000 -0.782  0.000
## genderf:I(age * age)  0.695  0.000 -0.979  0.000
## genderm:I(age * age)  0.000  0.726  0.000 -0.978  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -7.24451152 -0.53995990  0.04787477  0.60276234  5.76966302
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
## adding cubic terms of age
fit.lme.2 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                   I(age*age*age):gender, data=dat,
                    random = ~ 1 | id,method="ML")
summary(fit.lme.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   144944.3 145022.6 -72462.17
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept) Residual
## StdDev:     17.7227 11.34474
##
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                               Value Std.Error    DF    t-value p-value
## genderf                    64.45107  3.298411   721  19.540039       0
## genderm                    46.67971  3.472294   721  13.443478       0
## genderf:age                 7.49048  0.733378 17766  10.213667       0
## genderm:age                11.46857  0.774428 17766  14.809086       0
## genderf:I(age * age)       -0.60229  0.053070 17766 -11.348941       0
## genderm:I(age * age)       -0.85428  0.055854 17766 -15.294934       0
## genderf:I(age * age * age)  0.01204  0.001211 17766   9.941387       0
## genderm:I(age * age * age)  0.01687  0.001269 17766  13.292823       0
##  Correlation:
##                            gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                     0.000
## genderf:age                -0.945  0.000
## genderm:age                 0.000 -0.952  0.000
## genderf:I(age * age)        0.921  0.000 -0.991  0.000
## genderm:I(age * age)        0.000  0.926  0.000 -0.991  0.000
## genderf:I(age * age * age) -0.889  0.000  0.970  0.000 -0.993       0.000
## genderm:I(age * age * age)  0.000 -0.893  0.000  0.969  0.000      -0.993
##                            gndrf:I(*a*a)
```

```
## genderm
## genderf:age
## genderm:age
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -7.40275579 -0.54663718  0.04382553  0.60320809  5.72214673
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
## updating model - same as above
## fit.lme.2 <- update(fit.lme.1, fev ~ -1 + gender  + age:gender + I(age*age):gender +
##                   I(age*age*age):gender)

anova(fit.lme.0, fit.lme.1, fit.lme.2)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit.lme.0     1  6 145661.2 145708.1 -72824.58
## fit.lme.1     2  8 145213.8 145276.4 -72598.92 1 vs 2 451.3298  <.0001
## fit.lme.2     3 10 144944.4 145022.6 -72462.17 2 vs 3 273.4885  <.0001
```

Preliminary covariance structure G

```
#############################################################################################################
###


#############################################################################################################
### preliminary covariance G
## ------------------------------------------------------------------------------------------

fit.lm.0 <- lm(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                I(age*age*age):gender, data=dat)
summary(fit.lm.0)
```

```
##
## Call:
## lm(formula = fev ~ -1 + gender + age:gender + I(age * age):gender +
##     I(age * age * age):gender, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.375 -13.618   1.711  14.295  71.351
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## genderf                    69.493830   5.047281  13.769  < 2e-16 ***
## genderm                    54.768689   5.313181  10.308  < 2e-16 ***
## genderf:age                 7.209947   1.166087   6.183 6.42e-10 ***
```

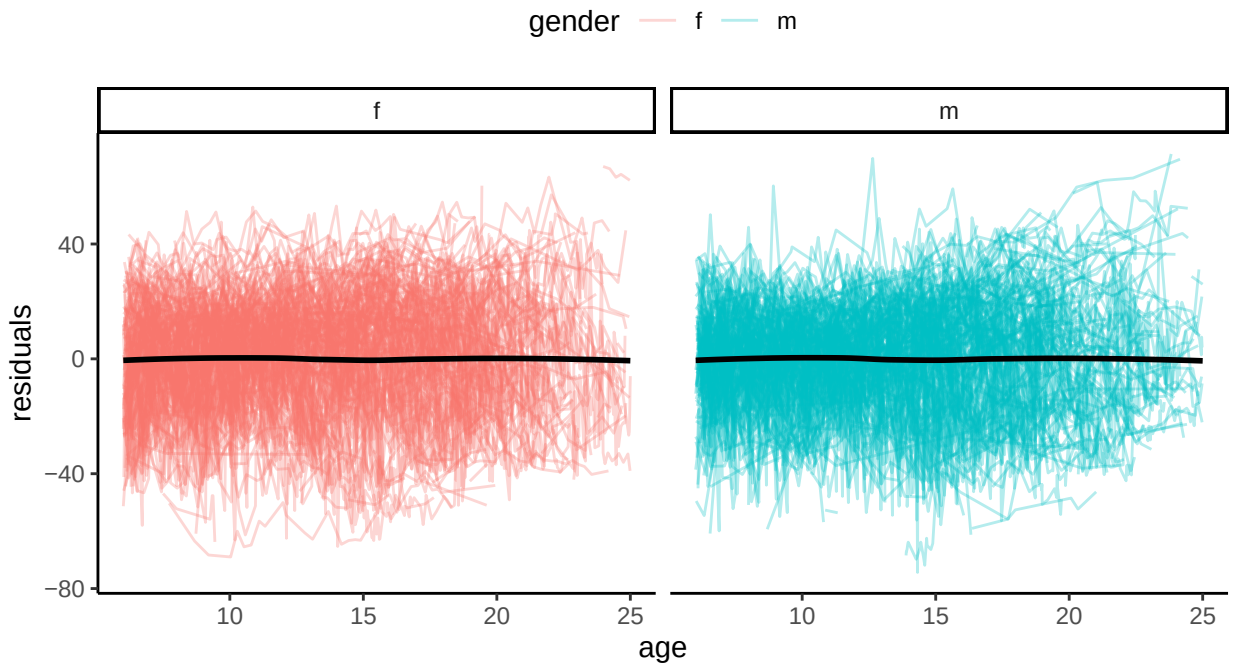
```
## genderm:age                10.670168   1.224638   8.713  < 2e-16 ***
## genderf:I(age * age)       -0.644283   0.083694  -7.698 1.45e-14 ***
## genderm:I(age * age)       -0.815717   0.087642  -9.307  < 2e-16 ***
## genderf:I(age * age * age)  0.013504   0.001889   7.147 9.19e-13 ***
## genderm:I(age * age * age)  0.015347   0.001969   7.795 6.75e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.28 on 18486 degrees of freedom
## Multiple R-squared:  0.9434, Adjusted R-squared:  0.9433
## F-statistic: 3.849e+04 on 8 and 18486 DF,  p-value: < 2.2e-16
```

```
## calculating residuals
dat$resid <- residuals(fit.lm.0)
head(dat)
```

```
##   id      Race Genotypes1 Genotypes2   age gender      fev      resid
## 1  1 Caucasian       F508      R553X 10.32      m 55.45775 -39.419474
## 2  1 Caucasian       F508      R553X 10.47      m 53.48924 -41.190974
## 3  1 Caucasian       F508      R553X 10.72      m 49.62416 -44.694488
## 4  1 Caucasian       F508      R553X 10.97      m 61.13242 -32.784394
## 5  1 Caucasian       F508      R553X 11.12      m 87.86685  -5.790132
## 6  2 Caucasian       F508       F508 14.04      m 52.75725 -33.500000
```

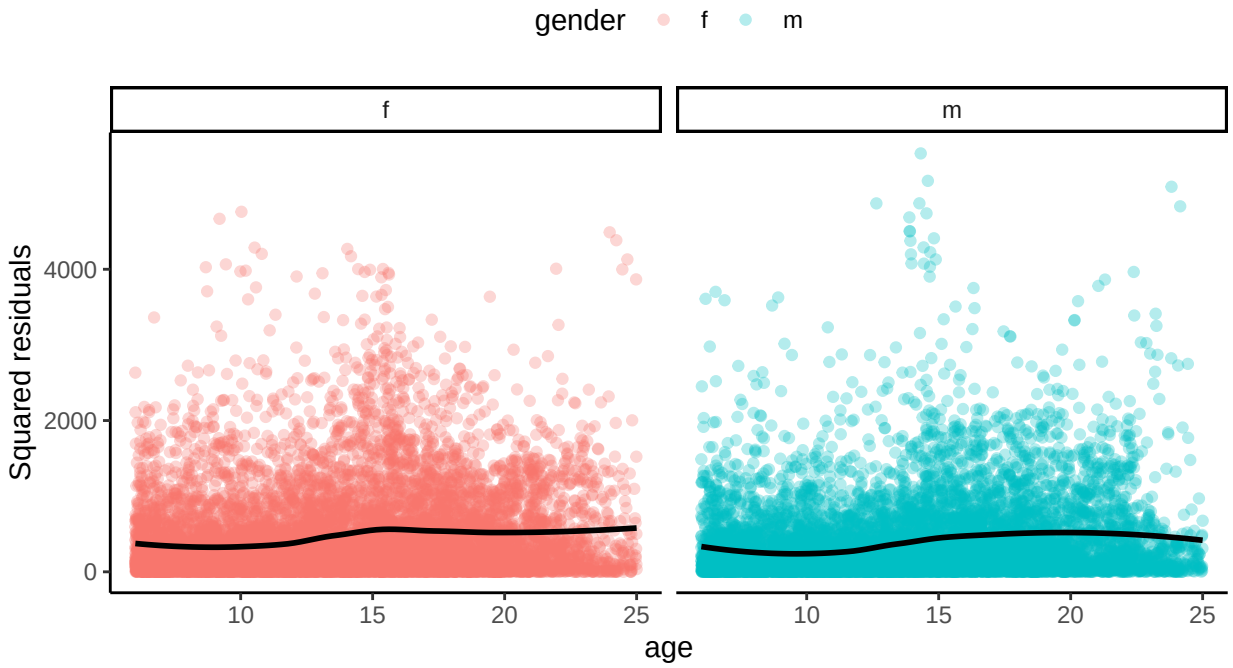
```
ggplot(dat, aes(x = age, y = resid, group = id)) +
    facet_wrap(~gender) +
    geom_line(aes(color = gender), alpha = 0.3) +
    geom_smooth(aes(group = 1), method="loess", color="black", se=FALSE) +
    theme_classic() +
    theme(legend.position = "top") +
    ylab("residuals")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(dat, aes(x = age, y = resid^2, group = id)) +
    facet_wrap(~gender) +
    geom_point(aes(color = gender), alpha = 0.3) +
    geom_smooth(aes(group = 1), method="loess", color="black", se=FALSE) +
    theme_classic() +
    theme(legend.position = "top") +
    ylab("Squared residuals")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
## model with quadratic random effects did not converge: adding "+ I(age*age)" to random effects
fit.lme.3 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                   I(age*age*age):gender, data=dat,
                    random = ~ age  | id, method="ML")

summary(fit.lme.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   144944.3 145022.6 -72462.17
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept) Residual
## StdDev:     17.7227 11.34474
##
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                                Value Std.Error    DF    t-value p-value
## genderf                    64.45107  3.298411   721  19.540039       0
## genderm                    46.67971  3.472294   721  13.443478       0
## genderf:age                 7.49048  0.733378 17766  10.213667       0
## genderm:age                11.46857  0.774428 17766  14.809086       0
```

```
## genderf:I(age * age)       -0.60229  0.053070 17766 -11.348941       0
## genderm:I(age * age)       -0.85428  0.055854 17766 -15.294934       0
## genderf:I(age * age * age)  0.01204  0.001211 17766   9.941387       0
## genderm:I(age * age * age)  0.01687  0.001269 17766  13.292823       0
##  Correlation:
##                            gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                     0.000
## genderf:age                -0.945  0.000
## genderm:age                 0.000 -0.952  0.000
## genderf:I(age * age)        0.921  0.000 -0.991  0.000
## genderm:I(age * age)        0.000  0.926  0.000 -0.991  0.000
## genderf:I(age * age * age) -0.889  0.000  0.970  0.000 -0.993       0.000
## genderm:I(age * age * age)  0.000 -0.893  0.000  0.969  0.000      -0.993
##                            gndrf:I(*a*a)
## genderm
## genderf:age
## genderm:age
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -7.40275579 -0.54663718  0.04382553  0.60320809  5.72214673
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
summary(fit.lme.3)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   141954.7 142048.6 -70965.37
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 25.051337 (Intr)
## age          1.841479 -0.761
## Residual    10.151809
##
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                               Value Std.Error    DF    t-value p-value
## genderf                    62.69593  3.524726   721  17.787463       0
## genderm                    47.38592  3.838589   721  12.344621       0
## genderf:age                 7.67477  0.764109 17766  10.044086       0
## genderm:age                11.24543  0.838210 17766  13.415997       0
## genderf:I(age * age)       -0.60303  0.056200 17766 -10.730110       0
## genderm:I(age * age)       -0.83348  0.060850 17766 -13.697347       0
## genderf:I(age * age * age)  0.01175  0.001306 17766   8.997012       0
## genderm:I(age * age * age)  0.01629  0.001389 17766  11.728614       0
##  Correlation:
```

```
##                             gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                      0.000
## genderf:age                 -0.924  0.000
## genderm:age                  0.000 -0.938  0.000
## genderf:I(age * age)         0.842  0.000 -0.975  0.000
## genderm:I(age * age)         0.000  0.867  0.000 -0.979  0.000
## genderf:I(age * age * age) -0.791  0.000  0.941  0.000 -0.989       0.000
## genderm:I(age * age * age)  0.000 -0.816  0.000  0.945  0.000      -0.989
##                             gndrf:I(*a*a)
## genderm
## genderf:age
## genderm:age
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.12207135 -0.51985113  0.04829077  0.58373373  6.63716072
##
## Number of Observations: 18494
## Number of Groups: 723
```

Residual covariance structure R

```
#############################################################################################################
###

## different variance for males and females
fit.lme.4 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                   I(age*age*age):gender, data=dat,
                    random = ~ age  | id, method="ML",
                    weights = varIdent(form = ~ 1 | gender))
summary(fit.lme.4)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##       AIC      BIC    logLik
##    141951 142052.8 -70962.51
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev    Corr
## (Intercept) 25.095132 (Intr)
## age          1.844172 -0.762
## Residual    10.288954
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | gender
##  Parameter estimates:
```

```
##         m         f
## 1.0000000 0.9745014
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                              Value Std.Error    DF    t-value p-value
## genderf                   62.62862  3.492402   721  17.932821       0
## genderm                   47.39750  3.880219   721  12.215162       0
## genderf:age                7.68376  0.755558 17766  10.169642       0
## genderm:age               11.24846  0.848374 17766  13.258845       0
## genderf:I(age * age)      -0.60343  0.055565 17766 -10.859823       0
## genderm:I(age * age)      -0.83392  0.061592 17766 -13.539287       0
## genderf:I(age * age * age)  0.01175  0.001291 17766   9.101948       0
## genderm:I(age * age * age)  0.01631  0.001406 17766  11.597094       0
##  Correlation:
##                           gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                    0.000
## genderf:age               -0.923  0.000
## genderm:age                0.000 -0.939  0.000
## genderf:I(age * age)       0.840  0.000 -0.975  0.000
## genderm:I(age * age)       0.000  0.868  0.000 -0.979  0.000
## genderf:I(age * age * age) -0.789  0.000  0.941  0.000 -0.989       0.000
## genderm:I(age * age * age)  0.000 -0.817  0.000  0.946  0.000      -0.989
##                           gndrf:I(*a*a)
## genderm
## genderf:age
## genderm:age
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.22395661 -0.51843057  0.04862607  0.58339514  6.72281456
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
R.b.1 <- getVarCov(fit.lme.4,type="conditional",individual=1)   # R_1; first male
R.b.2 <- getVarCov(fit.lme.4,type="conditional",individual=12)   # R_2; first female
R.b.1
```

```
## id 1
## Conditional variance covariance matrix
##        1      2      3      4      5
## 1 105.86   0.00   0.00   0.00   0.00
## 2   0.00 105.86   0.00   0.00   0.00
## 3   0.00   0.00 105.86   0.00   0.00
## 4   0.00   0.00   0.00 105.86   0.00
## 5   0.00   0.00   0.00   0.00 105.86
##   Standard Deviations: 10.289 10.289 10.289 10.289 10.289
```

```
R.b.1
```

```
## id 1
## Conditional variance covariance matrix
##        1      2      3      4      5
```

```
## 1 105.86   0.00   0.00   0.00   0.00
## 2   0.00 105.86   0.00   0.00   0.00
## 3   0.00   0.00 105.86   0.00   0.00
## 4   0.00   0.00   0.00 105.86   0.00
## 5   0.00   0.00   0.00   0.00 105.86
##   Standard Deviations: 10.289 10.289 10.289 10.289 10.289
```

```
## allowing for AR(1) exp
fit.lme.5 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                   I(age*age*age):gender, data=dat,
                    random = ~ age  | id, method="ML",
                   corExp(form = ~ age | id))
summary(fit.lme.5)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC    BIC    logLik
##   141218.2 141320 -70596.12
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 24.46315 (Intr)
## age          1.79206 -0.751
## Residual    10.16447
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##    range
## 0.016454
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                                Value Std.Error    DF    t-value p-value
## genderf                    63.54202  3.607706   721  17.612858       0
## genderm                    47.80098  3.895726   721  12.270107       0
## genderf:age                 7.58913  0.791732 17766   9.585486       0
## genderm:age                11.19460  0.857492 17766  13.055060       0
## genderf:I(age * age)       -0.60329  0.058474 17766 -10.317198       0
## genderm:I(age * age)       -0.83170  0.062426 17766 -13.323083       0
## genderf:I(age * age * age)  0.01189  0.001363 17766   8.729453       0
## genderm:I(age * age * age)  0.01629  0.001428 17766  11.408728       0
##  Correlation:
##                            gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                     0.000
## genderf:age                -0.929  0.000
## genderm:age                 0.000 -0.941  0.000
## genderf:I(age * age)        0.851  0.000 -0.977  0.000
## genderm:I(age * age)        0.000  0.873  0.000 -0.979  0.000
## genderf:I(age * age * age) -0.801  0.000  0.943  0.000 -0.989       0.000
## genderm:I(age * age * age)  0.000 -0.822  0.000  0.947  0.000      -0.989
##                            gndrf:I(*a*a)
## genderm
## genderf:age
## genderm:age
```

```
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.12800702 -0.52080597  0.04882077  0.58381724  6.36308204
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
anova(fit.lme.3, fit.lme.4, fit.lme.5)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit.lme.3     1 12 141954.7 142048.6 -70965.37
## fit.lme.4     2 13 141951.0 142052.8 -70962.51 1 vs 2 5.715667  0.0168
## fit.lme.5     3 13 141218.2 141320.0 -70596.12
```

Model reduction

```
## do we really random slopes? The variance of the random slopes, is not negligible (~1.8^2), so I woul


## testing if the cubic terms are needed in the mean form
fit.lme.5.1 <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender,
                   data=dat,
                   random = ~ age  | id, method="ML",
                   corExp(form = ~ age | id))
summary(fit.lme.5.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   141418.8 141504.9 -70698.42
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 25.092200 (Intr)
## age          1.800298 -0.752
## Residual    10.221047
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##      range
## 0.01674874
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender
##                         Value Std.Error    DF   t-value p-value
## genderf              88.58823 2.1959143   721  40.34230       0
## genderm              84.14385 2.2515080   721  37.37222       0
## genderf:age           1.08984 0.2653860 17768   4.10664       0
```

```
## genderm:age           1.95814 0.2787764 17768   7.02405       0
## genderf:I(age * age) -0.09911 0.0088488 17768 -11.20084       0
## genderm:I(age * age) -0.12818 0.0093783 17768 -13.66745       0
##  Correlation:
##                      gendrf gendrm gndrf: gndrm: gndrf:I(*a)
## genderm               0.000
## genderf:age          -0.869  0.000
## genderm:age           0.000 -0.884  0.000
## genderf:I(age * age)  0.660  0.000 -0.884  0.000
## genderm:I(age * age)  0.000  0.703  0.000 -0.905  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.00340961 -0.52031162  0.04900868  0.58466303  6.28583914
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
anova(fit.lme.5, fit.lme.5.1) ## we do seem to need cubic terms
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit.lme.5       1 13 141218.2 141320.0 -70596.12
## fit.lme.5.1     2 11 141418.8 141504.9 -70698.42 1 vs 2 204.6105  <.0001
```

```
## testing if there is an interaction between age and gender
fit.lme.5.2 <- lme(fev ~ -1 + gender  + age + I(age*age) + I(age*age*age),
                data=dat,
                random = ~ age  | id, method="ML",
                corExp(form = ~ age | id))
summary(fit.lme.5.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat
##        AIC      BIC    logLik
##   141227.1 141305.4 -70603.56
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 24.535511 (Intr)
## age          1.799443 -0.752
## Residual    10.166993
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##      range
## 0.01644333
## Fixed effects:  fev ~ -1 + gender + age + I(age * age) + I(age * age * age)
##                       Value Std.Error    DF   t-value p-value
## genderf            55.96694 2.7253822   721  20.53545       0
## genderm            57.35477 2.7206186   721  21.08152       0
## age                 9.17624 0.5809970 17769  15.79395       0
## I(age * age)       -0.70318 0.0426182 17769 -16.49963       0
```

```
## I(age * age * age)  0.01381 0.0009844 17769  14.02563       0
##  Correlation:
##                     gendrf gendrm age    I(*ag)
## genderm              0.885
## age                 -0.906 -0.909
## I(age * age)         0.835  0.837 -0.978
## I(age * age * age) -0.786 -0.787  0.944 -0.989
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.16400081 -0.52308413  0.04959683  0.58324360  6.38511326
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
anova(fit.lme.5, fit.lme.5.2) ## the trajectories of FEV are different for males and females
```

```
##             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit.lme.5       1 13 141218.2 141320.0 -70596.12
## fit.lme.5.2     2 10 141227.1 141305.4 -70603.56 1 vs 2 14.88494  0.0019
```

Final model

```
fit.lme.f.reml <- lme(fev ~ -1 + gender  + age:gender + I(age*age):gender +
                  I(age*age*age):gender, data=dat,
                   random = ~ age  | id, method="REML",
                  corExp(form = ~ age | id))

summary(fit.lme.f.reml)
```

```
## Linear mixed-effects model fit by REML
##   Data: dat
##        AIC      BIC    logLik
##   141257.4 141359.2 -70615.72
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 24.553727 (Intr)
## age          1.798503 -0.752
## Residual    10.165206
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##      range
## 0.01645754
## Fixed effects:  fev ~ -1 + gender + age:gender + I(age * age):gender + I(age *      age * age):gende
##                                Value Std.Error    DF    t-value p-value
## genderf                    63.52634  3.609925   721  17.597691       0
## genderm                    47.79245  3.898177   721  12.260206       0
## genderf:age                 7.59163  0.791846 17766   9.587254       0
## genderm:age                11.19524  0.857704 17766  13.052574       0
```

```
## genderf:I(age * age)       -0.60344  0.058482 17766 -10.318304       0
## genderm:I(age * age)       -0.83169  0.062440 17766 -13.319888       0
## genderf:I(age * age * age)  0.01190  0.001363 17766   8.729697       0
## genderm:I(age * age * age)  0.01629  0.001428 17766  11.405036       0
##  Correlation:
##                            gendrf gendrm gndrf: gndrm: gndrf:I(*a) gndrm:I(*a)
## genderm                     0.000
## genderf:age                -0.929  0.000
## genderm:age                 0.000 -0.941  0.000
## genderf:I(age * age)        0.851  0.000 -0.976  0.000
## genderm:I(age * age)        0.000  0.873  0.000 -0.979  0.000
## genderf:I(age * age * age) -0.800  0.000  0.943  0.000 -0.989       0.000
## genderm:I(age * age * age)  0.000 -0.822  0.000  0.946  0.000      -0.989
##                            gndrf:I(*a*a)
## genderm
## genderf:age
## genderm:age
## genderf:I(age * age)
## genderm:I(age * age)
## genderf:I(age * age * age)
## genderm:I(age * age * age)  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -8.12751471 -0.52056765  0.04833701  0.58375846  6.36308564
##
## Number of Observations: 18494
## Number of Groups: 723
```

```
## AR term: exp(-1/0.01645754)= 4.085217e-27

beta.f.reml <- fixed.effects(fit.lme.f.reml)  #  beta
sebeta.model.f.reml <-  sqrt(diag(fit.lme.f.reml$varFix))
b.f.reml <- random.effects(fit.lme.f.reml)    #  posterior modes bi
G.f.reml <- getVarCov(fit.lme.f.reml, type="random.effects")   #  G
G.f.reml
```

```
## Random effects variance covariance matrix
##             (Intercept)      age
## (Intercept)     602.890 -33.2050
## age             -33.205   3.2346
##   Standard Deviations: 24.554 1.7985
```

```
sigma2.f <- fit.lme.f.reml$sigma^2  # sigma^2
R.f <- getVarCov(fit.lme.f.reml,type="conditional",individual=1)  #  R_i
V.f <- getVarCov(fit.lme.f.reml,type="marginal",individual=1)   # V_i
R.f
```

```
## id 1
## Conditional variance covariance matrix
##            1          2          3          4          5
## 1 1.0333e+02 1.1374e-02 2.8755e-09 7.2698e-16 8.0021e-20
## 2 1.1374e-02 1.0333e+02 2.6124e-05 6.6045e-12 7.2698e-16
## 3 2.8755e-09 2.6124e-05 1.0333e+02 2.6124e-05 2.8755e-09
## 4 7.2698e-16 6.6045e-12 2.6124e-05 1.0333e+02 1.1374e-02
## 5 8.0021e-20 7.2698e-16 2.8755e-09 1.1374e-02 1.0333e+02
```

```
##   Standard Deviations: 10.165 10.165 10.165 10.165 10.165
```

```
V.f
```

```
## id 1
## Marginal variance covariance matrix
##        1      2      3      4      5
## 1 365.37 262.07 262.11 262.15 262.18
## 2 262.07 365.49 262.33 262.49 262.59
## 3 262.11 262.33 366.03 263.06 263.28
## 4 262.15 262.49 263.06 366.96 263.99
## 5 262.18 262.59 263.28 263.99 367.72
##   Standard Deviations: 19.115 19.118 19.132 19.156 19.176
```

```
## PA predicted values $X_i \times betahat$ are produced by level=0;
## SS predicted values X_i betahat + Z_i bhat_i are produced by level =  1;
## both are gotten by level = 0:1
fitted(fit.lme.f.reml, level=0:1)[1:5,]
```

```
##      fixed       id
## 1 92.65699 64.03209
## 2 92.53486 63.95716
## 3 92.29949 63.80044
## 4 92.02564 63.60526
## 5 91.84351 63.47032
```

```
dat$predi <- fitted(fit.lme.f.reml, level=1) ## individual SS predictions
head(dat)
```

```
##   id      Race Genotypes1 Genotypes2   age gender      fev      resid    predi
## 1  1 Caucasian       F508      R553X 10.32      m 55.45775 -39.419474 64.03209
## 2  1 Caucasian       F508      R553X 10.47      m 53.48924 -41.190974 63.95716
## 3  1 Caucasian       F508      R553X 10.72      m 49.62416 -44.694488 63.80044
## 4  1 Caucasian       F508      R553X 10.97      m 61.13242 -32.784394 63.60526
## 5  1 Caucasian       F508      R553X 11.12      m 87.86685  -5.790132 63.47032
## 6  2 Caucasian       F508       F508 14.04      m 52.75725 -33.500000 46.74487
```

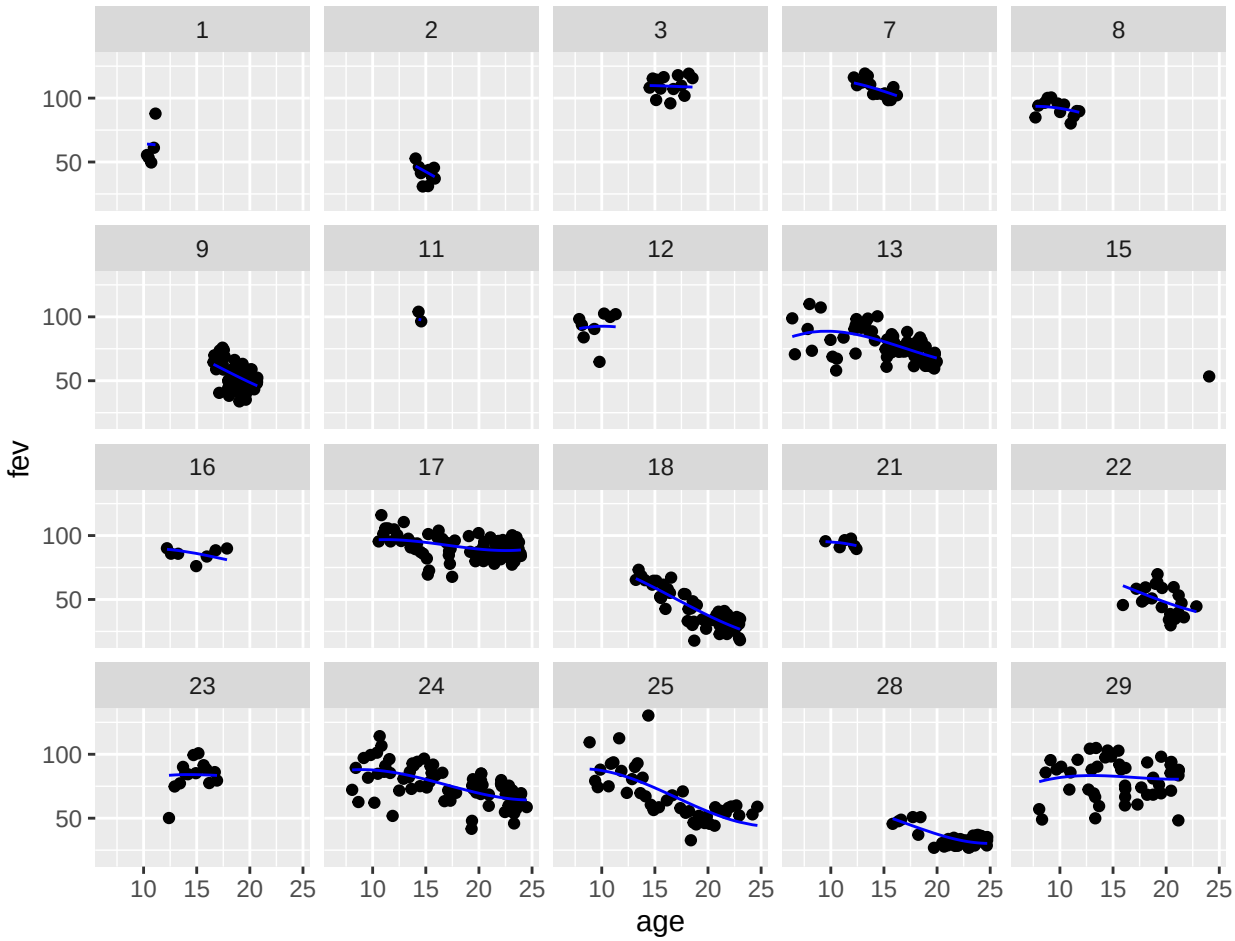
```
unique(dat$id)[1:10]
```

```
##  [1]  1  2  3  7  8  9 11 12 13 15
```

```
## selecting a few trajectories to plot
datp <- subset(dat, id < 30)

ggplot(data = datp, aes(x = age, y = fev)) + geom_point() +
    geom_line(aes(x = age, y = predi), alpha = 1, color="blue") +
    facet_wrap(~id)
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



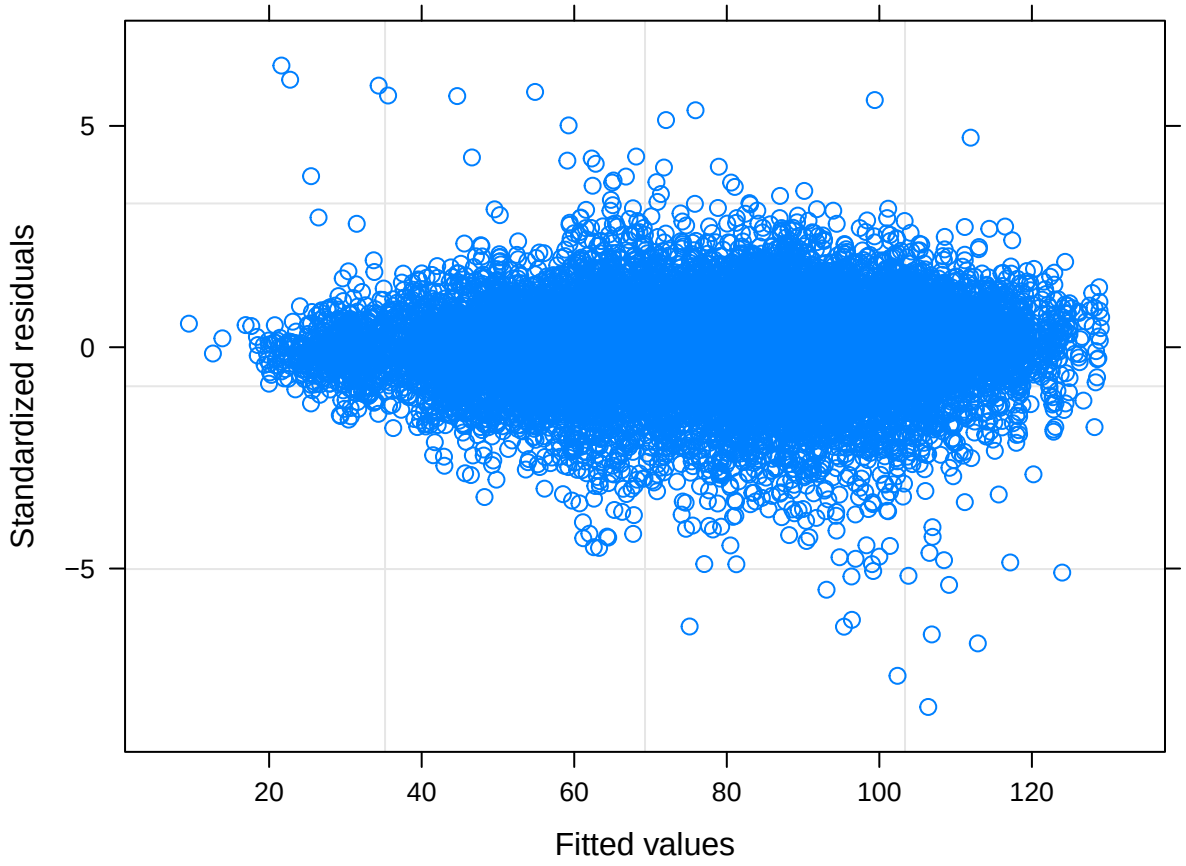

```
## Can also extract both marginal (PA) residuals  Y_i-X_i betahat,
## and conditional (SS) residuals Y_i-X_i betahat - Z_i bhat_i
## level = 0 gives PA, level = 1 gives SS (default), and level = 0:1 gives both;
## type = default is "raw" residuals; can get standardized "pearson" (type=pearson)
residuals(fit.lme.f.reml, level=0:1)[1:5, ]
```

```
##        fixed         id
## 1 -37.199235  -8.574335
## 2 -39.045622 -10.467918
## 3 -42.675327 -14.176281
## 4 -30.893224  -2.472838
## 5  -3.976658  24.396532
```

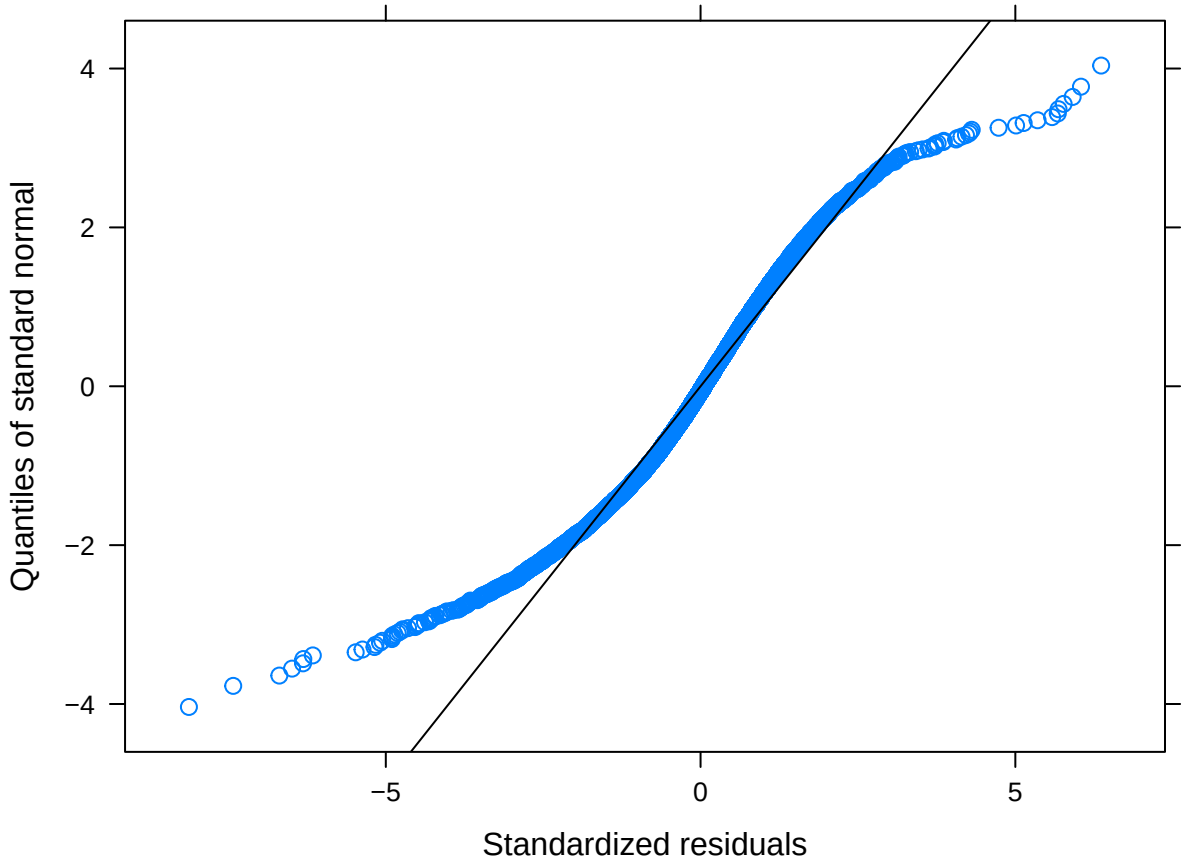
```
###   PA and SS pearson residuals -- the PA residuals differ from those
residuals(fit.lme.f.reml, level=0:1, type="pearson")[1:5, ]
```

```
##        fixed         id
## 1 -3.6594669 -0.8434984
## 2 -3.8411049 -1.0297792
## 3 -4.1981763 -1.3945887
## 4 -3.0391145 -0.2432649
## 5 -0.3912029  2.4000038
```

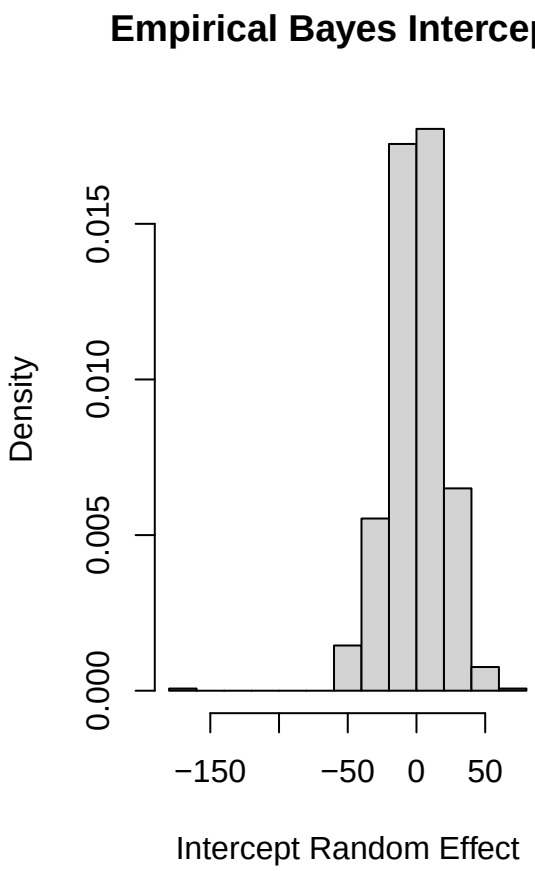
```
##  Plot SS residuals vs. predicted values
plot(fit.lme.f.reml, resid(. , type="p",level=1) ~ fitted(.,level=1) )
```

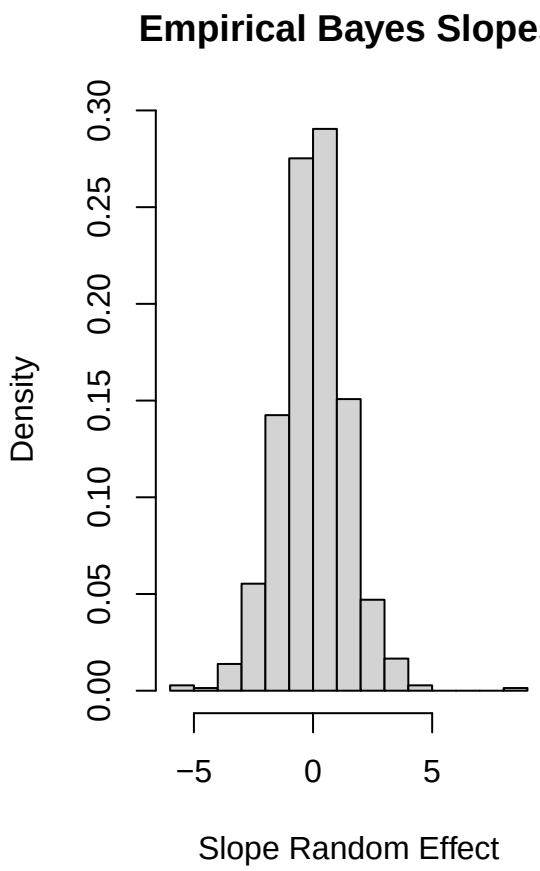



```
#  QQ plot of SS residuals
qqnorm(fit.lme.f.reml, ~ resid(. , type="p",level=1),abline=c(0,1))
```




```
##  histograms of random effects
par(mfrow=c(1,2))
hist(b.f.reml[,1],xlab="Intercept Random Effect",main="Empirical Bayes Intercepts",freq=FALSE)
hist(b.f.reml[,2],xlab="Slope Random Effect",main="Empirical Bayes Slopes",freq=FALSE)
```


Empirical Bayes Intercepts
Density
0.000
0.005
0.010
0.015
−150
−50
0
50
Intercept Random Effect


Empirical Bayes Slopes
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
−5
0
5
Slope Random Effect