Longitudinal Data Analysis

Yunsheng Lu

December 9, 2023

1 Introduction

This projects aims to analyze the following longitudinal data. The Six Cities Study of Air Pollution and Health was a longitudinal study designed to characterize lung growth as measured by changes in pulmonary function in children and adolescents, and the factors that influence lung function growth. A cohort of 13,379 children born on or after 1967 was enrolled in six communities across the U.S.: Watertown (Massachusetts), Kingston and Harriman (Tennessee), a section of St. Louis (Missouri), Steubenville (Ohio), Portage (Wisconsin), and Topeka (Kansas). Most children were enrolled in the first or second grade (between the ages of six and seven) and measurements of study participants were obtained annually until graduation from high school or loss to follow-up. At each annual examination, spirometry, the measurement of pulmonary function, was performed and a respiratory health questionnaire was completed by a parent or guardian.

The dataset contains a subset of the pulmonary function data collected in the Six Cities Study. The data consist of all measurements of FEV1, height and age obtained from a randomly selected subset of the female participants living in Topeka, Kansas. The random sample consists of 300 girls, with a minimum of one and a maximum of twelve observations over time:

Variable List:

Subject ID, Height, Age, Initial Height, Initial Age, Log(FEV1).

1	1.20	9.3415	1.20	9.3415	0.21511
1	1.28	10.3929	1.20	9.3415	0.37156
1	1.33	11.4524	1.20	9.3415	0.48858
1	1.42	12.4600	1.20	9.3415	0.75142
1	1.48	13.4182	1.20	9.3415	0.83291
1	1.50	15.4743	1.20	9.3415	0.89200
1	1.52	16.3723	1.20	9.3415	0.87129
2	1.13	6.5873	1.13	6.5873	0.30748
2	1.19	7.6496	1.13	6.5873	0.35066
2	1.49	12.7392	1.13	6.5873	0.75612
2	1.53	13.7741	1.13	6.5873	0.86710
2	1.55	14.6940	1.13	6.5873	1.04732
2	1.56	15.8220	1.13	6.5873	1.15373
2	1.57	16.6680	1.13	6.5873	0.92426
2	1.57	17.6318	1.13	6.5873	1.13462
3	1.18	6.9131	1.18	6.9131	0.43178
3	1.23	7.9754	1.18	6.9131	0.38526
3	1.30	8.9665	1.18	6.9131	0.59884
3	1.35	9.9877	1.18	6.9131	0.75142
3	1.47	11.0773	1.18	6.9131	0.96698
3	1.57	13.0678	1.18	6.9131	0.89609
3	1.59	14.1027	1.18	6.9131	1.01885
3	1.60	15.0801	1.18	6.9131	1.10526
3	1.60	16.0164	1.18	6.9131	1.08519

Figure 1: Dataset

Basically, we are interested in the following questions:

1. How does age affect the growth of the height for each object?
2. How doe age affect the change of lung functionality.

We will consider both the linear model and linear mixed effect model for both questions. For the first question, the response variable is height and the predictors include Age, Initial Age, Intial Height. For the second model, the response variable is log(FEV1), which is an indicator of the lung functionality, and the predictors share with the first question. The reason we don't include log(FEV1) for the first model and Height for the second model is that age is highly correlated with both (cor=0.887,0.897,respectively) and we are primarily interested in the effect of age.

We should notice that the measurements are unbalanced and irregular. Although most R packages can well handle the data of this type, this can bring extra difficulties, for instance, the outlier detection. In our analysis, we will first borrow tools from Functional PCA to perform exploratory data analysis, and then conduct the model fitting. Finally, a model diagnostic is included to further recognize the choice of our final models.

2 Exploratory (Functional) Data Analysis

Since we are interested in the effect of age on both log(FEV1) and Height, we will first visualize the paths of both responses over age for each subject:

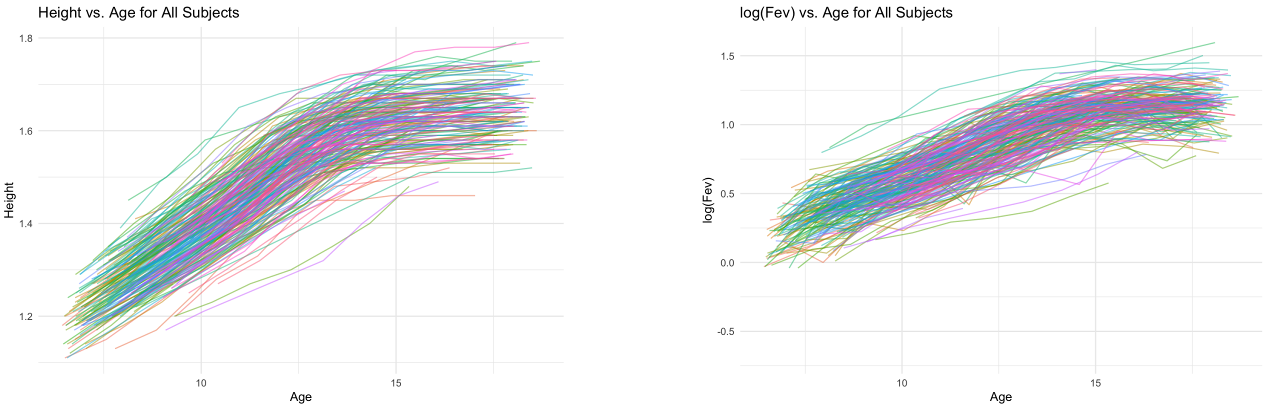
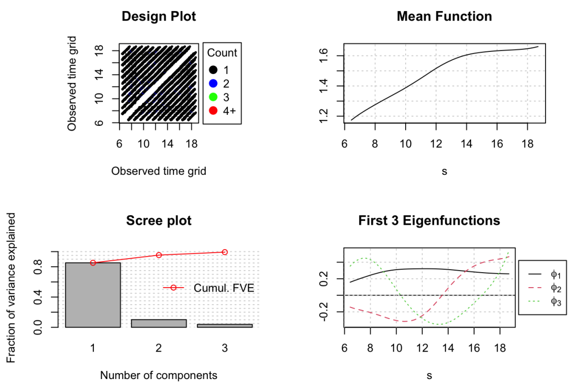
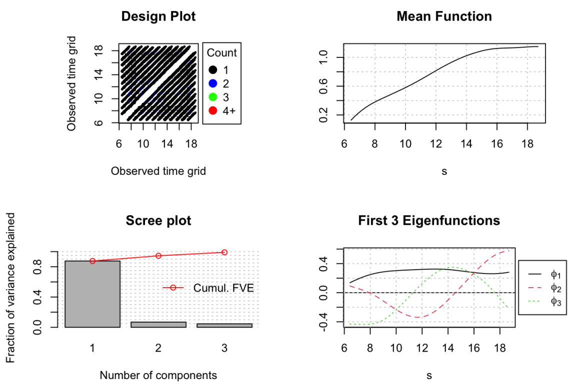


Figure 2: ggplot for Height and log(FEV1)

According to the visualizations, we can notice that there are potential outliers on the margins. We will treat these longitudinal data as functional data to capture outliers. We use **FPCA()** from library(*'fdapace'*) to first perform the functioal PCA, and the result is as the following:



(a) FPCA for Height



(b) FPCA for log(FEV1)

The default choice of the dimension K is the minimum that explains 99% percent of the data. To further recognize these choices, we also used **SelectK()** based on AIC, and the outputs are K=3 for both. Figure 4 is the projections of each subject on the first 3 components:

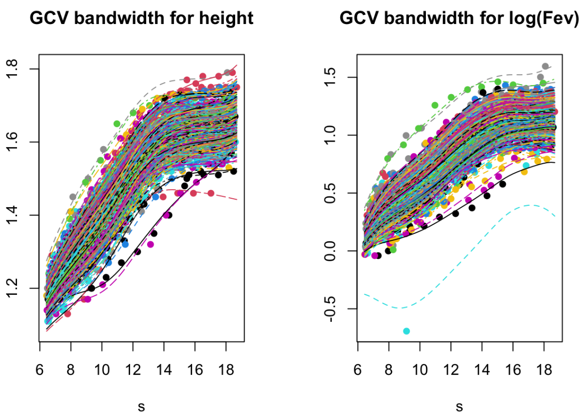
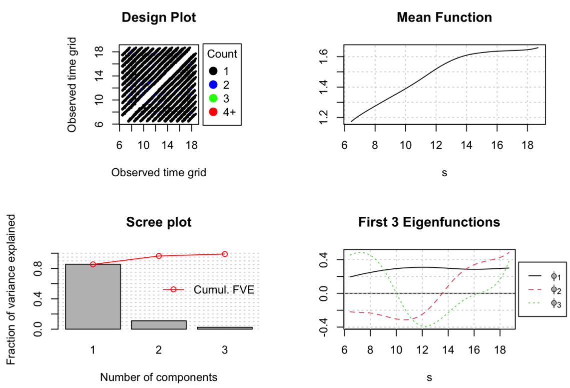
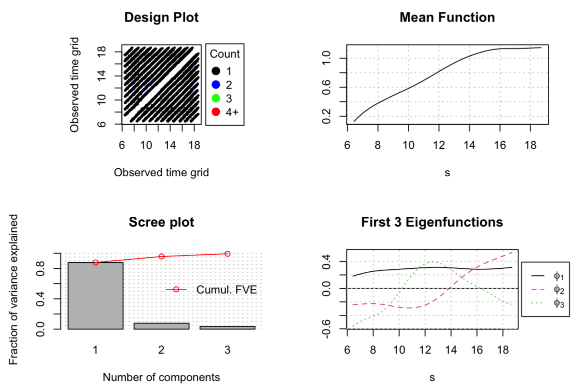


Figure 4: Dataset

We use **outliers.depth.pond()** to detect the outlier for functional data. This method is based on notion of depth, which measures the centrality of a data point, so a low depth indicates a potential outlier. A detailed introduction of this concept is in [1]. The results shows that the following subject IDs are potential outliers: $\{1, 10, 81, 108, 112, 139, 197, 207, 223, 230\}$, so we removed them, and we plot the smooth paths again based on the FPCA of the cleaned data:



(a) FPCA for Height (cleaned)



(b) FPCA for log(FEV1) (cleaned)

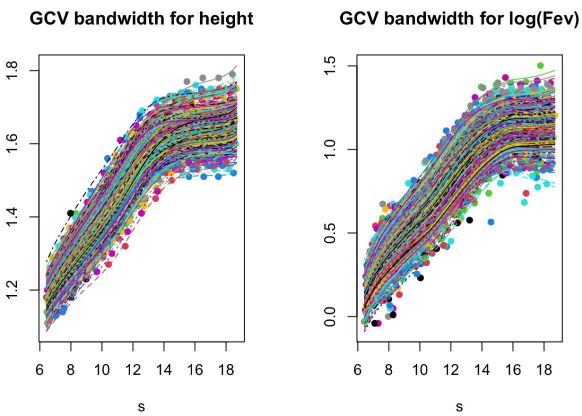


Figure 6: Dataset

3 Model Fitting

From Figure 2, we can notice that the increment of both Height and log(FEV1) first increase and then decrease, so we consider a quadratic regression. We specify the full linear model to be:

$$\text{Height}, \log(\text{FEV1}) \sim \text{Initial Height} * \text{Age}^C + \text{Initial Age} * \text{Age}^C + \text{Initial Height} * (\text{Age}^C)^2 + \text{Initial Age} * (\text{Age}^C)^2$$

where $\text{Age}^C = \text{Age} - 12.7$. Here 12.7 is the median of the mode (The full list of modes are in Appendix). Here we use this method to decorrelate the linear and quadratic term.

3.1 Covariance Structure Specification

Based on this full model, we use **gls()** to perform model fitting and the covariance structure selection. First of all, we assume the within-group homoscedasticity structure. However, unstructured covariance still failed to converge for both models, so we basically consider the following list of covariance structure: Variance Component, Compound Symmetry, and Continuous AR(1). We summarize the AIC, BIC and log likelihood as the following table for each covariance structure:

	Height			log(FEV1)		
	AIC	BIC	logLik	AIC	BIC	logLik
VC	-7845.866	-7790.239	3932.933	-2882.541	-2826.914	1451.21
CS	-8273.165	-8211.975	4147.582	-3975.476	-3914.286	1998.738
CAR(1)	-9659.351	-9598.161	4840.675	-4259.984	-4198.795	2140.992

Table 1: Linear Model Selection

We should note that in VC and CS, we specified the form to be $1|id$, and we specified $Age^C|id$ for CAR(1), because it admits non-integer input. All the indices indicate that we should specify the convariance structure to be continuous AR(1). This choice is reasonable because it admits the continuous time covariate.

3.2 Linear Model Selection

Based on CAR(1), we now perform the model selection for both models. Our method is *Backward Elimination*, where at each step, we remove the variable with p-value ≥ 0.05, and refit the model, so on and so forth, until every variable is significant. A detailed procedure is included in the appendix, and here we summarize our final linear model for both responses:

$$\text{Height} \sim \text{Initial Height} * \text{Age}^C + \text{Initial Height} * \text{Age}^C + \text{Initial Age} : (\text{Age}^C)^2$$

$$\log(\text{FEV1}) \sim \text{Intial Height} * \text{Age}^C + \text{Intial Age} * \text{Age}^C + \text{Intial Height} * (\text{Age}^C)^2$$

Estimation Details are demonstrated as the following:

```
## Coefficients:
##                                  Value   Std.Error   t-value p-value
## (Intercept)                  0.8060176 0.023824950  33.83082       0
## baseage                     -0.0500622 0.001881813 -26.60319       0
## age_centered                 0.0481952 0.005021929   9.59695       0
## baseht                       0.8946701 0.027296194  32.77637       0
## baseage:age_centered         0.0037196 0.000413654   8.99213       0
## age_centered:baseht         -0.0296846 0.005639317  -5.26387       0
## baseage:I(age_centered^2)   -0.0004942 0.000009015 -54.82659       0
```

(a) Height

```
## Coefficients:
##                                 Value  Std.Error    t-value p-value
## (Intercept)                -0.4756576 0.10341665  -4.599430  0.0000
## baseage                    -0.0972844 0.00701344 -13.871128  0.0000
## age_centered                0.0779349 0.01953055   3.990411  0.0001
## baseht                      1.6949025 0.11146382  15.205854  0.0000
## I(age_centered^2)           0.0108803 0.00425424   2.557513  0.0106
## baseage:age_centered        0.0083268 0.00161823   5.145630  0.0000
## age_centered:baseht        -0.0472213 0.02190225  -2.156001  0.0312
## baseht:I(age_centered^2)   -0.0142882 0.00337140  -4.238075  0.0000
```

(b) log(FEV1)

3.3 Linear Mixed Effect Model Selection

Now we try to see if adding random effects will improve the model fitting for the chose linear models. We use **lme()** to perform our model fitting. In general, we consider two cases:

- Add random intercept.
- Add random intercept + random slope

Unfortunately, CAR(1) failed to converge for adding both random intercept and random slope for both models. Thus, no random effects VS random intercept is made based on CAR(1), but random intercept VS random intercept + random slope is made based on VC.

Again, we summarize AIC, BIC and log-Likelihood for both height and log(FEV1) as the following table:

	Height			log(FEV1)		
	AIC	BIC	logLik	AIC	BIC	logLik
No Random Effect+CAR(1)	-9659.386	-9609.322	4838.693	-4261.969	-4206.342	2140.985
Random Intercept+CAR(1)	-9657.386	-9601.759	4838.693	-4324.731	-4263.542	2173.366
No Random Effect+VC	-7844.771	-7800.269	3930.385	-2883.46	-2833.395	1450.730
Random Intercept+VC	-8271.995	-8221.931	4144.997	-3975.552	-3919.926	1997.776
Random Intercept+Random Slope+VC	-8822.103	-8760.913	4422.051	-4052.125	-3985.373	2038.063

Table 2: Random Effect Selection

Notice that in the following context, when comparing the linear mixed effect models with nested random effects, suppose the parameters are q and $q+1$, respectively, then the likelihood ratio test is base on 50:50 mixture of χ^2_q-distribution and χ^2_{q+1}-distribution. We summarize the results of likelihood ratio test as the following table:

	Height		log(FEV1)	
	Test Statistics	p-value	Test Statistics	p-value
Random Intercept+CAR(1)	0	0.5	64.8	4.44e-16
Random Intercept+VC	429	0	1094	0
Random Slope+VC	554	0	81	0

Table 3: Likelihood Ratio Test Summary

If we simply consider the case of continous AR(1), then the model fitting for Height suggests that there is no need to include Random Intercept, while the model fitting for log(FEV1) indicates that a random intercept will significantly boost the performance.

Remark 1. The following are some interpretations of the results from the table above.

1. Notice that the identity likelihood between linear model and linear model with random intercept for continuous AR(1) is not a coincidence: Essentially speaking, adding a random intercept provides the extra flexibility for model fitting in the sense that every subject can start at different level. However, notice that the "Intial Height" variable serves the same purpose.

2. Notice that when we assume the covariance structure to be Variance Compound, we can see an improvement when we included random intercept. From some perspectives, this means that Variance Component is wrong and Continuous AR(1) indeed captures the underlying truth.

3. We still included VC, because we are concerned if adding a random slope will make a difference to the model fitting. However, suggested by AIC, BIC, and log(FEV1), even if a random slope is considered, for both responses, these more detailed linear mixed effect models under VC actually worse than the model with CAR(1) under any circumstances.

Finally, by taking random effect into consideration, we finalized our model selection for Height and log(FEV1):

$$\text{Height}_i \sim \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Age}_{ij} : (\text{Age}^C_{ij})^2$$

$$\log(\text{FEV1})_i \sim \text{Intial Height}_{ij} * \text{Age}^C_{ij} + \text{Intial Age}_{ij} * \text{Age}^C_{ij} + \text{Intial Height}_{ij} * (\text{Age}^C_{ij})^2 + b_i$$

The following is some fitting details for the log(FEV1) (We didn't update our Height Model):

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -4324.731 -4263.542 2173.366
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:   0.0765799 0.08299468
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.5543551
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *      I(age_centered^2)
##                               Value  Std.Error   DF    t-value p-value
## (Intercept)              -0.4617239 0.10565333 1630  -4.370179  0.0000
## baseage                  -0.0983031 0.00773473  287 -12.709319  0.0000
## age_centered              0.0894918 0.01495700 1630   5.983271  0.0000
## baseht                    1.6919647 0.11813115  287  14.322764  0.0000
## I(age_centered^2)         0.0111648 0.00368091 1630   3.033176  0.0025
## baseage:age_centered      0.0080803 0.00125115 1630   6.458302  0.0000
## age_centered:baseht      -0.0538530 0.01657942 1630  -3.248184  0.0012
## baseht:I(age_centered^2) -0.0145157 0.00292325 1630  -4.965610  0.0000
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.587
## age_centered              0.069  0.089
## baseht                   -0.938 -0.829 -0.087
## I(age_centered^2)        -0.389 -0.097 -0.064  0.313
## baseage:age_centered      0.026  0.024  0.413 -0.025  0.146
## age_centered:baseht      -0.063 -0.077 -0.904  0.076 -0.013 -0.760
## baseht:I(age_centered^2)  0.385  0.103  0.074 -0.314 -0.998 -0.158  0.012
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.53494531 -0.54680099  0.03047478  0.63011954  2.48426377
##
## Number of Observations: 1925
## Number of Groups: 290
```

Figure 8: Linear Mixed Effect Model for log(FEV1)

For detailed information for covariance matrix, we can use **getVarCov()** to get the variance-covariance matrix for random intercept and the marginal for each individual by specifying *"individuals"*, from which we can further calculate the residual covariance matrix. The details of computations are shown in the Appendix,and here we include the result for the second subject (id=1 is an outlier). Notice that the random intercept follows $\mathcal{N}(0, 5.86 \times 10^{-3})$, and:

```
Marginal variance covariance matrix
        1       2       3       4       5       6       7       8
1 0.01275 0.00955 0.00605 0.00596 0.00592 0.00589 0.00588 0.00587
2 0.00955 0.01275 0.00621 0.00605 0.00597 0.00592 0.00590 0.00588
3 0.00605 0.00621 0.01275 0.00961 0.00804 0.00698 0.00654 0.00625
4 0.00596 0.00605 0.00961 0.01275 0.00987 0.00792 0.00711 0.00657
5 0.00592 0.00597 0.00804 0.00987 0.01275 0.00941 0.00801 0.00708
6 0.00589 0.00592 0.00698 0.00792 0.00941 0.01275 0.01005 0.00823
7 0.00588 0.00590 0.00654 0.00711 0.00801 0.01005 0.01275 0.00977
8 0.00587 0.00588 0.00625 0.00657 0.00708 0.00823 0.00977 0.01275
  Standard Deviations: 0.113 0.113 0.113 0.113 0.113 0.113 0.113 0.113
```

(a) Marginal

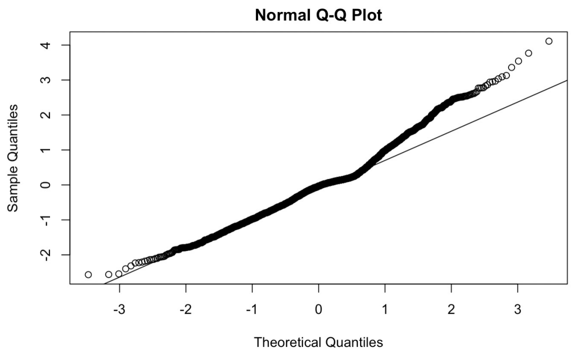
```
          1        2        3        4        5        6        7        8
1 6.89e-03 3.68e-03 0.000183 9.93e-05 5.77e-05 2.97e-05 1.80e-05 1.02e-05
2 3.68e-03 6.89e-03 0.000342 1.86e-04 1.08e-04 5.55e-05 3.37e-05 1.91e-05
3 1.83e-04 3.42e-04 0.006888 3.74e-03 2.17e-03 1.12e-03 6.78e-04 3.84e-04
4 9.93e-05 1.86e-04 0.003741 6.89e-03 4.00e-03 2.06e-03 1.25e-03 7.07e-04
5 5.77e-05 1.08e-04 0.002174 4.00e-03 6.89e-03 3.54e-03 2.15e-03 1.22e-03
6 2.97e-05 5.55e-05 0.001118 2.06e-03 3.54e-03 6.89e-03 4.18e-03 2.37e-03
7 1.80e-05 3.37e-05 0.000678 1.25e-03 2.15e-03 4.18e-03 6.89e-03 3.90e-03
8 1.02e-05 1.91e-05 0.000384 7.07e-04 1.22e-03 2.37e-03 3.90e-03 6.89e-03
```

(b) Residual

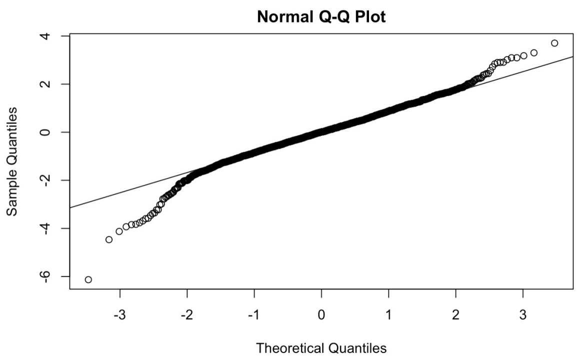
Furthermore, for each individual, we can use **ranef()** to extract the estimated random effect, and **fitted(model,level=0)** and **fitted(model,level=1)** to get the estimation of fixed effect and fixed effect+random effect, respectively. We print out the first several estimates in the appendix.

4 Model Diagnostic

We further conduct the model diagnostic to check the model assumptions. First we use the QQ-plot to check the normality assumption. Notice that in both of our finalized models, we assumed a continuous AR(1) covariance structure, thus we need to transform he residuals by multiplying the lower-triangular matrix in the Cholesky decomposition of the residual covariance matrix to decorrelate the residuals. The results for both are demonstrated as the following:

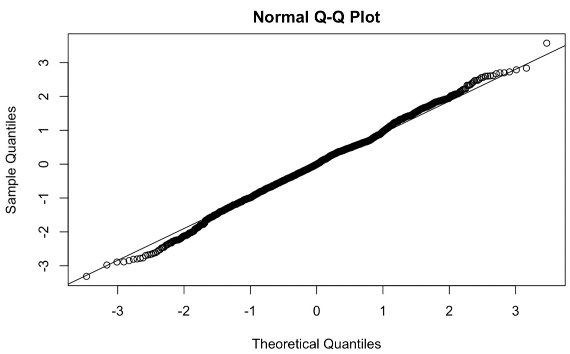


(a) Height+CAR(1)

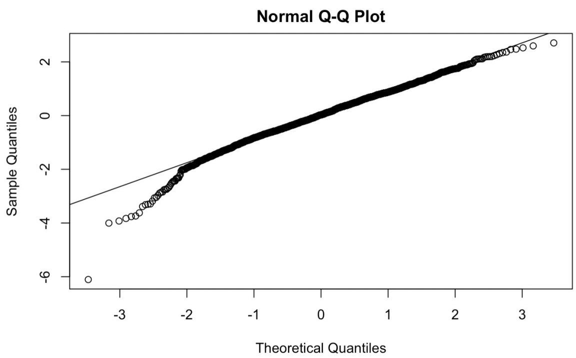


(b) log(FEV1)+CAR(1)

Unfortunately, we can see that especially for Height Model, it doesn't demonstrate the perfect normality. We would like to do a little bit investigations. The following is the QQ-plot of the residual under Variance Component covariance structure:



(a) Height+VC



(b) log(FEV1)+VC

In fact, we can see the scatter plots for both variables fit the lines very well. This is a little bit out of our expectation, because the previous analysis has shown that CAR(1) provides much better model fitting than VC according to AIC, BIC, and log-Likelihood. Thus my conjecture would be:

Remark 2. In general, I suspect the deviations in the first two QQ-plots actually come from the numerical Cholesky's decomposition.

1. First, our covariance matrix is large, because there are 1925 observations in total, and this can cause numerical instability.

2. It's **important** to note that our residual covariance matrix can possibly be almost singular. I say "possibly" because I didn't find a handy way to extract the whole covariance matrix from my output. However, if we simply investigate the residual covariance of the subject with id=2, the determinant of the Cholesky decomposition is 2.54×10^{-14}, which is included in the appendix.

5 Model Interpretation

First, recall that we are interested in the effect of age for both the height and lung growth for each subject.

5.1 Height

Now let's recall our final Height model and its summary:

$$\text{Height}_i \sim \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Age}_{ij} : (\text{Age}^C_{ij})^2$$

```
## Coefficients:
##                                Value    Std.Error   t-value p-value
## (Intercept)                0.8060176 0.023824950  33.83082       0
## baseage                   -0.0500622 0.001881813 -26.60319       0
## age_centered               0.0481952 0.005021929   9.59695       0
## baseht                     0.8946701 0.027296194  32.77637       0
## baseage:age_centered       0.0037196 0.000413654   8.99213       0
## age_centered:baseht       -0.0296846 0.005639317  -5.26387       0
## baseage:I(age_centered^2) -0.0004942 0.000009015 -54.82659       0
```

(a) Height

We need to be cautious when interpreting the baseline effect. Our response is Height, and "Initial Height" itself already includes all we want to know about the height. In principles, 1 unit change in initial height should lead to 1 unit change in the height at the baseline, but we obtain 0.895 change in the height in our estimation. Of course, the reason is simple: we are also trying to interpolating the baseline height using other variable for example, the baseline age. Thus, we will instead focus on explaining the time effect in our model.

Here we have both the linear time variable and quadratic time variable. Notice that if we take the derivative on both side:

$$\frac{d\text{Height}}{d\text{Age}^C} = 0.048 + 0.0037 \cdot \text{Initial Age} - 0.0297 \cdot \text{Initial Height} - 0.0005 \cdot \text{Initial Age} \cdot \text{Age}^C$$

Thus our interpretation of the effect of the age would be the following:

1. When the age of a subject increase by 1 unit, averagely speaking, we will expect an increasing of the subject's height by (0.048+0.0037 Initial Age - 0.0297 Initial Height -0.0005 Initial Age $\times$ Age_0), where Age_0 is the subject's current Age - 12.7 (don't forget the decorrelation at the beginning). Meanwhile, for the next unit in the increasing of age, we will expect (0.0005 Intial Age) decreasing in the increasing of the height compared with this time.

2. Suppose there are two individuals in the study, A and B, and A is one unit taller than B when entering the study. Now during each time interval of 1 unit, we will expect (in the sense of the average) that A will experience 0.0297 less increment in height compared with B.

3. Suppose there are two individuals in the study, A is a years old and B is b years old ($a > b$). After k years they enter the study, they are $a + k$, $b + k$ years old respectively, and A is likely to experience a:

$$0.0037(a - b) - 0.0005(a(a + k) - b(b + k))$$

more increasing in height than B at this kth year.

5.2 log(FEV1)

We also recall the final log(FEV1) model:

$$\log(\text{FEV1})_i \sim \text{Intial Height}_{ij} * \text{Age}^C_{ij} + \text{Intial Age}_{ij} * \text{Age}^C_{ij} + \text{Intial Height}_{ij} * (\text{Age}^C_{ij})^2 + b_i$$

The estimation is referred to Figure 8. The model interpretation of the fixed effect is essentially the same as the Height model. In addition, it has random effect: a random intercept follows $\mathcal{N}(0, 5.86 \times 10^{-3})$. The model interpretation of the random effect is as the following: suppose there is an individual in the study, x years old and y units tall when entering the study, then the interval $(-0.46 - 0.098x + 1.69y - 1.96 \cdot \sqrt{5.86 \times 10^{-3}}, -0.46 - 0.098x + 1.69y + 1.96 \cdot \sqrt{5.86 \times 10^{-3}})$ has 95% of probability to cover his or her true log(FEV1).

6 Conclusion

Notice that the measurements in our study is irregular and unbalanced. As we mentioned in the class, the identification of outlier is not easy for longitudinal data, and the nature of our data makes it more challenging. To solve this problem, we consider every variable to be continuous, and in particular, we assume the time to be continuous, and consequently, we incorporate functional data analysis techniques. We first smooth the sparse data and then perform functional PCA. We used the first 3 components for both responses, and then removed the outlier according to their depths.

When conducting the model fitting, there are several interesting phenomena. The first is that the log likelihood didn't change when we added the random intercept for the Height model, the reason is that we also included the variable "Initial Height," which already served as a "random intercept" to make it possible for different individuals to start at different point. The second is that although with full mean structure, we shows that continuous AR(1) is much better than Variance Compound, the QQ-plot in the model diagnostics actually shows the reverse. Our explanation is that the continuous AR(1) in our case is actually closed to singular (plus that it's a very large matrix), both resulted in an unstable numerical solution of the inverse of its Chloesky's decomposition.

Finally, we specified our model as the following:

$$\text{Height}_i \sim \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Height}_{ij} * \text{Age}^C_{ij} + \text{Initial Age}_{ij} : (\text{Age}^C_{ij})^2$$

$$\log(\text{FEV1})_i \sim \text{Intial Height}_{ij} * \text{Age}^C_{ij} + \text{Intial Age}_{ij} * \text{Age}^C_{ij} + \text{Intial Height}_{ij} * (\text{Age}^C_{ij})^2 + b_i$$

In our analysis, we have shown that Initial Height, Initial Age, and Age all have a significant impact on both the growth of Height and Lung Functionality of the subjects involved. Moreover, the effect of Age on both the Height and the Lung Functionality shows a quadratic trend, which means although both are increasing, the increasing rates are decreasing, and this should be intuitive. In fact, the model interpretation in our analysis is extremely challenging, which is due to three reasons:

1. Two continuous group effects have interaction with the quadratic time.
2. Initial Age and Age are two high related concepts.
3. We centralized the Age at the beginning to decorrelate the linear and the quadratic Age.

7 Discussion and Limits

First, unstructured Covariance Structure is excluded from consideration due to the failure to converge for our R packages. Similarly, when considering the random effects, the selected CAR(1) failed to converge when adding the random slope. Although we have shown that adding a random slope under variance compound is worse then simply a linear model with CAR(1), it's better to have a fair comparison.

Moreover, if our interpolation is true for the phenomenon in model diagnostics, then some advanced techniques is needed because both QQ-plots and Loess curve depends on the notion of "transformed residuals." However, if the problems come from data itself, then techniques like Box-Cox power transformation should be considered, but this will bring extra difficulties in model interpretation, which is already, very very challenging.

References

[1] Sara López-Pintado and Juan Romo. "On the concept of depth for functional data". In: *Journal of the American statistical Association* 104.486 (2009), pp. 718–734.

8 Appendix (Next Page)

6250FinalProject

Yunsheng Lu

2023-12-04

Preprocessing

Read the data

```
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```
library(haven)
data<- read_dta("fev1.dta")
```

Decorrelating linear and quadratic term

```
Modes <- function(x) {
  ux <- unique(x)
  tab <- tabulate(match(x, ux))
  ux[tab == max(tab)]
}
Modes(data$age)
```

```
## [1] 12.0712 12.7310 15.4114 14.3874 11.4908
```

```
#take the median=12.7
data$age_centered=data$age-12.7
```

Exploratary Data Analysis

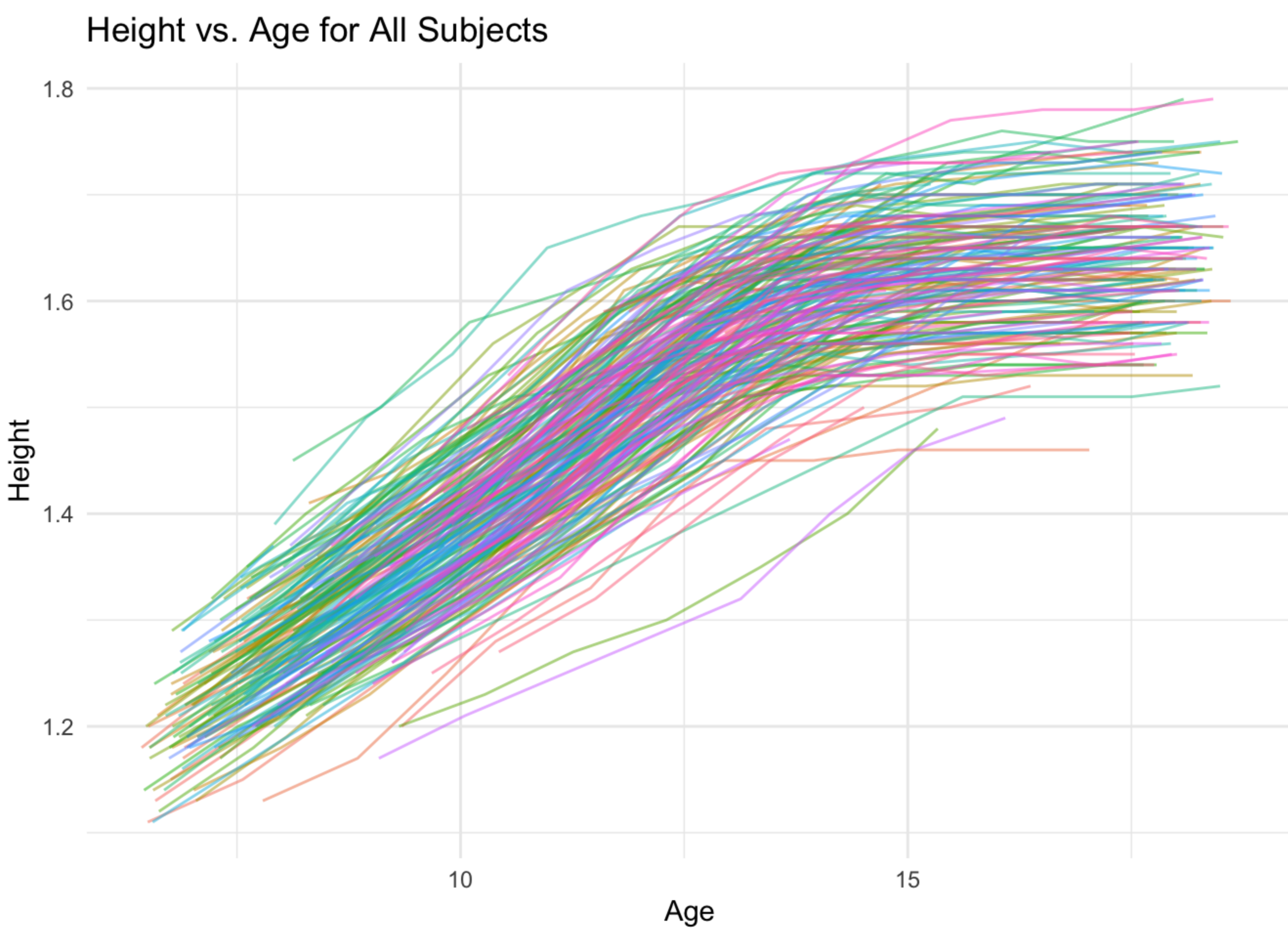
Data Visualization

```
library(ggplot2)
```

Height

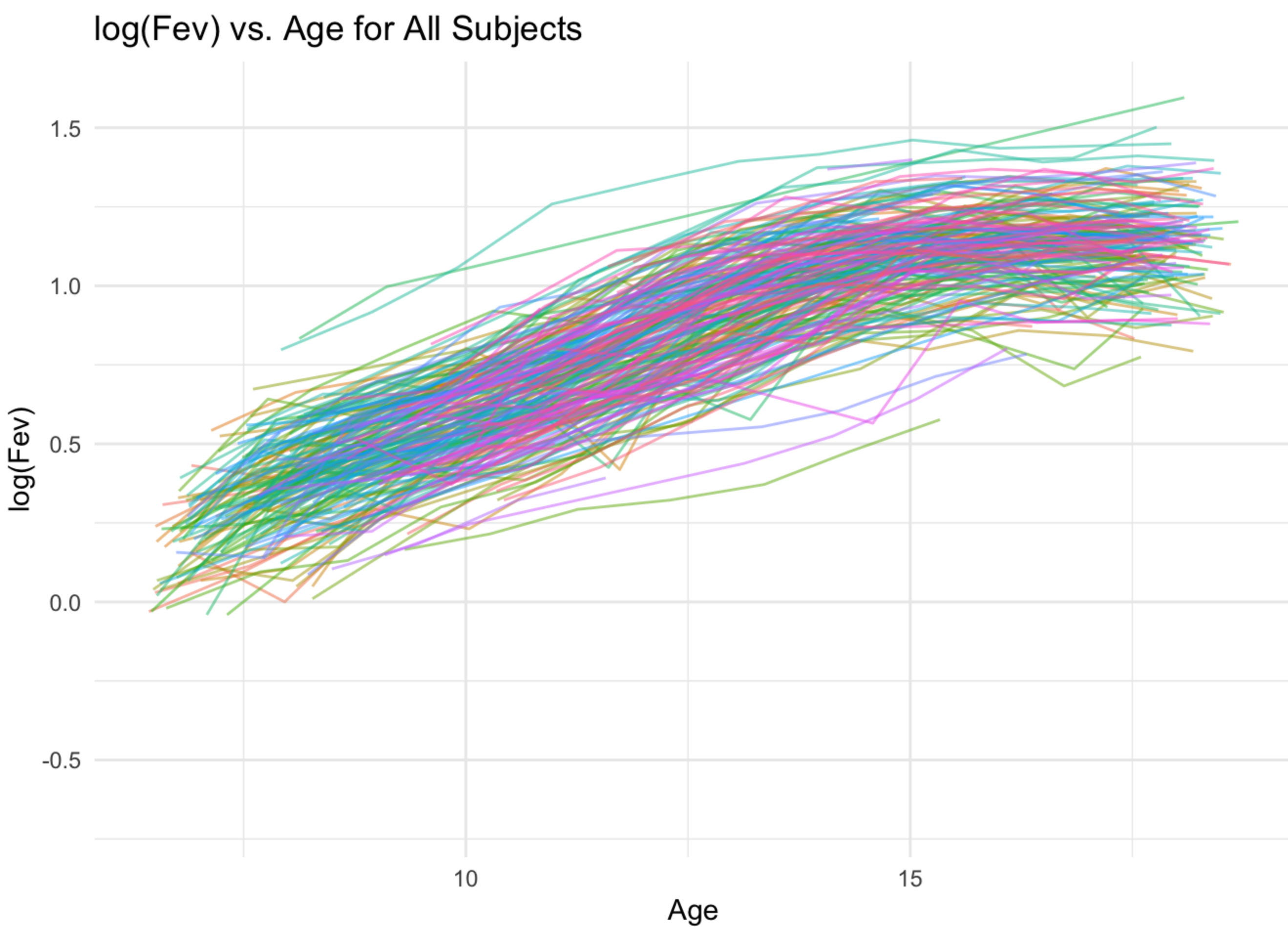
```
ggplot(data, aes(x = age, y = ht, group =id, color = factor(id))) +
  geom_line(alpha = 0.5, size = 0.5) +  # Adjust alpha and size as needed
  labs(x = "Age", y = "Height", title = "Height vs. Age for All Subjects") +
  theme_minimal() +
  scale_color_discrete(name = "Subject") +
  theme(legend.position = "none")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



log(Fev)

```
ggplot(data, aes(x = age, y = logfev1, group =id, color = factor(id))) +
  geom_line(alpha = 0.5, size = 0.5) +  # Adjust alpha and size as needed
  labs(x = "Age", y = "log(Fev)", title = "log(Fev) vs. Age for All Subjects") +
  theme_minimal() +
  scale_color_discrete(name = "Subject") +
  theme(legend.position = "none")
```



FPCA

FPCA for Height

```
library(fdapace)
library(fda.usc)
```

```
## Loading required package: fda
```

```
## Loading required package: splines
```

```
## Loading required package: fds
```

```
## Loading required package: rainbow
```

```
## Loading required package: MASS
```

```
## Loading required package: pcaPP
```

```
## Loading required package: RCurl
```

```
## Loading required package: deSolve
```

```
## Warning: package 'deSolve' was built under R version 4.2.3
```

```
##
## Attaching package: 'fda'
```

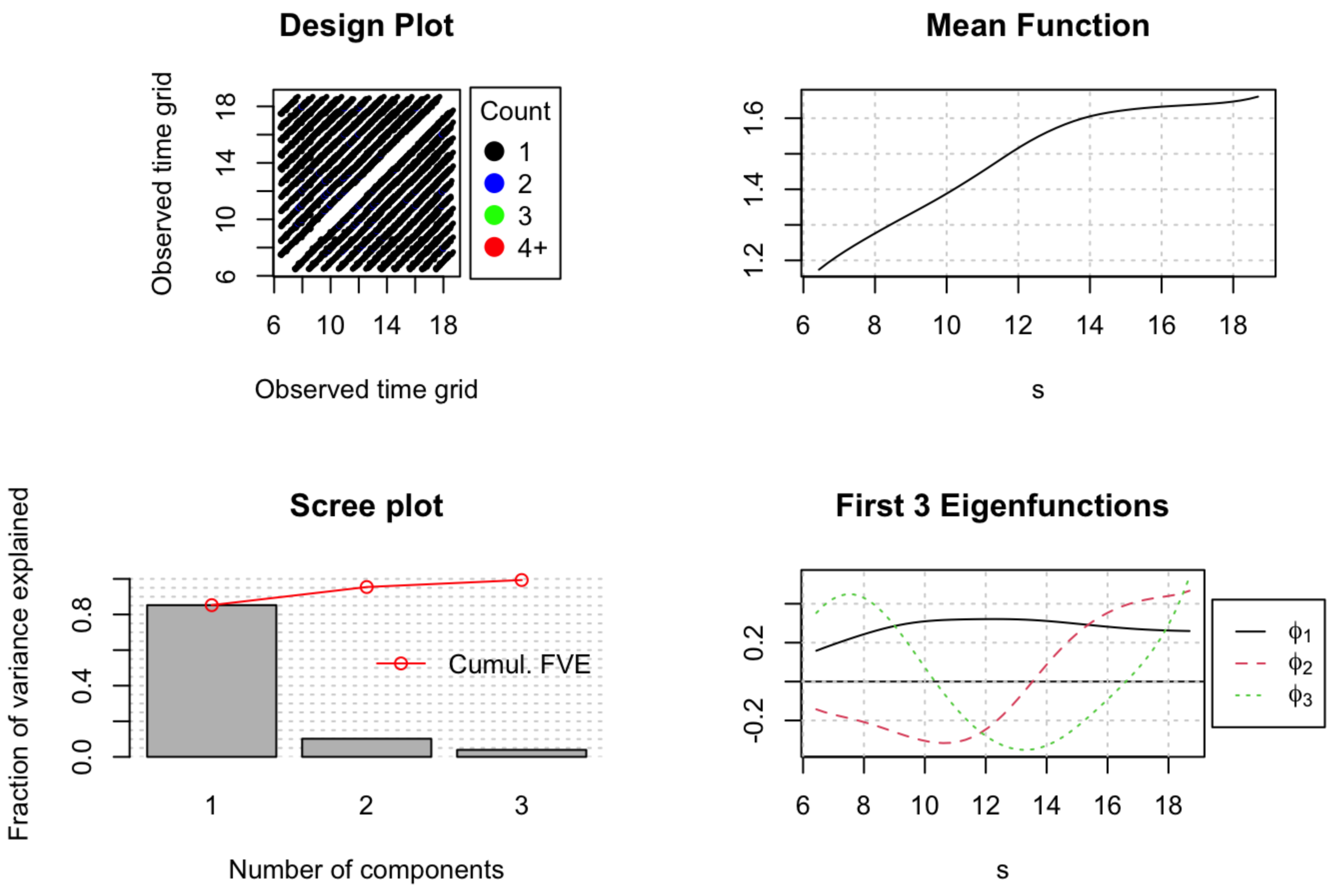
```
## The following object is masked from 'package:graphics':
##
##     matplot
```

```
## Loading required package: mgcv
```

```
## This is mgcv 1.8-41. For overview type 'help("mgcv-package")'.
```

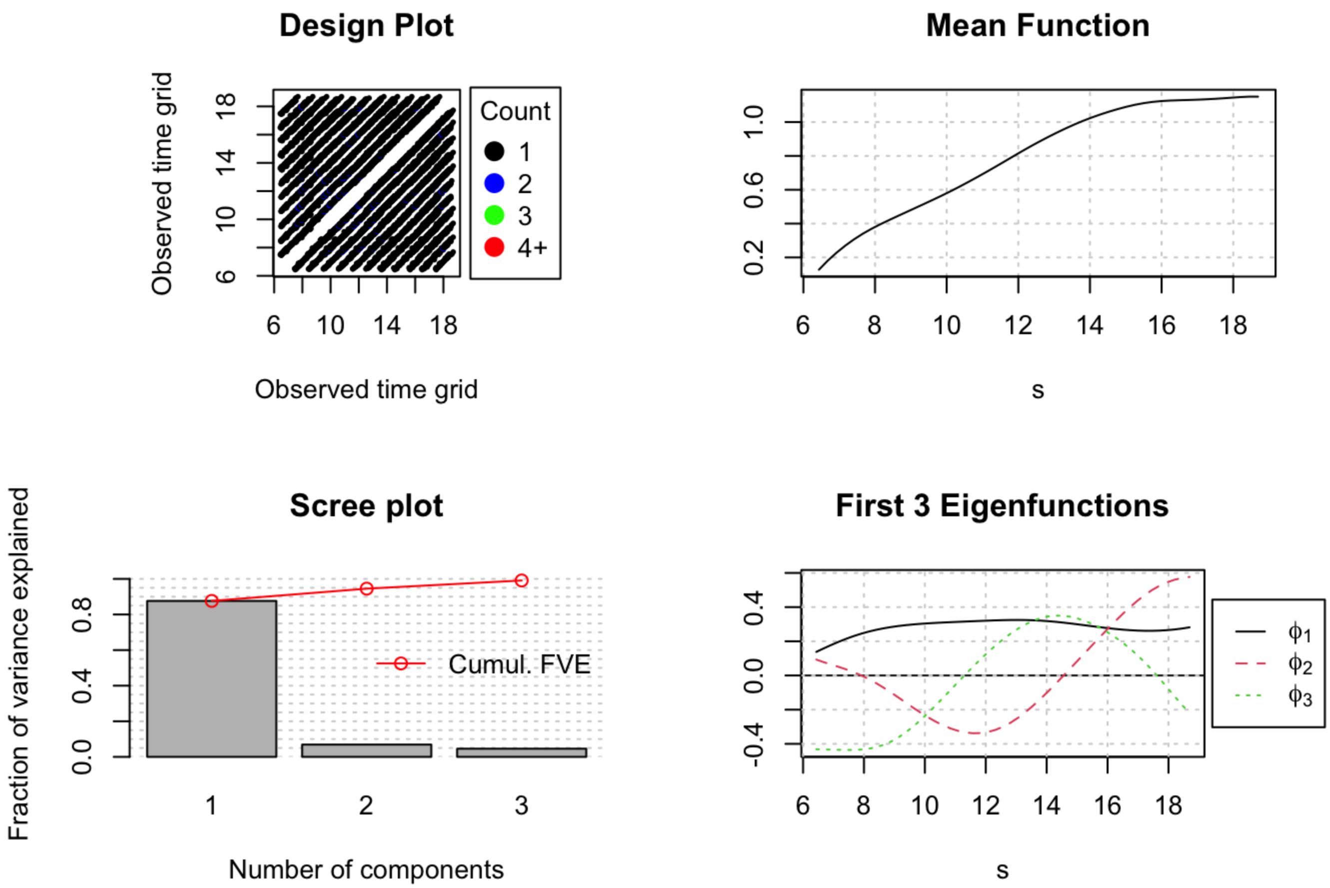
```
##  fda.usc is running sequentially usign foreach package
##  Please, execute ops.fda.usc() once to run in local parallel mode
##  Deprecated functions: min.basis, min.np, anova.hetero, anova.onefactor, anova.RPm
##  New functions: optim.basis, optim.np, fanova.hetero, fanova.onefactor, fanova.RPm
## ----------------------------------------------------------------------------------
```

```
Age <- split(data$age, data$id)
Height <- split(data$ht, data$id)
ID<- as.list(1:length(unique(data$id)))
FD1=list(Lid=ID,Ly=Height,Lt=Age)
fpcaheight=FPCA(FD1$Ly,FD1$Lt)
plot(fpcaheight)
```



FPCA for log(Fev)

```
Age <- split(data$age, data$id)
Fev <- split(data$logfev1, data$id)
ID<- as.list(1:length(unique(data$id)))
FD2=list(Lid=ID,Ly=Fev,Lt=Age)
fpcafev=FPCA(FD2$Ly,FD2$Lt)
plot(fpcafev)
```



K-Selection

```
SelectK(fpcaheight,criterion='BIC')
```

```
## $K
## [1] 3
##
## $criterion
## [1] -13579.63
```

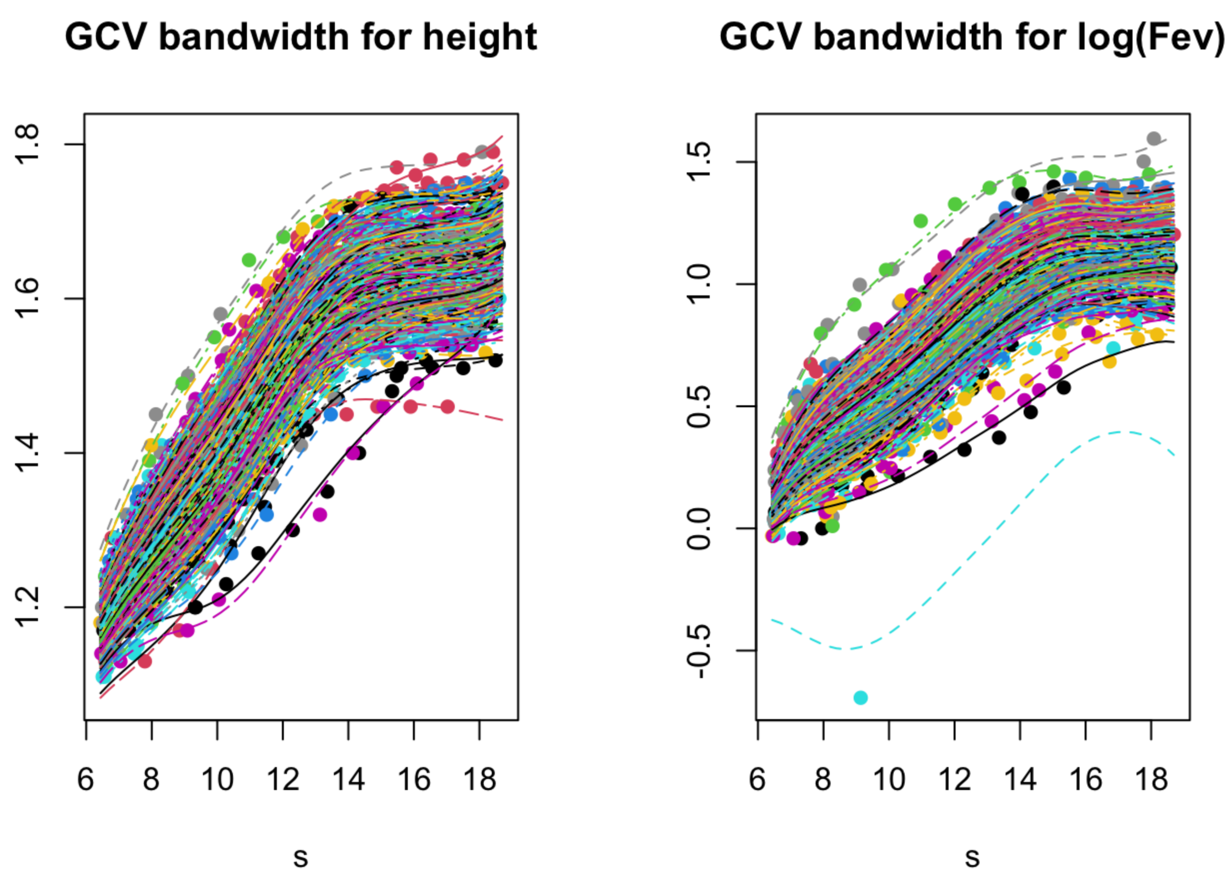
```
SelectK(fpcafev,criterion='AIC')
```

```
## $K
## [1] 3
##
## $criterion
## [1] -7931.743
```

EDA for Functional Data

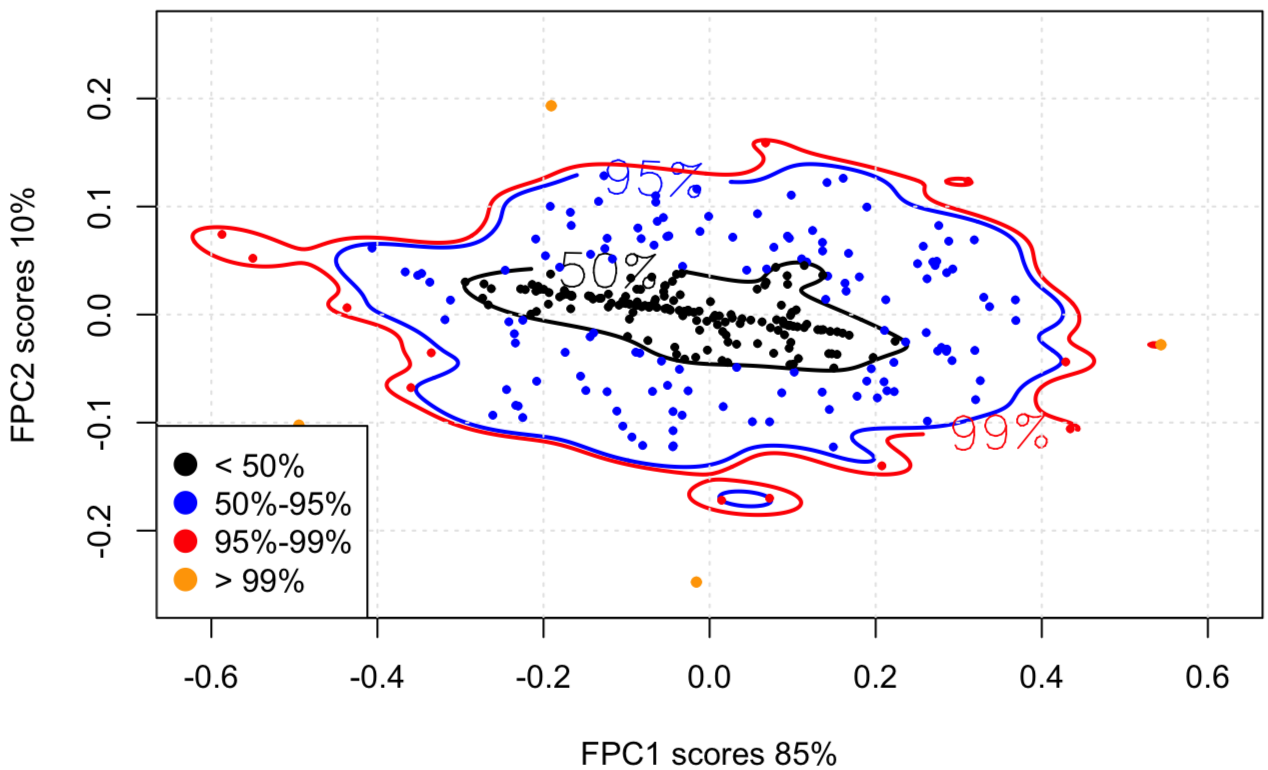
Smooth Paths Visualization

```
par(mfrow=c(1,2))
CreatePathPlot( fpcaheight, main = "GCV bandwidth for height", pch = 16)
CreatePathPlot( fpcafev, main = "GCV bandwidth for log(Fev)", pch = 16)
```

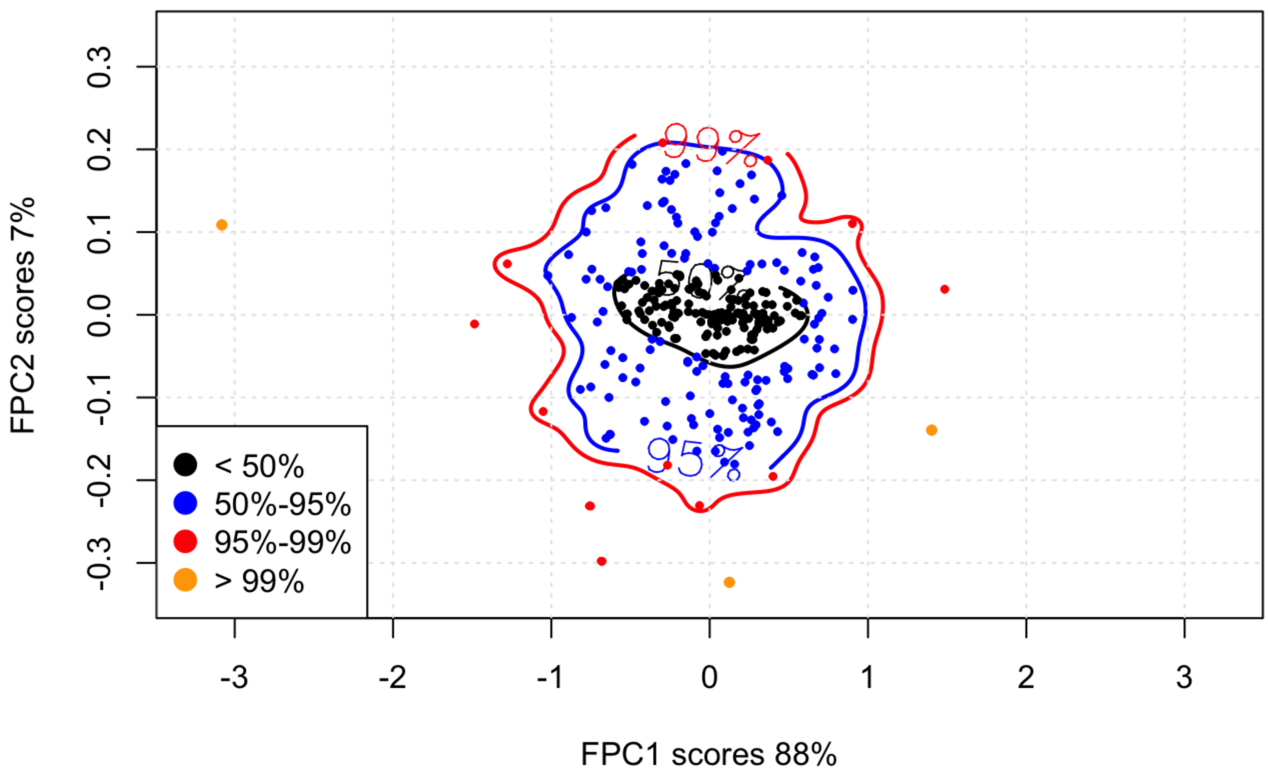


Outlier Visualization

```
par(mfrow=c(1,1))
CreateOutliersPlot(fpcaheight,optns=list(K=3,variant='KDE'))
```




```
CreateOutliersPlot(fpcafev,optns=list(K=3,variant='KDE'))
```



Outlier Detection

Height

```
fdaheight=fdata(t(fpcaheight$phi %*% t(fpcaheight$xiEst)))
outliers.depth.pond(
  fdaheight,
  nb = 200,
  smo = 0.05,
  quan = 0.5,
  dfunc = depth.mode
)
```

```
## $outliers
## [1] "10"  "81"  "108" "112" "139" "230" "1"
##
## $dep.out
## [1] 0.8615018 0.8952600 1.4221514 0.8747433 2.2970757 0.8153464 2.6175350
##
## $iteration
## [1] 1 1 1 1 1 1 2
##
## $quantile
##      50%
## 2.621135
##
## $Dep
##          1          2          3          4          5          6          7
##  2.9349154 20.2451921 34.3644596 31.4362225 10.7666760 22.8782326 28.9938325
##          8          9         10         11         12         13         14
## 34.4050724 26.7775896  0.8615018 26.2886847 43.1622951 33.8937919 35.5843168
##         15         16         17         18         19         20         21
## 38.0887774 20.9435679 42.2321891 37.8079961 42.0099149 39.3593371 12.6171347
##         22         23         24         25         26         27         28
## 19.9706185 18.8396936 43.1715428 30.2336417 42.0343783 25.4928003 13.6558534
##         29         30         31         32         33         34         35
## 12.5562180 28.4857490 37.3728811 24.2010581 38.4945186 39.3510176 14.5472584
##         36         37         38         39         40         41         42
## 38.7704589 26.7308099 23.5050285 15.2210980 34.5270649 15.4694317 40.4783525
##         43         44         45         46         47         48         49
## 18.2434405 25.4189278 28.0834591  9.1106579 10.1924145 35.0111321 30.9711406
##         50         51         52         53         54         55         56
## 31.5248765 34.9717411 43.1871489 40.5128146 21.1960535  4.8033191 37.1918169
##         57         58         59         60         61         62         63
## 35.0194862 42.4578664 36.0697375 37.3758633 20.6949161 38.9309186 33.4490339
##         64         65         66         67         68         69         70
## 33.4418323 41.6755827 18.1931145 23.0357662 24.6541330 42.2133979  8.7106066
##         71         72         73         74         75         76         77
## 31.3010542 31.1726048 33.7667888 18.7989091 18.3449071 32.7533624 21.0074906
##         78         79         80         81         82         83         84
## 35.6968758 32.8417376 37.8891129  0.8952600 11.2327170 22.5363246 23.3043113
##         85         86         87         88         89         90         91
## 28.0148995 40.3657424 22.9771716  7.9149609 23.0650137 39.4897905 32.1367039
##         92         93         94         95         96         97         98
## 24.4187803 13.9659531 27.9366806 30.2296604 43.0775722 19.4851372 14.1822304
##         99        100        101        102        103        104        105
## 29.0470435 35.5403380 12.4300152 14.9332119 21.0739943 37.8401240 12.3587206
##        106        107        108        109        110        111        112
##  6.7624525 41.9354268  1.4221514 13.3539393 30.5628942 18.1966743  0.8747433
##        113        114        115        116        117        118        119
## 32.4976399  9.0808945 25.4549624 37.1131618 16.8800933 22.8038472 35.1264282
##        120        121        122        123        124        125        126
## 39.1823371 35.4029125 40.6631150 24.6585086 33.0326049 36.1159451 39.8184557
##        127        128        129        130        131        132        133
## 35.3242591 11.0168277  5.3550810 41.7484739 16.2030806 40.3537422 18.8574146
```

```
##       134       135       136       137       138       139       140 
## 22.3157733 31.8302121 14.0474066 28.7452389 23.6979757  2.2970757 32.5338125 
##       141       142       143       144       145       146       147 
##  9.8195192 13.1259380 22.9071636 25.6089166 38.8297015 17.5433837 42.5446713 
##       148       149       150       151       152       153       154 
## 14.6152504 42.8298557 34.2988369 39.9463684 40.9232253 14.6302934  6.7499171 
##       155       156       157       158       159       160       161 
## 41.5963646 26.4303174 23.2206968 42.4640028 39.9529858 30.7761668 29.0874945 
##       162       163       164       165       166       167       168 
## 26.5858304 20.3954881 35.7549648  8.8937511 13.1599024 35.9906724 29.4192873 
##       169       170       171       172       173       174       175 
## 33.1421814 30.3590215 42.8660752 27.7818503 10.5045873 38.3940361  3.3381836 
##       176       177       178       179       180       181       182 
## 31.6684338 41.8966502 24.5333078 27.5438144 10.5670475 31.9242708 38.3501989 
##       183       184       185       186       187       188       189 
## 43.1530843 30.4094848 41.1858661 37.8546783 36.9439970 18.6675058 26.7473424 
##       190       191       192       193       194       195       196 
## 25.3543088 30.0559702 39.4586969 43.1068219 22.2978602 39.8541454 35.5844527 
##       197       198       199       200       201       202       203 
##  7.9015532 21.4564408 30.2683085 42.4267515 40.6095726 40.2148466 29.4361264 
##       204       205       206       207       208       209       210 
## 13.6574791 41.1308656 28.6494741 10.9326502 27.3298812 41.1930371 15.7632584 
##       211       212       213       214       215       216       217 
## 31.3708154 17.3765772 29.8319624  9.2671196 40.9611869 23.6109424 34.8534030 
##       218       219       220       221       222       223       224 
## 42.0402154 40.6957893 31.9872294 20.7138168 40.9037655 39.6068678 32.3303266 
##       225       226       227       228       229       230       231 
## 10.0802982 30.6660356 43.1852078 28.7618790  7.3825330  0.8153464 24.5968552 
##       232       233       234       235       236       237       238 
## 23.4841227  6.6460800 27.5282649 41.5468759 34.2607953 30.5459668 35.2467133 
##       239       240       241       242       243       244       245 
## 26.7184531 19.1919524 26.0665165 40.6419326 24.5513465 19.8724008 37.5589876 
##       246       247       248       249       250       251       252 
## 31.2913590 13.3399619 36.8020991 42.9165602 14.6106775 35.6270773 11.3061705 
##       253       254       255       256       257       258       259 
## 40.9229291 14.6455704 19.6001823 34.7042763 38.5976677 42.6619431 27.5089176 
##       260       261       262       263       264       265       266 
## 10.7177319 40.4986902 39.4902277 35.9593711 35.1350235 15.9395143  5.8265087 
##       267       268       269       270       271       272       273 
## 14.0401450 21.8187454 31.5276925 28.7275847 27.3483890 23.4370382 30.5129595 
##       274       275       276       277       278       279       280 
## 35.0622647 27.7584615 36.3749941 35.4658165 28.9796825  6.9755026 25.6140554 
##       281       282       283       284       285       286       287 
## 38.8780273 36.5754225 19.5260981 14.0236134 19.6591700 13.9419273 42.3038733 
##       288       289       290       291       292       293       294 
##  6.3360849 31.4515445 39.0326347 28.0183103  3.5466633 41.5443892 36.9880983 
##       295       296       297       298       299       300 
## 16.8272665 21.6623371 40.3294849 17.3842235 41.5796680 24.3393908
```

logfev

```
fdafev=fdata(t(fpcafev$phi %*% t(fpcafev$xiEst)))
outliers.depth.pond(
  fdafev,
  nb = 200,
  smo = 0.05,
  quan = 0.5,
  dfunc = depth.mode
)
```

```
## $outliers
## [1] "81"  "112" "139" "197" "207" "230" "223"
##
## $dep.out
## [1] 0.5930545 0.6283337 0.6350639 0.3989423 1.5693466 0.8410625 2.1419915
##
## $iteration
## [1] 1 1 1 1 1 1 2
##
## $quantile
##      50%
## 2.350116
##
## $Dep
##          1          2          3          4          5          6          7
##  8.6164486 27.3413089 24.1193427 20.7426221 11.1780738 35.6933140 18.5851999
##          8          9         10         11         12         13         14
## 41.5326577 16.8400021 20.1973065 30.8633191 35.5516368 36.1553579 40.5656960
##         15         16         17         18         19         20         21
## 23.4679595 31.6947417 36.2217688 40.3958228 31.3230409 11.8541104 35.8232146
##         22         23         24         25         26         27         28
## 32.5088220 34.9749268 35.5475173 41.5482888 32.2486358 15.3929019 25.9322162
##         29         30         31         32         33         34         35
## 33.4770308 28.7588553 31.7539884 17.1583012 17.6611819 39.6910905 10.0524406
##         36         37         38         39         40         41         42
## 22.3205302 33.2641978 30.4090764 21.5839863 26.2201869 37.6714224 22.4945768
##         43         44         45         46         47         48         49
## 33.2525999 32.2676983 11.6548139 14.8070318  5.0261295 32.4343818 33.2425020
##         50         51         52         53         54         55         56
## 31.7155976 27.7819411 24.8823016 35.8360370 13.3617511 10.6887968 38.6613217
##         57         58         59         60         61         62         63
## 25.4828352 40.2948421 36.0346619 25.6251010 23.4067973 39.4953631 15.9528469
##         64         65         66         67         68         69         70
## 33.7855039 23.7761953  9.0909866 21.4716940 36.9295842  5.7369485 21.2299430
##         71         72         73         74         75         76         77
## 28.3743559 34.6922879 12.8329869 29.5575068  7.3913340 32.2452830 24.8492838
##         78         79         80         81         82         83         84
## 22.4828682  4.2644491 12.5743464  0.5930545 16.2924016 40.4243456 40.9818195
##         85         86         87         88         89         90         91
##  6.0954155 31.1078168 27.1218251  9.8344824 17.4424843 33.0710958 41.3026009
##         92         93         94         95         96         97         98
## 23.0549630 28.8164646 26.6499951 23.2763292 41.2605951 26.9731980 31.9539808
##         99        100        101        102        103        104        105
## 39.2070280 40.6679380 27.9373982 21.5146059 18.3304539 28.5422959 31.7926510
##        106        107        108        109        110        111        112
## 30.9277835 36.6026511  9.7225585 31.0999556 12.9330533 27.2463277  0.6283337
##        113        114        115        116        117        118        119
## 15.1070199 25.8846667 32.7797537 24.6202487 16.7369671 22.1857214 31.4619545
##        120        121        122        123        124        125        126
## 25.4152134 39.3284942 39.0742194 28.5420930 18.2500989 33.9704252 31.1818788
##        127        128        129        130        131        132        133
## 34.4646127 17.4818600 14.5459357 40.1809596 25.0201305 36.4887043 23.1160461
```

```
##       134       135       136       137       138       139       140 
## 23.5938653 19.3123912  3.7338471 21.4899375 31.9248742  0.6350639  4.3525022 
##       141       142       143       144       145       146       147 
##  9.5159327 17.0627999 31.4329876 29.8186151 13.0523309 11.2695803 19.0701404 
##       148       149       150       151       152       153       154 
## 23.9825133 40.3265381 27.3253412 25.6140600 40.9950995 32.0727066  9.7932627 
##       155       156       157       158       159       160       161 
## 31.2029555 32.9015250 12.3228988 33.5728295 41.4640659 38.3444463 40.6539568 
##       162       163       164       165       166       167       168 
## 25.1186836 30.7468756 22.7283499 17.9737003 12.9134996 36.4770546 32.2121951 
##       169       170       171       172       173       174       175 
## 20.2591227 26.2392013 32.8811644 16.3059183 31.9679143 23.0471683 17.0914059 
##       176       177       178       179       180       181       182 
## 18.3007103 36.7567672 33.7461520 35.1335115 18.4073839 28.4036403 39.8013629 
##       183       184       185       186       187       188       189 
## 41.3629758 16.9349441 35.8996902 26.4476545 28.1469955 41.0606742 34.4347950 
##       190       191       192       193       194       195       196 
## 30.8541165 40.1331629 32.3600486 41.4177729 27.8134250 35.0505950 30.3694387 
##       197       198       199       200       201       202       203 
##  0.3989423 20.8228647 12.3579249 39.8342269 38.3116688 32.3773404 29.1952770 
##       204       205       206       207       208       209       210 
## 12.7559094 41.4776487 20.9808991  1.5693466 35.2203532 40.1811071 36.7929883 
##       211       212       213       214       215       216       217 
## 24.5736572 37.5582645 38.7187769 35.2028251 38.7664468  8.7779147 34.2026376 
##       218       219       220       221       222       223       224 
## 33.7914898 26.6725028 35.9705707 35.2327924 40.7297365  2.5342026 23.3982174 
##       225       226       227       228       229       230       231 
##  4.3770084 26.5036772 37.4730687 32.2799069 19.5851679  0.8410625 18.0486088 
##       232       233       234       235       236       237       238 
## 31.7547262 21.2055717 29.1856831 30.8877473 21.7812515 19.5493611 11.9524932 
##       239       240       241       242       243       244       245 
## 17.3052940 25.0340362 26.0525989 28.9008375 41.2223884 22.9481369 37.1176176 
##       246       247       248       249       250       251       252 
##  7.0377089 38.6817680 32.4246102 41.2356336 22.8384519 39.4129947 33.7174744 
##       253       254       255       256       257       258       259 
## 28.9150851 28.5175009 32.0320377 18.3755418 40.0442837 20.3137176 29.0078216 
##       260       261       262       263       264       265       266 
## 11.3084103 30.7176299 39.7642006 36.4163163 40.0405619 37.3280324 17.9797387 
##       267       268       269       270       271       272       273 
## 41.6318500 28.8462513 35.6291680  7.1808483 11.1667361 26.1275668 23.3594417 
##       274       275       276       277       278       279       280 
## 38.1425484 23.1503254 28.9709775 40.6610508 21.7657875 40.5582763 33.3977896 
##       281       282       283       284       285       286       287 
## 39.9136960 29.6246280 36.6593427 40.6665450 39.4958297 40.2988029 30.1174804 
##       288       289       290       291       292       293       294 
## 25.1106338 25.9908984 34.9000927 24.7594077  8.1138836  9.6352881  8.5776323 
##       295       296       297       298       299       300 
##  4.7430188 11.2203153 37.0999963  7.8217906 33.4770782 28.4557466 
```

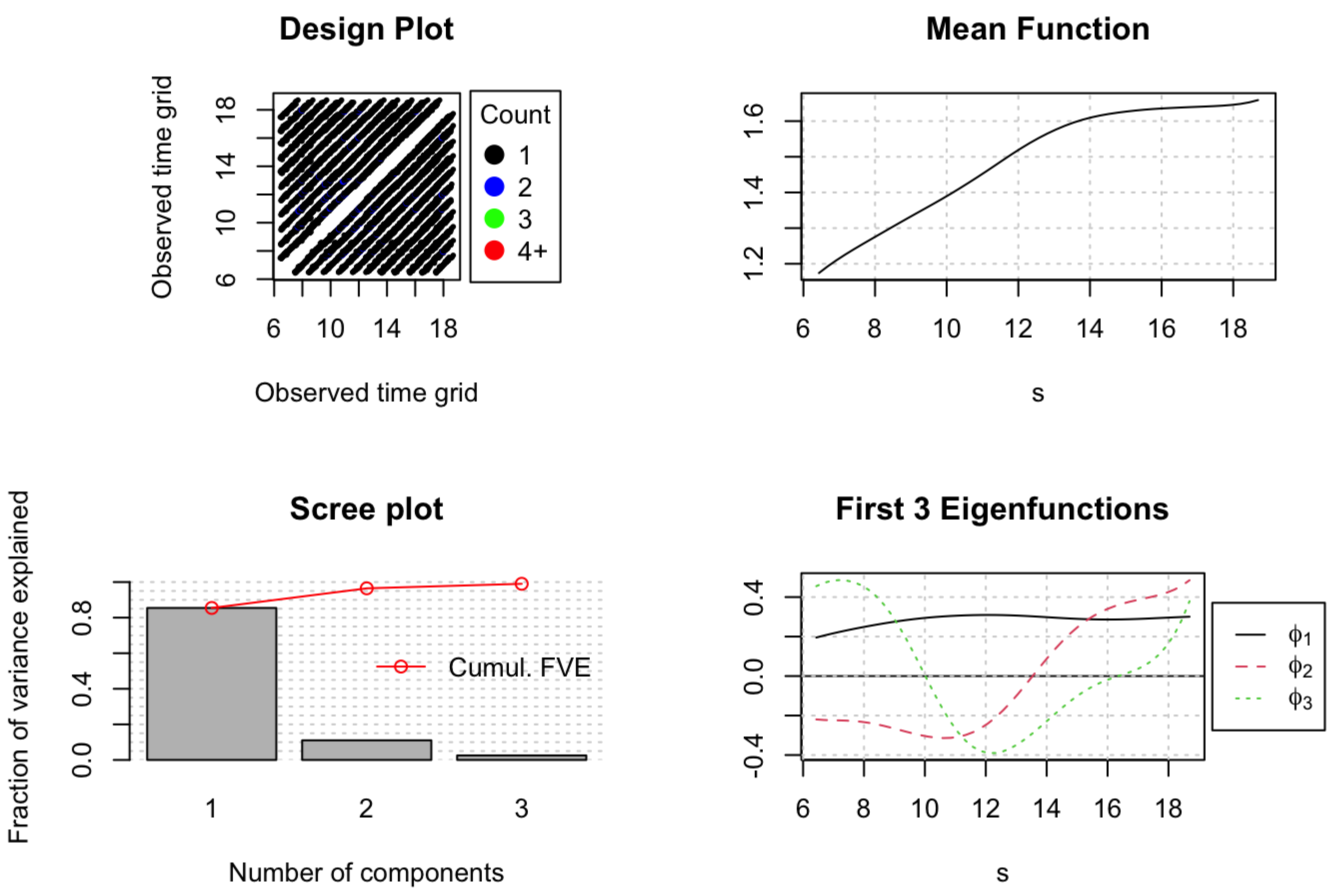
Data Cleaning

```
ol=c(1,10,81,108,112,139,197,207,223,230)
data=data[!data$id %in% ol, ]
```

FPCA again

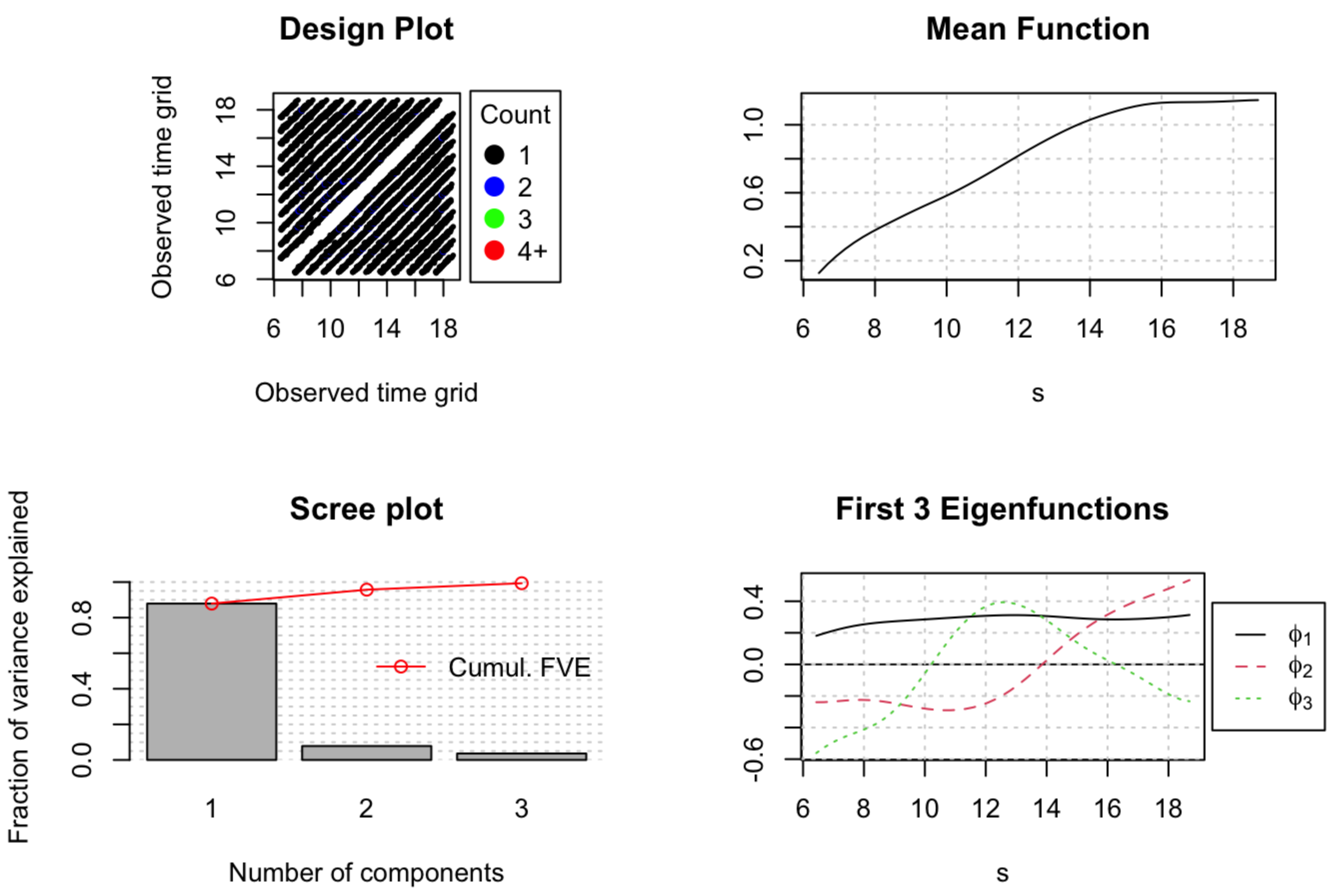
FPCA for Height

```
library(fdapace)
library(fda.usc)
Age <- split(data$age, data$id)
Height <- split(data$ht, data$id)
ID<- as.list(1:length(unique(data$id)))
FD1=list(Lid=ID,Ly=Height,Lt=Age)
fpcaheight=FPCA(FD1$Ly,FD1$Lt)
plot(fpcaheight)
```



FPCA for log(Fev)

```
Age <- split(data$age, data$id)
Fev <- split(data$logfev1, data$id)
ID<- as.list(1:length(unique(data$id)))
FD2=list(Lid=ID,Ly=Fev,Lt=Age)
fpcafev=FPCA(FD2$Ly,FD2$Lt)
plot(fpcafev)
```



K-Selection

```
SelectK(fpcaheight,criterion='AIC')
```

```
## $K
## [1] 3
##
## $criterion
## [1] -13388.55
```

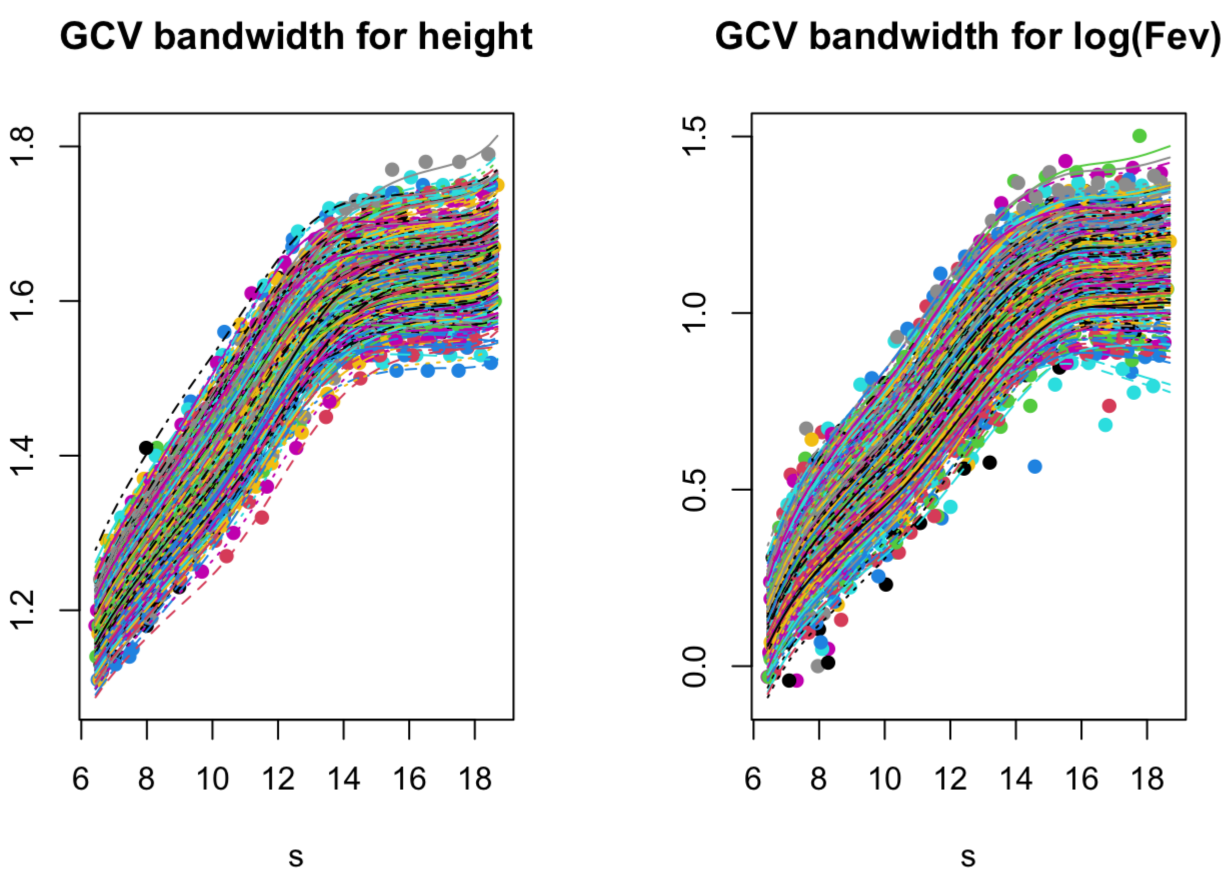
```
SelectK(fpcafev,criterion='AIC')
```

```
## $K
## [1] 3
##
## $criterion
## [1] -7849.175
```

EDA for Functional Data

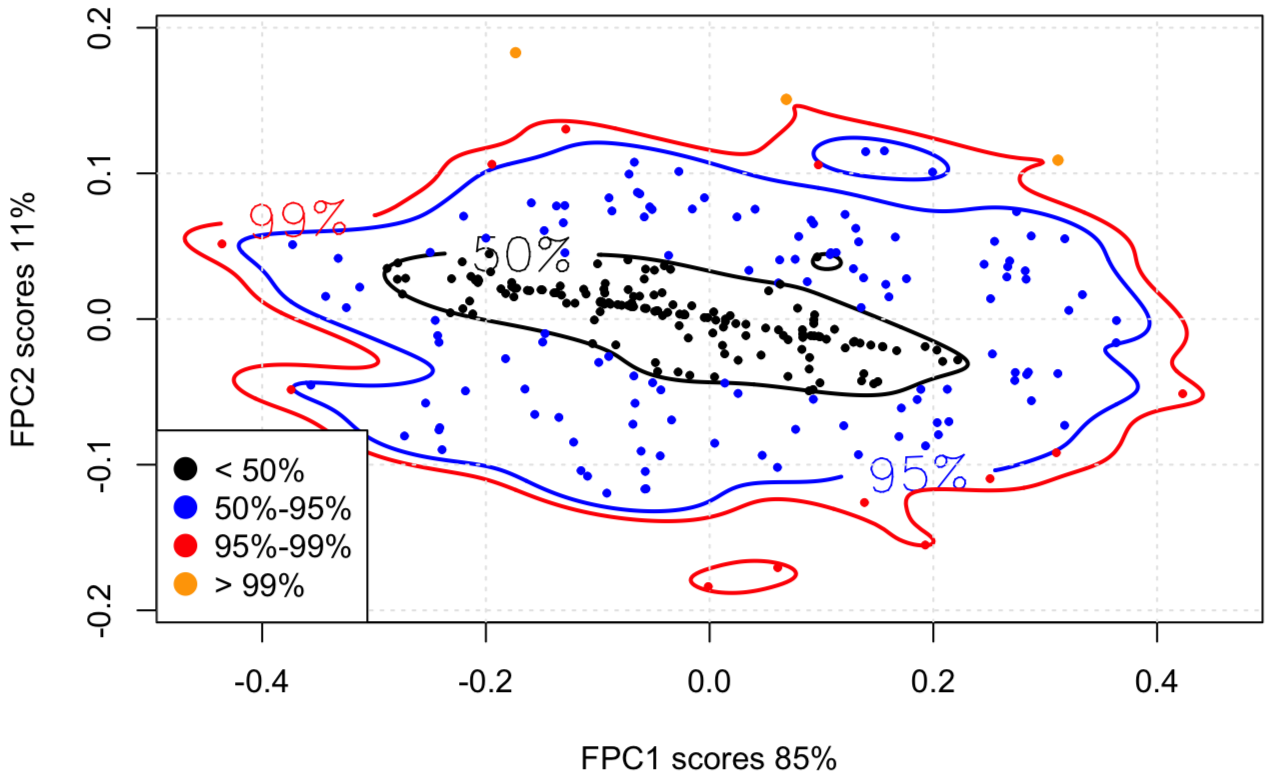
Smooth Paths Visualization

```
par(mfrow=c(1,2))
CreatePathPlot( fpcaheight, main = "GCV bandwidth for height", pch = 16)
CreatePathPlot( fpcafev, main = "GCV bandwidth for log(Fev)", pch = 16)
```

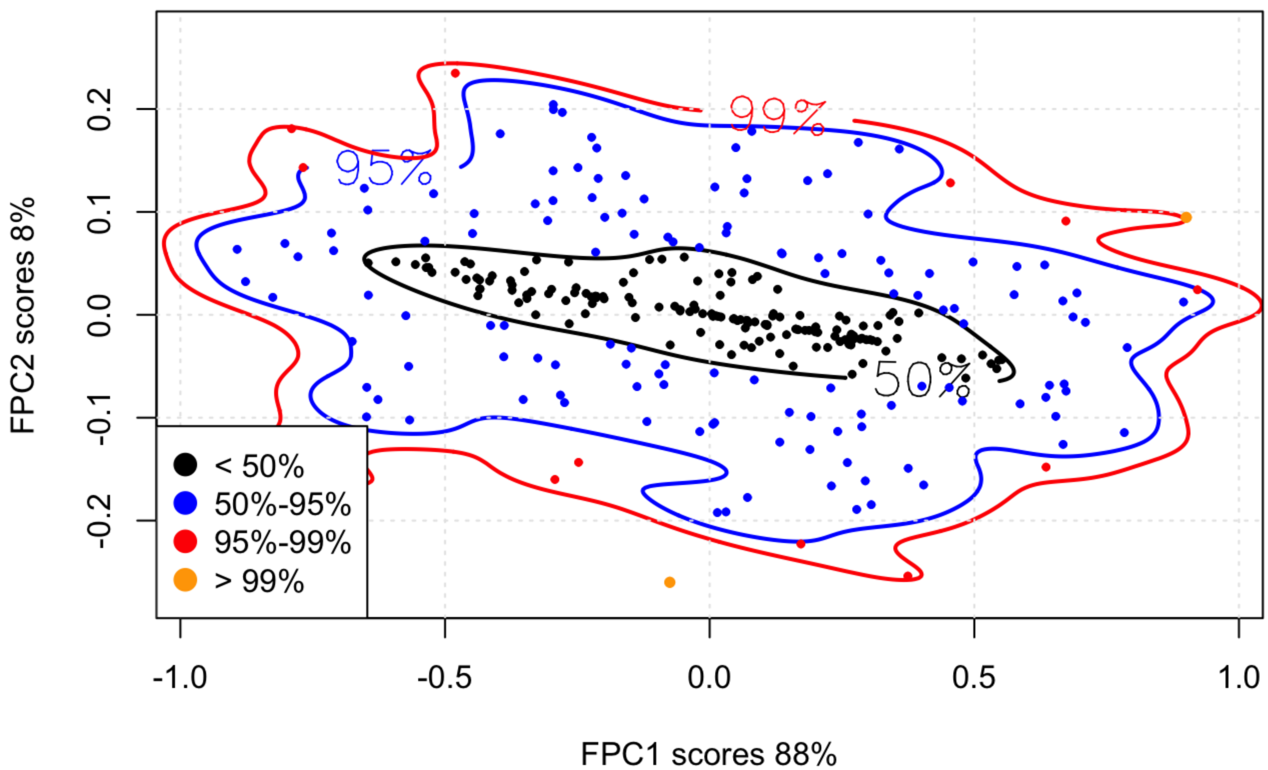


Outlier Visualization

```
CreateOutliersPlot(fpcaheight,optns=list(K=3,variant='KDE'))
```



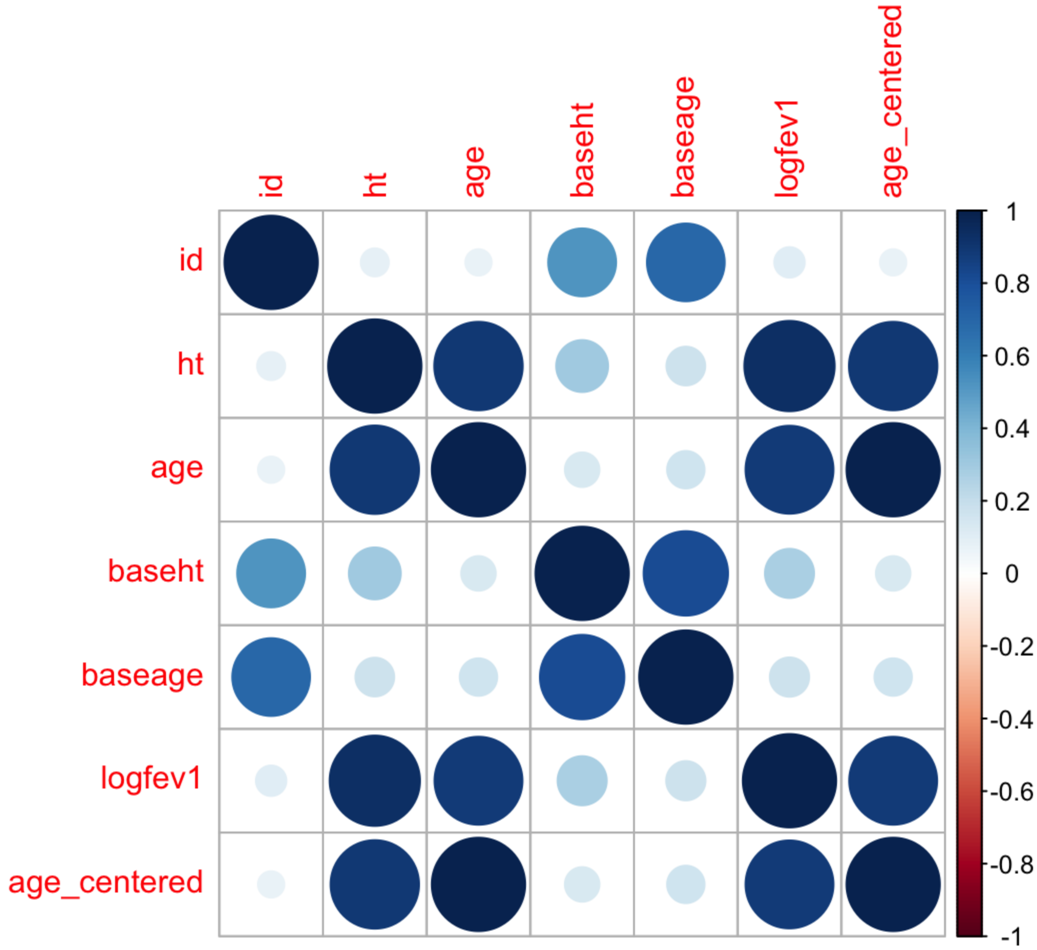

```
CreateOutliersPlot(fpcafev,optns=list(K=3,variant='KDE'))
```




```r
library("corrplot")
```

```
## corrplot 0.92 loaded
```

```r
corrplot(cor(data))
```



Model Fitting

Height

Covariance Structure Specification

Variance Component

```
model1.1=gls(ht~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+baseh
t*I(age_centered^2),data=data,method='ML')
summary(model1.1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage *      I(age_c
entered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -7845.866 -7790.239 3932.933
##
## Coefficients:
##                                Value   Std.Error   t-value p-value
## (Intercept)                0.7704280 0.019133231  40.26649  0.0000
## baseage                   -0.0526035 0.001474126 -35.68453  0.0000
## age_centered               0.0567178 0.004126722  13.74402  0.0000
## baseht                     0.9424045 0.021812199  43.20539  0.0000
## I(age_centered^2)          0.0022231 0.001311819   1.69466  0.0903
## baseage:age_centered       0.0036568 0.000346493  10.55366  0.0000
## age_centered:baseht       -0.0350411 0.004518550  -7.75495  0.0000
## baseage:I(age_centered^2) -0.0003128 0.000109954  -2.84485  0.0045
## baseht:I(age_centered^2)  -0.0031903 0.001486100  -2.14674  0.0319
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.596
## age_centered              -0.051  0.007
## baseht                    -0.937 -0.836  0.029
## I(age_centered^2)         -0.716 -0.428  0.002  0.674
## baseage:age_centered       0.034  0.030  0.375 -0.032  0.048
## age_centered:baseht        0.020 -0.017 -0.899 -0.006 -0.019 -0.741
## baseage:I(age_centered^2) -0.412 -0.696  0.072  0.577  0.479 -0.174  0.022
## baseht:I(age_centered^2)   0.685  0.619 -0.027 -0.734 -0.917  0.036  0.002
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.788
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.24433643 -0.64671424 -0.01222602  0.60342349  3.65996545
##
## Residual standard error: 0.03136632
## Degrees of freedom: 1925 total; 1916 residual
```

Compound Symmetry

```
model1.2=gls(ht~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+baseh
t*I(age_centered^2),data=data,corr=corCompSymm(form=~1|id),method='ML')
summary(model1.2)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage *      I(age_c
entered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -8273.165 -8211.975 4147.582
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | id
##  Parameter estimate(s):
##       Rho
## 0.3305408
##
## Coefficients:
##                               Value   Std.Error   t-value p-value
## (Intercept)               0.7858155 0.025471371  30.85093  0.0000
## baseage                  -0.0526574 0.001990480 -26.45460  0.0000
## age_centered              0.0532659 0.003568305  14.92752  0.0000
## baseht                    0.9311428 0.029425281  31.64431  0.0000
## I(age_centered^2)         0.0023617 0.001119472   2.10962  0.0350
## baseage:age_centered      0.0038010 0.000302971  12.54569  0.0000
## age_centered:baseht      -0.0331933 0.003873067  -8.57029  0.0000
## baseage:I(age_centered^2) -0.0003893 0.000093954  -4.14321  0.0000
## baseht:I(age_centered^2)  -0.0028613 0.001261449  -2.26829  0.0234
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.623
## age_centered               0.010  0.032
## baseht                    -0.940 -0.852 -0.023
## I(age_centered^2)         -0.463 -0.282 -0.023  0.434
## baseage:age_centered       0.057  0.073  0.339 -0.066  0.055
## age_centered:baseht       -0.035 -0.057 -0.892  0.048 -0.004 -0.726
## baseage:I(age_centered^2) -0.279 -0.442  0.075  0.376  0.460 -0.200  0.034
## baseht:I(age_centered^2)   0.451  0.403 -0.010 -0.477 -0.914  0.043 -0.013
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.779
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.26654724 -0.68155132 -0.01265035  0.55361962  3.79292750
##
## Residual standard error: 0.03103683
## Degrees of freedom: 1925 total; 1916 residual
```

Continuous AR(1)

```
model1.3=gls(ht~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+baseh
t*I(age_centered^2),data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage *      I(age_c
entered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -9659.351 -9598.161 4840.675
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8025287
##
## Coefficients:
##                                Value  Std.Error    t-value p-value
## (Intercept)                0.7725592 0.02978999  25.933516  0.0000
## baseage                   -0.0513122 0.00232850 -22.036577  0.0000
## age_centered               0.0479007 0.00503373   9.515960  0.0000
## baseht                     0.9286359 0.03451869  26.902405  0.0000
## I(age_centered^2)          0.0023713 0.00121898   1.945317  0.0519
## baseage:age_centered       0.0037796 0.00041881   9.024608  0.0000
## age_centered:baseht       -0.0297572 0.00564286  -5.273424  0.0000
## baseage:I(age_centered^2) -0.0004334 0.00010101  -4.290139  0.0000
## baseht:I(age_centered^2)  -0.0022471 0.00138240  -1.625511  0.1042
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.634
## age_centered               0.052  0.056
## baseht                    -0.941 -0.857 -0.060
## I(age_centered^2)         -0.594 -0.373 -0.016  0.558
## baseage:age_centered       0.072  0.094  0.452 -0.086  0.041
## age_centered:baseht       -0.070 -0.081 -0.911  0.082 -0.003 -0.779
## baseage:I(age_centered^2) -0.365 -0.571  0.056  0.489  0.494 -0.124  0.012
## baseht:I(age_centered^2)   0.575  0.521 -0.009 -0.610 -0.919  0.020 -0.004
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.794
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.46211449 -0.69910796 -0.02351251  0.63190728  3.65093380
##
## Residual standard error: 0.02972835
## Degrees of freedom: 1925 total; 1916 residual
```

Summary: Based on the AIC and BIC, we will choose continuous AR(1) to be our covariance structure.

Model Selection for Linear Model

first

```
model1.3=gls(ht~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+baseh
t*I(age_centered^2),data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage *      I(age_c
entered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -9659.351 -9598.161 4840.675
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8025287
##
## Coefficients:
##                               Value  Std.Error    t-value p-value
## (Intercept)               0.7725592 0.02978999  25.933516  0.0000
## baseage                  -0.0513122 0.00232850 -22.036577  0.0000
## age_centered              0.0479007 0.00503373   9.515960  0.0000
## baseht                    0.9286359 0.03451869  26.902405  0.0000
## I(age_centered^2)         0.0023713 0.00121898   1.945317  0.0519
## baseage:age_centered      0.0037796 0.00041881   9.024608  0.0000
## age_centered:baseht      -0.0297572 0.00564286  -5.273424  0.0000
## baseage:I(age_centered^2) -0.0004334 0.00010101  -4.290139  0.0000
## baseht:I(age_centered^2)  -0.0022471 0.00138240  -1.625511  0.1042
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.634
## age_centered               0.052  0.056
## baseht                    -0.941 -0.857 -0.060
## I(age_centered^2)         -0.594 -0.373 -0.016  0.558
## baseage:age_centered       0.072  0.094  0.452 -0.086  0.041
## age_centered:baseht       -0.070 -0.081 -0.911  0.082 -0.003 -0.779
## baseage:I(age_centered^2) -0.365 -0.571  0.056  0.489  0.494 -0.124  0.012
## baseht:I(age_centered^2)   0.575  0.521 -0.009 -0.610 -0.919  0.020 -0.004
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.794
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.46211449 -0.69910796 -0.02351251  0.63190728  3.65093380
##
## Residual standard error: 0.02972835
## Degrees of freedom: 1925 total; 1916 residual
```

second

$age^2_{centered}$ is just a little bit above the threshold. We first see what if we remove baseht:$age^2_{centered}$

```
model1.3=gls(ht~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2),data=
data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage *      I(age_c
entered^2)
##   Data: data
##         AIC       BIC   logLik
##   -9658.698 -9603.071 4839.349
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##      Phi
## 0.802494
##
## Coefficients:
##                                Value   Std.Error   t-value p-value
## (Intercept)                0.8004205 0.024373471  32.83983  0.0000
## baseage                   -0.0493405 0.001988215 -24.81647  0.0000
## age_centered               0.0478294 0.005035478   9.49849  0.0000
## baseht                     0.8944003 0.027357259  32.69334  0.0000
## I(age_centered^2)          0.0005494 0.000479624   1.14552  0.2521
## baseage:age_centered       0.0037935 0.000418887   9.05606  0.0000
## age_centered:baseht       -0.0297898 0.005645002  -5.27720  0.0000
## baseage:I(age_centered^2) -0.0005638 0.000061391  -9.18323  0.0000
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.478
## age_centered               0.070  0.071
## baseht                    -0.910 -0.797 -0.083
## I(age_centered^2)         -0.201  0.317 -0.061 -0.008
## baseage:age_centered       0.074  0.097  0.453 -0.092  0.153
## age_centered:baseht       -0.083 -0.093 -0.911  0.100 -0.017 -0.779
## baseage:I(age_centered^2)  0.186 -0.303  0.080  0.008 -0.989 -0.178  0.016
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.49764493 -0.70751428 -0.01413286  0.64566934  3.55960874
##
## Residual standard error: 0.02974693
## Degrees of freedom: 1925 total; 1917 residual
```

third

Now p-value for $age^2_{centered}$ is 0.252, and thus we remove it, but keep the interaction between $age^2_{centered}$ and baseage. Everything is significant now.

```
model1.3=gls(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),data=
data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_centered
^2)
##   Data: data
##         AIC       BIC   logLik
##   -9659.386 -9609.322 4838.693
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8017266
##
## Coefficients:
##                                  Value   Std.Error   t-value p-value
## (Intercept)                  0.8060176 0.023824950  33.83082       0
## baseage                     -0.0500622 0.001881813 -26.60319       0
## age_centered                 0.0481952 0.005021929   9.59695       0
## baseht                       0.8946701 0.027296194  32.77637       0
## baseage:age_centered         0.0037196 0.000413654   8.99213       0
## age_centered:baseht         -0.0296846 0.005639317  -5.26387       0
## baseage:I(age_centered^2) -0.0004942 0.000009015 -54.82659       0
##
##  Correlation:
##                             (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                      0.583
## age_centered                 0.059  0.095
## baseht                      -0.931 -0.837 -0.084
## baseage:age_centered         0.108  0.052  0.468 -0.092
## age_centered:baseht         -0.088 -0.092 -0.914  0.100 -0.785
## baseage:I(age_centered^2) -0.092  0.070  0.134  0.001 -0.186 -0.010
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.50715974 -0.70733519 -0.01249239  0.63528182  3.47071042
##
## Residual standard error: 0.02971482
## Degrees of freedom: 1925 total; 1918 residual
```

Summary: This results in the following model:

```
model=gls(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),data=dat
a,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_centered
^2)
##   Data: data
##         AIC       BIC   logLik
##   -9659.386 -9609.322 4838.693
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8017266
##
## Coefficients:
##                                Value   Std.Error   t-value p-value
## (Intercept)                0.8060176 0.023824950  33.83082       0
## baseage                   -0.0500622 0.001881813 -26.60319       0
## age_centered               0.0481952 0.005021929   9.59695       0
## baseht                     0.8946701 0.027296194  32.77637       0
## baseage:age_centered       0.0037196 0.000413654   8.99213       0
## age_centered:baseht       -0.0296846 0.005639317  -5.26387       0
## baseage:I(age_centered^2) -0.0004942 0.000009015 -54.82659       0
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.583
## age_centered               0.059  0.095
## baseht                    -0.931 -0.837 -0.084
## baseage:age_centered       0.108  0.052  0.468 -0.092
## age_centered:baseht       -0.088 -0.092 -0.914  0.100 -0.785
## baseage:I(age_centered^2) -0.092  0.070  0.134  0.001 -0.186 -0.010
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.50715974 -0.70733519 -0.01249239  0.63528182  3.47071042
##
## Residual standard error: 0.02971482
## Degrees of freedom: 1925 total; 1918 residual
```

Linear Mixed Effect Model

Now we are considering to add a random effect to the model we selected from the previous model, and we are wondering random intercept, random age, which one would be the best.

Random Intercept for CAR(1)

```
model=lme(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),random=~
1|id,data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -9657.386 -9601.759 4838.693
##
## Random effects:
##  Formula: ~1 | id
##          (Intercept)   Residual
## StdDev: 6.592544e-07 0.02971482
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8017266
## Fixed effects:  ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_c
entered^2)
##                                Value   Std.Error   DF   t-value p-value
## (Intercept)                0.8060176 0.023824950 1631  33.83082       0
## baseage                   -0.0500622 0.001881813  287 -26.60319       0
## age_centered               0.0481952 0.005021929 1631   9.59695       0
## baseht                     0.8946701 0.027296194  287  32.77637       0
## baseage:age_centered       0.0037196 0.000413654 1631   8.99213       0
## age_centered:baseht       -0.0296846 0.005639317 1631  -5.26387       0
## baseage:I(age_centered^2) -0.0004942 0.000009015 1631 -54.82659       0
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.583
## age_centered               0.059  0.095
## baseht                    -0.931 -0.837 -0.084
## baseage:age_centered       0.108  0.052  0.468 -0.092
## age_centered:baseht       -0.088 -0.092 -0.914  0.100 -0.785
## baseage:I(age_centered^2) -0.092  0.070  0.134  0.001 -0.186 -0.010
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.50715973 -0.70733519 -0.01249239  0.63528182  3.47071041
##
## Number of Observations: 1925
## Number of Groups: 290
```

Random Slope for CAR(1) (diverge)

```
#model=lme(ht~baseage*age_centered+baseht*age_centered+baseage*baseage:I(age_centered^
2),random=~age_centered|id,data=data,corr=corCAR1(form=~age_centered|id),method='ML')
```

Variance Component

```
model2.1=gls(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),data=
data,method='ML')
summary(model2.1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_centered
^2)
##   Data: data
##         AIC       BIC   logLik
##   -7844.771 -7800.269 3930.385
##
## Coefficients:
##                               Value   Std.Error   t-value p-value
## (Intercept)               0.7956633 0.013295854  59.84296       0
## baseage                  -0.0502904 0.001036776 -48.50651       0
## age_centered              0.0563197 0.004122000  13.66319       0
## baseht                    0.9080605 0.014823978  61.25619       0
## baseage:age_centered      0.0037316 0.000339377  10.99531       0
## age_centered:baseht      -0.0351481 0.004518187  -7.77924       0
## baseage:I(age_centered^2) -0.0005446 0.000009983 -54.55681       0
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.511
## age_centered              -0.065  0.062
## baseht                    -0.921 -0.802  0.014
## baseage:age_centered       0.080 -0.091  0.397 -0.009
## age_centered:baseht        0.012 -0.006 -0.904 -0.006 -0.749
## baseage:I(age_centered^2) -0.116  0.089  0.181 -0.001 -0.248 -0.016
##
## Standardized residuals:
##          Min           Q1          Med           Q3          Max
## -3.313881846 -0.654537752 -0.008381967  0.613382205  3.573507587
##
## Residual standard error: 0.03140786
## Degrees of freedom: 1925 total; 1918 residual
```

Random Intercept for Variance Component

```
model2.1=lme(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),rando
m=~1|id,data=data,method='ML')
summary(model2.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -8271.995 -8221.931 4144.997
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:  0.01784644 0.02543214
##
## Fixed effects:  ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_c
entered^2)
##                                 Value   Std.Error   DF   t-value p-value
## (Intercept)                 0.8121249 0.022512427 1631  36.07452       0
## baseage                    -0.0508754 0.001771149  287 -28.72449       0
## age_centered                0.0532102 0.003559552 1631  14.94856       0
## baseht                      0.8993186 0.025858771  287  34.77809       0
## baseage:age_centered        0.0038241 0.000294651 1631  12.97844       0
## age_centered:baseht        -0.0332972 0.003873176 1631  -8.59688       0
## baseage:I(age_centered^2) -0.0005501 0.000008438 1631 -65.18721       0
##  Correlation:
##                            (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                     0.596
## age_centered                0.005  0.061
## baseht                     -0.934 -0.841 -0.032
## baseage:age_centered        0.078  0.006  0.370 -0.051
## age_centered:baseht        -0.038 -0.049 -0.899  0.047 -0.738
## baseage:I(age_centered^2) -0.072  0.054  0.185  0.003 -0.257 -0.014
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.47867786 -0.67797762  0.00818003  0.69124584  3.14528885
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
#Residual
sigma(model2.1)^2
```

```
## [1] 0.0006467938
```

```
# Random Effect
getVarCov(model2.1,type="random")
```

```
## Random effects variance covariance matrix
##             (Intercept)
## (Intercept)   0.0003185
##   Standard Deviations: 0.017846
```

```
# Marginal covariance
getVarCov(model2.1,type="marginal",individuals = 1)
```

```
## id 2
## Marginal variance covariance matrix
##            1          2          3          4          5          6          7
## 1 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
## 2 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
## 3 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850
## 4 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850
## 5 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850
## 6 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850
## 7 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529
## 8 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
##            8
## 1 0.00031850
## 2 0.00031850
## 3 0.00031850
## 4 0.00031850
## 5 0.00031850
## 6 0.00031850
## 7 0.00031850
## 8 0.00096529
##   Standard Deviations: 0.031069 0.031069 0.031069 0.031069 0.031069 0.031069 0.031069
0.031069
```

```
getVarCov(model2.1,type="marginal",individuals = 2)
```

```
## id 3
## Marginal variance covariance matrix
##            1          2          3          4          5          6          7
## 1 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
## 2 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
## 3 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850 0.00031850
## 4 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850 0.00031850
## 5 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850 0.00031850
## 6 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529 0.00031850
## 7 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00096529
## 8 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
## 9 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850 0.00031850
##            8          9
## 1 0.00031850 0.00031850
## 2 0.00031850 0.00031850
## 3 0.00031850 0.00031850
## 4 0.00031850 0.00031850
## 5 0.00031850 0.00031850
## 6 0.00031850 0.00031850
## 7 0.00031850 0.00031850
## 8 0.00096529 0.00031850
## 9 0.00031850 0.00096529
##   Standard Deviations: 0.031069 0.031069 0.031069 0.031069 0.031069 0.031069 0.031069
0.031069 0.031069
```

Random Slope for Variance Component

```
model2.2=lme(ht~baseage*age_centered+baseht*age_centered+baseage*baseage:I(age_centered^
2),random=~age_centered|id,data=data,method='ML')
summary(model2.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -8822.103 -8760.913 4422.051
##
## Random effects:
##  Formula: ~age_centered | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev     Corr
## (Intercept)  0.01905488 (Intr)
## age_centered 0.00457483 0.899
## Residual     0.02068321
##
## Fixed effects:  ht ~ baseage * age_centered + baseht * age_centered + baseage *
baseage:I(age_centered^2)
##                                Value   Std.Error   DF   t-value p-value
## (Intercept)                0.7974547 0.024052741 1631  33.15442       0
## baseage                   -0.0508070 0.001894515  287 -26.81796       0
## age_centered               0.0554258 0.006153774 1631   9.00681       0
## baseht                     0.9092496 0.027885716  287  32.60629       0
## baseage:age_centered       0.0037436 0.000500671 1631   7.47709       0
## age_centered:baseht       -0.0347789 0.007074666 1631  -4.91597       0
## baseage:I(age_centered^2) -0.0005485 0.000007076 1631 -77.50912       0
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.627
## age_centered               0.728  0.489
## baseht                    -0.939 -0.855 -0.699
## baseage:age_centered       0.485  0.724  0.559 -0.638
## age_centered:baseht       -0.708 -0.659 -0.928  0.760 -0.826
## baseage:I(age_centered^2) -0.040  0.032  0.088 -0.002 -0.117 -0.007
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.73839429 -0.61558972 -0.02930903  0.63957439  2.95072155
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
#Residual
sigma(model2.2)^2
```

```
## [1] 0.0004277953
```

```
# Random Effect
getVarCov(model2.2,type="random")
```

```
## Random effects variance covariance matrix
##              (Intercept) age_centered
## (Intercept)   3.6309e-04   7.8409e-05
## age_centered  7.8409e-05   2.0929e-05
##   Standard Deviations: 0.019055 0.0045748
```

```
# Marginal covariance
getVarCov(model2.2,type="marginal",individuals = 1)
```

```
## id 2
## Marginal variance covariance matrix
##             1           2           3           4           5           6
## 1  0.00061432  1.3392e-04 -1.1814e-04 -1.6940e-04 -2.1495e-04 -0.00027082
## 2  0.00013392  5.3272e-04 -3.3978e-05 -6.2221e-05 -8.7326e-05 -0.00011811
## 3 -0.00011814 -3.3978e-05  7.9706e-04  4.5126e-04  5.2415e-04  0.00061352
## 4 -0.00016940 -6.2221e-05  4.5126e-04  9.8347e-04  6.4848e-04  0.00076228
## 5 -0.00021495 -8.7326e-05  5.2415e-04  6.4848e-04  1.1868e-03  0.00089452
## 6 -0.00027082 -1.1811e-04  6.1352e-04  7.6228e-04  8.9452e-04  0.00148450
## 7 -0.00031271 -1.4120e-04  6.8054e-04  8.4763e-04  9.9616e-04  0.00117830
## 8 -0.00036044 -1.6750e-04  7.5691e-04  9.4487e-04  1.1119e-03  0.00131680
##             7           8
## 1 -0.00031271 -0.00036044
## 2 -0.00014120 -0.00016750
## 3  0.00068054  0.00075691
## 4  0.00084763  0.00094487
## 5  0.00099616  0.00111190
## 6  0.00117830  0.00131680
## 7  0.00174270  0.00147050
## 8  0.00147050  0.00207330
##   Standard Deviations: 0.024785 0.023081 0.028232 0.03136 0.03445 0.038529 0.041745
0.045534
```

```
getVarCov(model2.2,type="marginal",individuals = 2)
```

```
## id 3
## Marginal variance covariance matrix
##             1           2          3          4           5           6
## 1  5.8427e-04  1.1111e-04 6.8785e-05 2.5174e-05 -2.1358e-05 -1.0636e-04
## 2  1.1111e-04  5.1716e-04 6.9072e-05 4.8166e-05  2.5859e-05 -1.4892e-05
## 3  6.8785e-05  6.9072e-05 4.9714e-04 6.9616e-05  6.9910e-05  7.0448e-05
## 4  2.5174e-05  4.8166e-05 6.9616e-05 5.1951e-04  1.1530e-04  1.5838e-04
## 5 -2.1358e-05  2.5859e-05 6.9910e-05 1.1530e-04  5.9153e-04  2.5220e-04
## 6 -1.0636e-04 -1.4892e-05 7.0448e-05 1.5838e-04  2.5220e-04  8.5139e-04
## 7 -1.5056e-04 -3.6079e-05 7.0728e-05 1.8078e-04  2.9820e-04  5.1271e-04
## 8 -1.9230e-04 -5.6089e-05 7.0992e-05 2.0193e-04  3.4164e-04  5.9687e-04
## 9 -2.3228e-04 -7.5257e-05 7.1245e-05 2.2220e-04  3.8326e-04  6.7749e-04
##             7           8           9
## 1 -1.5056e-04 -1.9230e-04 -2.3228e-04
## 2 -3.6079e-05 -5.6089e-05 -7.5257e-05
## 3  7.0728e-05  7.0992e-05  7.1245e-05
## 4  1.8078e-04  2.0193e-04  2.2220e-04
## 5  2.9820e-04  3.4164e-04  3.8326e-04
## 6  5.1271e-04  5.9687e-04  6.7749e-04
## 7  1.0520e-03  7.2957e-04  8.3047e-04
## 8  7.2957e-04  1.2827e-03  9.7495e-04
## 9  8.3047e-04  9.7495e-04  1.5411e-03
##   Standard Deviations: 0.024172 0.022741 0.022297 0.022793 0.024321 0.029179 0.032435
0.035815 0.039257
```

Model Selection for Linear Mixed Effect Model

Random Intercept+CAR(1)

```
library("emdbook")
1-pchibarsq(2*(4838.693-4838.693),df=1,mix=0.5)
```

```
## [1] 0.5
```

Random Intercept+VC

```
1-pchibarsq(2*(4144.997-3930.385),df=2,mix=0.5)
```

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

Random Slope+VC

```
1-pchibarsq(2*(4422.051-4144.997),df=2,mix=0.5)
```

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

Summary: With continuous AR(1) covariance structure, we would like to argue that it's not necessary to add a random intercept, because the log likelihood didn't increase while the AIC and BIC index both decrease. We also considered to add a random slope to everyone, but lme() function in R was unable to run this with continuous AR(1) covariance structure, so we simply consider the default case, when the covariance for the residuals to be Variance Component. In this case, we can see that adding a random slope significantly improve the model according to likelihood ratio test.

Discussion: I suspect that adding a random slope could improve the quadratic regression with continuous AR(1), but R fails to converge for this setting.

log(Fev)

Covariance Structure Specification

Variance Component

```
model1.1=gls(logfev1~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+
baseht*I(age_centered^2),data=data,method='ML')
summary(model1.1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseage *      I
(age_centered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC  logLik
##   -2882.541 -2826.914 1451.27
##
## Coefficients:
##                               Value  Std.Error    t-value p-value
## (Intercept)              -0.4441416 0.06944965  -6.395160  0.0000
## baseage                  -0.0862636 0.00535077 -16.121724  0.0000
## age_centered              0.0944858 0.01497914   6.307826  0.0000
## baseht                    1.5992041 0.07917374  20.198668  0.0000
## I(age_centered^2)         0.0079281 0.00476163   1.664991  0.0961
## baseage:age_centered      0.0081826 0.00125770   6.506031  0.0000
## age_centered:baseht      -0.0571591 0.01640140  -3.485014  0.0005
## baseage:I(age_centered^2) -0.0004141 0.00039911  -1.037587  0.2996
## baseht:I(age_centered^2)  -0.0092563 0.00539423  -1.715953  0.0863
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.596
## age_centered              -0.051  0.007
## baseht                    -0.937 -0.836  0.029
## I(age_centered^2)         -0.716 -0.428  0.002  0.674
## baseage:age_centered       0.034  0.030  0.375 -0.032  0.048
## age_centered:baseht        0.020 -0.017 -0.899 -0.006 -0.019 -0.741
## baseage:I(age_centered^2) -0.412 -0.696  0.072  0.577  0.479 -0.174  0.022
## baseht:I(age_centered^2)   0.685  0.619 -0.027 -0.734 -0.917  0.036  0.002
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.788
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -4.10346214 -0.63694049  0.01964762  0.70601772  3.05432781
##
## Residual standard error: 0.1138532
## Degrees of freedom: 1925 total; 1916 residual
```

Compound Symmetry

```
model1.2=gls(logfev1~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+
baseht*I(age_centered^2),data=data,corr=corCompSymm(form=~1|id),method='ML')
summary(model1.2)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseage *      I
(age_centered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -3975.476 -3914.286 1998.738
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | id
##  Parameter estimate(s):
##       Rho
## 0.5862531
##
## Coefficients:
##                                Value  Std.Error    t-value p-value
## (Intercept)               -0.4135877 0.10411417  -3.972444  0.0001
## baseage                   -0.0947613 0.00818545 -11.576791  0.0000
## age_centered               0.1010489 0.01047978   9.642276  0.0000
## baseht                     1.6337953 0.12068819  13.537325  0.0000
## I(age_centered^2)          0.0086339 0.00327407   2.637061  0.0084
## baseage:age_centered       0.0083467 0.00089433   9.332823  0.0000
## age_centered:baseht       -0.0633117 0.01133396  -5.586017  0.0000
## baseage:I(age_centered^2) -0.0003810 0.00027529  -1.383987  0.1665
## baseht:I(age_centered^2)  -0.0102029 0.00368117  -2.771659  0.0056
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.627
## age_centered               0.027  0.035
## baseht                    -0.939 -0.854 -0.034
## I(age_centered^2)         -0.336 -0.205 -0.040  0.314
## baseage:age_centered       0.056  0.074  0.324 -0.066  0.055
## age_centered:baseht       -0.046 -0.060 -0.889  0.057  0.008 -0.719
## baseage:I(age_centered^2) -0.206 -0.320  0.073  0.274  0.452 -0.218  0.044
## baseht:I(age_centered^2)   0.329  0.293  0.002 -0.348 -0.913  0.051 -0.026
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.775
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -4.21079402 -0.68933311 -0.02238891  0.67670859  2.95954273
##
## Residual standard error: 0.1133853
## Degrees of freedom: 1925 total; 1916 residual
```

Continuous AR(1)

```
model1.3=gls(logfev1~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+
baseht*I(age_centered^2),data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseage *      I
(age_centered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -4259.984 -4198.795 2140.992
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.7644989
##
## Coefficients:
##                                Value  Std.Error    t-value p-value
## (Intercept)               -0.4704213 0.11173894  -4.210003  0.0000
## baseage                   -0.0966416 0.00872368 -11.078087  0.0000
## age_centered               0.0777927 0.01956916   3.975273  0.0001
## baseht                     1.6867668 0.12938135  13.037171  0.0000
## I(age_centered^2)          0.0105790 0.00490060   2.158726  0.0310
## baseage:age_centered       0.0083519 0.00163126   5.119932  0.0000
## age_centered:baseht       -0.0472485 0.02190863  -2.156616  0.0312
## baseage:I(age_centered^2) -0.0000504 0.00040648  -0.123900  0.9014
## baseht:I(age_centered^2)  -0.0137400 0.00556294  -2.469919  0.0136
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.633
## age_centered               0.042  0.051
## baseht                    -0.941 -0.856 -0.052
## I(age_centered^2)         -0.617 -0.387 -0.009  0.581
## baseage:age_centered       0.069  0.089  0.446 -0.081  0.043
## age_centered:baseht       -0.062 -0.075 -0.910  0.074 -0.009 -0.776
## baseage:I(age_centered^2) -0.378 -0.594  0.059  0.507  0.496 -0.124  0.010
## baseht:I(age_centered^2)   0.597  0.541 -0.015 -0.634 -0.920  0.019  0.001
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.795
##
## Standardized residuals:
##          Min           Q1          Med           Q3          Max
## -4.207879314 -0.672505130  0.004005873  0.698147655  3.034908058
##
## Residual standard error: 0.1134055
## Degrees of freedom: 1925 total; 1916 residual
```

Summary: Indicated by the AIC and BIC, we will again choose continuous AR(1) to be the covariance structure.

Model Selection for Linear Model

first

```
model1.3=gls(logfev1~baseage*age_centered+baseht*age_centered+baseage*I(age_centered^2)+
baseht*I(age_centered^2),data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseage *      I
(age_centered^2) + baseht * I(age_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -4259.984 -4198.795 2140.992
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.7644989
##
## Coefficients:
##                                Value  Std.Error    t-value p-value
## (Intercept)               -0.4704213 0.11173894  -4.210003  0.0000
## baseage                   -0.0966416 0.00872368 -11.078087  0.0000
## age_centered               0.0777927 0.01956916   3.975273  0.0001
## baseht                     1.6867668 0.12938135  13.037171  0.0000
## I(age_centered^2)          0.0105790 0.00490060   2.158726  0.0310
## baseage:age_centered       0.0083519 0.00163126   5.119932  0.0000
## age_centered:baseht       -0.0472485 0.02190863  -2.156616  0.0312
## baseage:I(age_centered^2) -0.0000504 0.00040648  -0.123900  0.9014
## baseht:I(age_centered^2)  -0.0137400 0.00556294  -2.469919  0.0136
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.633
## age_centered               0.042  0.051
## baseht                    -0.941 -0.856 -0.052
## I(age_centered^2)         -0.617 -0.387 -0.009  0.581
## baseage:age_centered       0.069  0.089  0.446 -0.081  0.043
## age_centered:baseht       -0.062 -0.075 -0.910  0.074 -0.009 -0.776
## baseage:I(age_centered^2) -0.378 -0.594  0.059  0.507  0.496 -0.124  0.010
## baseht:I(age_centered^2)   0.597  0.541 -0.015 -0.634 -0.920  0.019  0.001
##                           bsg:I(_^2)
## baseage
## age_centered
## baseht
## I(age_centered^2)
## baseage:age_centered
## age_centered:baseht
## baseage:I(age_centered^2)
## baseht:I(age_centered^2)  -0.795
##
## Standardized residuals:
##          Min           Q1          Med           Q3          Max
## -4.207879314 -0.672505130  0.004005873  0.698147655  3.034908058
##
## Residual standard error: 0.1134055
## Degrees of freedom: 1925 total; 1916 residual
```

Second

From the output above, we will exclude the interaction between the baseage and $age^2_{centered}$

```
model1.3=gls(logfev1~baseage*age_centered+ht*age_centered+ht*I(age_centered^2),data=dat
a,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1.3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + ht * age_centered + ht * I(age_centered^
2)
##   Data: data
##         AIC       BIC   logLik
##   -4439.59 -4383.963 2229.795
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.7767047
##
## Coefficients:
##                           Value  Std.Error    t-value p-value
## (Intercept)          -1.2145062 0.11087912 -10.953426  0.0000
## baseage              -0.0021143 0.00384085  -0.550470  0.5821
## age_centered         -0.0539286 0.02683715  -2.009476  0.0446
## ht                    1.3815324 0.06772420  20.399391  0.0000
## I(age_centered^2)    -0.0027501 0.00316780  -0.868132  0.3854
## baseage:age_centered  0.0024013 0.00099215   2.420276  0.0156
## age_centered:ht       0.0438506 0.01693051   2.590031  0.0097
## ht:I(age_centered^2) -0.0004766 0.00194778  -0.244684  0.8067
##
##  Correlation:
##                      (Intr) baseag ag_cnt ht     I(_^2) bsg:g_ ag_cn:
## baseage              -0.330
## age_centered          0.025  0.031
## ht                   -0.955  0.042 -0.026
## I(age_centered^2)    -0.064  0.115  0.720  0.042
## baseage:age_centered  0.128 -0.155 -0.297 -0.085 -0.208
## age_centered:ht       0.040  0.014 -0.946 -0.058 -0.651  0.000
## ht:I(age_centered^2) -0.035 -0.109 -0.602  0.058 -0.977  0.201  0.512
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -4.5347083 -0.7124173  0.0335971  0.7128241  2.8764129
##
## Residual standard error: 0.1103691
## Degrees of freedom: 1925 total; 1917 residual
```

Summary: This results in the following linear model:

```
model=gls(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),data
=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *      I(a
ge_centered^2)
##   Data: data
##         AIC       BIC   logLik
##   -4261.969 -4206.342 2140.985
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.7645114
##
## Coefficients:
##                                Value  Std.Error    t-value p-value
## (Intercept)               -0.4756576 0.10341665  -4.599430  0.0000
## baseage                   -0.0972844 0.00701344 -13.871128  0.0000
## age_centered               0.0779349 0.01953055   3.990411  0.0001
## baseht                     1.6949025 0.11146382  15.205854  0.0000
## I(age_centered^2)          0.0108803 0.00425424   2.557513  0.0106
## baseage:age_centered       0.0083268 0.00161823   5.145630  0.0000
## age_centered:baseht       -0.0472213 0.02190225  -2.156001  0.0312
## baseht:I(age_centered^2) -0.0142882 0.00337140  -4.238075  0.0000
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                    0.548
## age_centered               0.070  0.107
## baseht                    -0.939 -0.800 -0.095
## I(age_centered^2)         -0.534 -0.132 -0.044  0.440
## baseage:age_centered       0.023  0.018  0.458 -0.021  0.121
## age_centered:baseht       -0.063 -0.086 -0.912  0.080 -0.016 -0.781
## baseht:I(age_centered^2)  0.528  0.140  0.053 -0.441 -0.998 -0.133  0.015
##
## Standardized residuals:
##          Min           Q1          Med           Q3          Max
## -4.211315612 -0.672744687  0.002213593  0.698225883  3.036833234
##
## Residual standard error: 0.1134081
## Degrees of freedom: 1925 total; 1917 residual
```

Linear Mixed Effect Model

Now we are considering to add a random effect to the model we selected from the previous model, and we are wondering random intercept, random age, which one would be the best.

Random Intercept for CAR(1)

```
model=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),rand
om=~1|id,data=data,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -4324.731 -4263.542 2173.366
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:   0.0765799 0.08299468
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.5543551
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *
I(age_centered^2)
##                              Value  Std.Error   DF    t-value p-value
## (Intercept)             -0.4617239 0.10565333 1630  -4.370179  0.0000
## baseage                 -0.0983031 0.00773473  287 -12.709319  0.0000
## age_centered             0.0894918 0.01495700 1630   5.983271  0.0000
## baseht                   1.6919647 0.11813115  287  14.322764  0.0000
## I(age_centered^2)        0.0111648 0.00368091 1630   3.033176  0.0025
## baseage:age_centered     0.0080803 0.00125115 1630   6.458302  0.0000
## age_centered:baseht     -0.0538530 0.01657942 1630  -3.248184  0.0012
## baseht:I(age_centered^2) -0.0145157 0.00292325 1630  -4.965610  0.0000
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.587
## age_centered              0.069  0.089
## baseht                   -0.938 -0.829 -0.087
## I(age_centered^2)        -0.389 -0.097 -0.064  0.313
## baseage:age_centered      0.026  0.024  0.413 -0.025  0.146
## age_centered:baseht      -0.063 -0.077 -0.904  0.076 -0.013 -0.760
## baseht:I(age_centered^2)  0.385  0.103  0.074 -0.314 -0.998 -0.158  0.012
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.53494531 -0.54680099  0.03047478  0.63011954  2.48426377
##
## Number of Observations: 1925
## Number of Groups: 290
```

Random Slope for CAR(1) (diverge)

```
#model=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),ran
dom=~age_centered|id,data=data,corr=corCAR1(form=~age_centered|id),method='ML')
```

Variance Component

```
model2.1=gls(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),d
ata=data,method='ML')
summary(model2.1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *      I(a
ge_centered^2)
##   Data: data
##        AIC       BIC  logLik
##   -2883.46 -2833.395 1450.73
##
## Coefficients:
##                              Value  Std.Error    t-value p-value
## (Intercept)             -0.4738447 0.06327640  -7.488491  0.0000
## baseage                 -0.0901256 0.00384416 -23.444821  0.0000
## age_centered             0.0956038 0.01494064   6.398910  0.0000
## baseht                   1.6466066 0.06466447  25.463855  0.0000
## I(age_centered^2)        0.0102943 0.00418006   2.462727  0.0139
## baseage:age_centered     0.0079561 0.00123863   6.423308  0.0000
## age_centered:baseht     -0.0567812 0.01639768  -3.462761  0.0005
## baseht:I(age_centered^2) -0.0136651 0.00332307  -4.112196  0.0000
##
##  Correlation:
##                         (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                  0.472
## age_centered            -0.023  0.079
## baseht                  -0.940 -0.741 -0.015
## I(age_centered^2)       -0.649 -0.150 -0.037  0.554
## baseage:age_centered    -0.042 -0.128  0.394  0.085  0.151
## age_centered:baseht      0.031 -0.002 -0.903 -0.023 -0.034 -0.749
## baseht:I(age_centered^2)  0.642  0.161  0.048 -0.556 -0.998 -0.166  0.032
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -4.12198765 -0.63858732  0.01483588  0.70796713  3.06505003
##
## Residual standard error: 0.1138852
## Degrees of freedom: 1925 total; 1917 residual
```

Random Intercept for general Variance Component

```
model2.1=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),r
andom=~1|id,data=data,method='ML')
summary(model2.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -3975.552 -3919.926 1997.776
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:  0.08680447 0.07297617
##
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *
I(age_centered^2)
##                              Value  Std.Error   DF    t-value p-value
## (Intercept)             -0.4433566 0.10184657 1630  -4.353182   0e+00
## baseage                 -0.0983801 0.00775375  287 -12.688058   0e+00
## age_centered             0.1021039 0.01045498 1630   9.766049   0e+00
## baseht                   1.6796296 0.11602430  287  14.476532   0e+00
## I(age_centered^2)        0.0106807 0.00292211 1630   3.655134   3e-04
## baseage:age_centered     0.0080776 0.00087318 1630   9.250764   0e+00
## age_centered:baseht     -0.0626195 0.01132650 1630  -5.528583   0e+00
## baseht:I(age_centered^2) -0.0141522 0.00232661 1630  -6.082751   0e+00
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.605
## age_centered              0.043  0.062
## baseht                   -0.938 -0.841 -0.057
## I(age_centered^2)        -0.278 -0.072 -0.082  0.222
## baseage:age_centered      0.011  0.005  0.349 -0.007  0.176
## age_centered:baseht      -0.038 -0.049 -0.896  0.047 -0.013 -0.728
## baseht:I(age_centered^2)  0.274  0.076  0.094 -0.222 -0.998 -0.191  0.012
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.10506112 -0.56954571  0.02822295  0.63480024  2.70618931
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
#Residual
sigma(model2.1)^2
```

```
## [1] 0.005325522
```

```
# Random Effect
getVarCov(model2.1,type="random")
```

```
## Random effects variance covariance matrix
##             (Intercept)
## (Intercept)    0.007535
##   Standard Deviations: 0.086804
```

```
# Marginal covariance
getVarCov(model2.1,type="marginal",individuals = 1)
```

```
## id 2
## Marginal variance covariance matrix
##          1        2        3        4        5        6        7        8
## 1 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 2 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 3 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535
## 4 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535
## 5 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535
## 6 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535
## 7 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535
## 8 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861
##   Standard Deviations: 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134
```

```
getVarCov(model2.1,type="marginal",individuals = 2)
```

```
## id 3
## Marginal variance covariance matrix
##          1        2        3        4        5        6        7        8
## 1 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 2 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 3 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535
## 4 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535
## 5 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535
## 6 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535
## 7 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535
## 8 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861
## 9 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
##          9
## 1 0.007535
## 2 0.007535
## 3 0.007535
## 4 0.007535
## 5 0.007535
## 6 0.007535
## 7 0.007535
## 8 0.007535
## 9 0.012861
##   Standard Deviations: 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134
```

Random Slope for general Variance Component

```
model2.2=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),r
andom=~age_centered|id,data=data,method='ML')
summary(model2.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -4052.125 -3985.373 2038.063
##
## Random effects:
##  Formula: ~age_centered | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev      Corr
## (Intercept)  0.086697392 (Intr)
## age_centered 0.008296027 0.116
## Residual     0.067808588
##
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *
I(age_centered^2)
##                              Value  Std.Error   DF    t-value p-value
## (Intercept)             -0.4601934 0.10168830 1630  -4.525530   0e+00
## baseage                 -0.0987906 0.00777499  287 -12.706195   0e+00
## age_centered             0.0965854 0.01445811 1630   6.680362   0e+00
## baseht                   1.6947600 0.11622103  287  14.582215   0e+00
## I(age_centered^2)        0.0123473 0.00293586 1630   4.205689   0e+00
## baseage:age_centered     0.0085317 0.00118108 1630   7.223664   0e+00
## age_centered:baseht     -0.0610615 0.01608639 1630  -3.795847   2e-04
## baseht:I(age_centered^2) -0.0154507 0.00233701 1630  -6.611324   0e+00
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.611
## age_centered              0.137  0.120
## baseht                   -0.939 -0.845 -0.145
## I(age_centered^2)        -0.258 -0.072 -0.070  0.208
## baseage:age_centered      0.078  0.114  0.450 -0.100  0.148
## age_centered:baseht      -0.133 -0.139 -0.912  0.149 -0.010 -0.775
## baseht:I(age_centered^2)  0.256  0.075  0.078 -0.208 -0.998 -0.158  0.009
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.56508528 -0.53298621  0.03636637  0.59691753  2.60740166
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
#Residual
sigma(model2.2)^2
```

```
## [1] 0.004598005
```

```
# Random Effect
getVarCov(model2.2,type="random")
```

```
## Random effects variance covariance matrix
##              (Intercept) age_centered
## (Intercept)   7.5164e-03   8.3788e-05
## age_centered  8.3788e-05   6.8824e-05
##   Standard Deviations: 0.086697 0.008296
```

```
# Marginal covariance
getVarCov(model2.2,type="marginal",individuals = 1)
```

```
## id 2
## Marginal variance covariance matrix
##           1         2         3         4         5         6         7
## 1 0.0136620 0.0087058 0.0069911 0.0066424 0.0063325 0.0059524 0.0056674
## 2 0.0087058 0.0130240 0.0070829 0.0068099 0.0065673 0.0062697 0.0060465
## 3 0.0069911 0.0070829 0.0121210 0.0076126 0.0076922 0.0077897 0.0078629
## 4 0.0066424 0.0068099 0.0076126 0.0123740 0.0079209 0.0080988 0.0082322
## 5 0.0063325 0.0065673 0.0076922 0.0079209 0.0127220 0.0083735 0.0085605
## 6 0.0059524 0.0062697 0.0077897 0.0080988 0.0083735 0.0133080 0.0089631
## 7 0.0056674 0.0060465 0.0078629 0.0082322 0.0085605 0.0089631 0.0138630
## 8 0.0053427 0.0057923 0.0079463 0.0083842 0.0087736 0.0092509 0.0096090
##           8
## 1 0.0053427
## 2 0.0057923
## 3 0.0079463
## 4 0.0083842
## 5 0.0087736
## 6 0.0092509
## 7 0.0096090
## 8 0.0146150
##   Standard Deviations: 0.11688 0.11412 0.1101 0.11124 0.11279 0.11536 0.11774 0.12089
```

```
getVarCov(model2.2,type="marginal",individuals = 2)
```

```
## id 3
## Marginal variance covariance matrix
##           1         2         3         4         5         6         7
## 1 0.0134490 0.0085174 0.0082057 0.0078846 0.0075419 0.0069159 0.0065904
## 2 0.0085174 0.0128590 0.0080218 0.0077753 0.0075123 0.0070318 0.0067820
## 3 0.0082057 0.0080218 0.0124480 0.0076733 0.0074846 0.0071399 0.0069607
## 4 0.0078846 0.0077753 0.0076733 0.0121660 0.0074561 0.0072513 0.0071449
## 5 0.0075419 0.0075123 0.0074846 0.0074561 0.0120240 0.0073702 0.0073413
## 6 0.0069159 0.0070318 0.0071399 0.0072513 0.0073702 0.0121850 0.0077003
## 7 0.0065904 0.0067820 0.0069607 0.0071449 0.0073413 0.0077003 0.0124850
## 8 0.0062830 0.0065461 0.0067915 0.0070443 0.0073141 0.0078069 0.0080632
## 9 0.0059886 0.0063201 0.0066293 0.0069480 0.0072880 0.0079091 0.0082320
##           8         9
## 1 0.0062830 0.0059886
## 2 0.0065461 0.0063201
## 3 0.0067915 0.0066293
## 4 0.0070443 0.0069480
## 5 0.0073141 0.0072880
## 6 0.0078069 0.0079091
## 7 0.0080632 0.0082320
## 8 0.0129030 0.0085370
## 9 0.0085370 0.0134270
##   Standard Deviations: 0.11597 0.1134 0.11157 0.1103 0.10965 0.11039 0.11174 0.11359
0.11588
```

Model Selection for Linear Mixed Effect Model

Random Intercept+CAR(1)

```
library("emdbook")
1-pchibarsq(2*(2173.366-2140.985),df=1,mix=0.5)
```

```
## [1] 4.440892e-16
```

Random Intercept+VC

```
1-pchibarsq(2*(1997.776-1450.730),df=1,mix=0.5)
```

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

Random Slope+VC

```
1-pchibarsq(2*(2038.063-1997.776),df=2,mix=0.5)
```

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

Summary : We see for log(fev1), it's necessary to add a random intercept under Continuous AR(1), and under general unstructured covariance, random slope is also needed.

Model Diagnostic

Selected Model

Height

```
model1=gls(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),data=da
ta,corr=corCAR1(form=~age_centered|id),method='ML')
summary(model1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_centered
^2)
##   Data: data
##         AIC       BIC   logLik
##   -9659.386 -9609.322 4838.693
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.8017266
##
## Coefficients:
##                                Value   Std.Error   t-value p-value
## (Intercept)                0.8060176 0.023824950  33.83082       0
## baseage                   -0.0500622 0.001881813 -26.60319       0
## age_centered               0.0481952 0.005021929   9.59695       0
## baseht                     0.8946701 0.027296194  32.77637       0
## baseage:age_centered       0.0037196 0.000413654   8.99213       0
## age_centered:baseht       -0.0296846 0.005639317  -5.26387       0
## baseage:I(age_centered^2) -0.0004942 0.000009015 -54.82659       0
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.583
## age_centered               0.059  0.095
## baseht                    -0.931 -0.837 -0.084
## baseage:age_centered       0.108  0.052  0.468 -0.092
## age_centered:baseht       -0.088 -0.092 -0.914  0.100 -0.785
## baseage:I(age_centered^2) -0.092  0.070  0.134  0.001 -0.186 -0.010
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -3.50715974 -0.70733519 -0.01249239  0.63528182  3.47071042
##
## Residual standard error: 0.02971482
## Degrees of freedom: 1925 total; 1918 residual
```

log(Fev)

```
model2=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),ran
dom=~1|id,corr=corCAR1(form=~age_centered|id),data=data,method='ML')
summary(model2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -4324.731 -4263.542 2173.366
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:   0.0765799 0.08299468
##
## Correlation Structure: Continuous AR(1)
##  Formula: ~age_centered | id
##  Parameter estimate(s):
##       Phi
## 0.5543551
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *
I(age_centered^2)
##                               Value  Std.Error   DF    t-value p-value
## (Intercept)              -0.4617239 0.10565333 1630  -4.370179  0.0000
## baseage                  -0.0983031 0.00773473  287 -12.709319  0.0000
## age_centered              0.0894918 0.01495700 1630   5.983271  0.0000
## baseht                    1.6919647 0.11813115  287  14.322764  0.0000
## I(age_centered^2)         0.0111648 0.00368091 1630   3.033176  0.0025
## baseage:age_centered      0.0080803 0.00125115 1630   6.458302  0.0000
## age_centered:baseht      -0.0538530 0.01657942 1630  -3.248184  0.0012
## baseht:I(age_centered^2) -0.0145157 0.00292325 1630  -4.965610  0.0000
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.587
## age_centered              0.069  0.089
## baseht                   -0.938 -0.829 -0.087
## I(age_centered^2)        -0.389 -0.097 -0.064  0.313
## baseage:age_centered      0.026  0.024  0.413 -0.025  0.146
## age_centered:baseht      -0.063 -0.077 -0.904  0.076 -0.013 -0.760
## baseht:I(age_centered^2)  0.385  0.103  0.074 -0.314 -0.998 -0.158  0.012
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.53494531 -0.54680099  0.03047478  0.63011954  2.48426377
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
ran=getVarCov(model2,type="random")
mar=getVarCov(model2,type="margin")
# Marginal covariance matrix
mar
```

```
## id 2
## Marginal variance covariance matrix
##           1         2         3         4         5         6         7
## 1 0.0127530 0.0095451 0.0060473 0.0059637 0.0059222 0.0058941 0.0058825
## 2 0.0095451 0.0127530 0.0062065 0.0060502 0.0059724 0.0059200 0.0058982
## 3 0.0060473 0.0062065 0.0127530 0.0096051 0.0080385 0.0069820 0.0065429
## 4 0.0059637 0.0060502 0.0096051 0.0127530 0.0098677 0.0079223 0.0071137
## 5 0.0059222 0.0059724 0.0080385 0.0098677 0.0127530 0.0094052 0.0080140
## 6 0.0058941 0.0059200 0.0069820 0.0079223 0.0094052 0.0127530 0.0100460
## 7 0.0058825 0.0058982 0.0065429 0.0071137 0.0080140 0.0100460 0.0127530
## 8 0.0058747 0.0058836 0.0062487 0.0065720 0.0070818 0.0082326 0.0097654
##           8
## 1 0.0058747
## 2 0.0058836
## 3 0.0062487
## 4 0.0065720
## 5 0.0070818
## 6 0.0082326
## 7 0.0097654
## 8 0.0127530
##   Standard Deviations: 0.11293 0.11293 0.11293 0.11293 0.11293 0.11293 0.11293 0.1129
3
```

```
# Random Intercept covariance matrix
ran
```

```
## Random effects variance covariance matrix
##             (Intercept)
## (Intercept)   0.0058645
##   Standard Deviations: 0.07658
```

```
# residual covariance matrix
mar$`2`-matrix(as.numeric(ran)*rep(1,64), nrow = 8, ncol = 8)
```

```
##              1            2            3            4            5            6
## 1 6.888116e-03 3.680666e-03 0.0001827722 9.925587e-05 5.768552e-05 2.965241e-05
## 2 3.680666e-03 6.888116e-03 0.0003420457 1.857506e-04 1.079545e-04 5.549246e-05
## 3 1.827722e-04 3.420457e-04 0.0068881162 3.740645e-03 2.173988e-03 1.117507e-03
## 4 9.925587e-05 1.857506e-04 0.0037406453 6.888116e-03 4.003235e-03 2.057805e-03
## 5 5.768552e-05 1.079545e-04 0.0021739881 4.003235e-03 6.888116e-03 3.540736e-03
## 6 2.965241e-05 5.549246e-05 0.0011175070 2.057805e-03 3.540736e-03 6.888116e-03
## 7 1.800133e-05 3.368827e-05 0.0006784143 1.249249e-03 2.149504e-03 4.181626e-03
## 8 1.019454e-05 1.907838e-05 0.0003842004 7.074760e-04 1.217310e-03 2.368144e-03
##              7            8
## 1 1.800133e-05 1.019454e-05
## 2 3.368827e-05 1.907838e-05
## 3 6.784143e-04 3.842004e-04
## 4 1.249249e-03 7.074760e-04
## 5 2.149504e-03 1.217310e-03
## 6 4.181626e-03 2.368144e-03
## 7 6.888116e-03 3.900886e-03
## 8 3.900886e-03 6.888116e-03
```

```
head(ranef(model2))
```

```
##   (Intercept)
## 2  0.04786341
## 3  0.09190034
## 4 -0.01704989
## 5 -0.05905019
## 6  0.01738473
## 7 -0.10142058
```

```
head(fitted(model2,level=0),20)
```

```
##         2         2         2         2         2         2         2         2
## 0.1065108 0.2555904 0.8058450 0.8845327 0.9450574 1.0071744 1.0450148 1.0789878
##         3         3         3         3         3         3         3         3
## 0.1821028 0.3355973 0.4666671 0.5894623 0.7067663 0.8844962 0.9582293 1.0161362
##         3         4         4         4
## 1.0609223 0.1362790 0.2870981 0.4165719
```

```
head(fitted(model2,level=1),20)
```

```
##         2         2         2         2         2         2         2         2
## 0.1543742 0.3034538 0.8537085 0.9323961 0.9929209 1.0550378 1.0928782 1.1268512
##         3         3         3         3         3         3         3         3
## 0.2740031 0.4274976 0.5585674 0.6813626 0.7986666 0.9763965 1.0501297 1.1080365
##         3         4         4         4
## 1.1528226 0.1192291 0.2700482 0.3995220
```

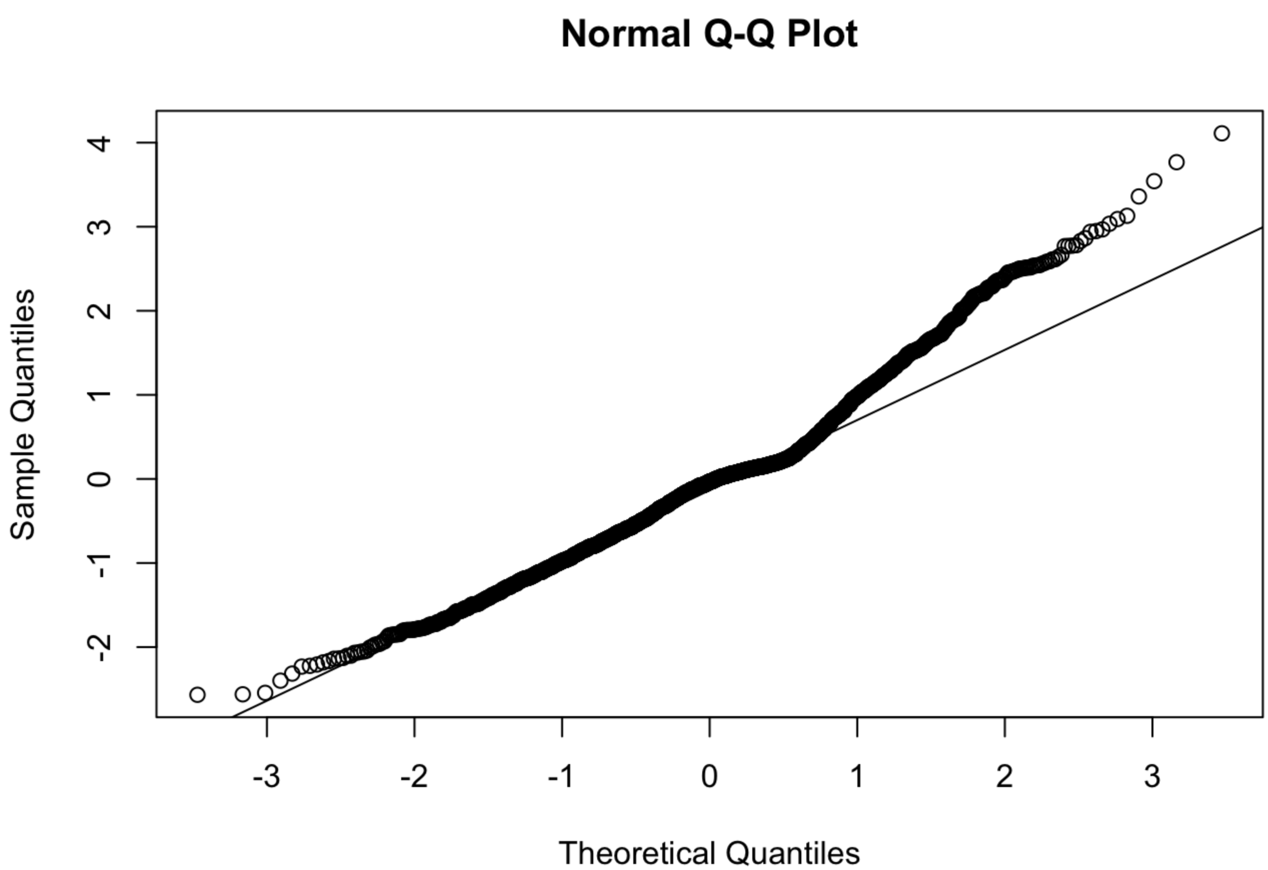
QQ-Plot

residuals

```
residuals1=residuals(model1,type="normalized")
residuals2=residuals(model2,type="normalized")
```

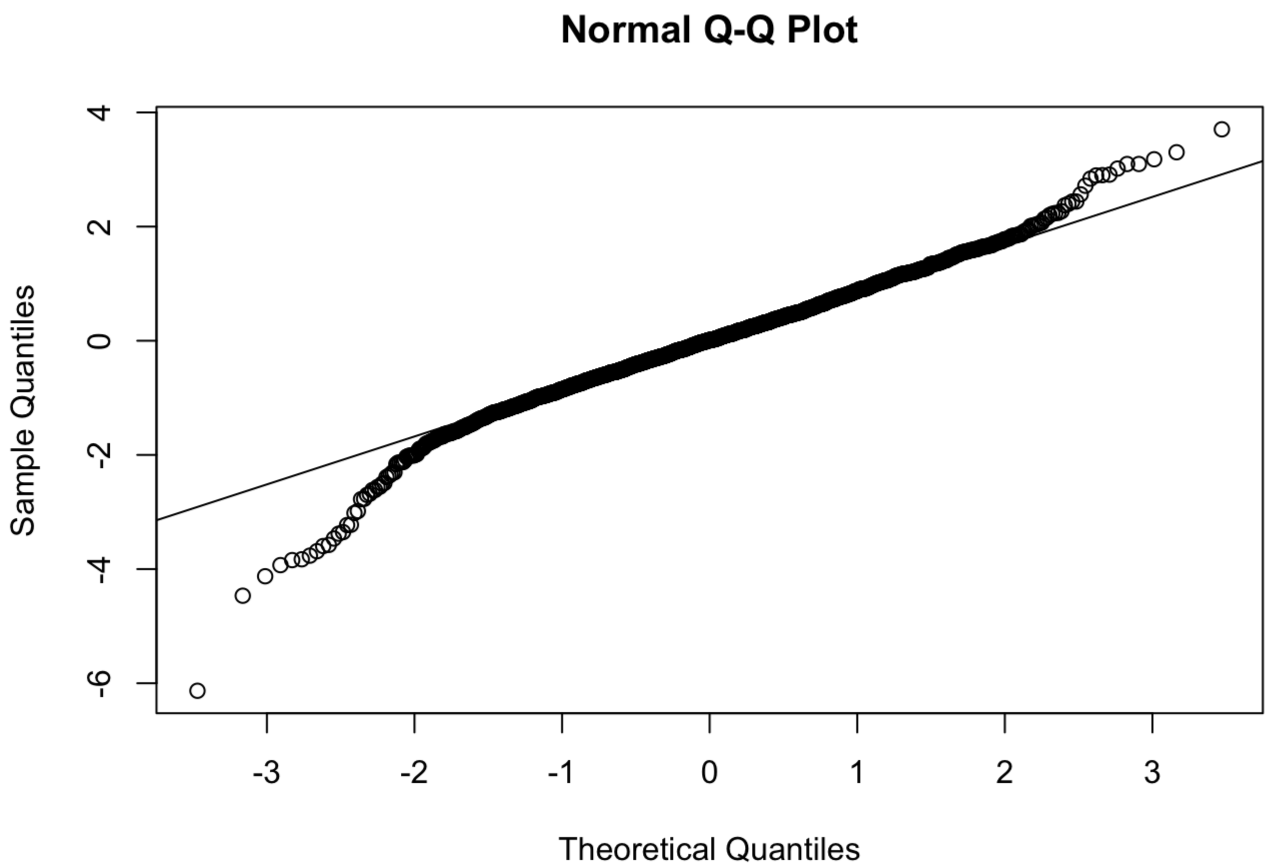
Height

```
qqnorm(residuals1)
qqline(residuals1)
```



log(Fev1)

```
qqnorm(residuals2)
qqline(residuals2)
```



Calculate the determinant of id=2 in the first model

```
# model1 is from gls() and we don't really need to specify the type as lme()
x=getVarCov(model1)
y=as.vector(x)
z=matrix(y,nrow=8)
det(chol(z))
```

```
## [1] 2.539926e-14
```

Try variance compound

Height

```
model1=gls(ht~baseage*age_centered+baseht*age_centered+baseage:I(age_centered^2),data=da
ta,method='ML')
summary(model1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: ht ~ baseage * age_centered + baseht * age_centered + baseage:I(age_centered
^2)
##   Data: data
##         AIC       BIC   logLik
##   -7844.771 -7800.269 3930.385
##
## Coefficients:
##                              Value   Std.Error   t-value p-value
## (Intercept)              0.7956633 0.013295854  59.84296       0
## baseage                 -0.0502904 0.001036776 -48.50651       0
## age_centered             0.0563197 0.004122000  13.66319       0
## baseht                   0.9080605 0.014823978  61.25619       0
## baseage:age_centered     0.0037316 0.000339377  10.99531       0
## age_centered:baseht     -0.0351481 0.004518187  -7.77924       0
## baseage:I(age_centered^2) -0.0005446 0.000009983 -54.55681       0
##
##  Correlation:
##                           (Intr) baseag ag_cnt baseht bsg:g_ ag_cn:
## baseage                    0.511
## age_centered              -0.065  0.062
## baseht                    -0.921 -0.802  0.014
## baseage:age_centered       0.080 -0.091  0.397 -0.009
## age_centered:baseht        0.012 -0.006 -0.904 -0.006 -0.749
## baseage:I(age_centered^2) -0.116  0.089  0.181 -0.001 -0.248 -0.016
##
## Standardized residuals:
##          Min           Q1          Med           Q3          Max
## -3.313881846 -0.654537752 -0.008381967  0.613382205  3.573507587
##
## Residual standard error: 0.03140786
## Degrees of freedom: 1925 total; 1918 residual
```

log(Fev)

```
model2=lme(logfev1~baseage*age_centered+baseht*age_centered+baseht*I(age_centered^2),ran
dom=~1|id,data=data,method='ML')
summary(model2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -3975.552 -3919.926 1997.776
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:  0.08680447 0.07297617
##
## Fixed effects:  logfev1 ~ baseage * age_centered + baseht * age_centered + baseht *
I(age_centered^2)
##                                Value  Std.Error   DF    t-value p-value
## (Intercept)               -0.4433566 0.10184657 1630  -4.353182   0e+00
## baseage                   -0.0983801 0.00775375  287 -12.688058   0e+00
## age_centered               0.1021039 0.01045498 1630   9.766049   0e+00
## baseht                     1.6796296 0.11602430  287  14.476532   0e+00
## I(age_centered^2)          0.0106807 0.00292211 1630   3.655134   3e-04
## baseage:age_centered       0.0080776 0.00087318 1630   9.250764   0e+00
## age_centered:baseht       -0.0626195 0.01132650 1630  -5.528583   0e+00
## baseht:I(age_centered^2) -0.0141522 0.00232661 1630  -6.082751   0e+00
##  Correlation:
##                          (Intr) baseag ag_cnt baseht I(_^2) bsg:g_ ag_cn:
## baseage                   0.605
## age_centered              0.043  0.062
## baseht                   -0.938 -0.841 -0.057
## I(age_centered^2)        -0.278 -0.072 -0.082  0.222
## baseage:age_centered      0.011  0.005  0.349 -0.007  0.176
## age_centered:baseht      -0.038 -0.049 -0.896  0.047 -0.013 -0.728
## baseht:I(age_centered^2)  0.274  0.076  0.094 -0.222 -0.998 -0.191  0.012
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.10506112 -0.56954571  0.02822295  0.63480024  2.70618931
##
## Number of Observations: 1925
## Number of Groups: 290
```

```
ran=getVarCov(model2,type="random")
mar=getVarCov(model2,type="margin")
# Marginal covariance matrix
mar
```

```
## id 2
## Marginal variance covariance matrix
##          1        2        3        4        5        6        7        8
## 1 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 2 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535
## 3 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535 0.007535
## 4 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535 0.007535
## 5 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535 0.007535
## 6 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535 0.007535
## 7 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861 0.007535
## 8 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.007535 0.012861
##   Standard Deviations: 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134 0.1134
```

```
# Random Intercept covariance matrix
ran
```

```
## Random effects variance covariance matrix
##             (Intercept)
## (Intercept)    0.007535
##   Standard Deviations: 0.086804
```

```
# residual covariance matrix
mar$`2`-matrix(as.numeric(ran)*rep(1,64), nrow = 8, ncol = 8)
```

```
##             1           2           3           4           5           6
## 1 0.005325522 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## 2 0.000000000 0.005325522 0.000000000 0.000000000 0.000000000 0.000000000
## 3 0.000000000 0.000000000 0.005325522 0.000000000 0.000000000 0.000000000
## 4 0.000000000 0.000000000 0.000000000 0.005325522 0.000000000 0.000000000
## 5 0.000000000 0.000000000 0.000000000 0.000000000 0.005325522 0.000000000
## 6 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.005325522
## 7 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## 8 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
##             7           8
## 1 0.000000000 0.000000000
## 2 0.000000000 0.000000000
## 3 0.000000000 0.000000000
## 4 0.000000000 0.000000000
## 5 0.000000000 0.000000000
## 6 0.000000000 0.000000000
## 7 0.005325522 0.000000000
## 8 0.000000000 0.005325522
```

```
head(ranef(model2))
```

```
##   (Intercept)
## 2  0.04355109
## 3  0.10716242
## 4 -0.03941513
## 5 -0.07166441
## 6  0.02562829
## 7 -0.11441950
```

```
head(fitted(model2,level=0),20)
```

```
##          2          2          2          2          2          2          2 
## 0.09125924 0.24406468 0.80987210 0.89125697 0.95404724 1.01877265 1.05844683 
##          2          3          3          3          3          3          3 
## 1.09438098 0.17051999 0.32702001 0.46078150 0.58623670 0.70625177 0.88859581 
##          3          3          3          4          4          4 
## 0.96455499 1.02445562 1.07105272 0.12268453 0.27693201 0.40947003
```

```
head(fitted(model2,level=1),20)
```

```
##         2         2         2         2         2         2         2         2 
## 0.1348103 0.2876158 0.8534232 0.9348081 0.9975983 1.0623237 1.1019979 1.1379321 
##         3         3         3         3         3         3         3         3 
## 0.2776824 0.4341824 0.5679439 0.6933991 0.8134142 0.9957582 1.0717174 1.1316180 
##         3         4         4         4 
## 1.1782151 0.0832694 0.2375169 0.3700549
```

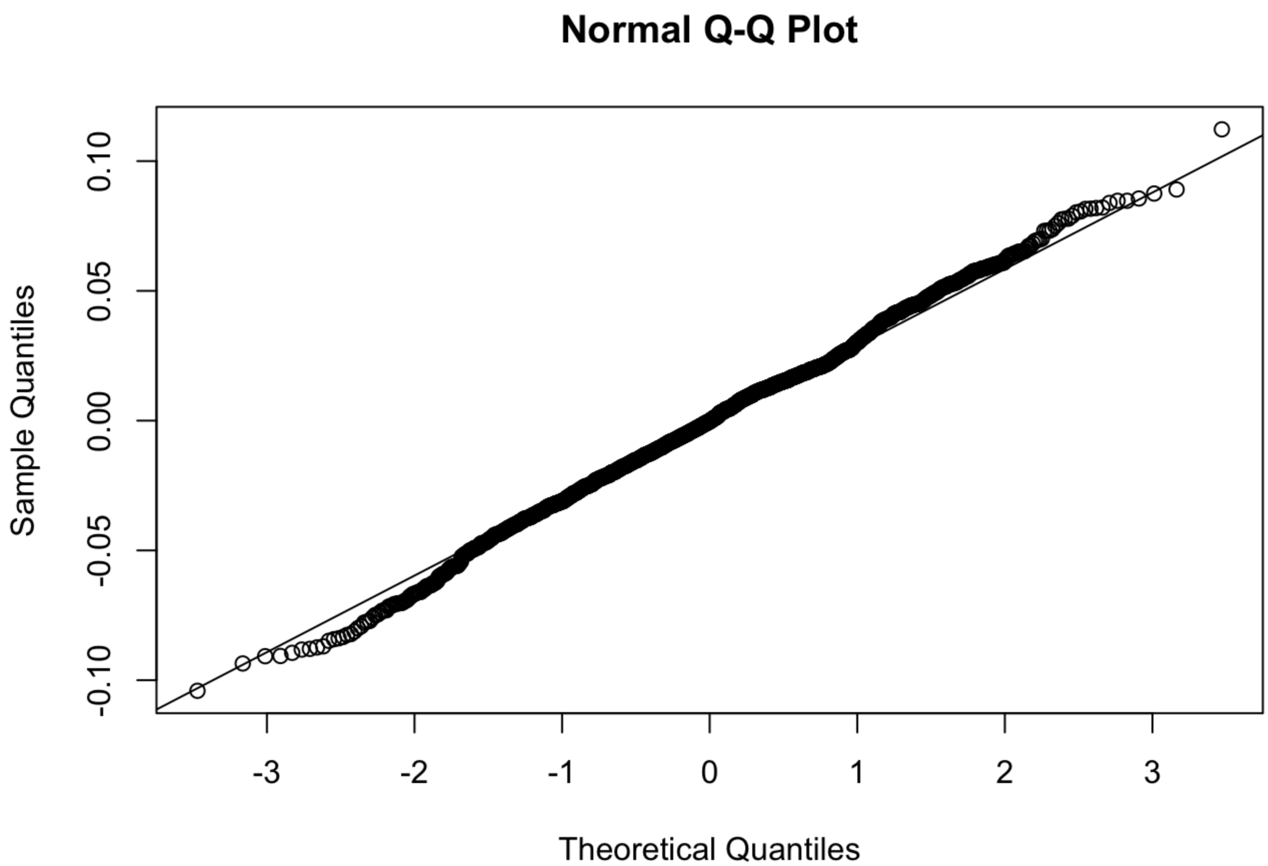
QQ-Plot

residuals

```
residuals1=residuals(model1)
residuals2=residuals(model2)
```

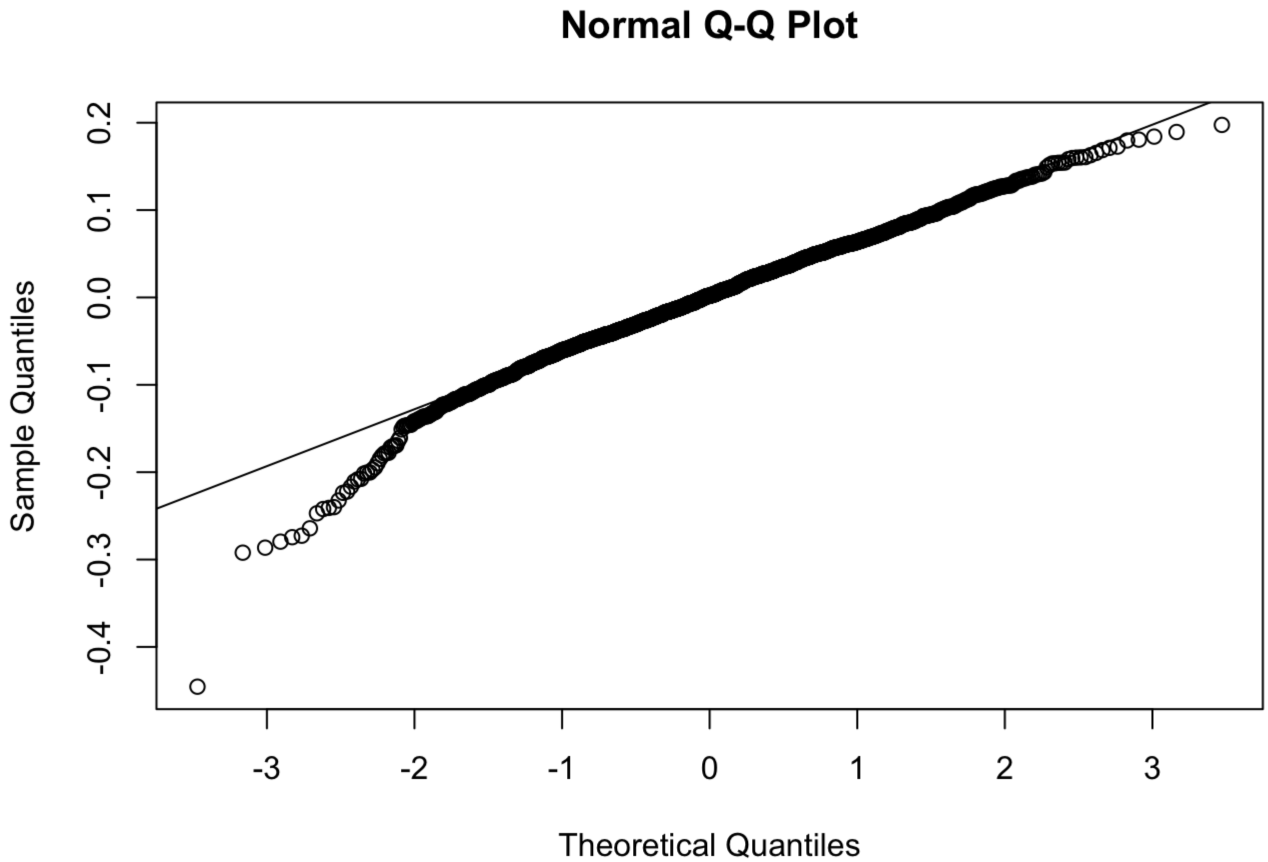
Height

```
qqnorm(residuals1)
qqline(residuals1)
```



log(Fev1)

```
qqnorm(residuals2)
qqline(residuals2)
```


Normal Q-Q Plot
Sample Quantiles
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-3
-2
-1
0
1
2
3
Theoretical Quantiles