6. Flexible extensions of the LLMM

Sonja Greven

Summer Term 2017

We have so far made several modeling assumptions in our longitudinal linear mixed model (LLMM) (3.5).

- While model (3.5) assumes a general variance-covariance matrix $\boldsymbol{\Sigma}_i$ for the errors $\boldsymbol{\varepsilon}_i$, we have typically assumed $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_{n_i}$ in the following.
- We have assumed a linear relationship between $\boldsymbol{Y}_i$ and the (transformed) covariates in $\boldsymbol{X}_i$.
- We have assumed a normal distribution for the random effects $\boldsymbol{b}_i$.

What if these assumptions are violated? This chapter discusses more flexible extensions of the LLMM.

Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Penalized splines in a nutshell

- Consider the **nonparametric regression problem**

$$Y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

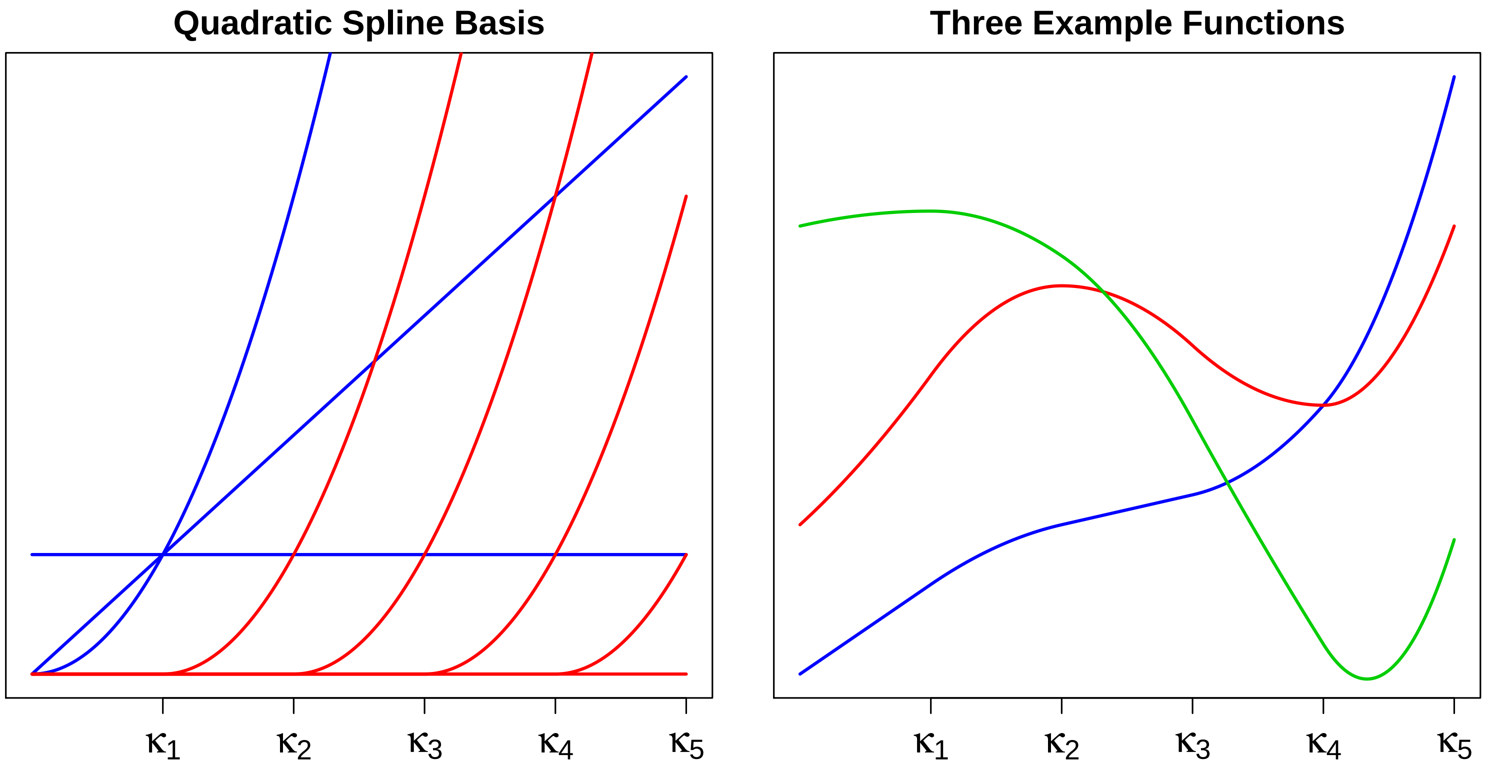
$\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, with unknown **smooth** function $m(\cdot)$.

- Approximate $m(\cdot)$ by a linear combination of (many) **spline basis functions**, e.g. truncated polynomials,

$$m(x) \approx \sum_{j=0}^{p-1} \beta_j x^j + \sum_{k=1}^{q} b_k (x - \kappa_k)_+^{(p-1)},$$

with $\kappa_1 < \cdots < \kappa_q$ a sequence of knots, and $u_+^{(p-1)} = (\max\{0, u\})^{(p-1)}$.

Penalized splines in a nutshell


Quadratic Spline Basis

Three Example Functions

Penalized splines in a nutshell

- This approximation gives a piecewise polynomial function of degree $(p-1)$ with certain smoothness properties (continuity of $(p-2)$th derivative. The b_k correspond to jumps in the $(p-1)$th derivative.)

- Estimation using a **regularization penalty**, to avoid overly wiggly function:

$$\min_{\boldsymbol{\beta},\boldsymbol{b}}\{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{b})+\lambda\boldsymbol{b}^T\boldsymbol{b}\}, \qquad (6.1)$$

  - $\boldsymbol{X}$ and $\boldsymbol{Z}$ design matrices for the spline basis functions x^j resp. $(x-\kappa_k)_+^{(p-1)}$
  - $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^T$, $\boldsymbol{b}=(b_1,\ldots,b_q)^T$
  - λ a smoothing parameter.

- After dividing by $(-2\sigma^2)$, (6.1) is equivalent to the penalized log-likelihood (4.2) $-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zb})^T\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zb})-\frac{1}{2}\boldsymbol{b}^T\boldsymbol{G}^{-1}\boldsymbol{b}$ and thus to obtaining the BLUE and BLUP for $\boldsymbol{\beta}$ and $\boldsymbol{b}$ in the LMM with

  - **fixed effects** $\beta_0,\ldots,\beta_{(p-1)}$: subspace of polynomial functions (degree $(p-1)$),
  - **random effects** $b_1,\ldots,b_q \overset{iid}{\sim} N(0,d^2)$: deviations from subspace,

  with $d^2 = \sigma^2/\lambda$. We can thus estimate the smoothing parameter $\lambda = \sigma^2/d^2$ data-driven using (RE)ML in this LMM.

- Analogous for other basis choices (e.g. B-Splines with a difference penalty) and for spatial effects, interaction surfaces, varying coefficients, $\ldots \rightarrow$ can combine these with random effects for the longitudinal data in a general LMM (additive mixed model).

- Examples: smooth mean over time, smooth dose-response functions for covariates, smoothly time-varying effects, spatial surfaces (spatial information on subjects), ...

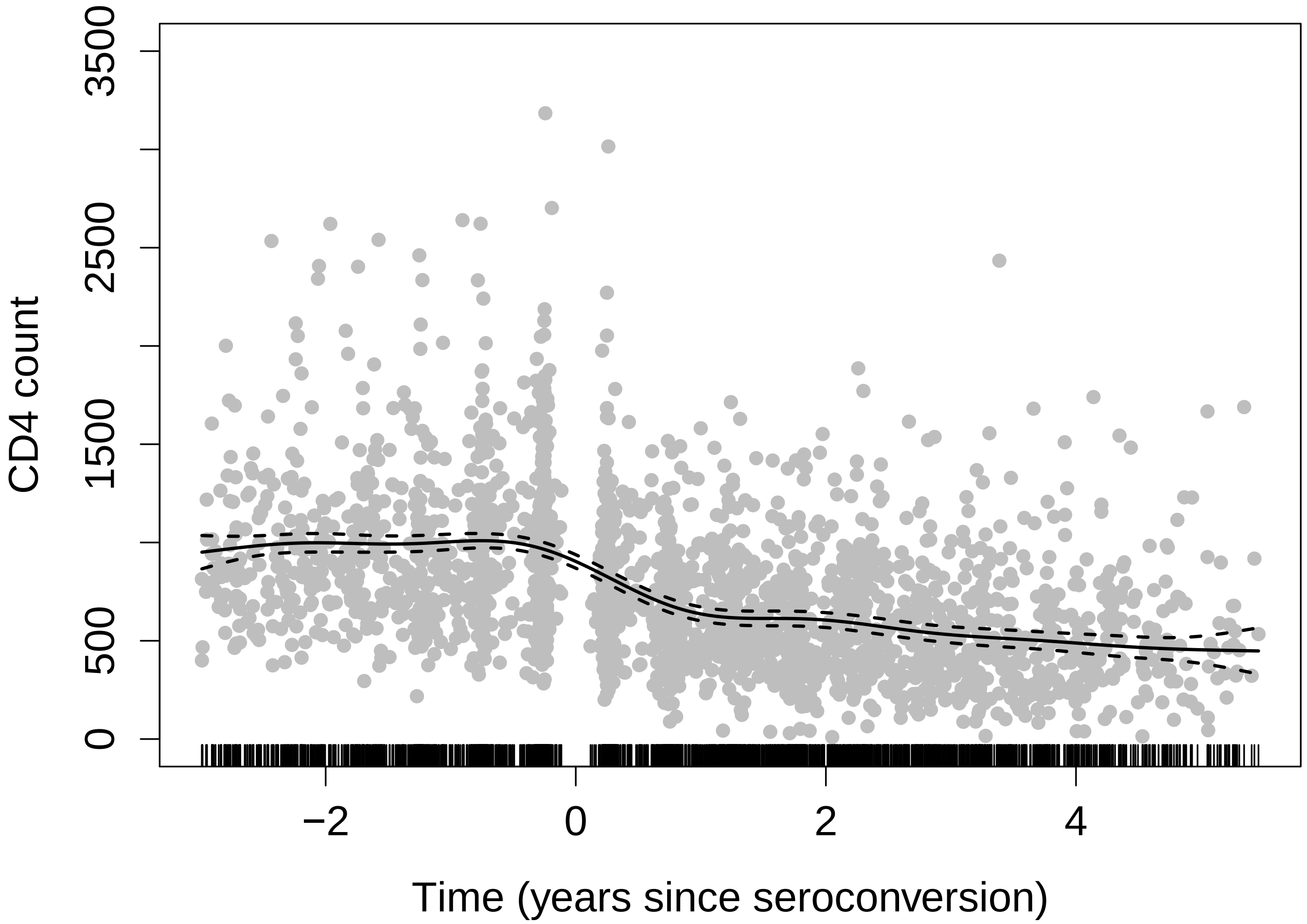
- More detailed information: Lexture **Mixed Models** and books Ruppert, Wand & Carroll, 2003; Wood, 2006 (2nd edition forthcoming).

- In R: packages `mgcv` and `gamm4`.

- Example smooth mean for the CD4 data (and random intercept per subject), fit with the `mgcv` package:

```
gamm(CD4 ~ s(Time), random = list(ID = ~ 1), data = cd4,
                    method = "REML") #ML is default
```

Example: The CD4 data



Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Introduction serial correlation

Remember the **longitudinal linear mixed model** (3.5):

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ \mathbf{b}_1, \ldots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_N \text{ are independent.} \end{cases}$$

The implied marginal covariance matrix for $\boldsymbol{Y}_i$ is

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$$

and the implied conditional covariance is

$$\text{Cov}(\mathbf{Y}_i|\boldsymbol{b}_i) = \boldsymbol{\Sigma}_i.$$

Introduction serial correlation

If $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_{n_i}$, this implies **conditional independence** of the Y_{ij}. When $\boldsymbol{b}_i$ consists of a random intercept, this implies compound symmetry or **equal marginal correlation** between any two measurements on the same subject.

A more general assumption is to decompose

$$\boldsymbol{\epsilon}_i = \boldsymbol{\epsilon}_{(1)i} + \boldsymbol{\epsilon}_{(2)i}, \quad \boldsymbol{\epsilon}_{(1)i} \text{ independent of } \boldsymbol{\epsilon}_{(2)i},$$

with $\boldsymbol{\epsilon}_{(2)i}$ representing serial correlation and $\boldsymbol{\epsilon}_{(1)i}$ an additional i.i.d. error. If Cov$(\boldsymbol{\epsilon}_{(2)i}) = \tau^2 \mathbf{H}_i$, this implies the covariance decomposition

$$\boldsymbol{\Sigma}_i = \tau^2 \mathbf{H}_i + \sigma^2 \boldsymbol{I}_{n_i}.$$

The model with i.i.d. error terms $(\tau^2 = 0)$ and the model with serial correlation only $(\sigma^2 = 0)$ occur as special cases.

Serial correlation

We are now interested in modeling the serial correlation, i.e. the covariance structure Cov$(\boldsymbol{\epsilon}_{(2)i}) = \tau^2\mathbf{H}_i$.

One typially assumes a structure that depends only on the temporal distance, with the (j,k)th element of $\mathbf{H}_i$ equal to

$$h_{ijk} = g(|t_{ij} - t_{ik}|)$$

and g a decreasing function with $g(0) = 1$ and $g(u) \to 0, u \to \infty$.

Models for the serial correlation

Two commonly used covariance structures are

$$\tau^2 g(u) = \tau^2 \exp(-\phi u) \qquad \text{(exponential)}$$

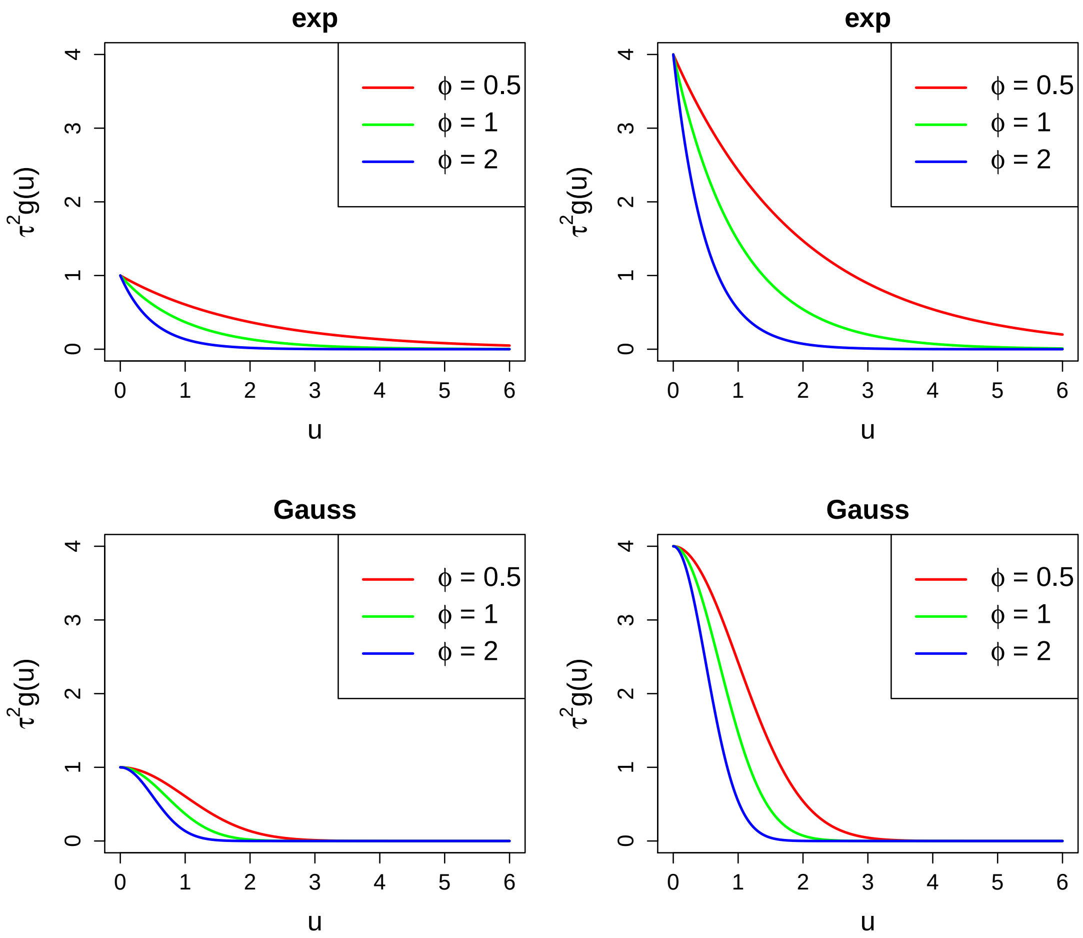
$$\tau^2 g(u) = \tau^2 \exp(-\phi u^2) \qquad \text{(Gaussian)}$$

- Large differences for small u
- Only two parameters: τ and ϕ
- Criterion for choosing between the two g: Maximum likelihood
- Special cases of the two parameter **power exponential family**

$$g(u) = \exp(-\phi |u|^\kappa), \quad \phi > 0, 0 < \kappa \leq 2.$$

Alternatives: e.g. the flexible **Matérn family**, AR(1), ...

Models for the serial correlation



Correlation structures in `R`

`lme` allows general covariances $\mathbf{\Sigma}_i$ through the argument `correlation`. The `corStruct` object assumes $\mathbf{\Sigma}_i = \tau^2 \boldsymbol{H}_i$, i.e. $\sigma^2 = 0$, no additional i.i.d. error term. Some of the available `corClasses`:

- **corCompSymm** and **corSymm**: constant respectively general unstructured correlation
- **corAR1** and **corARMA**: AR(1)- resp. ARMA(p,q)-correlation, time variable must be an integer
- **corCAR1**: “continuous time AR(1)”
- **corExp** and **corGaus**: Exponential and Gaussian correlation functions

From the `nlme` manual: “Users may define their own corStruct classes by specifying a constructor function [...]”.

A note of caution

In applications, the effect of serial correlation is often dominated by the combination of random effects and measurement error (also causing convergence problems when all are present in the model).

The total number of covariance parameters in $\boldsymbol{\alpha}$ should be relatively small unless there are many repeated observations per subject.

Example orthodont data: a model with random intercept and slope, i.i.d. error plus serial correlation yields at least $3 + 1 + 2 = 6$ parameters for the $10 = 4(4 + 1)/2$ disjunct entries in $\boldsymbol{V}_i$. For $n_i \equiv 4$ observations per subject it is difficult to distinguish between different serial correlations.

Example orthodont data

```
mycorr1 <- corExp(form =~ 1|Subject)
mycorr2 <- corGaus(form =~ 1|Subject)

mymodel0 <- lme(distance ~ age + Sex, data = Orthodont,
                random=~1|Subject)

mymodel1 <- lme(distance ~ age + Sex, data = Orthodont,
                random=~1|Subject, corr=mycorr1)

mymodel2 <- lme(distance ~ age + Sex, data = Orthodont,
                random=~1|Subject, corr=mycorr2)

anova(mymodel0, mymodel1, mymodel2)
```

yields:

	Model	df	logLik
mymodel0	0	5	-218.7562
mymodel1	1	6	-218.6984
mymodel2	2	6	-218.7156

Do we need serial correlation? The semi-variogram

We can assess serial correlation using the **(semi-)variogram**.

We can view the $Y_{ij} = Y_i(t_{ij}), j = 1, \ldots, n_i$, as samples from underlying continuous-time **stochastic processes** $\{Y_i(t), t \in \mathbb{R}\}$.

The variogram can be used to characterize the covariance of a stationary stochastic process $\{Y(t), t \in \mathbb{R}\}$ (i.e. the joint probability distributions at several time points do not change when shifted in time):

$$v(u) = \frac{1}{2}\mathsf{E}[\{Y(t) - Y(t-u)\}^2],\ u \geq 0.$$

The semi-variogram for the random intercept model

In the random intercept model

$$Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + b_i + \epsilon_{(1)ij} + \epsilon_{(2)ij} =: \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + Y_{ij}^c$$

with i.i.d. $\epsilon_{(1)ij}$ and serially correlated $\epsilon_{(2)ij}$, $\mathsf{E}(Y_{ij}^c) = 0$ and

$$\begin{aligned} \mathsf{Var}(Y_{ij}^c) &= d^2 + \sigma^2 + \tau^2 \\ \mathsf{Cov}(Y_{ij}^c, Y_{ik}^c) &= d^2 + \tau^2 g(|t_{ij} - t_{ik}|), \quad j \neq k. \end{aligned}$$

As the first two moments are shift-invariant and these determine the joint distributions for Gaussian processes, Y_{ij}^c thus is stationary.

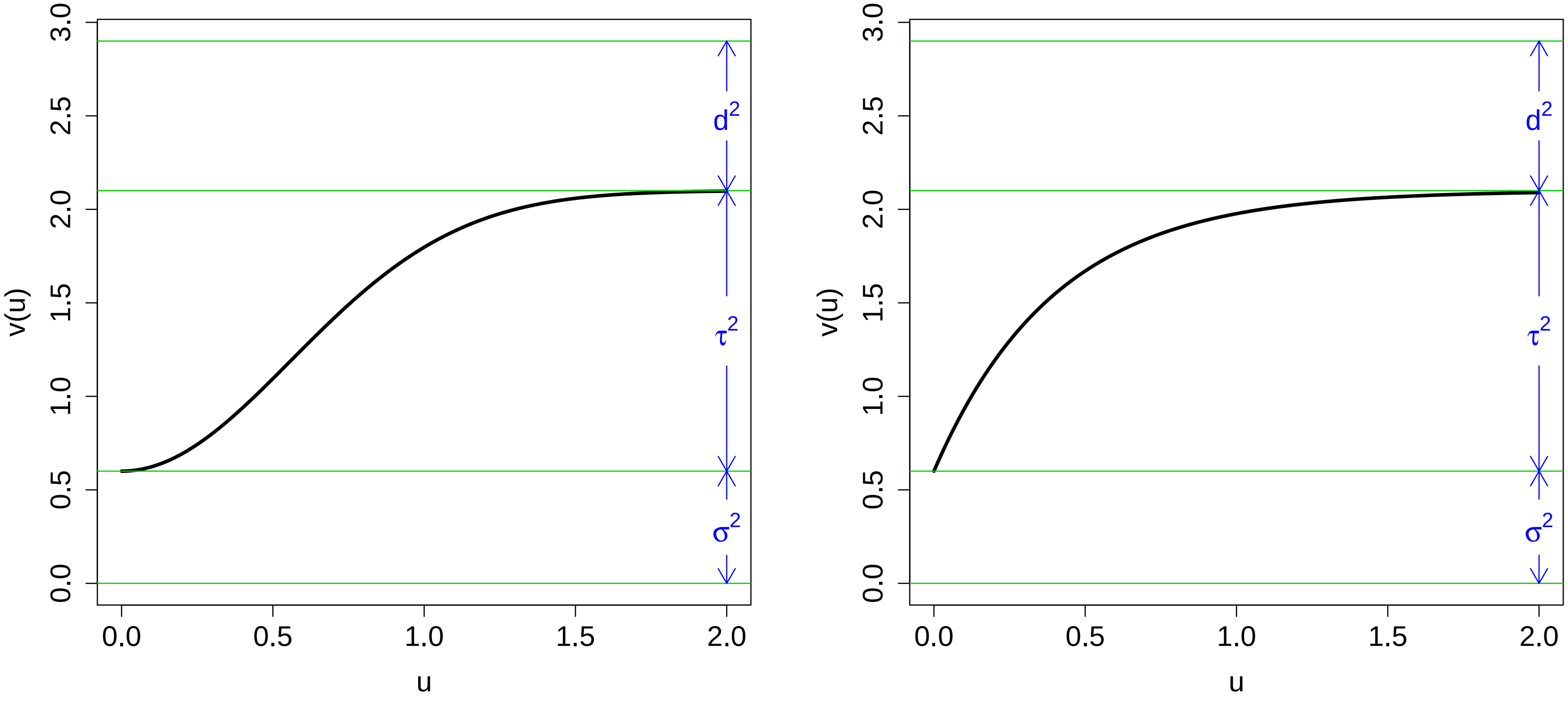
- For $i = 1, \ldots, N$ and $j \neq k$,

$$\begin{aligned} v(|t_{ij} - t_{ik}|) &= \frac{1}{2}\mathsf{E}[(Y_{ij}^c - Y_{ik}^c)^2] \\ &= \frac{1}{2}[\mathsf{Var}(Y_{ij}^c) + \mathsf{Var}(Y_{ik}^c) - 2\mathsf{Cov}(Y_{ij}^c, Y_{ik}^c)] \\ &= [d^2 + \sigma^2 + \tau^2] - [d^2 + \tau^2 g(|t_{ij} - t_{ik}|)] \\ &= \sigma^2 + \tau^2(1 - g(|t_{ij} - t_{ik}|)) \\ \Rightarrow v(u) &= \sigma^2 + \tau^2(1 - g(u)). \end{aligned}$$

- $v(0) = \sigma^2$ ("nugget effect"), as $g(0) = 1$

- $\lim_{u\to\infty} v(u) = \sigma^2 + \tau^2 \leq \mathsf{Var}(Y_{ij}^c) = d^2 + \sigma^2 + \tau^2$, as $g(u) \to 0, u \to \infty$.

The semi-variogram



Which shape does $g(.)$ have in these two plots?

The empirical semi-variogram

The function $v(\cdot)$ has to be estimated from the residuals $r_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T \hat{\boldsymbol{\beta}}$.

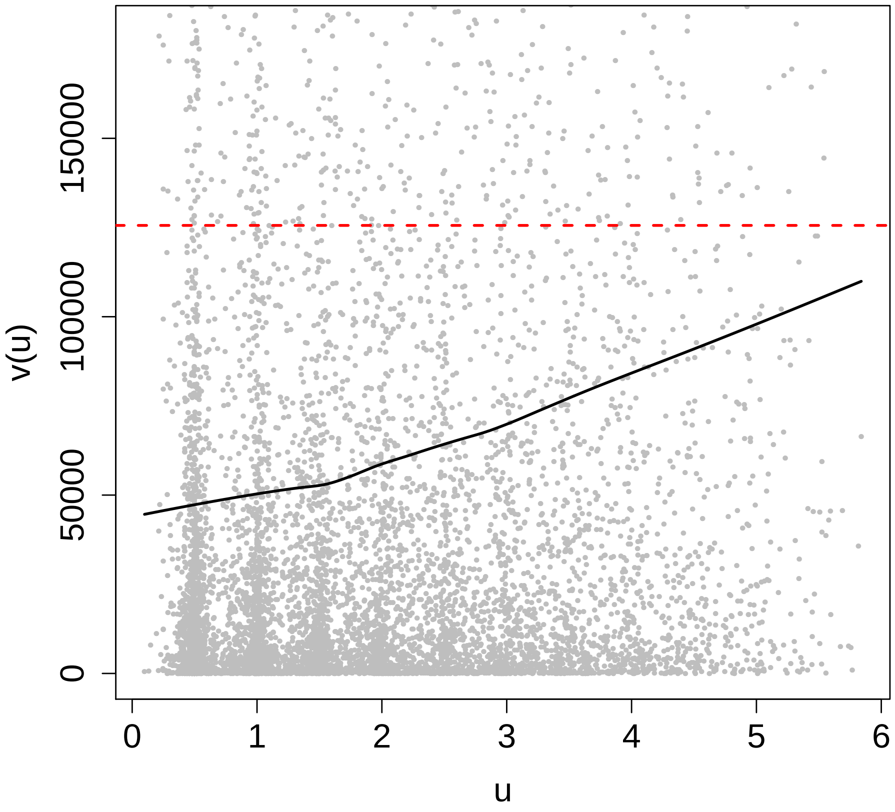
- For pairs on the same subject, consider the half-squared differences

$$v_{ijk} = (r_{ij} - r_{ik})^2/2 \ \text{ for } i = 1, \ldots, N, \ j \neq k.$$

- Plotting and smoothing $(|t_{ij} - t_{ik}|, v_{ijk})$ gives the **empirical semi-variogram** $\hat{v}(\cdot)$.
- Pairs on different subjects estimate the total variance $\sigma^2 + \tau^2 + d^2$,

$$\widehat{\sigma}^2 + \widehat{\tau}^2 + \widehat{d}^2 = \frac{1}{(n^2 - \sum_{i=1}^N n_i^2)} \sum_{i \neq l} \sum_{j=1}^{n_i} \sum_{k=1}^{n_l} (r_{ij} - r_{lk})^2/2.$$

The empirical semi-variogram for the CD4 data



The empirical semi-variogram

- The comparison of the empirical semi-variogram with the variogram based on a fitted covariance structure gives an informal plausibility check for the assumed covariance structure.
- In R: Function `Variogram` in package `nlme` or own code.
- If there are **additional random effects** in the model, Y_{ij}^c is no longer stationary.

  See Verbeke & Molenberghs, 2000, Section 10.4.4., for a discussion of alternatives using transformed residuals.

Overview Chapter 6 - Flexible extensions of the LLMM

6.1 Smooth models for the mean

6.2 Serial correlation

6.3 Non-normal random effects

Non-normal random effects

To allow for non-normally distributed random effects, approaches exist that assume instead e.g. t-distributions or mixture distributions

$$f(\boldsymbol{b}_i|\boldsymbol{\pi}, \boldsymbol{\phi}) = \sum_{k=1}^{K} \pi_k f(\boldsymbol{b}_i|\boldsymbol{\phi}_k),$$

with

- weights $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$, $\sum_{k=1}^{K} \pi_k = 1$,
- parametric families of distributions $f\left(\boldsymbol{b}_i|\boldsymbol{\phi}_k\right)$ with parameters $\boldsymbol{\phi}_k$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ (often assumed as multivariate normal $f\left(\boldsymbol{b}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)$ with mean $\boldsymbol{\mu}_k$ and homogeneous covariances $\boldsymbol{\Sigma}$).

Non-normal random effects

Estimation is usually done using Bayesian approaches or an EM-algorithm, see e.g. Chapter 12 in Verbeke & Molenberghs (2000) for a discussion of the latter approach.

Verbeke & Molenberghs (2000) call the mixture model the **heterogeneity model**, as it assumes heterogeneous subgroups in the population.