Chapter 3

The Linear Mixed Effects Model

3.1 The Linear Mixed Model

- In the previous chapter we focused on the *multivariate regression model*

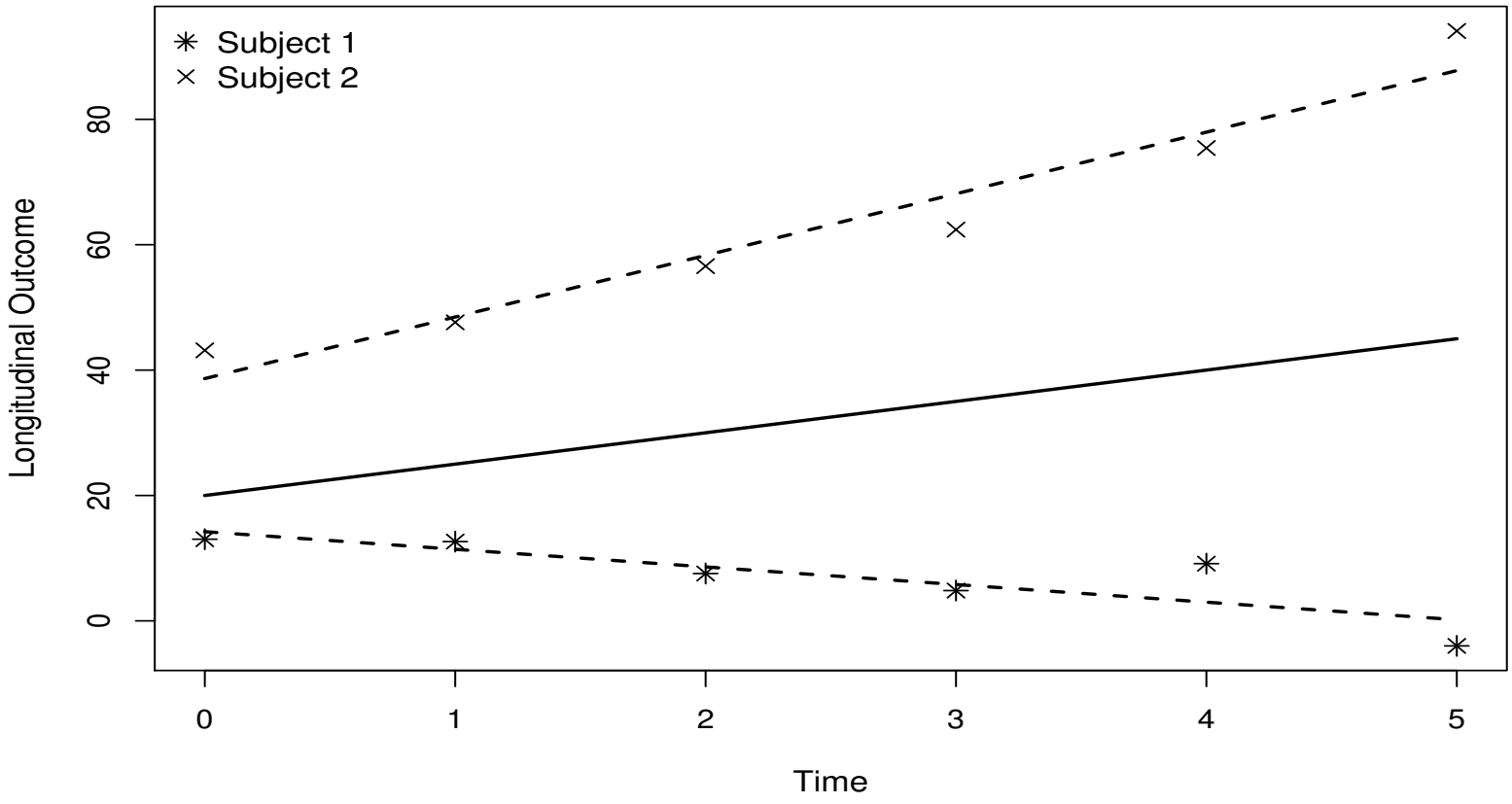
$$y_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, V_i),$$

where

- ▷ y_i the vector of responses for the ith subject
- ▷ X_i design matrix describing structural component
- ▷ V_i covariance matrix describing the correlation structure

- **Alternative intuitive approach:** Each subject in the population has her own subject-specific mean response profile over time

3.1 The Linear Mixed Model (cont'd)



3.1 The Linear Mixed Model (cont'd)

- The evolution of each subject in time can be described by a linear model

$$y_{ij} = \tilde{\beta}_{i0} + \tilde{\beta}_{i1} t_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2),$$

  where

  - ▷ y_{ij} the jth response of the ith subject
  - ▷ $\tilde{\beta}_{i0}$ is the intercept and $\tilde{\beta}_{i1}$ the slope for subject i

- **Assumption:** Subjects are randomly sampled from a population $\Rightarrow$ subject-specific regression coefficients are also sampled from a population of regression coefficients

$$\tilde{\beta}_i \sim \mathcal{N}(\beta, D)$$

3.1 The Linear Mixed Model (cont'd)

- We can reformulate the model as

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \varepsilon_{ij},$$

  where

  - ▷ βs are known as the *fixed effects*
  - ▷ b_is are known as the *random effects*

- In accordance for the random effects we assume

$$b_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim \mathcal{N}(0, D)$$

3.1 The Linear Mixed Model (cont'd)

- Put in a general form

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i, \\ \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathsf{I}_{n_i}), \end{cases}$$

with

- ▷ X design matrix for the fixed effects β
- ▷ Z design matrix for the random effects b_i
- ▷ $b_i \perp\!\!\!\perp \varepsilon_i$

3.2 Interpretation

- Fixed and random effects:
  - ▷ β_j denotes the change in the average y_i when x_j is increased by one unit
  - ▷ b_i are interpreted in terms of how a subset of the regression parameters for the ith subject deviates from those in the population

- Advantageous feature: population + subject-specific predictions
  - ▷ β describes mean response changes in the population
  - ▷ $\beta + b_i$ describes individual response trajectories

3.2 Interpretation (cont'd)

- Example: We fit a linear mixed model for the AIDS dataset assuming
  - ▷ different average longitudinal evolutions per treatment group (fixed part)
  - ▷ random intercepts & random slopes (random part)

$$\begin{cases} y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \{\texttt{ddI}_i \times t_{ij}\} + b_{i0} + b_{i1} t_{ij} + \varepsilon_{ij}, \\ \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

- Note: We did not include a main effect for treatment due to randomization

3.2 Interpretation (cont'd)

	Value	Std.Err.	t-value	p-value
β_0	7.189	0.222	32.359	< 0.001
β_1	-0.163	0.021	-7.855	< 0.001
β_2	0.028	0.030	0.952	0.342

- No evidence of differences in the average longitudinal evolutions between the two treatments

3.2 Interpretation (cont'd)

- Interaction & nonlinear terms: As we have seen in the previous chapter (see pp. 57–62), often
  - ▷ the effect of some predictors may be nonlinear (e.g., time effect), and/or
  - ▷ the effect of some predictors on the outcome may be influenced from other predictors (e.g., different average longitudinal evolutions per treatment group)

- In such cases, we need to consider more elaborate models that contain terms to capture these features, namely
  - ▷ polynomials or splines to model nonlinearities
  - ▷ interaction effects

3.2 Interpretation (cont'd)

- When such terms are included in the model, the interpretation of the parameters can become quite complicated

- To understand a complex mixed model we can visualize it using **effect plots**

- Example: We fit a model to the PBC dataset for serum bilirubin that contains
  - ▷ *fixed effects:*
    - * nonlinear time effect with splines, main effect of sex, age and baseline prothrombin
    - * interaction effects of sex with nonlinear time, age and baseline prothrombin
  - ▷ *random effects:* nonlinear time effect

3.2 Interpretation (cont'd)

- The model has the form:

$$\begin{aligned}\log(\texttt{serBilir}_{ij}) = {} & \beta_0 + \beta_1 N(\texttt{Time}_{ij})_1 + \beta_2 N(\texttt{Time}_{ij})_2 + \beta_3 \texttt{Female}_i + \beta_4 \texttt{Age}_i + \\ & \beta_5 \texttt{basePro}_i + \beta_6 \{\texttt{Female}_i \times \texttt{Age}_i\} + \\ & \beta_7 \{\texttt{Female}_i \times \texttt{basePro}_i\} + \beta_8 \{\texttt{Female}_i \times N(\texttt{Time}_{ij})_1\} + \\ & \beta_9 \{\texttt{Female}_i \times N(\texttt{Time}_{ij})_2\} + b_{i0} + b_{i1} N(\texttt{Time}_{ij})_1 + \\ & b_{i2} N(\texttt{Time}_{ij})_2 + \varepsilon_{ij}\end{aligned}$$

where

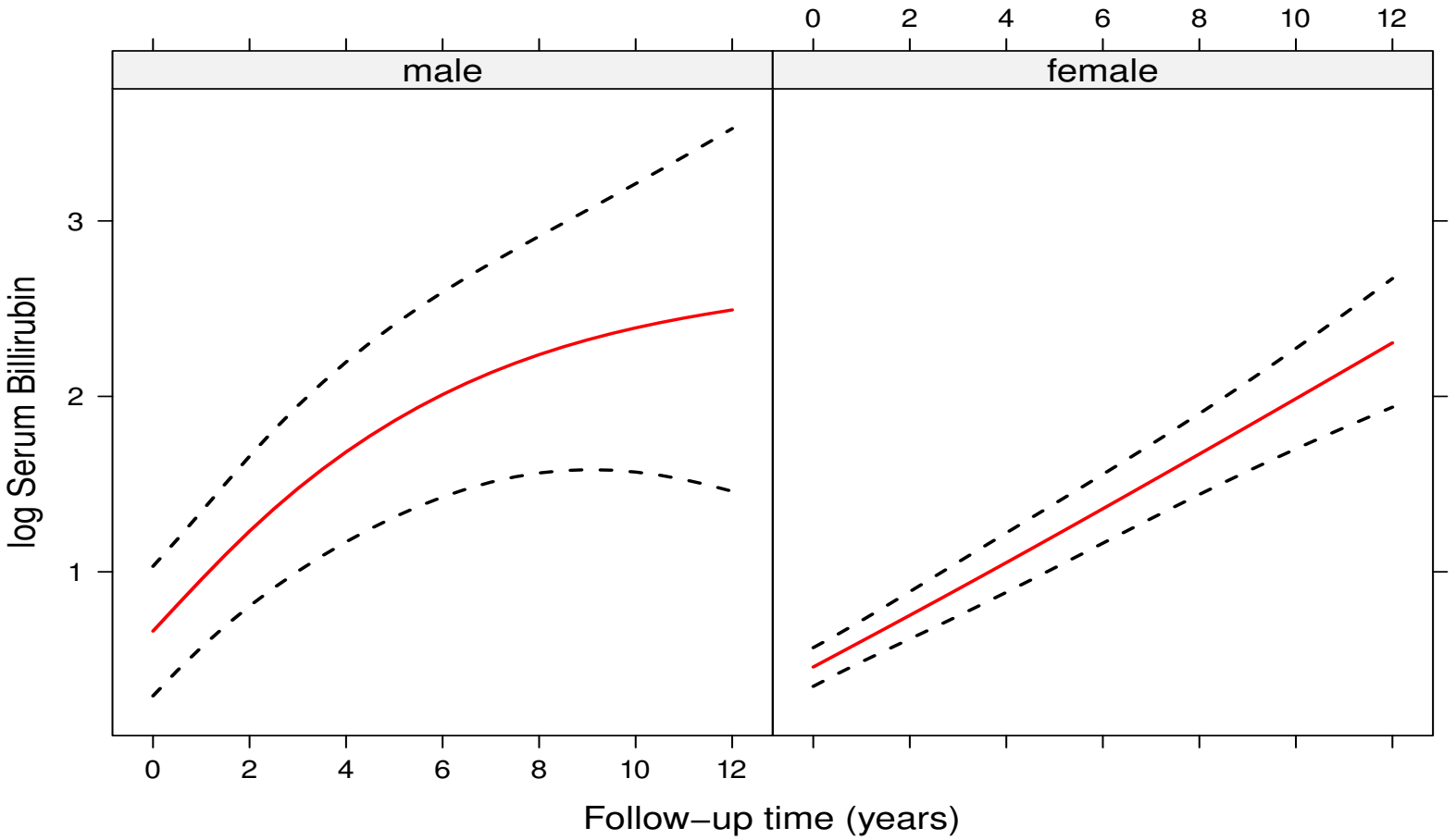
  - ▷ the terms $N(\texttt{Time}_{ij})_1$ and $N(\texttt{Time}_{ij})_2$ denote the basis for a natural spline with two degrees of freedom
  - ▷ $b_i \sim \mathcal{N}(0, D)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

3.2 Interpretation (cont'd)

- In this model not all coefficients have a direct interpretation in isolation

- Hence to understand the model we depict
  - ▷ how the average longitudinal profiles evolve over time time,
  - ▷ separately for males and females, and prothrombin time of 10.6 sec
  - ▷ for the average age of 49 years old
  - ▷ including also the corresponding 95% pointwise confidence intervals
  - ▷ (in the app different ages and prothrombin times can be selected)

3.2 Interpretation (cont'd)



3.3 Hierarchical vs Marginal

- How do the random effects capture correlation:
  - ▷ Given the random effects, the measurements of each subject are independent (*conditional independence assumption*)

$$p(y_i \mid b_i) = \prod_{j=1}^{n_i} p(y_{ij} \mid b_i)$$

  - ▷ Marginally (integrating out the random effects), the measurements of each subject are correlated

$$p(y_i) = \int p(y_i \mid b_i)\, p(b_i)\, db_i \quad \Rightarrow \quad y_i \sim \mathcal{N}(X_i\beta,\ Z_i D Z_i^\top + \sigma^2 \mathsf{I}_{n_i})$$

3.3 Hierarchical vs Marginal (cont'd)

- Hence, with random effects we again model the correlations in the repeated measurements of each subject

- Notes: In using random effects for modeling the covariance matrix

  - ▷ The more random effects we include the more flexibly we capture the correlations
  - ▷ By using random effects we also directly allow for heteroscedasticity (i.e., non-constant variances in time)
  - ▷ Nevertheless, we do assume a particular type of structure for the correlations and the variances – they are **not** allowed completely free
  - ▷ Random effects work equally well with balanced or unbalanced data

3.3 Hierarchical vs Marginal (cont'd)

- Let's try the app...

3.3 Hierarchical vs Marginal (cont'd)

- Hierarchical formulation
  - ▷ a model for y_i given b_i, and a model for b_i
  - ▷ D is the covariance matrix of the random effects $\Rightarrow$ **needs to be positive definite**

- Marginal formulation
  - ▷ a model for y_i, and a specific form of the marginal covariance matrix $V_i = Z_i D Z_i^\top + \sigma^2 \mathsf{I}_{n_i}$
  - ▷ only V_i needs to be positive definite
  - ▷ V_i **can be positive definite without** D **being positive definite**

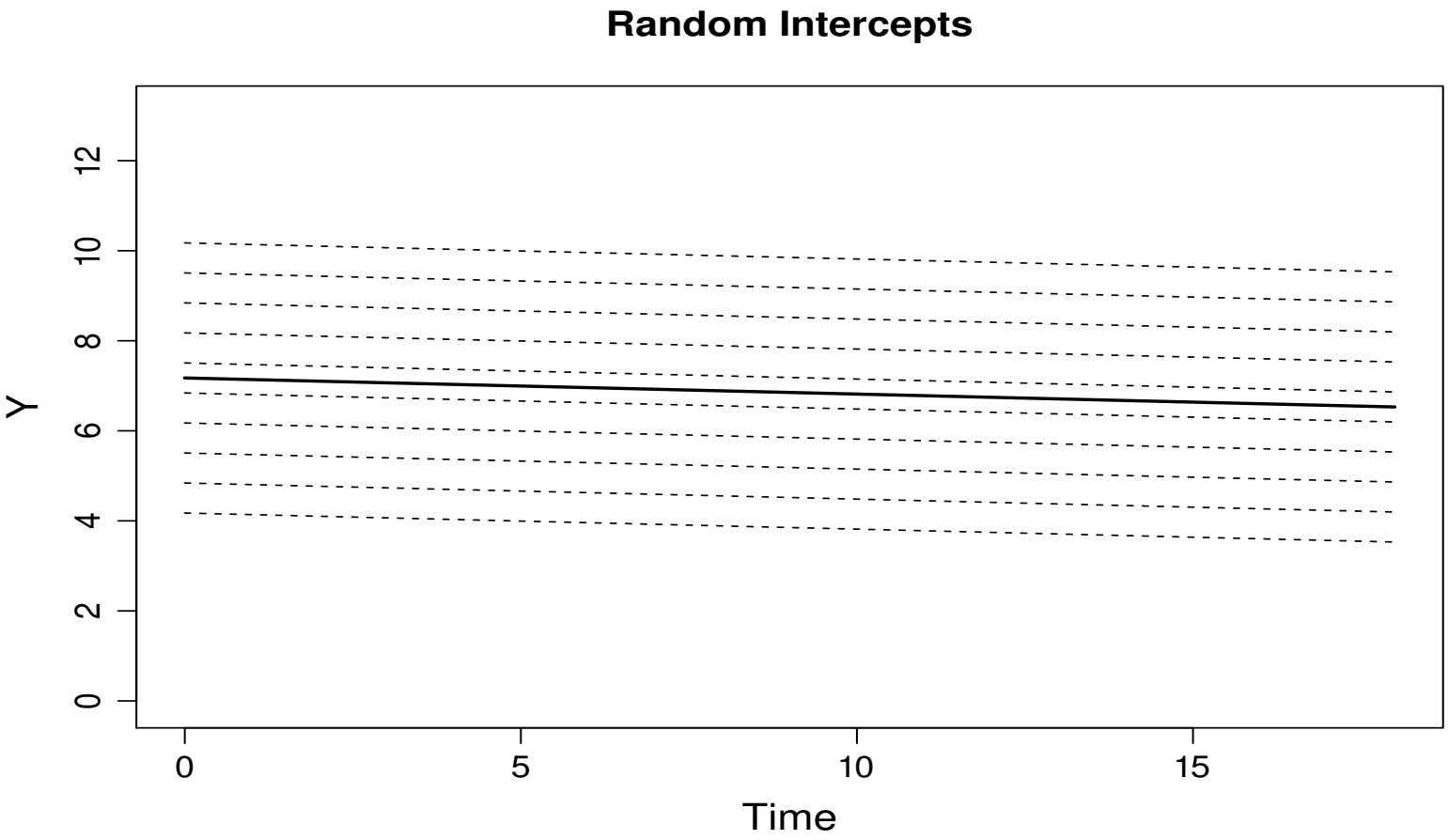
3.3 Hierarchical vs Marginal (cont'd)

The hierarchical model implies the marginal one, not vice versa

- A simple example: Random-intercepts model

$$\begin{cases} y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + \varepsilon_{ij}, \\ \\ b_{i0} \sim \mathcal{N}(0, \sigma_b^2), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \end{cases}$$

3.3 Hierarchical vs Marginal (cont'd)



3.3 Hierarchical vs Marginal (cont'd)

- Implied marginal covariance matrix has the form

$$V_i = \sigma_b^2 1_{n_i} 1_{n_i}^\top + \sigma^2 \mathsf{I}_{n_i}$$

it assumes

- ▷ constant variance $\sigma_b^2 + \sigma^2$ over time, and
- ▷ equal positive correlation $\rho = \sigma_b^2/(\sigma_b^2 + \sigma^2)$ between the measurements of any two time points (aka *intra-class correlation*)
- ▷ it is known as the *compound symmetric* covariance matrix

3.3 Hierarchical vs Marginal (cont'd)

- Note that we could also have a compound symmetric covariance matrix with negative intra-class correlation
  - ▷ such a matrix could never have come from a mixed model

Random intercepts **imply** compound symmetry
but
Compound symmetry **does not imply** random intercepts

3.3 Hierarchical vs Marginal (cont'd)

- What are the implications of this?

- Statistical software that fit mixed models under ML actually fit the implied marginal model
  - ▷ we can construct examples where two mixed models have exactly the same implied marginal model
  - ▷ based on the fitted model we **cannot** say under which model the data have been generated

- We can only do it under a Bayesian approach (because there we actually fit the hierarchical model)

3.4 Estimation

- Fixed effects: For known marginal covariance matrix $V_i = Z_i D Z_i^\top + \sigma^2 \mathsf{I}_{n_i}$, the fixed effects are estimated using generalized least squares

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i^\top V_i^{-1} X_i \right)^{-1} \sum_{i=1}^{n} X_i^\top V_i^{-1} y_i$$

- Variance Components: The unique parameters in V_i are estimated based on either maximum likelihood (ML) or restricted maximum likelihood (REML)
  - ▷ REML provides unbiased estimates for the variance components in small samples

3.4 Estimation (cont'd)

- Two-step iterative procedure
  - ▷ Step 0: Set initial values for D and σ^2
  - ▷ Step 1: Calculate the covariance matrix $\widehat{V}_i^{it=k}$ and following the fixed effects $\hat{\beta}^{it=k}$
  - ▷ Step 2: Update $\widehat{V}_i^{it=k+1}$ using REML or ML
  - ▷ Step 3: Check convergence criterion, if not satisfied return to Step 1

Steps 1–3 are repeated until convergence is attained

3.4 Estimation (cont'd)

- Estimation of random effects
  - ▷ based on a fitted mixed model, estimates for the random effects are based on the posterior distribution:

$$p(b_i \mid y_i; \theta) = \frac{p(y_i \mid b_i; \theta)\, p(b_i; \theta)}{p(y_i; \theta)}$$

$$\propto p(y_i \mid b_i; \theta)\, p(b_i; \theta),$$

in which θ is replaced by its MLE $\hat{\theta}$

3.4 Estimation (cont'd)

- This is a whole distribution
  - ▷ in the linear mixed model we have seen, this posterior distribution has a closed-form:

$$[b_i \mid y_i; \theta] \sim \mathcal{N}\Big\{DZ_i^\top V_i^{-1}(y_i - X_i\beta),\ DZ_i^\top K Z_i D\Big\},$$

with

$$K = V_i^{-1} - V_i^{-1}X_i\left(\sum_{i=1}^{n} X_i^\top V_i^{-1} X_i\right)^{-1} X_i^\top V_i^{-1}$$

3.4 Estimation (cont'd)

- To obtain estimates for the random effects we typically use measures of location from this posterior distribution (e.g., mean or mode)

- Due to the fact that in linear mixed models we obtain a normal distribution, in which the mean and mode coincide, we use estimates of the random effects the means

$$\widehat{b}_i = DZ_i^\top V_i^{-1}(y_i - X_i\beta)$$

- These estimates are called the *empirical Bayes* estimates of the random effects

3.4 Estimation (cont'd)

- Estimates of the random effects are very useful in prediction
  - ▷ in this context there is an important difference between the marginal models we have seen in Chapter 2 and the mixed models of this chapter

- In particular, the predictions from a marginal model are

$$\widehat{y}_i^{marg} = X_i\widehat{\beta}$$

whereas as from the mixed model

$$\widehat{y}_i^{subj} = X_i\widehat{\beta} + Z_i\widehat{b}_i$$

3.4 Estimation (cont'd)

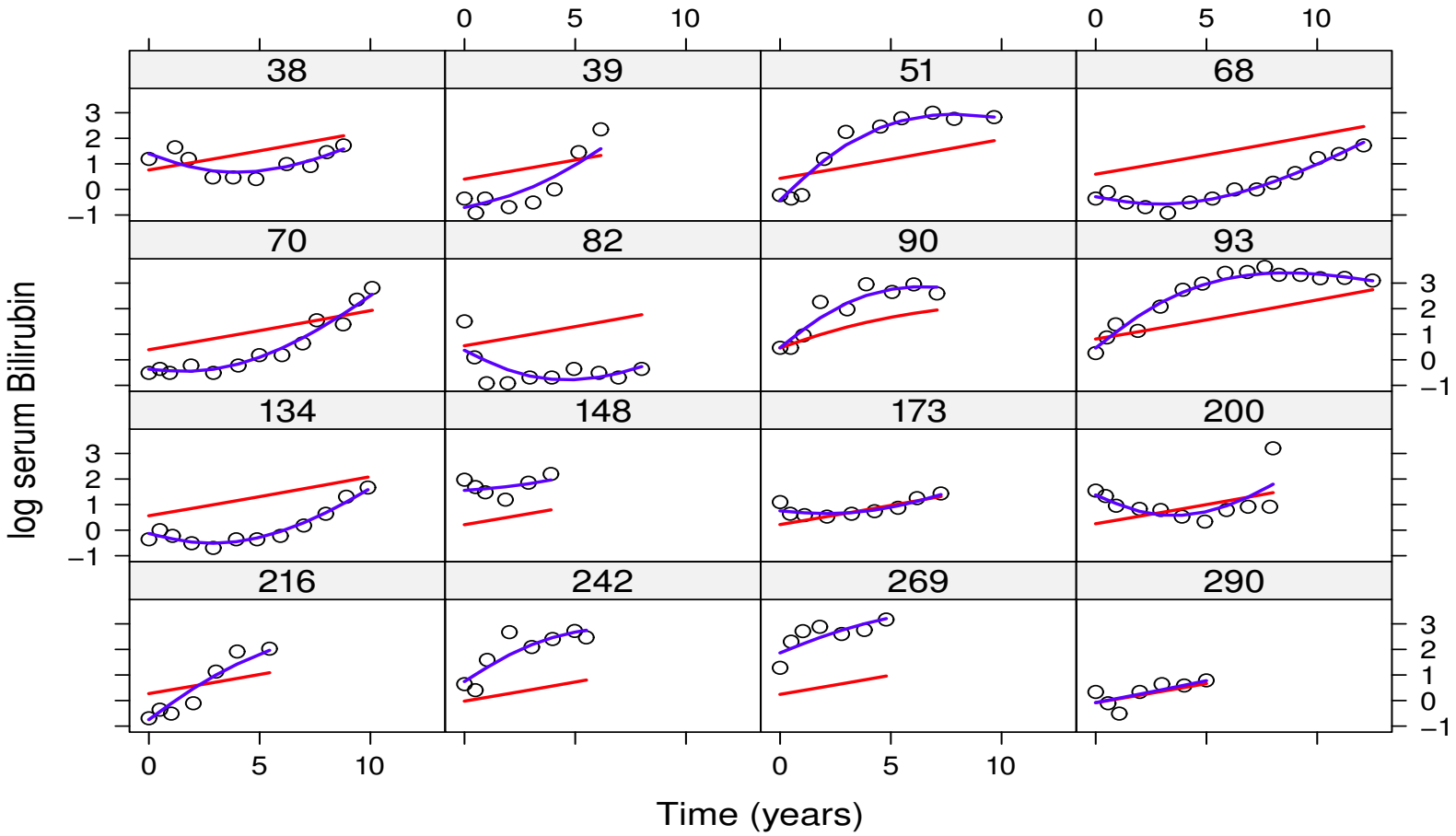
- The difference is that
  - ▷ from the marginal model we obtain predictions for the *'average'* patient having characteristics X_i (i.e., age, sex, etc.)
  - ▷ from the mixed model we obtain predictions for the *'average'* patient that has characteristics X_i and observed data y_i (i.e., they have a subject-specific nature)

- The predictions $X_i\widehat{\beta} + Z_i\widehat{b}_i$ we obtain from the mixed model are called the *Best Linear Unbiased Predictions (BLUPs)*
  - ▷ 'linear' because they are a linear combination of $\widehat{\beta}$ and $\widehat{b}_i$
  - ▷ 'unbiased' because their average equals the true subject-specific mean
  - ▷ 'best' because they have the smallest variance of all linear predictors

3.4 Estimation (cont'd)

- Example: To see an example of the difference between the marginal and subject-specific predictions, we compare the two sets of predictions for the complex linear mixed model we have seen in Section 3.2 (pp.142–145) for 16 randomly selected patients
  - ▷ red lines denote the marginal predictions,
  - ▷ blue lines denote the subject-specific predictions
  - ▷ black circles the observed data

3.4 Estimation (cont'd)



3.4 Estimation (cont'd)

- We clearly observe that the subject-specific prediction are much closer to the data of each individual patient than the marginal predictions

3.5 Mixed-Effects Models in R

R> There are two primary packages in R for mixed models analysis:

- Package **nlme**
  - * fits linear & nonlinear mixed effects models, and marginal models for normal data
  - * allows for both random effects & correlated error terms
  - * several options for covariances matrices and variance functions

- Package **lme4**
  - * fits linear, nonlinear & generalized mixed effects models
  - * uses only random effects
  - * allows for nested and crossed random-effects designs

3.5 Mixed-Effects Models in R (cont'd)

R> The basic function to fit linear mixed models in the **nlme** package is `lme()`, and has three basic arguments

- ▷ `fixed`: a formula specifying the response vector and the fixed-effects structure
- ▷ `random`: a formula specifying the random-effects structure
- ▷ `data`: a data frame containing all the variables

3.5 Mixed-Effects Models in R (cont'd)

R> The data frame that contains all variables should be in the *long format*

Subject	y	time	gender	age
1	5.1	0.0	male	45
1	6.3	1.1	male	45
2	5.9	0.1	female	38
2	6.9	0.9	female	38
2	7.1	1.2	female	38
2	7.3	1.5	female	38
⋮	⋮	⋮	⋮	⋮

3.5 Mixed-Effects Models in R (cont'd)

R> The code used to fit the linear mixed model for the AIDS dataset (pp.139) is as follows

```
lmeFit <- lme(CD4 ~ obstime + obstime:drug, data = aids,
    random = ~ obstime | patient)

summary(lmeFit)
```

3.5 Mixed-Effects Models in R (cont'd)

R> The same fixed-effects structure but only random intercepts

```
lme(CD4 ~ obstime + obstime:drug, data = aids,
    random = ~ 1 | patient)
```

R> The same fixed-effects structure, random intercepts & random slopes, with a diagonal covariance matrix (using the `pdDiag()` function)

```
lme(CD4 ~ obstime + obstime:drug, data = aids,
    random = list(patient = pdDiag(form = ~ obstime)))
```

3.5 Mixed-Effects Models in R (cont'd)

R> The basic function to fit linear mixed models in the **lme4** package is `lmer()`, and has two basic arguments

- ▷ `formula`: a formula specifying the response vector, the fixed- and random-effects structure
- ▷ `data`: a data frame containing all the variables

R> Again the data should be in the long format

3.5 Mixed-Effects Models in R (cont'd)

R> The analogous code to fit the linear mixed model for the AIDS dataset (pp.139) is as follows

```
lmerFit <- lmer(CD4 ~ obstime + obstime:drug + (obstime | patient),
                data = aids)

summary(lmerFit)
```

3.5 Mixed-Effects Models in R (cont'd)

R> To fit the same model but with a diagonal matrix for the random effects the call becomes:

```
lmerFit2 <- lmer(CD4 ~ obstime + obstime:drug + (1 | patient) +
                    (0 + obstime | patient),
                 data = aids)

summary(lmerFit2)
```

3.6 Nested and Crossed Random Effects*

- In the examples we have seen so far the type of correlated data we have seen is longitudinal data
  - ▷ correlations stems from the fact that we measure *the same* outcome repeatedly in time for each subject

- Another commonly encountered feature that induces correlation is clustering, e.g.,
  - ▷ patients are clustered within hospitals
  - ▷ children are clustered within schools or families
  - ▷ ...

3.6 Nested and Crossed Random Effects* (cont'd)

- Example: In the Glaucoma data we have a multilevel clustered design (see pp.11)
  - ▷ each location is nested within the hemifield
  - ▷ each hemifield is nested within the eye
  - ▷ each eye is nested within the patient

Measurements in the same cluster are expected to be **(positively) correlated**

3.6 Nested and Crossed Random Effects* (cont'd)

- To account for the correlations in each level of the multilevel structure we can include level-specific random effects

- Continuing in the Glaucoma data example, we focus (for simplicity) in the higher two levels, namely the patient and the eye

  ▷ we fit a linear mixed model with a separate random effect per level

$$\begin{cases} \texttt{VF}_{ijk} = \beta_0 + \beta_1 \texttt{Time}_{ijk} + b_i + u_{ij} + \varepsilon_{ijk} \\ b_i \sim \mathcal{N}(0, \sigma^2_{patient}), \qquad u_{ij} \sim \mathcal{N}(0, \sigma^2_{eye}), \\ \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

3.6 Nested and Crossed Random Effects* (cont'd)

- where
  - ▷ $\texttt{VF}_{ijk}$ denotes the k-th visual field sensitivity measurement for the j-th eye of the i-th patient
  - ▷ $\texttt{Time}_{ijk}$ denotes the corresponding time point this measurement was taken
  - ▷ b_i is the random effect for the patients – the measurements of the i-th patient are correlated because all these measurements share the *same* random effect b_i
  - ▷ u_{ij} is the random effect for the eye within the patient – the measurements of the j-th eye of the i-th patient are more correlated than the measurements of the j'-th eye because they share the *same* random effect u_{ij}

3.6 Nested and Crossed Random Effects* (cont'd)

- The estimated variance components from the Glaucoma data are:
  - ▷ $\sigma_{patient} = 4.3$
  - ▷ $\sigma_{eye} = 5.8$
  - ▷ $\sigma = 7.9$

- Based on these variance components we can compute the corresponding correlations, i.e.,
  - ▷ measurements in the same eye have correlation

$$\frac{\sigma^2_{patient} + \sigma^2_{eye}}{\sigma^2_{patient} + \sigma^2_{eye} + \sigma^2} = 0.46$$

3.6 Nested and Crossed Random Effects* (cont'd)

▷ and measurements from different eyes

$$\frac{\sigma^2_{patient}}{\sigma^2_{patient} + \sigma^2_{eye} + \sigma^2} = 0.16$$

- It goes without saying, that if the correlations in the data are more complex we could additional random effects

- Example: Continuing in the Glaucoma example, by including only random intercepts terms we assume that the correlations are constant in time

  ▷ as we have previously discussed, this may be a simplistic assumption for longitudinal data

3.6 Nested and Crossed Random Effects* (cont'd)

- We extend the model by including a random slopes terms in the patient level, i.e.,

$$\begin{cases} \texttt{VF}_{ijk} &= \beta_0 + \beta_1 \texttt{Time}_{ijk} + b_i + u_{ij} + \varepsilon_{ijk} \\ b_i &\sim \mathcal{N}(0, D_{patient}), \qquad u_{ij} \sim \mathcal{N}(0, \sigma^2_{eye}), \\ \varepsilon_{ijk} &\sim \mathcal{N}(0, \sigma^2) \end{cases}$$

  ▷ now, in the patient level we have a covariance matrix D

3.6 Nested and Crossed Random Effects* (cont'd)

- The estimated variance components from the Glaucoma data are:
  - ▷ $\sigma_{patient,int} = 4.7$
  - ▷ $\sigma_{patient,slp} = 0.4$
  - ▷ $\text{corr}_{patient,int-slp} = -0.4$
  - ▷ $\sigma_{eye} = 5.8$
  - ▷ $\sigma = 7.8$

3.6 Nested and Crossed Random Effects* (cont'd)

- The examples we have seen so far in this section refer to settings in which the measurements of one level are *nested* within another level
  - ▷ due to this feature, the random effects we have used in the previous examples of the Glaucoma data are called *nested random effects*

- However, there are also settings in which we have different types of groupings of measurements that are not nested
  - ▷ in these cases we use *crossed random effects*

3.6 Nested and Crossed Random Effects* (cont'd)

- Example: One feature of visual field sensitivity measurements is that they exhibit the so-called *Global Visit Effect* (see pp.12)
  - ▷ in particular, for some visits some patients showed strangely low sensitivity to the stimuli
  - ▷ in the next visit, their sensitivity levels improved
  - ▷ it is not possible this low sensitivity to be due to Glaucoma because it is an irreparable disease
  - ▷ hence, the low sensitivity measurements are attributed to other reasons (e.g., tiredness)

- To capture this Global Visit Effect we can include a random effect for each visit
  - ▷ this random effect is **not** nested to the previously used random effects

3.6 Nested and Crossed Random Effects* (cont'd)

- Hence, our model now becomes

$$\begin{cases} \texttt{VF}_{ijk} = \beta_0 + \beta_1 \texttt{Time}_{ijk} + b_i + \texttt{v}_k + \varepsilon_{ijk} \\ b_i \sim \mathcal{N}(0, \sigma^2_{patient}), \qquad \texttt{v}_k \sim \mathcal{N}(0, \sigma^2_{visit}), \\ \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

- The estimated variance components from the Glaucoma data are:
  - $\sigma_{patient} = 5.9$
  - $\sigma_{visit} = 0.8$
  - $\sigma = 8.9$

3.7 Mixed Models with Correlated Errors

- We have seen two classes of models for longitudinal data, namely
  - ▷ *Marginal Models*

$$y_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, V_i), \quad \text{and}$$

  - ▷ *Conditional Models*

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i, \\ \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathsf{I}_{n_i}) \end{cases}$$

3.7 Mixed Models with Correlated Errors (cont'd)

- It is also possible to combine the two approaches and obtain a linear mixed model with correlated error terms

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i, \\ \\ b_i \sim \mathcal{N}(0, D), \quad \varepsilon_i \sim \mathcal{N}(0, \Sigma_i), \end{cases}$$

where, as in marginal models, we can consider different forms for Σ_i

- The corresponding marginal model is of the form

$$y_i \sim \mathcal{N}(X_i\beta,\ Z_i D Z_i^\top + \Sigma_i)$$

3.7 Mixed Models with Correlated Errors (cont'd)

- Features
  - ▷ both b_i and Σ_i try to capture the correlation in the observed responses y_i
  - ▷ this model does not assume conditional independence

- Choice between the two approaches is to a large extent philosophical
  - ▷ *Random Effects*: trajectory of a subject dictated by time-independent random effects $\Rightarrow$ the shape of the trajectory is an inherent characteristic of this subject
  - ▷ *Serial Correlation*: attempts to more precisely capture features of the trajectory by allowing subject-specific trends to vary in time

3.7 Mixed Models with Correlated Errors (cont'd)

Erasmus MC
Universitair Medisch Centrum Rotterdam

Often in practice it is **not** possible to include both a serial correlation term and many random effects because of numerical problems

- Example: In the AIDS dataset we investigate the fit of a mixed model with exponential serial correlation and increasing number of random effects – in particular:
  - ▷ Model I: random intercepts
  - ▷ Model II: random intercepts & random slopes
  - ▷ Model III: random intercepts, random linear slopes & random quadratic slopes

  the fixed-effects part includes linear and quadratic slopes and their interaction with treatment

3.7 Mixed Models with Correlated Errors (cont'd)

	Model I	Model II	Model III
Intercept	7.173	7.214	7.218
$\texttt{Time}_{ij}$	−0.247	−0.251	−0.258
$\texttt{Time}_{ij}^2$	0.007	0.007	0.008
$\texttt{ddI}_i \times \texttt{Time}_{ij}$	0.186	0.154	0.158
$\texttt{ddI}_i \times \texttt{Time}_{ij}^2$	−0.013	−0.010	−0.011

- We observe small differences in the estimated fixed effects

3.7 Mixed Models with Correlated Errors (cont'd)

	Model I	Model II	Model III
ϕ	2.29	0.52	0.33
95% CI	(1.62; 3.23)	(0.08; 3.46)	(0.02; 5.89)

- However, we observe a more profound effect in the estimated parameter of the exponential serial correlation structure
  - ▷ as we include more random effects, less information is available for estimating the serial correlation structure – note length of 95% CIs

- *Numerical problems:*
  - ▷ The model is fitted with the exponential serial correlation structure,
  - ▷ but if you instead tried the Gaussian serial correlation structure, then Model III does not appropriately converge (Hessian matrix of the MLEs is not positive-definite)

3.8 Time-Varying Covariates*

- Up to now all, and with the exception of the time variable, all covariates we have included in the model were fixed from baseline

- However, often we may also be interested in assessing how a longitudinal outcome is associated with a covariate whose value changes over time

  ▷ such covariates are called *time-varying covariates*

- Example: In the PBC dataset we are interested in the effect of prothrombin time on serum bilirubin – prothrombin time has also been collected longitudinally during follow-up

3.8 Time-Varying Covariates* (cont'd)

- The handling of time-varying covariates poses some *important challenges*:

  1. Not always the longitudinal outcome and the time-varying covariate are collected at the same time points

  2. The longitudinal outcome at a particular time point t may depend not only on the value of the covariate at the same time point but also at other time points

  3. There are two types of time-varying covariates, *endogenous* and *exogenous*
     - ▷ a time-varying covariate is *exogenous* if its distribution at time t is conditionally independent of all preceding outcomes
     - ▷ a time-varying covariate is *endogenous* if it is not exogenous

3.8 Time-Varying Covariates* (cont'd)

- The formal definitions of *exogenous* and *endogenous* time-varying covariates are:

$$p\{x_i(t) \mid \mathcal{H}_i^Y(t), \mathcal{H}_i^X(t)\} = p\{x_i(t) \mid \mathcal{H}_i^X(t)\}$$

$$p\{x_i(t) \mid \mathcal{H}_i^Y(t), \mathcal{H}_i^X(t)\} \neq p\{x_i(t) \mid \mathcal{H}_i^X(t)\}$$

where

  - $\mathcal{H}_i^Y(t) = \{y_i(t_{i1}), \ldots, y_i(t_{ik}); t_{ik} \leq t\}$ denotes the set of longitudinal measurements up to time t
  - $\mathcal{H}_i^X(t) = \{x_i(t_{i1}), \ldots, x_i(t_{ik}); t_{ik} \leq t\}$ denotes the set of covariate measurements up to time t

3.8 Time-Varying Covariates* (cont'd)

- These features complicate postulating an appropriate model with such covariates

- A procedure to follow when working with time-varying covariates
  - ▷ Determine if the covariate is *endogenous* or *exogenous*
    * if it is exogenous, then
      - we can proceed by postulating a standard mixed (or marginal) model, and
      - the longitudinal outcome at time t can only be associated with past covariate measurements, i.e., $\mathcal{H}_i^X(t)$

    * if it is endogenous, then more complicated types of analysis are required (joint models or marginal structural models) that fall outside the scope of this course

3.8 Time-Varying Covariates* (cont'd)

Erasmus MC
Universitair Medisch Centrum Rotterdam
Zafung

▷ Next, determine how to link the time-varying covariate to the longitudinal outcome (association structure)

* the longitudinal outcome at t is associated to the covariate at which time points (the same, previous time points, etc.)

* Note: If the scientific interest is focused on a particular type of association structure but in reality the longitudinal outcomes is differently associated to the time-varying covariate, then the estimated association of interest may be diluted (biased) unless a specific type of analysis is followed (a marginal model with independent error terms, i.e., linear regression and corrected standard errors using the sandwich estimator)

3.8 Time-Varying Covariates* (cont'd)

▷ Depending on the chosen association structure in the previous step, and if the time-varying covariate is not measured at the same time points as the longitudinal outcome, then a form of interpolation may be required

- Example: In the PBC dataset we are interested in the effect of prothrombin time on serum bilirubin

$$\begin{aligned}\log(\texttt{serBilir}_{ij}) = {} & \beta_0 + \beta_1 N(\texttt{Time}_{ij})_1 + \beta_2 N(\texttt{Time}_{ij})_2 + \beta_3 \texttt{Female}_i + \beta_4 \texttt{Age}_i + \\ & \beta_5 \texttt{Prothr}_{ij} + b_{i0} + b_{i1} N(\texttt{Time}_{ij})_1 + b_{i2} N(\texttt{Time}_{ij})_2 + \varepsilon_{ij}\end{aligned}$$

the covariance matrix of the random effects is assumed to be diagonal

3.8 Time-Varying Covariates* (cont'd)

	Value	Std.Err.	t-value	p-value
β_0	0.347	0.366	0.948	0.343
β_1	1.772	0.139	12.738	< 0.001
β_2	1.266	0.197	6.422	< 0.001
β_3	-0.233	0.184	-1.263	0.207
β_4	-0.000	0.006	-0.080	0.936
β_5	0.036	0.008	4.675	< 0.001

- Log serum bilirubin at time t is strongly related with the prothrombin time at the same time point – a unit increase of prothrombin time at follow-up time t increases the expected log serum bilirubin at the same follow-up time by 0.036

3.8 Time-Varying Covariates* (cont'd)

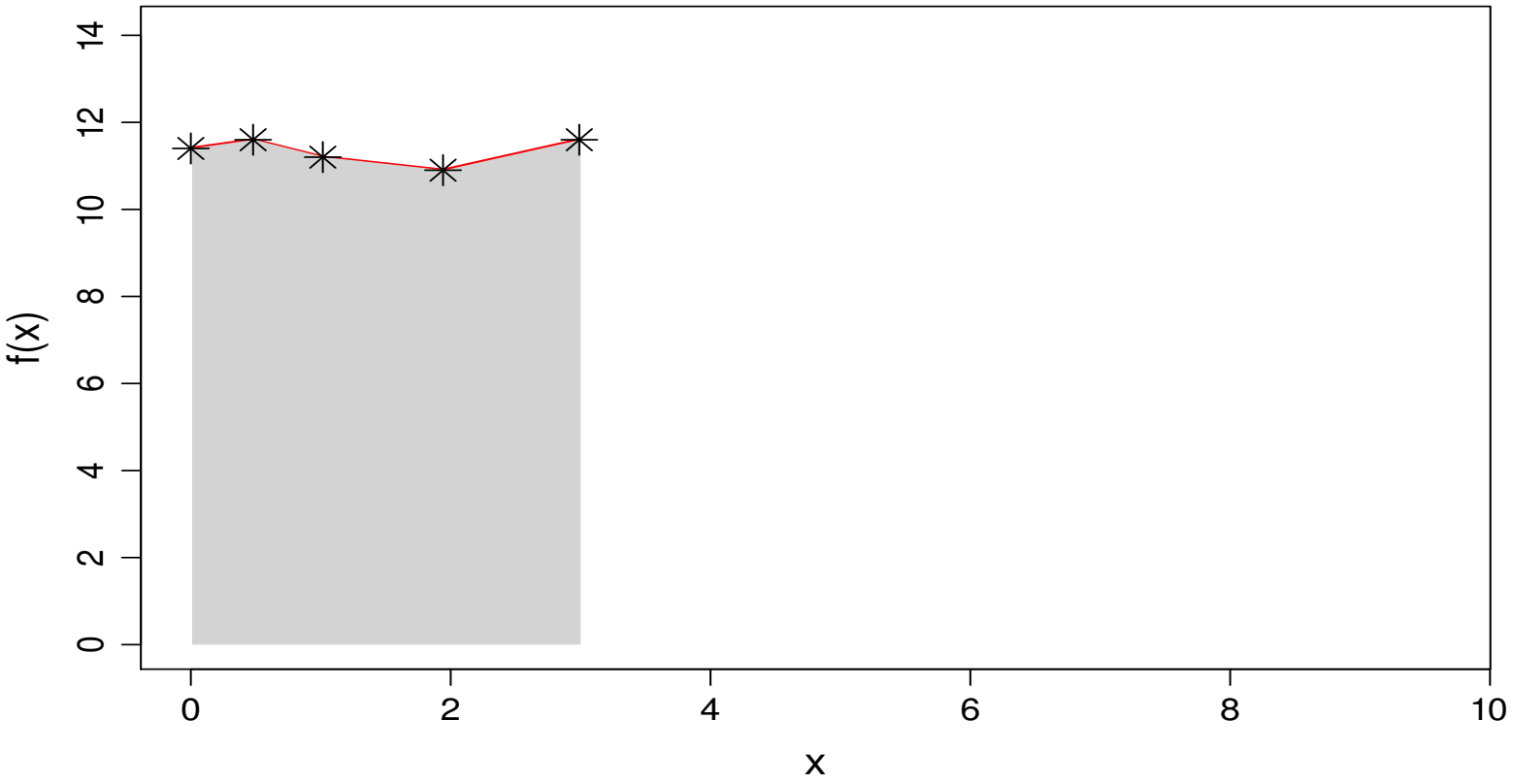
- We continue on the same example, but now we allow the log serum bilirubin at time t to be associated with the prothrombin time at previous time points as well – in particular:

$$\begin{aligned}\log(\texttt{serBilir}_{ij}) = {} & \beta_0 + \beta_1 N(\texttt{Time}_{ij})_1 + \beta_2 N(\texttt{Time}_{ij})_2 + \beta_3 \texttt{Female}_i + \beta_4 \texttt{Age}_i + \\ & \beta_5 \texttt{CumProthr}_{ij} + b_{i0} + b_{i1} N(\texttt{Time}_{ij})_1 + b_{i2} N(\texttt{Time}_{ij})_2 + \varepsilon_{ij}\end{aligned}$$

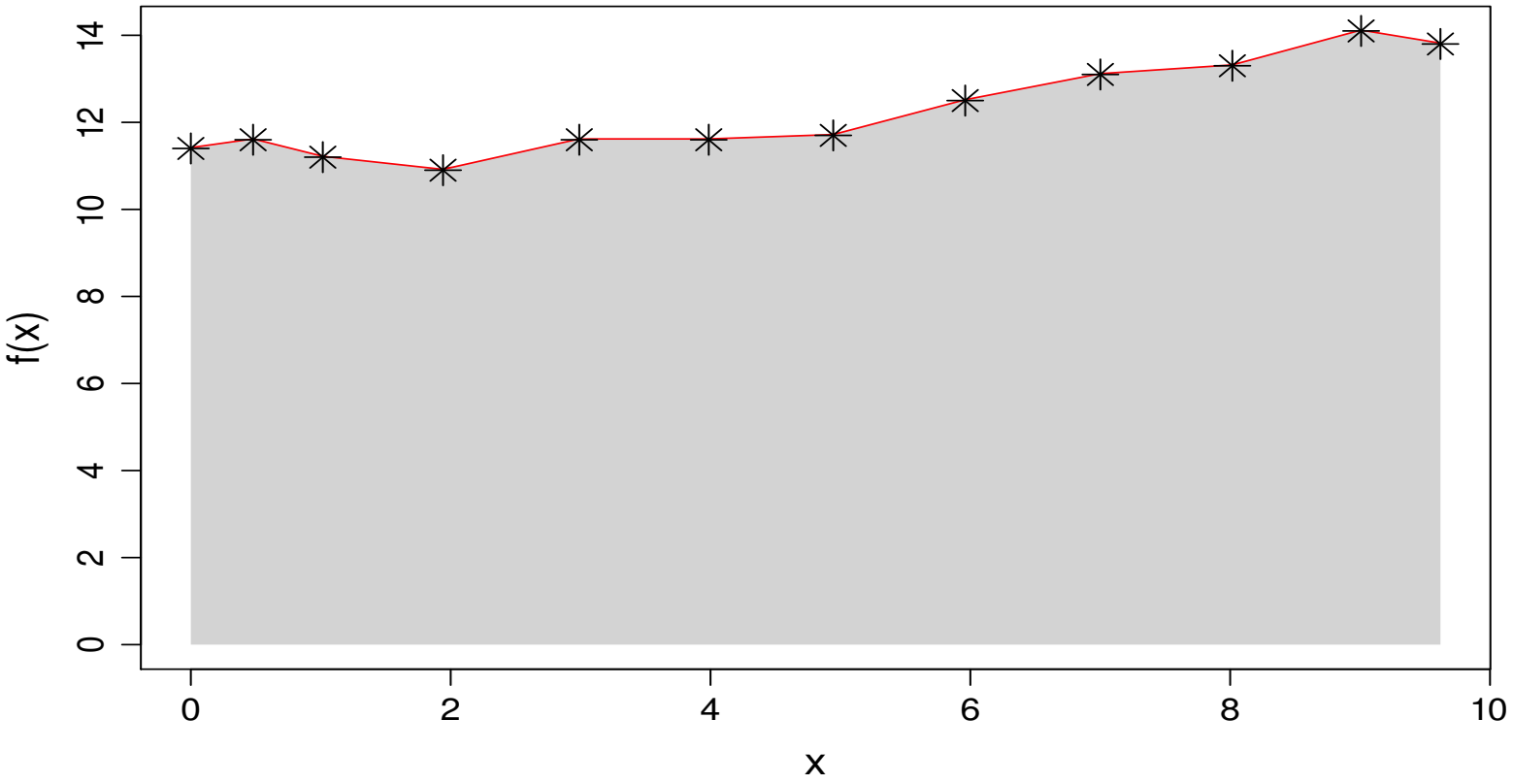
  the covariance matrix of the random effects is assumed to be diagonal

- $\texttt{CumProthr}_{ij}$ denotes the cumulative effect of prothrombin time
  - ▷ for Patient 21 and at two different follow-up times this effect is:

3.8 Time-Varying Covariates* (cont'd)



3.8 Time-Varying Covariates* (cont'd)



3.8 Time-Varying Covariates* (cont'd)

	Value	Std.Err.	t-value	p-value
β_0	0.728	0.361	2.015	0.044
β_1	1.700	0.159	10.686	< 0.001
β_2	1.256	0.210	5.974	< 0.001
β_3	-0.245	0.187	-1.311	0.191
β_4	-0.000	0.006	-0.007	0.995
β_5	0.009	0.004	2.462	0.014

- Log serum bilirubin at time t is strongly related with the cumulative prothrombin time up to the same time point – a unit increase of the cumulative prothrombin time up to follow-up time t increases the expected log serum bilirubin at the same follow-up time by 0.009

3.9 Model Building

- Mixed models consist of two parts, namely
  - ▷ *fixed effects* that describe how specific covariates influence the average longitudinal evolutions
  - ▷ *random effects* that describe how specific regression coefficients deviate from the over all mean described by the fixed effects
    - * the random effects also model the correlations in the repeated measurements

- Interest can either be
  - ▷ on the fixed-effects part alone (e.g., does treatment influence the average evolutions) or
  - ▷ on both parts (e.g., to obtain subject specific predictions)

3.9 Model Building (cont'd)

- The general model building strategy we have seen in the previous chapter for marginal models also applies in the case of mixed models – more specifically:

1. Put all the covariates of interest in the fixed-effects part, considering possible nonlinear terms and/or interactions between them – **do NOT** remove the ones that are not significant

2. Then select an appropriate random-effects structure that adequately describes the correlations in the repeated measurements
   * typically we start from random intercepts and include each time an additional random effect term to see if we improve the fit (i.e., random slopes, quadratic random slopes, etc.)
   * you should be a bit anti-conservative, i.e., do not favor a simpler covariance matrix if the p-value is just non-significant

3.9 Model Building (cont'd)

3. Finally, return to the mean part and exclude non significant covariates
   * first start by testing the nonlinear & interaction terms

3.10 Hypothesis Testing

- ***

3.11 Residuals

- As we have similarly done for marginal models in Chapter 2, before extracting conclusions from mixed models, we will first need to validate the underlying assumptions they make

- To do this we can use the residuals of the model

- In the setting of mixed models we have two types of residuals
  - ▷ *Marginal residuals:* These are based on the implied marginal model behind a linear mixed model (see pp.146)
  - ▷ *Conditional residuals:* These are based on the hierarchical representation of the mixed model and utilize the empirical Bayes estimates of the random effects (see pp.159)

3.11 Residuals (cont'd)

- The exact definitions are as follows:
  - ▷ *Marginal residuals:*

$$\begin{cases} y_i & = X_i\beta + \varepsilon_i^*, \quad \varepsilon_i^* \sim \mathcal{N}(0, Z_i D Z_i^\top + \sigma^2 \mathsf{I}_{n_i}) \\ r_i^{marg} & = y_i - X_i\widehat{\beta} \end{cases}$$

  - ▷ These residuals predict the marginal errors ε_i^*
  - ▷ They can be used to
    - * investigate misspecification of the mean structure $X_i\beta$
    - * validate the assumptions for the within-subjects covariance structure $Z_i D Z_i^\top + \sigma^2 \mathsf{I}_{n_i}$

3.11 Residuals (cont'd)

▷ *Conditional residuals*

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad b_i \sim \mathcal{N}(0, D), \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathsf{I}_{n_i}) \\ \\ r_i^{cond} = y_i - X_i\widehat{\beta} - Z_i\widehat{b}_i \end{cases}$$

▷ These residuals predict the conditional errors ε_i

▷ They can be used to

* investigate misspecification of the hierarchical mean structure $X_i\beta + Z_i b_i$
* validate the assumptions for the within-subjects variance structure σ^2

3.11 Residuals (cont'd)

- Example: We evaluate the assumptions behind the following model fitted to the Prothro dataset:

$$\begin{cases} \texttt{pro}_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})N(\texttt{Time}_{ij})_1 + (\beta_2 + b_{i2})N(\texttt{Time}_{ij})_2 + \\ \quad (\beta_3 + b_{i3})N(\texttt{Time}_{ij})_3 + \beta_4\texttt{predn}_i + \beta_5\{\texttt{predn}_i \times N(\texttt{Time}_{ij})_1\} + \\ \quad \beta_6\{\texttt{predn}_i \times N(\texttt{Time}_{ij})_2\} + \beta_7\{\texttt{predn}_i \times N(\texttt{Time}_{ij})_3\} + \varepsilon_{ij}, \\ \\ b_i \sim \mathcal{N}(0, \text{diag}\{D\}), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

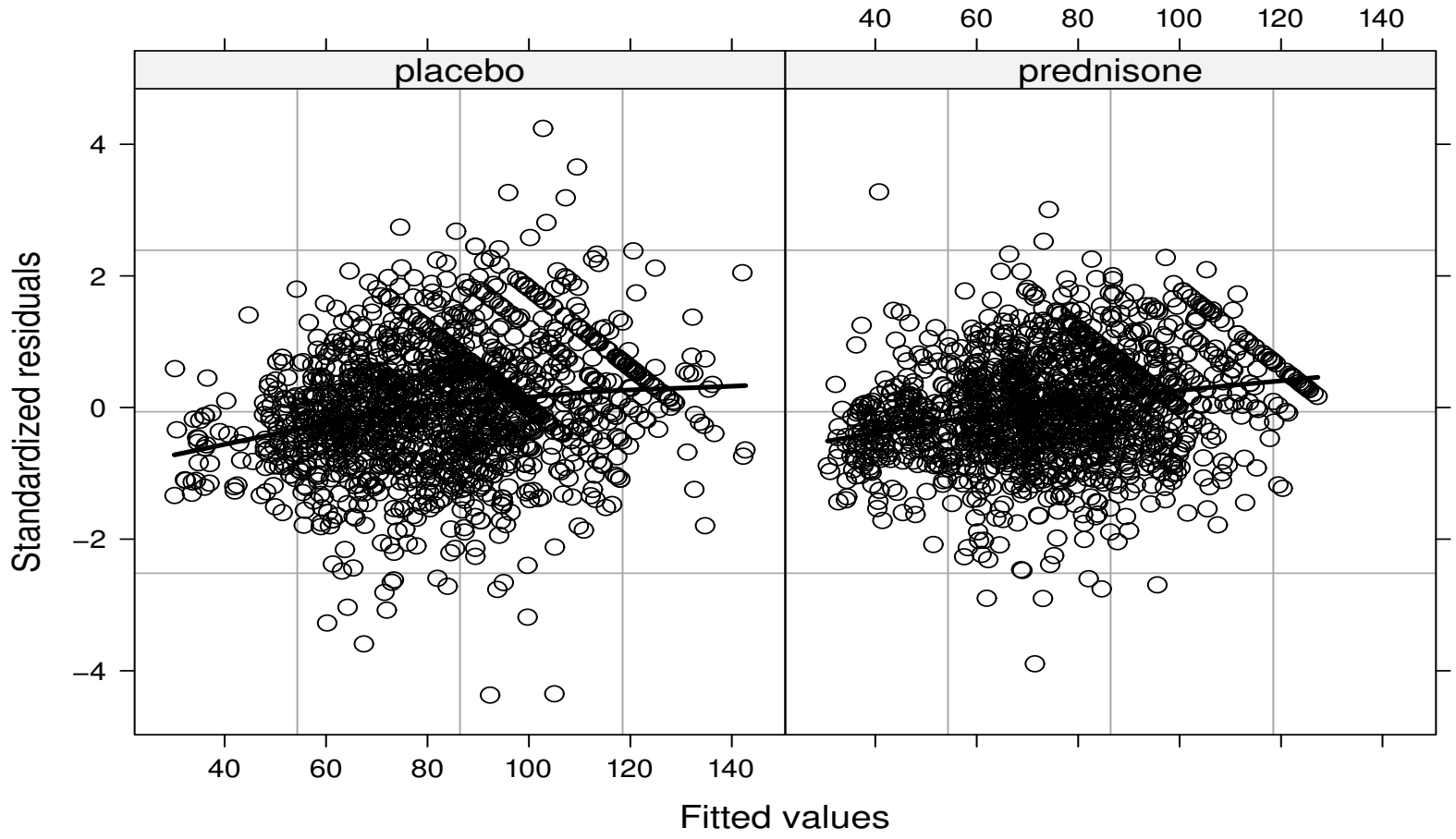
$N(\cdot)$ denotes a natural cubic spline basis

3.11 Residuals (cont'd)

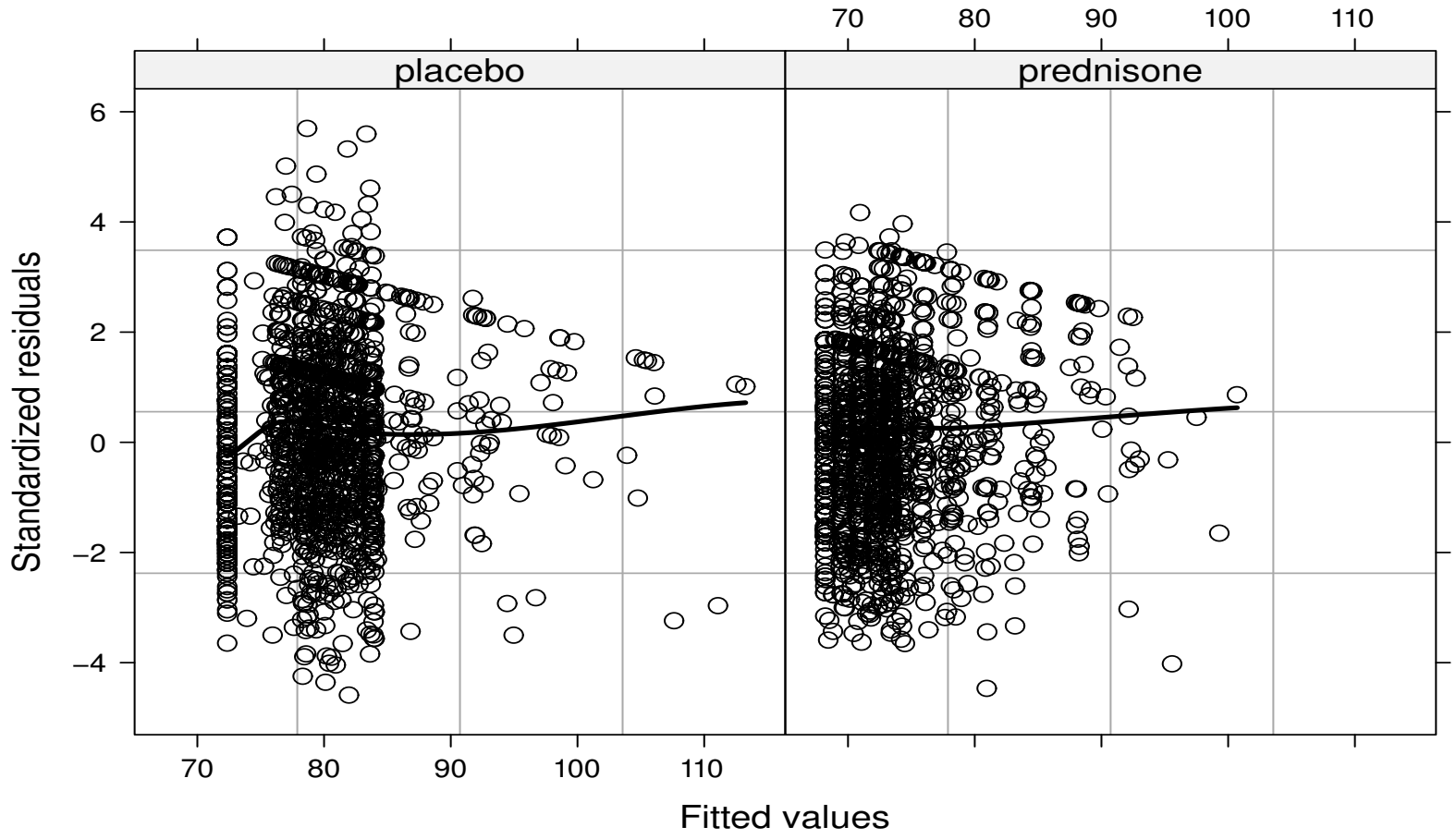
by plotting

- ▷ the standardized marginal residuals versus fitted values per treatment group
- ▷ the standardized conditional residuals versus fitted values per treatment group
- ▷ QQ-plot of the standardized marginal residuals
- ▷ QQ-plot of the standardized conditional residuals

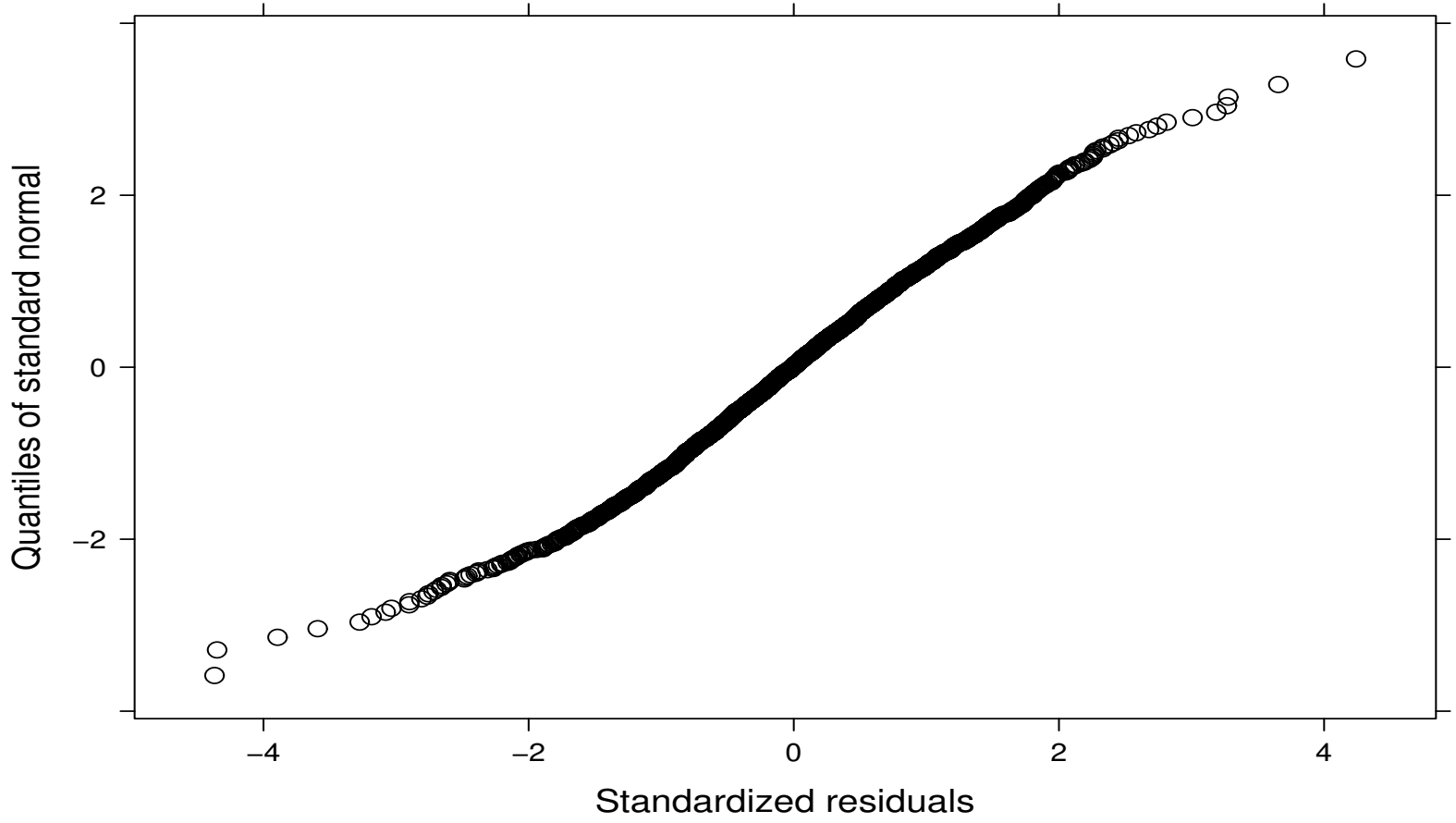
3.11 Residuals (cont'd)



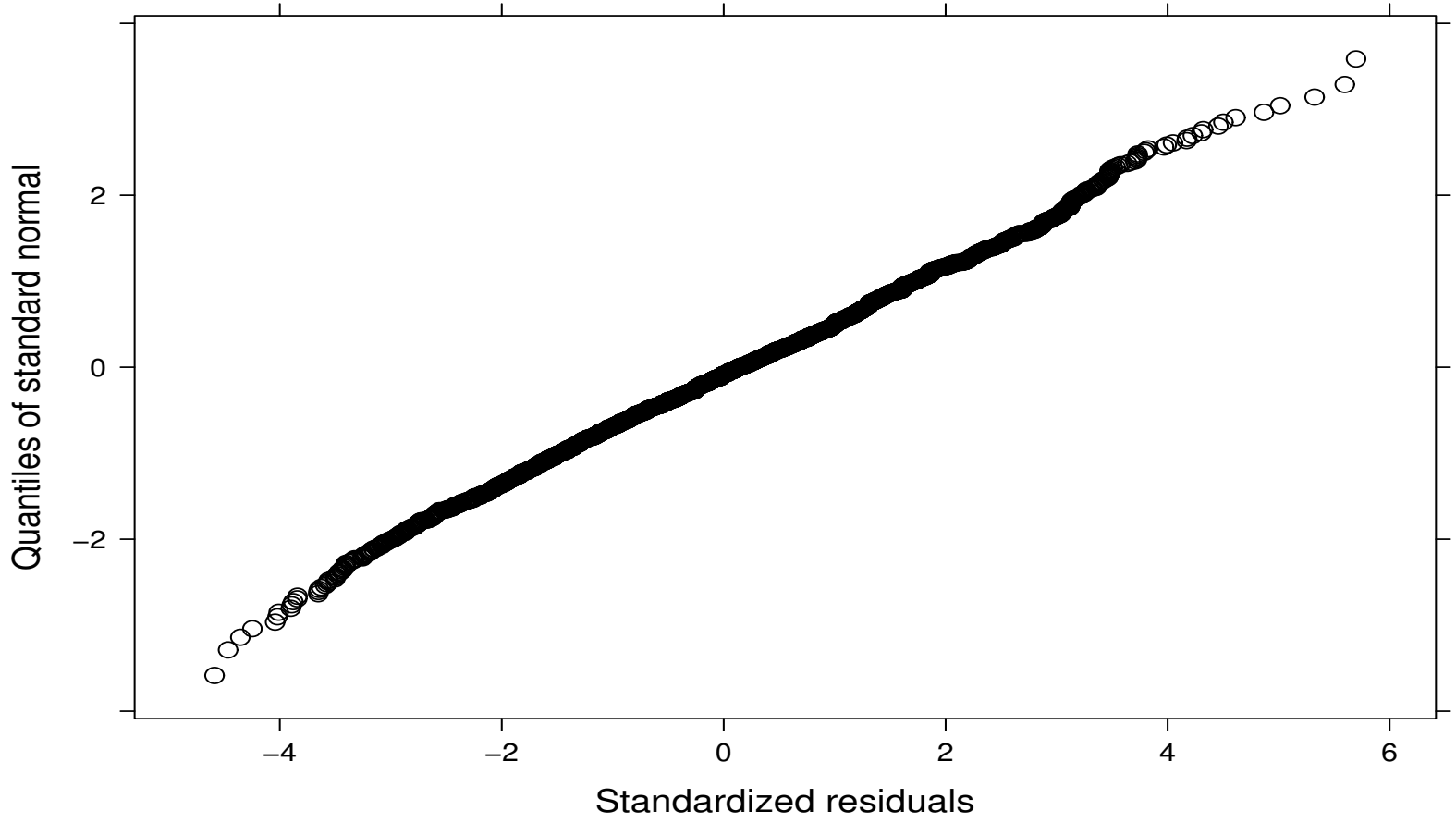
3.11 Residuals (cont'd)



3.11 Residuals (cont'd)



3.11 Residuals (cont'd)



3.11 Residuals (cont'd)

- Observations
  - ▷ the plots of the residuals versus the fitted values do show a slightly systematic behavior
  - ▷ the QQ-plots do not show big discrepancies from normality

3.12 Review of Key Points

- Mixed effects models constitute an alternative modeling framework for analyzing grouped/cluster data
  - ▷ basic idea: sample units in the same group/cluster share the same random effects
  - ▷ the random effects are *unobserved* variables that induce correlation

- From a practical viewpoint mixed models provide a more flexible manner to model correlations when
  - ▷ we have unbalanced data and/or
  - ▷ the correlation structure has a complicated form (e.g., multilevel designs)

3.12 Review of Key Points (cont'd)

- The random effects can be estimated using empirical Bayes methodology
  - ▷ mixed models provide subject-specific predictions that are more accurate than marginal predictions

- Mixed models can be extended to include correlated error terms
  - ▷ this is in the same spirit as the marginal models of Chapter 2
  - ▷ however, this extension often makes the model computationally unstable

3.12 Review of Key Points (cont'd)

- Hypothesis testing
  - ▷ for the covariance structure and for nested models likelihood ratio tests are most often used, for non-nested models AIC/BIC
  - ▷ for the mean structure t and F tests with appropriate degrees of freedom

- Residuals
  - ▷ standard residuals plots are used to check the model assumptions
  - ▷ marginal and conditional residuals available