Covariance Models (*)

Mixed Models Laird & Ware (1982)

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

$\boldsymbol{Y}_i$: $(n_i \times 1)$ response vector

$\boldsymbol{X}_i$: $(n_i \times p)$ design matrix for fixed effects

$\boldsymbol{\beta}$: $(p \times 1)$ regression coefficient for fixed effects

<u>Note</u>: see pg. 60 for specific examples.

<u>Note</u>: FLW Appendix A = "Gentle Intro to Matrices"

Covariance Models (*)

Mixed Models Laird & Ware (1982)

$$\begin{array}{lll} \boldsymbol{Z}_i & : & (n_i \times q) \text{ design matrix for random effects} \\ \boldsymbol{b}_i & : & (q \times 1) \text{ vector of random effects} \\ \boldsymbol{e}_i & : & (n_i \times 1) \text{ vector of errors} \end{array}$$

For the random components of the model we typically assume:

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$$

$$\boldsymbol{e}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_i)$$

Laird & Ware

Chair, Dept. Biostatistics HSPH 1990-1999

Associate Dean HSPH

LMM and components of variation (*)

This yields a covariance structure:

$$\text{cov}(\boldsymbol{Y}_i) = \underbrace{\boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T}_{} + \underbrace{\boldsymbol{R}_i}_{}$$

between-cluster var + **within-cluster var**

- We assume that observations on different subjects are independent.

- Note: This is a matrix (compact) way of writing the covariance for any possibe pair Y_{ij}, Y_{ik}, and represents the variance and covariance details that we presented on pp. 60-1 and 60-2.

LMM and components of variation

Within-Subject: Independence Model :

$$\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}$$

or general diagonal matrix

Then, assuming normal errors we have that $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,n_i})$ are conditionally independent given $\boldsymbol{b}_i$.

- This model assumes that the within-subject errors do not have any serial correlation.

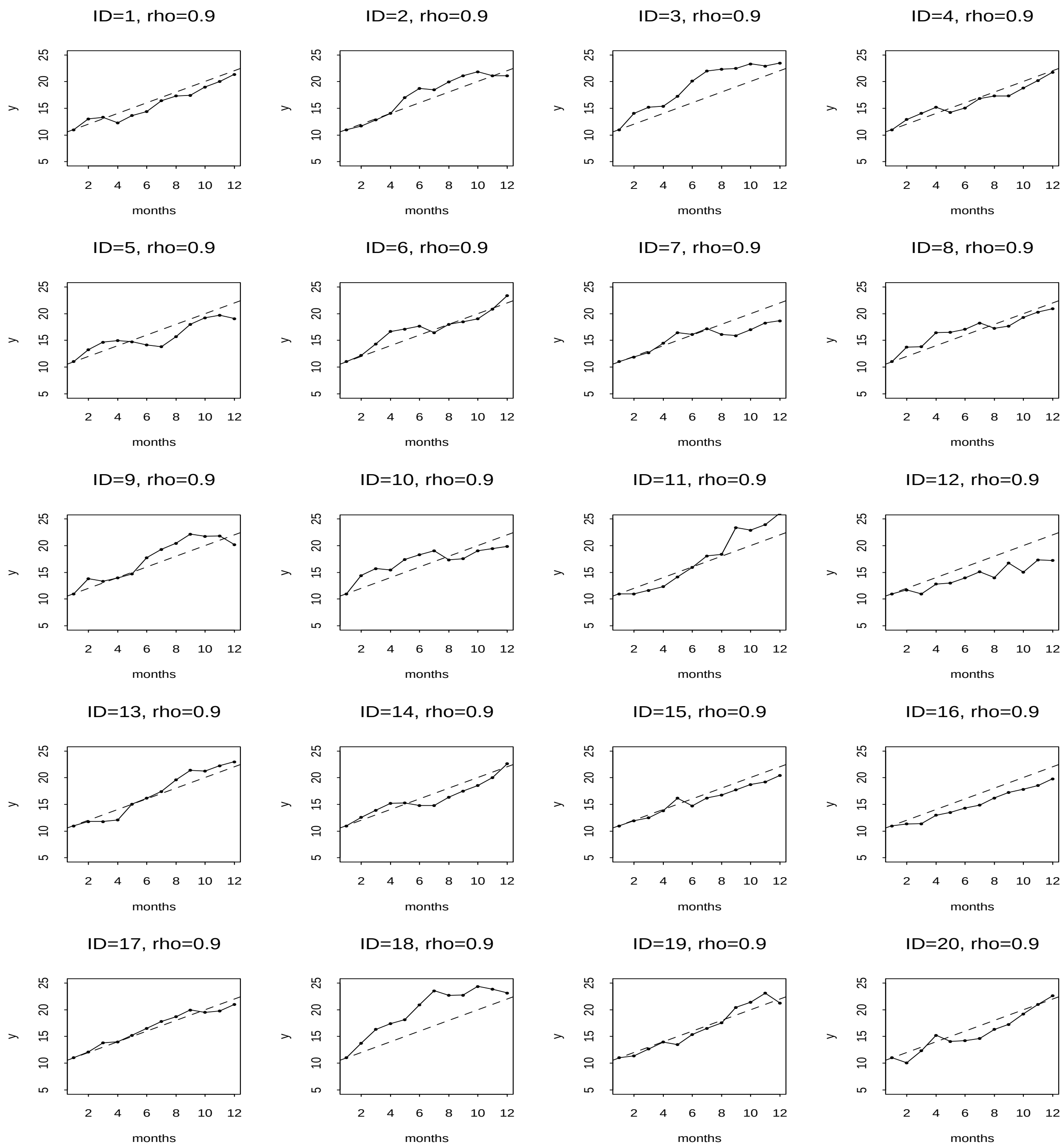
More on Covariance Models

Within-Subject: Serial Models

- Linear mixed models assume that each subject follows his/her own line. In some situations the dependence is more **local** meaning that observations close in time are more similar than those far apart in time.

- One model that we introduced is called the **autoregressive** model where:

$$\text{cov}(e_{ij}, e_{ik}) = \sigma^2 \rho^{|j-k|}$$


ID=1, rho=0.9
ID=2, rho=0.9
ID=3, rho=0.9
ID=4, rho=0.9
ID=5, rho=0.9
ID=6, rho=0.9
ID=7, rho=0.9
ID=8, rho=0.9
ID=9, rho=0.9
ID=10, rho=0.9
ID=11, rho=0.9
ID=12, rho=0.9
ID=13, rho=0.9
ID=14, rho=0.9
ID=15, rho=0.9
ID=16, rho=0.9
ID=17, rho=0.9
ID=18, rho=0.9
ID=19, rho=0.9
ID=20, rho=0.9
y
5 10 15 20 25
months
2 4 6 8 10 12

More on Covariance Models

Autoregressive Correlation **Assume** $t_{ij} = j, n_i \equiv 4$:

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

More on Covariance Models

Autoregressive Correlation **Assume $t_{ij} = j$:**

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{(n-1)} \\ \rho & 1 & \rho & \dots & \rho^{(n-2)} \\ \rho^2 & \rho & 1 & \dots & \rho^{(n-3)} \\ \vdots & & & \ddots & \vdots \\ \rho^{(n-1)} & \rho^{(n-2)} & \rho^{(n-3)} & \dots & 1 \end{bmatrix}$$

More on Covariance Models

Autoregressive Correlation **Assume t_{ij} unique**:

$$\text{corr}(\boldsymbol{e}_i) = \begin{bmatrix} 1 & \rho^{|t_{i1}-t_{i2}|} & \rho^{|t_{i1}-t_{i3}|} & \dots & \rho^{|t_{i1}-t_{in}|} \\ \rho^{|t_{i2}-t_{i1}|} & 1 & \rho^{|t_{i2}-t_{i3}|} & \dots & \rho^{|t_{i2}-t_{in}|} \\ \rho^{|t_{i3}-t_{i1}|} & \rho^{|t_{i3}-t_{i2}|} & 1 & \dots & \rho^{|t_{i3}-t_{in}|} \\ \vdots & & \ddots & & \vdots \\ \rho^{|t_{in}-t_{i1}|} & \rho^{|t_{in}-t_{i2}|} & \rho^{|t_{in}-t_{i3}|} & \dots & 1 \end{bmatrix}$$

More on Covariance Models

Mixed + Serial

- Diggle (1988) proposed the following model

$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_{i,0} + W_i(t_{ij}) + \epsilon_{ij}$$

Covariance Models

Mixed + Serial

The most general type of covariance model will combine some **random effects** with some additional aspects that characterize within-subject **serial correlation**.

One such model contains three sources of random variation:

random intercept	$b_{i,0}$
serial process	$W_i(t_{ij})$
measurement error	ϵ_{ij}

We assume:

$$\begin{aligned} \text{var}(b_{i,0}) &= \nu^2 \\ \text{cov}[W(s), W(t)] &= \sigma^2 \rho^{|s-t|} \\ \text{var}(\epsilon_{ij}) &= \tau^2 \end{aligned}$$

Then:

$$\textbf{Total Variance} = \nu^2 + \sigma^2 + \tau^2$$

$$\textbf{Covariance}(Y_{ij}, Y_{ik}) = \nu^2 + \sigma^2 \rho^{|t_{ij} - t_{ik}|}$$

Covariance Models

Mixed + Serial **Q**: How to biologically interpret these three sources of variation?

- **random intercept**: This represents a "trait" of the subject.
  - ▷ FEV1 – child "size" not captured by age and height.
  - ▷ CD4 – subject's "normal" steady-state level.
- **serial variation**: This represents a "state" for the subject.
  - ▷ FEV1 – child current health status (infected with PseudoA)
  - ▷ CD4 – subject's current immune status (diet? treatment?)
- **measurement error**: This represents the instrumentation or process used to generate the final quantitative measurement.
  - ▷ FEV1 – result of only one "trial" with expiration.
  - ▷ CD4 – blood sample, lab processing.

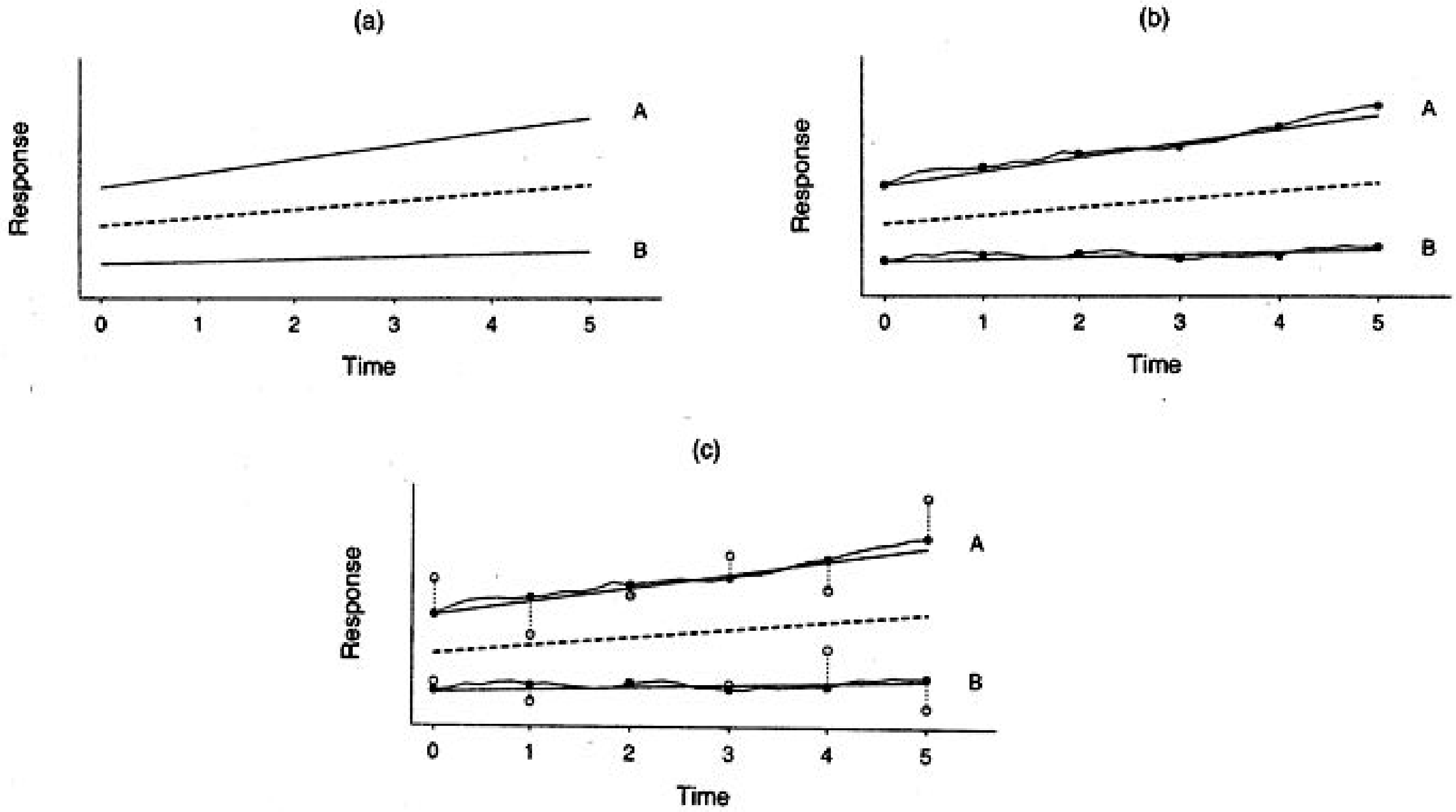


Fig. 2.4 Graphical representation of the cumulative impact of three sources of variability in longitudinal data: (a) between-individual heterogeneity, (b) within-individual biological variation (where • denotes repeated measure free of measurement error), and (c) measurement error (where ∘ denotes observed repeated measure with measurement error).

EDA for Covariance Structure

Numerical Summaries

- Empirical covariance & correlation

Variogram

Define:

$$\begin{aligned} R_{ij} &= Y_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta} \\ &= b_{i,0} + W(t_{ij}) + \epsilon_{ij} \end{aligned}$$

Note:

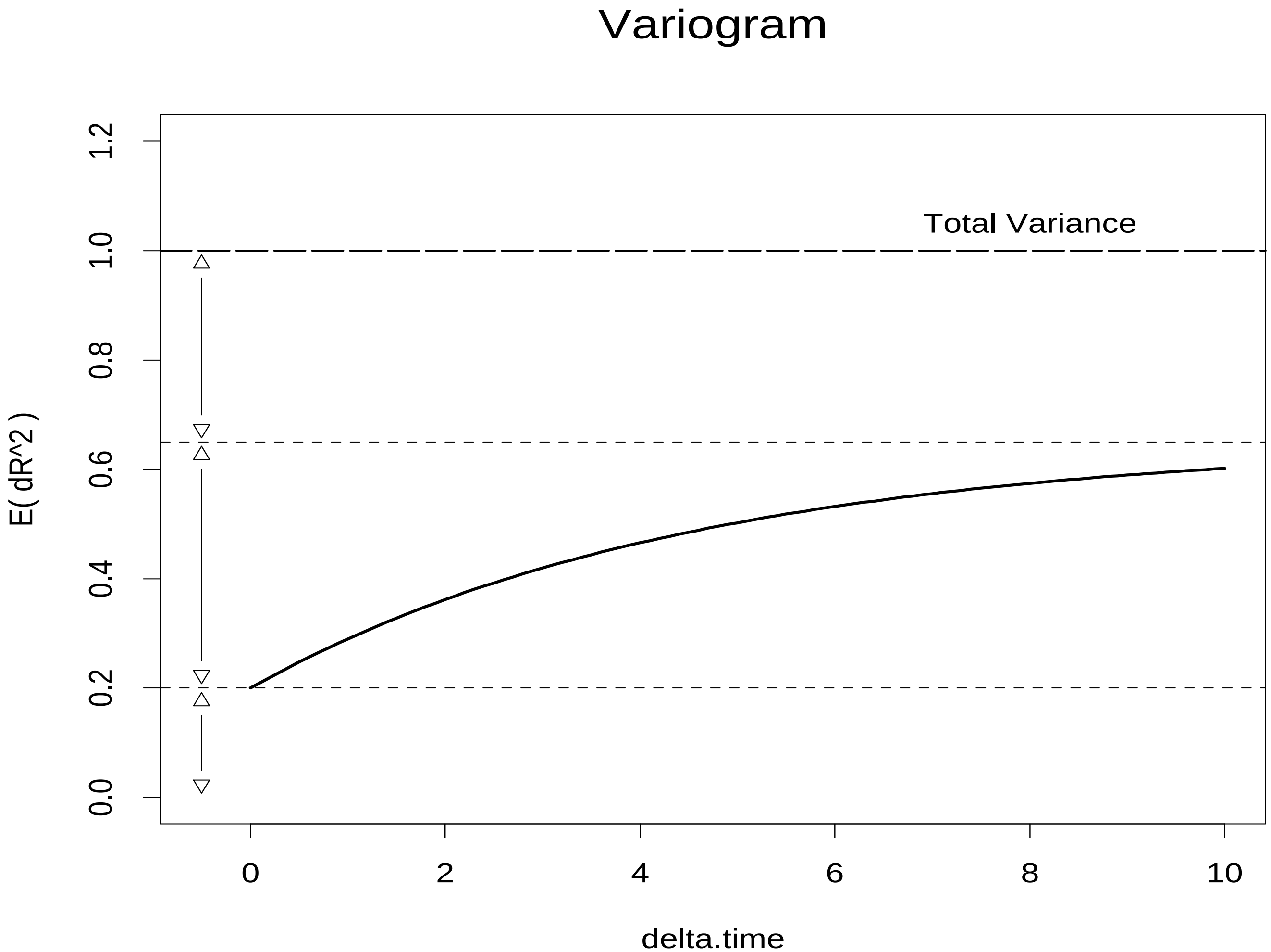
$$\text{var}(R_{ij}) = \nu^2 + \sigma^2 + \tau^2$$

$$E\left[\frac{1}{2}(R_{ij} - R_{ik})^2\right] = \sigma^2 \cdot (1 - \rho^{|t_{ij} - t_{ik}|}) + \tau^2$$

$$\begin{aligned} R_{ij} - R_{ik} &= (b_{i,0} + W_i(t_{ij}) + \epsilon_{ij}) \; - \\ &\quad (b_{i,0} + W_i(t_{ik}) + \epsilon_{ik}) \\ &= [W_i(t_{ij}) - W_i(t_{ik})] + [\epsilon_{ij} - \epsilon_{ik}] \end{aligned}$$

Plot:

$$\frac{1}{2}(\widehat{R}_{ij} - \widehat{R}_{ik})^2 \quad \text{versus} \quad |t_{ij} - t_{ik}|$$


Variogram
Total Variance
E(dR^2)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
delta.time

Variogram: Key Features

When $t_{ij} = t_{ik}$:

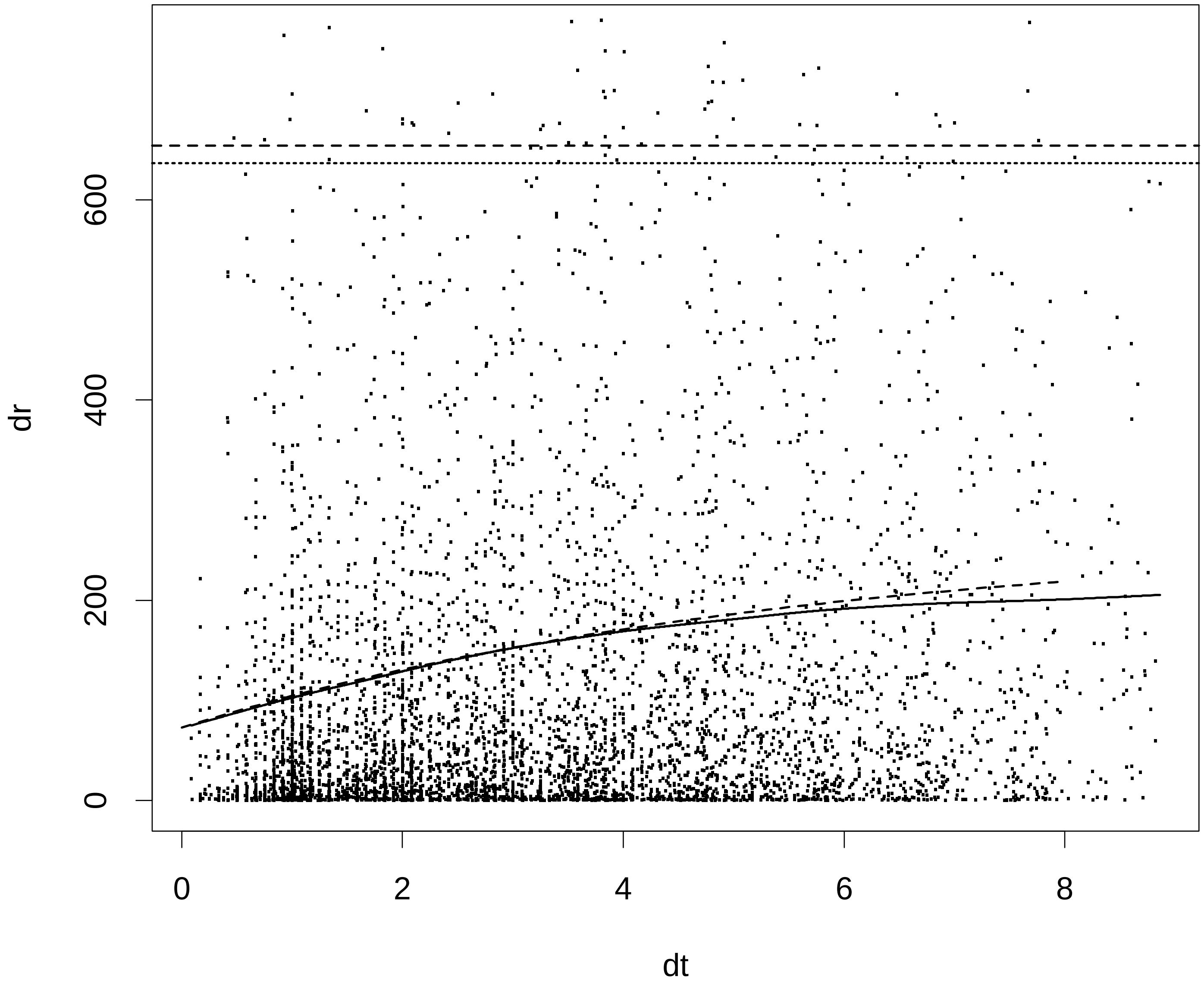
$$
\begin{aligned}
E\left[\frac{1}{2}(R_{ij} - R_{ik})^2\right] &= \sigma^2 \cdot (1 - \rho^{|t_{ij} - t_{ik}|}) + \tau^2 \\
&= \sigma^2 \cdot (1 - \rho^0) + \tau^2 \\
&= \tau^2 \\
&= \textbf{measurement error variance}
\end{aligned}
$$

Variogram: Key Features

When $t_{ij} >> t_{ik}$: (large time separation)

$$
\begin{aligned}
E\left[\frac{1}{2}(R_{ij} - R_{ik})^2\right] &= \sigma^2 \cdot (1 - \rho^{|t_{ij} - t_{ik}|}) + \tau^2 \\
&= \sigma^2 \cdot (1 - \rho^{\infty}) + \tau^2 \\
&= \sigma^2 + \tau^2 \\
&= \text{serial and measurement error variances}
\end{aligned}
$$

FEV1 residual variogram



Recall Simple Linear Regression (**)

- In simple linear regression we fit the model

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$$

- We can write the estimate of the slope, β_1 as follows:

$$\widehat{\beta}_1 = \frac{1}{\sum_i (X_i - \overline{X})^2} \sum_i (X_i - \overline{X}) \cdot (Y_i - \overline{Y})$$

- This method is sometimes called "ordinary least squares", or OLS.

Recall Simple Linear Regression (**)

- In some applications we still want to fit the regression model:

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$$

- But now we want to assign weights, w_i, to each observation.
- Using the weights leads to "weighted least squares" (WLS).
- We can write the estimate of the slope, β_1 as follows:

$$\widehat{\beta}_1(w) = \frac{1}{\sum_i w_i \cdot (X_i - \overline{X})^2} \sum_i (X_i - \overline{X}) \cdot w_i \cdot (Y_i - \overline{Y})$$

- With longitudinal data we have a method of estimation that generalizes this to allow <u>covariance</u> weights.

Estimation of $\boldsymbol{\beta}$ with known $\boldsymbol{\Sigma}_i$ (**)

Weighted least squares:

In univariate regression, WLS yields estimates of $\boldsymbol{\beta}$ that minimize the objective function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{N} w_i (Y_i - \boldsymbol{X}_i \boldsymbol{\beta})^2$$

Analogously, the multivariate version of WLS finds the value of the parameter $\boldsymbol{\beta}(W)$ that minimizes

$$Q_W(\boldsymbol{\beta}) = \sum_{i=1}^{N} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^T \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$$

where $\boldsymbol{W}_i$ is an $(n_i \times n_i)$ positive definite symmetric matrix.

Estimation of $\boldsymbol{\beta}$ with known $\boldsymbol{\Sigma}_i$ (**)

It's straight forward to see that

$$\boldsymbol{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} Q_W(\boldsymbol{\beta}) = -2 \sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$$

- This is a general way to statistically define the regression estimator – a solution to equations.

- In general $\boldsymbol{W}_i$ is chosen as the inverse of $\boldsymbol{\Sigma}_i$.

The solution to the minimization solves $\boldsymbol{U}(\boldsymbol{\beta}) = 0$ and yields

$$\widehat{\boldsymbol{\beta}}(W) = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i\right)$$

Properties of $\boldsymbol{\beta}(W)$ (**)

Given $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \boldsymbol{X}_N$ and $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots \boldsymbol{W}_N$

$$
\begin{aligned}
E\left[\widehat{\boldsymbol{\beta}}(W)\right] &= \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i E[\boldsymbol{Y}_i]\right) \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i \boldsymbol{\beta}\right) \\
&= \boldsymbol{\beta}
\end{aligned}
$$

- Notice that the estimate $\widehat{\boldsymbol{\beta}}(W)$ is unbiased no matter what weighting scheme is used.

Properties of $\boldsymbol{\beta}(W)$ (**)

1 When W_i is correctly specified as the inverse of the variance of $\boldsymbol{Y}_i$ then:

$$\boldsymbol{W}_i = \boldsymbol{\Sigma}_i^{-1} \quad \Rightarrow$$

$$\text{var}\left[\widehat{\boldsymbol{\beta}}(\Sigma^{-1})\right] \quad = \quad \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i \right)^{-1}$$

- When we use `gllamm`, SAS PROC MIXED, or S+ `lme` this is what is returned to provide standard errors for the estimated regression coefficients.

Properties of $\beta(W)$ (**)

2 When W_i is not the inverse of the variance of $\boldsymbol{Y}_i$ then:

$$\boldsymbol{W}_i \neq \boldsymbol{\Sigma}_i^{-1} \quad \Rightarrow$$

$$\text{var}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{W})\right] = \overbrace{\boldsymbol{A}^{-1}}^{\text{bread}} \underbrace{\left(\sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{\Sigma}_i \boldsymbol{W}_i \boldsymbol{X}_i\right)}_{\text{cheese}} \overbrace{\boldsymbol{A}^{-1}}^{\text{bread}}$$

$$\boldsymbol{A} = \sum_i \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i$$

- More on this "sandwich" later...

Likelihood Estimation for Linear Mixed Models (**)

Parameters:

$\boldsymbol{\beta}$: regression parameter, fixed effects coefficient

$\boldsymbol{\alpha}$: variance components

$$\boldsymbol{\alpha} \Rightarrow \quad \boldsymbol{D}(\boldsymbol{\alpha}) \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{\alpha})$$

where $\text{cov}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T + \boldsymbol{R}_i$

Normality:

$$
\begin{aligned}
E(\boldsymbol{Y}_i) &= \boldsymbol{X}_i\boldsymbol{\beta} \\
\mathrm{cov}(\boldsymbol{Y}_i) &= \boldsymbol{\Sigma}(\boldsymbol{\alpha})
\end{aligned}
$$

$$
\begin{aligned}
f(\boldsymbol{Y}_i;\boldsymbol{\beta},\boldsymbol{\alpha}) &= |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-n_i/2} \times \\
& \quad \exp\left[-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right]
\end{aligned}
$$

Maximum Likelihood:

Find the values for the regression coefficients, $\boldsymbol{\beta}$, and the variance components that maximizes the likelihood – e.g. put the highest available probability on the observed data.

R.A. Fisher

ML versus REML

- There is a variant of ML estimation known as **REML**.
  - ▷ "Residual" ML
  - ▷ "Restricted" ML
- REML is used to provide slightly less biased estimates of variance components.
- However, be careful using REML when you change the covariates in your model since one can not use changes in REML log likelihoods to test for fixed effects.
- Useful for a single fitted model, or to compare covariance models with a fixed regression model.

Inference in the Linear Mixed Model

$\boxed{\star}$ In practice:

(1) "Saturated mean model" & explore the covariance.

(2) Fix the covariance & explore the mean.

Likelihood Ratio Tests – Fixed Effects

Standard likelihood theory can be applied to test

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$$

where

$$\begin{aligned} E[\boldsymbol{Y}] &= [\boldsymbol{X}_1, \boldsymbol{X}_2] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \\ &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 \end{aligned}$$

[1]**Full Model**: $E[\boldsymbol{Y}] = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2$

[0]**Reduced Model**: $E[\boldsymbol{Y}] = \boldsymbol{X}_1\boldsymbol{\beta}_1$

Likelihood Ratio Tests – Fixed Effects

In this case we have (when null hypothesis is true):

$$\text{Likelihood Ratio} = \frac{\mathcal{L}_{ML}(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\alpha}}; \text{ML using model 1})}{\mathcal{L}_{ML}(\widehat{\boldsymbol{\beta}}_1, \mathbf{0}, \widehat{\boldsymbol{\alpha}}; \text{ML using model 0})}$$

$$\begin{aligned} \texttt{LRstatistic} &= 2\log \text{ Likelihood Ratio} \\ &= 2\log \mathcal{L}_{ML,1} - 2\log \mathcal{L}_{ML,0} \\ &\sim \chi^2(q) \end{aligned}$$

Where q is the number of coefficients that are set to zero in the reduced model.

Other Tests – Fixed Effects (*)

We also have for a general linear contrast $\boldsymbol{A}$ and a hypothesis

$$H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{0}$$

Wald Test:

$$(\boldsymbol{A}\widehat{\boldsymbol{\beta}})^T \left(\boldsymbol{A}\mathsf{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{A}^T\right)^{-1} (\boldsymbol{A}\widehat{\boldsymbol{\beta}}) \quad \sim \quad \chi^2(q)$$

F Test:

$$\begin{aligned} F &= \frac{(\boldsymbol{A}\widehat{\boldsymbol{\beta}})^T \left(\boldsymbol{A}\ \mathsf{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{A}^T\right)^{-1} (\boldsymbol{A}\widehat{\boldsymbol{\beta}})}{\mathsf{rank}(\boldsymbol{A})} \\ &\sim F(ndf = \mathsf{rank}(\boldsymbol{A}),\ ddf) \end{aligned}$$

⋆ A Model that fits the data

∘ Compare the fitted covariance to the empirical assessment of it:

$$\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{Z}_i \widehat{\boldsymbol{D}} \boldsymbol{Z}_i^T + \boldsymbol{R}_i(\widehat{\boldsymbol{\alpha}}) \quad \text{versus} \quad \text{cov}(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}_i})$$

$$\widehat{\gamma}(\Delta) = \widehat{\tau}^2 + \widehat{\sigma}^2[1 - \widehat{\rho}(\Delta)] \quad \text{versus} \quad \underline{\text{empirical variogram}}$$

$$\widehat{\text{var}}(Y_{ij}) = \widehat{\tau}^2 + \widehat{\sigma}^2 + \widehat{\nu}^2 \quad \text{versus} \quad \underline{\text{empirical variance}}$$

◦ Look at the maximized likelihood:

- Compare $-2 \log \mathcal{L}$

- AIC, BIC

◦ Don't lose sight of the goals of analysis. If covariance selection is to obtain valid model based standard errors then we can assess the impact on $\widehat{\boldsymbol{\beta}}$ and $s.e.$'s. We can also calculate an empirical (sandwich) variance estimate.

Likelihood Ratio Tests – Variance Components

We may want to test whether we have random intercepts and slopes, or just random intercepts.

$$H_0 : \boldsymbol{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{versus} \quad H_1 : \boldsymbol{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

Q: What is the distribution of the likelihood ratio statistic

$$\texttt{LRstat} = 2 \cdot \log \frac{\mathcal{L}_{ML}(\widehat{\boldsymbol{\theta}}_{ML,\text{model 1}})}{\mathcal{L}_{ML}(\widehat{\boldsymbol{\theta}}_{ML,\text{model 0}})}$$

LR Testing for Variance Components (*)

$\boxed{\circ}$ $D_{22} = 0$ is on the boundary of the parameter space!!!

$\Rightarrow$ This violates the standard assumption that we use to justify the $\chi^2(p_1 - p_0)$ distribution of the LR statistic.

$\boxed{\star}$ We appeal to results in Stram and Lee (1994) that build upon results in Self & Liang (1987) showing that `LR stat` is a mixture of χ^2.

Note: For a fixed mean strucure we can use the LR based on either ML or REML. (Why?)

See: Verbeke and Molenberghs (1997) pages 108-111.

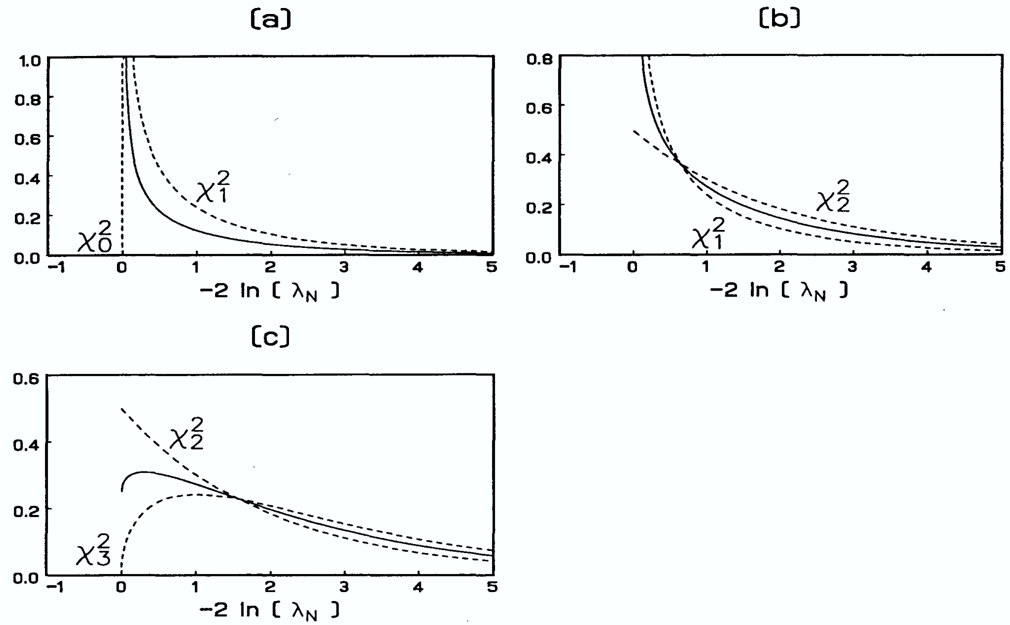


FIGURE 3.4. *Graphical representation of the asymptotic null distribution of the likelihood ratio statistic for testing the significance of random effects in a linear mixed model, for three different types of hypotheses. For each case, the distribution (solid line) is a mixture of two chi-squared distributions (dashed lines), with both weights equal to 0.5:*
(a) Case 1: no random effects versus one random effect.
(b) Case 2: one random effect versus two random effects.
(c) Case 3: two random effects versus three random effects.

Case 1: No Random Effects versus One Random Effect

For testing $H_0 : D = 0$ versus $H_1 : D = d_{11}$, where d_{11} is a non-negative scalar, we have that the asymptotic null distribution of $-2 \ln \lambda_N$ is a mixture of χ^2_1 and χ^2_0 with equal weights 0.5. The χ^2_0 distribution is the distribution which gives probability mass one to the value 0. The mixture is shown in panel (a) of Figure 3.4. Note that if the classical null distribution would be used, all p-values would be overestimated. Therefore, the null hypothesis would be accepted too often, resulting in incorrectly simplifying the covariance structure of the model, which may seriously invalidate inferences, as shown by Altham (1984).

CASE 2: ONE VERSUS TWO RANDOM EFFECTS

In the case one wishes to test

$$H_0 : D = \begin{pmatrix} d_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

for some strictly positive d_{11}, versus H_1 that D is a (2×2) positive semi-definite matrix, we have that the asymptotic null distribution of $-2 \ln \lambda_N$ is a mixture with equal weights 0.5 for χ^2_2 and χ^2_1, shown in Figure 3.4(b). Similar to case 1, we have that ignoring the boundary problems may result in too parsimonious covariance structures.

CASE 3: q VERSUS $q + 1$ RANDOM EFFECTS

For testing the hypothesis

$$H_0 : D = \begin{pmatrix} D_{11} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}, \quad (3.26)$$

in which D_{11} is a $(q \times q)$ positive definite matrix, versus H_1 that D is a general $((q+1) \times (q+1))$ positive semi-definite matrix, the large-sample behavior of the null distribution of $-2 \ln \lambda_N$ is a mixture of χ^2_{q+1} and χ^2_q, again with equal weights 0.5. A graphical representation for the case of testing two random effects ($q = 2$) versus three random effects is given in the third panel of Figure 3.4. Again, we have that the correction due to the boundary problems reduces the p-values in order to protect against the use of oversimplified covariance structures.

S+ LMM Program:

```
#
# cfkids-CDA-NewLMM.q
#
# ------------------------------------------------------------
#
# PURPOSE:  Use linear mixed models to characterize longitudinal
#            change by gender and genotype.
#
# AUTHOR:  P. Heagerty
#
# DATE:  00/07/10  Revised 14Feb2002
#
# ------------------------------------------------------------
#
#
#####
##### Read data
#####
#
source("cfkids-read.q")
#
#
#####
##### Trellis plots of individuals and groups
#####
```

```
#
# Create Grouped Data Set
#
ntotal <-  cumsum( unlist( lapply( split( cfkids$id, cfkids$id) , length ) ) )
cf.subset <- groupedData(
    fev1 ~ age | id, outer = ~ factor(f508)*female,
    data = cfkids[ 1:ntotal[(8*4*1)], ] )
#
cfkids <- groupedData(
    fev1 ~ age | id, outer = ~ factor(f508)*female,
    data = cfkids )
#
# trellis plot, by id, first 1 pages, 8x4
#
postscript( file="cfkids-trellis.1.ps", horiz=F )
plot( cf.subset, layout = c(4,8) )
graphics.off()
postscript( file="cfkids-trellis.2.ps", horiz=T )
par( pch="." )
plot( cfkids, outer = ~ factor(f508)*factor(female), layout=c(3,2),
     aspect=1 )
graphics.off()
#
#####
##### Linear Mixed Models
#####
#
options( contrasts=c("contr.treatment","contr.helmert") )
```

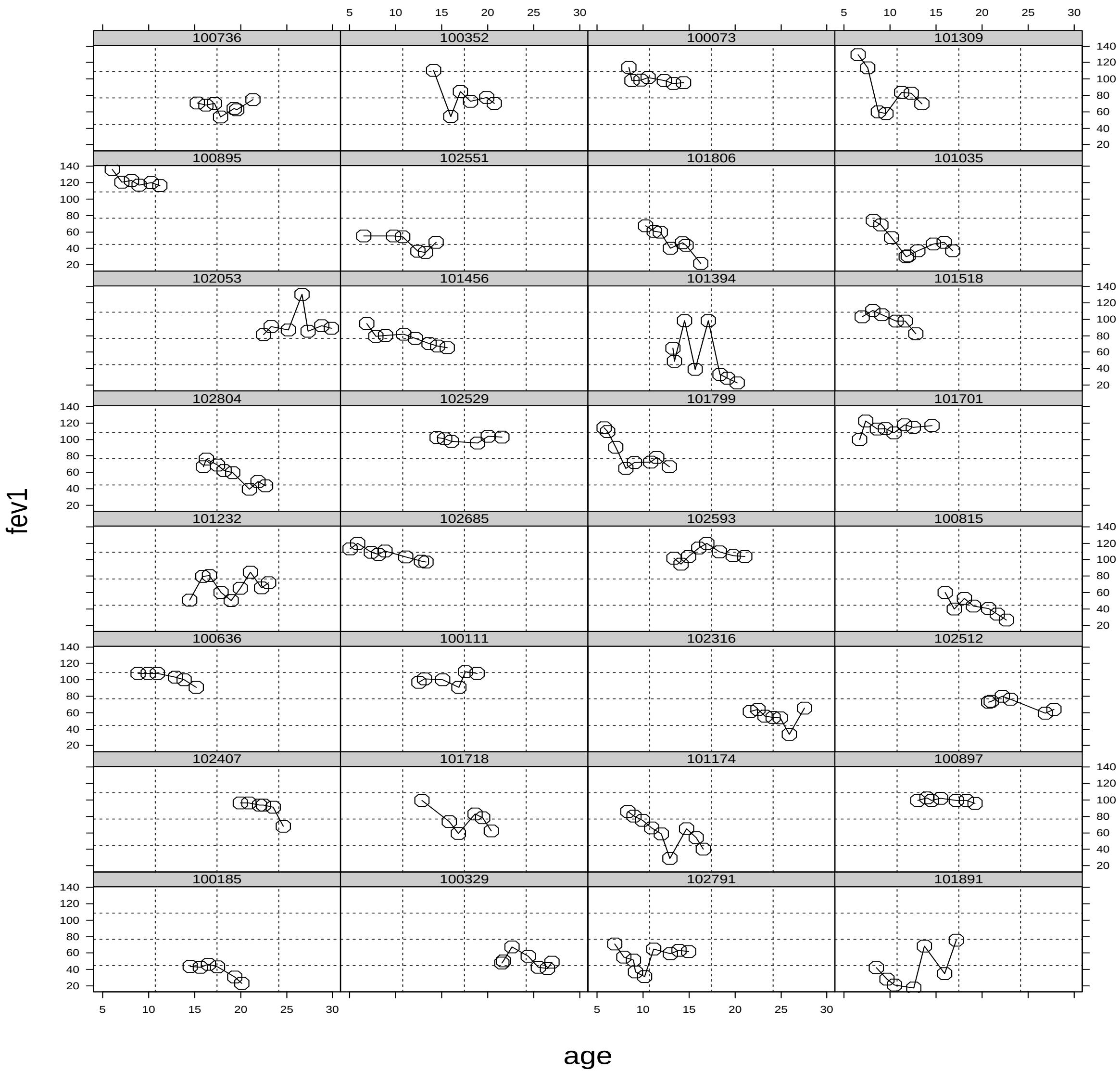
```
#

### Intercept only

fit0 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            data = cfkids )
summary( fit0 )

### Intercept plus Slope

fit1 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 + ageL | id, pdClass="pdSymm", REML=F),
            data = cfkids )
summary( fit1 )

### EDA for serial correlation

postscript( file="cfkids-Variogram.ps", horiz=T )
plot( Variogram(  fit0, form = ~ age | id , resType="response" ) )
graphics.off()


### Intercept plus AR(1)

fit2a <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
```

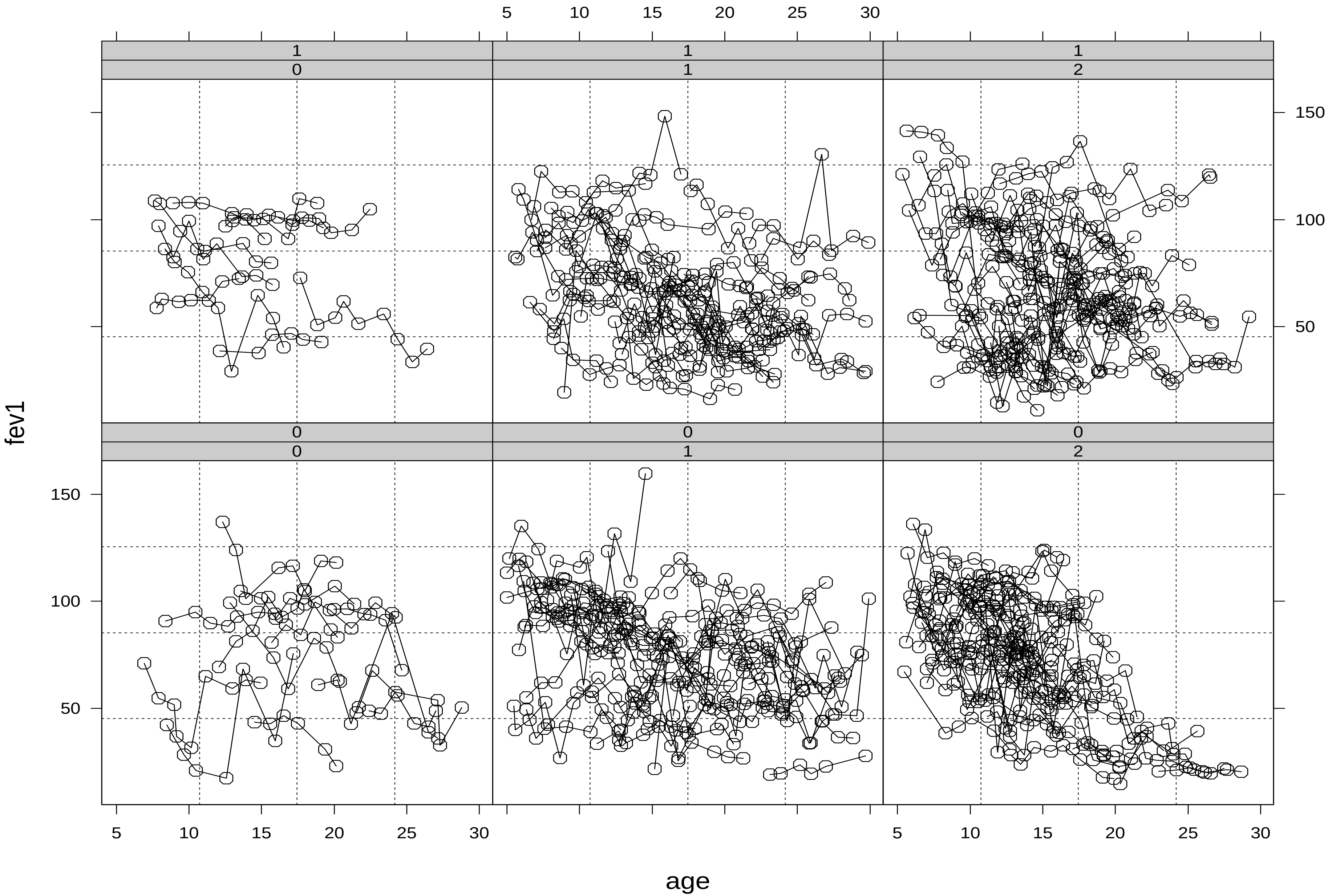
```
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corAR1( form = ~ 1 | id ),
            data = cfkids )
summary( fit2a )

### another way

fit2b <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=F),
            data = cfkids )
summary( fit2b )

### another way

fit2c <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corCAR1( form = ~ ageL | id ),
            data = cfkids )
summary( fit2c )

fit2 <- fit2b

### Intercept plus AR(1) plus measurement error
```

```
fit3 <- lme( fev1 ~ age0 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=T),
            data = cfkids )
summary( fit3 )

#
##### compare these models
#
anova( fit0, fit1, fit2, fit3 )

#
#####
##### Residual Analysis -- using fit3
#####
#
pop.res <- resid( fit3, level=0 )
cluster.res <- resid( fit3, level=1 )
print( var( pop.res ) )
print( var( cluster.res ) )
#
postscript( file="cfkids-NewResiduals.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, pop.res, pch="." )
lines( smooth.spline( cfkids$age0, pop.res, df=5 ) )
title("Residuals (pop) vs Age0")
abline( h=0, lty=2 )
```

```
plot( cfkids$ageL, pop.res, pch="." )
lines( smooth.spline( cfkids$ageL, pop.res, df=5 ) )
title("Residuals (pop) vs AgeL")
abline( h=0, lty=2 )
graphics.off()
#
postscript( file="cfkids-NewResiduals2.ps", horiz=F )
par( mfrow=c(2,1) )
plot( cfkids$age0, cluster.res, pch="." )
lines( smooth.spline( cfkids$age0, cluster.res, df=5 ) )
abline( h=0, lty=2 )
title("Residuals (cluster) vs Age0")
b0 <- unlist( fit2$coefficients$random )
age0 <- unlist( lapply( split( cfkids$age0, cfkids$id ), min ) )
plot( age0, b0 )
lines( smooth.spline( age0, b0, df=5 ) )
abline( h=0, lty=2 )
title("EB b0 versus Age0")
graphics.off()

#
##### Do we need a quadratic age0???
#
fit4 <- lme( fev1 ~ age0 + age0^2 + ageL + female*ageL + factor(f508)*ageL,
            method = "ML",
            random = reStruct( ~ 1 | id, pdClass="pdSymm", REML=F),
            correlation = corExp( form = ~ ageL | id, nugget=T),
            data = cfkids )
```

```
summary( fit4 )
#
anova( fit3, fit4 )
#
# end-of-file...
```

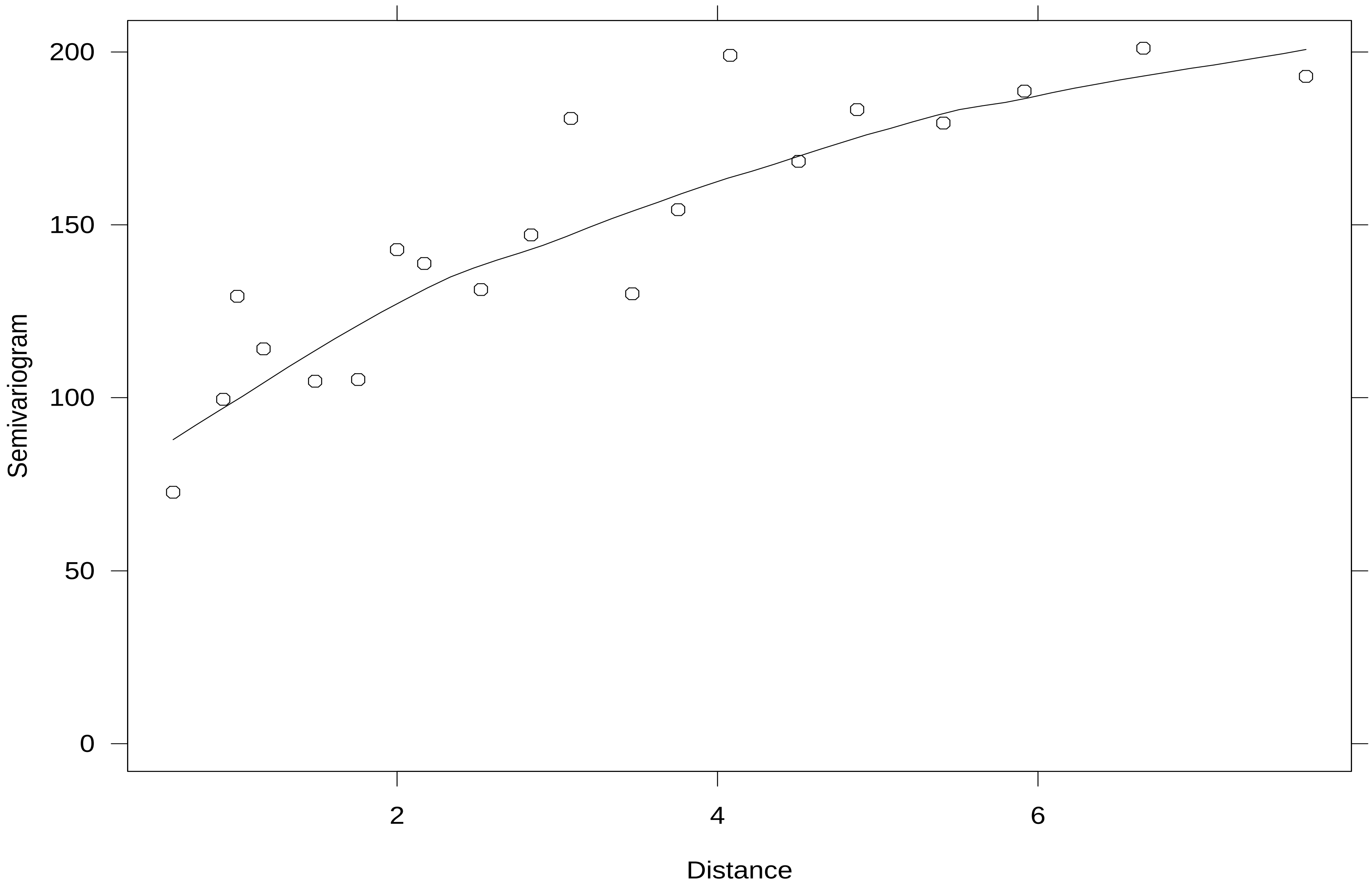

100736
100352
100073
101309
100895
102551
101806
101035
102053
101456
101394
101518
102804
102529
101799
101701
101232
102685
102593
100815
100636
100111
102316
102512
102407
101718
101174
100897
100185
100329
102791
101891
fev1
age



fev1
age
1
0
1
1
1
2
0
0
0
1
0
2
5
10
15
20
25
30
50
100
150

Fit 0 **Random Intercepts**

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12532.01 12590.55 -6255.005

Random effects:
 Formula:  ~ 1 | id
        (Intercept) Residual
StdDev:    22.30435 12.17881

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  103.8074  6.644683 1309  15.62262  <.0001
             age0   -1.8553  0.331087  195  -5.60372  <.0001
             ageL   -0.5881  0.393625 1309  -1.49417  0.1354
           female   -1.1620  3.367634  195  -0.34505  0.7304
    factor(f508)1   -4.2821  5.561343  195  -0.76998  0.4422
    factor(f508)2   -6.7427  5.592591  195  -1.20564  0.2294
      female:ageL   -0.8257  0.249786 1309  -3.30577  0.0010
ageLfactor(f508)1   -0.4873  0.423076 1309  -1.15191  0.2496
ageLfactor(f508)2   -0.6568  0.421964 1309  -1.55655  0.1198

Number of Observations: 1513
Number of Groups: 200
```


200
150
100
50
0
Semivariogram
2
4
6
Distance

Fit 1 Random Intercepts and Slopes

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12430.34 12499.52 -6202.168

Random effects:                                              StdDev   Corr
 Formula:  ~ 1 + ageL | id                       (Intercept) 22.330642 (Inter
 Structure: General positive-definite                   ageL  2.087337 -0.156
                                                    Residual 10.865196

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.5162  6.581633 1309  15.87997  <.0001
             age0   -1.9104  0.327870  195  -5.82672  <.0001
             ageL   -0.6019  0.585611 1309  -1.02777  0.3042
           female   -1.2969  3.338418  195  -0.38847  0.6981
    factor(f508)1   -4.2382  5.511575  195  -0.76896  0.4428
    factor(f508)2   -6.6550  5.542114  195  -1.20081  0.2313
      female:ageL   -0.7636  0.378110 1309  -2.01965  0.0436
ageLfactor(f508)1   -0.5003  0.630694 1309  -0.79323  0.4278
ageLfactor(f508)2   -0.7451  0.629441 1309  -1.18370  0.2367

Number of Observations: 1513
Number of Groups: 200
```

Fit 2a Random Intercepts + AR(1) errors

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12425.01 12488.87 -6200.503

Random effects:                       Correlation Structure: AR(1)
 Formula:  ~ 1 | id                    Formula:  ~ 1 | id
        (Intercept) Residual           Parameter estimate(s):
StdDev:    21.57121 13.25308                  Phi
                                        0.3760331

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.2928  6.704807 1309  15.55493  <.0001
             age0   -1.8599  0.328983  195  -5.65355  <.0001
             ageL   -0.6754  0.507081 1309  -1.33188  0.1831
           female   -1.2954  3.428658  195  -0.37781  0.7060
    factor(f508)1   -4.6714  5.664678  195  -0.82465  0.4106
    factor(f508)2   -6.8660  5.695078  195  -1.20560  0.2294
      female:ageL   -0.8151  0.325773 1309  -2.50206  0.0125
ageLfactor(f508)1   -0.3865  0.546674 1309  -0.70696  0.4797
ageLfactor(f508)2   -0.6224  0.545237 1309  -1.14152  0.2539

Number of Observations: 1513         Number of Groups: 200
```

Fit 2b Random Intercepts + corExp errors

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12412.76 12476.62 -6194.378

Random effects:                Correlation Structure: Exponential spatial corr
 Formula:  ~ 1 | id             Formula:  ~ ageL | id
        (Intercept) Residual    Parameter estimate(s):
StdDev:    21.70338 13.08926        range
                                0.9136573

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.1874  6.700086 1309  15.55015  <.0001
             age0   -1.8560  0.329094  195  -5.63975  <.0001
             ageL   -0.5882  0.499669 1309  -1.17720  0.2393
           female   -1.2584  3.430314  195  -0.36684  0.7141
    factor(f508)1   -4.7019  5.662454  195  -0.83036  0.4073
    factor(f508)2   -6.7593  5.691155  195  -1.18769  0.2364
      female:ageL   -0.8559  0.321526 1309  -2.66206  0.0079
ageLfactor(f508)1   -0.4242  0.539225 1309  -0.78677  0.4316
ageLfactor(f508)2   -0.7042  0.537553 1309  -1.30999  0.1904

Number of Observations: 1513         Number of Groups: 200
```

Fit 2c Random Intercepts + CAR(1) errors

```
Linear mixed-effects model fit by maximum likelihood

 Data: cfkids
       AIC      BIC    logLik
  12412.76 12476.62 -6194.378

Random effects:                 Correlation Structure: Continuous AR(1)
 Formula:  ~ 1 | id              Formula:  ~ ageL | id
        (Intercept) Residual     Parameter estimate(s):
StdDev:    21.69983 13.09153           Phi
                                 0.3350188

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.1877  6.699496 1309  15.55158  <.0001
             age0   -1.8560  0.329056  195  -5.64039  <.0001
             ageL   -0.5883  0.499812 1309  -1.17704  0.2394
           female   -1.2584  3.430073  195  -0.36686  0.7141
    factor(f508)1   -4.7025  5.662058  195  -0.83053  0.4073
    factor(f508)2   -6.7597  5.690753  195  -1.18784  0.2363
      female:ageL   -0.8559  0.321620 1309  -2.66136  0.0079
ageLfactor(f508)1   -0.4241  0.539381 1309  -0.78625  0.4319
ageLfactor(f508)2   -0.7041  0.537708 1309  -1.30942  0.1906

Number of Observations: 1513         Number of Groups: 200
```

Fit 3 Random Intercepts + corExp + meas. error

```
Linear mixed-effects model fit by maximum likelihood
 Data: cfkids
       AIC      BIC    logLik
  12384.92 12454.11 -6179.461

Random effects:                Correlation Structure: Exponential spatial corr
 Formula:  ~ 1 | id             Formula:  ~ ageL | id
        (Intercept) Residual    Parameter estimate(s):
StdDev:    19.69916 15.86913       range    nugget
                                5.116653 0.2878059

Fixed effects: fev1 ~ age0 + ageL + female * ageL + factor(f508) * ageL
                      Value Std.Error   DF   t-value p-value
      (Intercept)  104.7466  6.760173 1309  15.49466  <.0001
             age0   -1.8739  0.328085  195  -5.71153  <.0001
             ageL   -0.7095  0.577371 1309  -1.22886  0.2193
           female   -1.2089  3.486725  195  -0.34671  0.7292
    factor(f508)1   -5.0121  5.753669  195  -0.87112  0.3848
    factor(f508)2   -7.0993  5.782222  195  -1.22778  0.2210
      female:ageL   -0.8259  0.372683 1309  -2.21612  0.0269
ageLfactor(f508)1   -0.3159  0.622931 1309  -0.50710  0.6122
ageLfactor(f508)2   -0.5817  0.621290 1309  -0.93627  0.3493

Number of Observations: 1513
Number of Groups: 200
```

ANOVA

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
fit0	1	11	12532.01	12590.55	-6255.005			
fit1	2	13	12430.34	12499.52	-6202.168	1 vs 2	105.6739	<.0001
fit2	3	12	12412.76	12476.62	-6194.378	2 vs 3	15.5802	1e-04
fit3	4	13	12384.92	12454.11	-6179.461	3 vs 4	29.8354	<.0001

LMM Summary

- Observe $\boldsymbol{Y}_i$, $i = 1, 2, \ldots, m$ independent clusters.

- **Model**: (Laird & Ware, 1982)

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

- $\boldsymbol{\beta}$ is the coefficient that is common to all clusters (fixed across clusters).

- $\boldsymbol{b}_i$ is the deviation of the coefficient that varies from cluster to cluster (random across clusters).

- $(\beta_j + b_{j,i})$ is the coefficient of $X_{i,j}$ for cluster i.

$$\begin{aligned} \boldsymbol{b}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \text{ between-cluster} \\ \boldsymbol{e}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_i) \text{ within-cluster} \end{aligned}$$

- **Estimation/Inference**: WLS, ML
  - ▷ Covariance model choice leads to WLS – but estimated regression coefficient is **unbiased** for any choice of weight (covariance).
  - ▷ Covariance model choice determines the **standard error** estimates for the regression coefficients – **correct** covariance model is needed for **correct** standard errors.