REML Estimation and Linear Mixed Models

3. Analysis of longitudinal data

Sue Welham

Rothamsted Research
Harpenden UK AL5 2JQ

November 18, 2008

Introduction

Longitudinal data is data arising from several measurements made on a set of subjects over time.

The amount of structure in the data varies between two extreme cases:

- Balanced repeated measurements: treatments are allocated to subjects as a designed experiment, these remain constant throughout the study and measurements are made on all subjects at a common set of time-points

- Observational data: subjects are observed at certain intervals over time. Each subject may be measured a different number of times, at different intervals to other subjects. There may be many background covariates that have to be accounted for, which may change over time and will usually not be independent (collinearity) and will often be confounded with the covariates of interest.

More typically, a data set might consist of experimental data, with treatments allocated (possibly changing) according to a pre-planned design, which is not quite balanced and with some additional background covariates to account for.

We will look initially at the balanced case and then consider unbalanced data.

Missing data

One important aspect of longitudinal data is the status of missing data: this be missing and uninformative:

- a scientist forgot to measure a plant
- a patient forgot to turn up for an assessment

but may be informative:

- the allocated treatment killed the plant
- the allocated treatment made the patient too sick to attend the assessment

Much work has been done on the analysis of missing data in longitudinal studies (see eg Verbeke & Molenberghs, 2000) but this will not be covered here.

Aims of analysis

- In longitudinal data/ repeated measurements, the aim is usually to model covariance across times within subjects in order to get better estimates of treatment effects and SEDs
- Efficient identification of a important fixed terms is usually the primary objective
- But identification of a good variance model can aid this process - this is our focus
- We will consider two approaches to this:
  - modelling the covariance structure directly using different covariance structures
  - modelling the covariance structure indirectly via random coefficient regression

Simple ANOVA model

General form of data:

- p replicates of g treatments allocated to $N = pg$ subjects as a designed experiment (may include blocking)
- repeated measurements made on subjects at r time points $\boldsymbol{t} = (t_1 \dots t_r)'$

Returning to Brien & Bailey method of model determination (assume no blocking):

<u>Tier 2</u>		<u>Tier 1</u>
Treatment	$\longrightarrow$	Subject
Time	$\dashrightarrow$	Measurement w/i Subjects

$\dashrightarrow$ indicates that times cannot be randomized: time 1 always comes first

- Simplest model

$$y_{ij} = \mu + T_v + \beta_j + (T\beta)_{vj} + s_i + e_{ij}$$

  - y_{ij} is measurement on subject i ($i = 1 \dots N$) at time j ($j = 1 \dots r$)
  - $v = v(i)$ is the treatment allocated to subject i
  - T_k fixed effect of treatment k, β_j fixed effect of measurement time j, $(T\beta)_{kj}$ fixed interaction of treatment k with time j
  - s_i random effect of ith subject, e_{ij} residual error for subject i at time j

Simple model: uniform correlations

Simple variance components model with

- $\text{var}\,(\boldsymbol{s}) = \sigma^2\gamma_s\boldsymbol{I}$ for $\boldsymbol{s} = (s_1 \ldots s_N)'$
- $\text{var}\,(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$ for $\boldsymbol{e} = (e_{11}\; e_{12} \ldots e_{Nr})'$

gives covariance model

$$\text{cov}\,(y_{ij}, y_{kl}) = \begin{cases} \sigma^2(\gamma_s + 1) & i = k, j = l \\ \sigma^2\gamma_s & i = k, j \neq l \\ 0 & \text{otherwise} \end{cases}$$

- *i.e.* uniform correlation across time within subjects, independence between subjects (= compound symmetry model)

Alternative specification:

$$y_{ij} = \mu + T_v r + \beta_j + (T\beta)_{vj} + \epsilon_{ij}$$

with uniform correlation structure applied directly to $\boldsymbol{e}$:

$$\text{cov}\,(\epsilon_{ij}, \epsilon_{kl}) = \begin{cases} \sigma_e^2 & i = k, j = l \\ \sigma_e^2\theta & i = k, j \neq l \\ 0 & \text{otherwise} \end{cases}$$

Simple model: uniform correlations (2)

Equivalence between models:

$$\sigma_e^2 = \sigma^2(\gamma_s + 1); \quad \theta = \frac{\gamma_s}{\gamma_s + 1}$$

In symbolic form, write these random models as

1. variance components form: subject + subject.time
2. covariance model form: subject.uniform(time)

In GenStat, model specification

```
1. vcomp [fixed=Tmt*Time] random=Subject/Time

2. vcomp [fixed=Tmt*Time] random=Subject.Time
   vstructure [term=Subject.Time] factor=Subject,Time; \
     model=identity,uniform
```

Matrix forms of model

Model 1:

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}$$

where

- $\boldsymbol{y} = (y_{11}\ y_{12} \ldots y_{1r}\ y_{21} \ldots y_{Nr})'$
- $\boldsymbol{\tau} = (\mu\ T_1 \ldots T_g\ \beta_1 \ldots \beta_r\ (T\beta)_{11} \ldots (T\beta)_{gr})'$ is the combined set of fixed effects with $(Nr \times (r+1)(g+1))$ design matrix $\boldsymbol{X}$ defining treatment (and time) allocations to units
- $\boldsymbol{u} = (s_1 \ldots s_N)'$ is the set of subject effects with $(Nr \times N)$ design matrix $\boldsymbol{Z} = \boldsymbol{I}_N \otimes \boldsymbol{1}_r$ defining the allocation of units to subjects with $\text{var}\,(\boldsymbol{s}) = \sigma^2\gamma_s \boldsymbol{I}$
- $\text{var}\,(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$ for $\boldsymbol{e} = (e_{11}\ e_{12} \ldots e_{Nr})'$

Hence

$$\text{var}\,(\boldsymbol{y}) = \sigma^2(\gamma \boldsymbol{ZZ}' + \boldsymbol{I}_n) = \sigma^2 \left\{\gamma(\boldsymbol{I}_N \otimes \boldsymbol{1}_r\boldsymbol{1}_r') + \boldsymbol{I}_n)\right\}$$

where $n = Nr$ is the total number of observations.

Note in this form $\sigma_s^2 \geq 0$.

Matrix forms of model

Model 2:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{\epsilon}$$

where

- $\mathrm{var}\,(\boldsymbol{\epsilon}) = \sigma_e^2(\boldsymbol{I}_N \otimes \boldsymbol{C})$ for $\boldsymbol{\epsilon} = (\epsilon_{11}\ \epsilon_{12} \ldots \epsilon_{Nr})'$ and
- $\boldsymbol{C} = [c_{ij}]$ has $c_{ii} = 1$ for $i = 1 \ldots r$, $c_{ij} = \theta$ for $i \neq j$, $i, j = 1 \ldots r$

Hence

$$\mathrm{var}\,(\boldsymbol{y}) = \mathrm{var}\,(\boldsymbol{\epsilon}) = \sigma_e^2(\boldsymbol{I}_N \otimes \boldsymbol{C}).$$

In this parameterization, the only constraint on the correlation parameter θ is $|\theta| < \sigma^2$.

Negative correlations are allowed - but may only be meaningful in certain specific circumstances:

- successive harvesting: a larger harvest in one period might result in less produce available in the next period
- weight loss: a person losing a lot of weight in one period may be less vigilant in the next period

In both these cases, we might argue that the cumulative variable is more interesting than period-wise increments (cf height vs growth per period).

General linear mixed model

We can generalise our definition of the linear mixed model to accommodate both forms of the model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

As previously, the fixed and random effects may be partitioned into terms associated with explanatory variables:

- $\boldsymbol{X} = [\, \boldsymbol{X}_1 \; \boldsymbol{X}_2 \ldots \boldsymbol{X}_b \,]$
- where $\boldsymbol{X}_i$ is an $n \times p_i$ design matrix for the ith fixed term, $\sum_i p_i = p$
- $\boldsymbol{Z} = [\, \boldsymbol{Z}_1 \; \boldsymbol{Z}_2 \ldots \boldsymbol{Z}_c \,]$
- where $\boldsymbol{Z}_j$ is an $n \times q_j$ design matrix for the jth random term, $\sum_j q_j = q$
- $\boldsymbol{\tau}, \boldsymbol{u}$ are partitioned conformally
  - $\boldsymbol{\tau} = (\boldsymbol{\tau}_1' \; \ldots \boldsymbol{\tau}_b)'$
  - $\boldsymbol{u} = (\boldsymbol{u}_1' \; \ldots \boldsymbol{u}_c)'$
  - with $\boldsymbol{u}_i \sim N(\boldsymbol{0}_{q_i}, \sigma^2 \boldsymbol{G}_i)$ for some valid covariance matrix $\boldsymbol{G}_i$ and cov$(\boldsymbol{u}_i, \boldsymbol{u}_j) = \boldsymbol{0}$
  - $\boldsymbol{e} \sim N(\boldsymbol{0}_n, \sigma^2 \boldsymbol{R})$ for some valid covariance matrix $\boldsymbol{R}$, with cov$(\boldsymbol{e}, \boldsymbol{u}_j) = \boldsymbol{0}$

General linear mixed model (2)

The variance matrix of the data takes the form

$$\begin{aligned} \text{var}(\boldsymbol{y}) &= \sigma^2(\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}) \\ &= \sigma^2(\sum_{i=1}^{c} \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i' + \boldsymbol{R}) \\ &= \sigma^2\boldsymbol{H} \end{aligned}$$

for $\boldsymbol{G} = \oplus\boldsymbol{G}_i$.

In general, both $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\psi})$ and $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\phi})$ may be functions of unknown parameters which are to be estimated via REML.

Results written in terms of a general value of $\boldsymbol{H}$ previously still hold in the model general model, but some further generalization is required.

For example, if we write $\boldsymbol{\kappa} = (\boldsymbol{\psi}', \boldsymbol{\phi}')'$, then the full set of variance parameters takes the form $(\sigma^2, \boldsymbol{\kappa}')'$.

General linear mixed model (3)

The log-likelihood function $\ell_2 = \ell(\sigma^2, \boldsymbol{\kappa}; \boldsymbol{y}_2)$ still takes the same form:

$$\ell_2 = -\frac{1}{2}\left\{c(\boldsymbol{X}) + (n-p)\log(\sigma^2) + \log|\boldsymbol{H}| + \log|(\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})| + \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}/\sigma^2\right\}$$

although now

$$\boldsymbol{H}^{-1} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}$$

The derivative of ℓ_2 with respect to σ^2 is unchanged, but derivatives with respect to elements of $\boldsymbol{\kappa}$ must also consider the form of

$$\frac{\partial \boldsymbol{H}}{\partial \psi_i} = \boldsymbol{Z}\frac{\partial \boldsymbol{G}}{\partial \psi_i}\boldsymbol{Z}' \;; \quad \frac{\partial \boldsymbol{H}}{\partial \phi_j} = \frac{\partial \boldsymbol{R}}{\partial \phi_j}$$

The form of $\hat{\boldsymbol{\tau}}$ is unchanged, with

$$\tilde{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{y} = (\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\tau}})$$

and the mixed model equations are extended to take account of $\boldsymbol{R}$ as

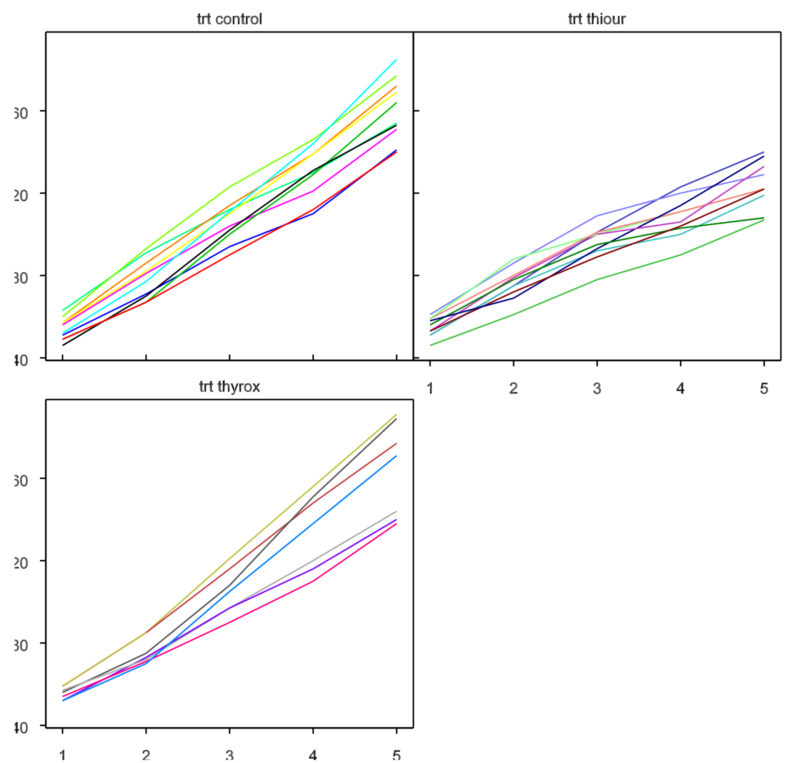
$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}$$

Finally

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\tau}} - \boldsymbol{Z}\tilde{\boldsymbol{u}} = \boldsymbol{R}\boldsymbol{P}\boldsymbol{y}$$

Example

- Repeated measurements of rat weights
- 27 rats allocated to 3 treatment groups: 10 control, 10 treated with chemical thiour, 7 treated with thyrox
- Measurements taken weekly over 5 weeks



Estimated variance parameters

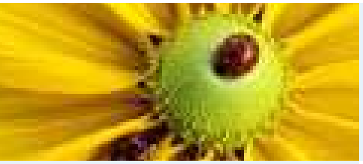
Introduction
Balanced repeated measurements
ANOVA model
General mixed model
Example
AR model
Composite models
Het. AR model
Ante-dependence
AIC & BIC
Model selection
Alternative approach
Unbalanced repeated measurements

```
Model 1
=======

Estimated variance components                        ( 75.54+51.47 = 127.0 )
-----------------------------                        ( 75.54/127.0 = 0.5948 )
Random term                component       s.e.
rat                            75.54      24.82
Residual variance model
-----------------------
Term             Factor        Model(order)  Parameter        Estimate       s.e.
rat.week                       Identity      Sigma2              51.47       7.43


Model 2
=======

Residual variance model
-----------------------
Term             Factor        Model(order)  Parameter        Estimate       s.e.
rat.week                                     Sigma2              127.0       25.5
                 rat           Identity      -                       -          -
                 week          Uniform       theta1             0.5948     0.0884
```



Estimated variance parameters

What happens if we put in both forms of uniform correlation?

```
    vcomp [fixed=Tmt*Time] random=Subject/Time
     vstructure [term=Subject.Time] factor=Subject,Time; model=identity,uniform

Estimated variance components
-----------------------------
Random term                component       s.e.
rat                            63.50     aliased
Residual variance model
-----------------------
Term             Factor        Model(order)  Parameter        Estimate      s.e.
rat.week                                     Sigma2              63.50     12.74
                 rat           Identity      -                       -         -
                 week          Uniform       theta1             0.1896    0.1768
```

- there are only two independent parameters in the model (dependence is not linear)
- no information left after two parameters have been fitted
- aliased indicates that the parameter cannot be optimised (usually sticks at starting position - here $\hat{\gamma}_s = 1$ hence $\hat{\sigma}_s^2 = \hat{\sigma}^2$)

Estimated variance parameters


```
    vcomp [fixed=Tmt*Time] random=Subject/Time
     vstructure [term=Subject.Time] factor=Subject,Time; model=identity,uniform

Estimated variance components
-----------------------------
Random term                component       s.e.
rat                            63.50     aliased
Residual variance model
-----------------------
Term             Factor        Model(order)  Parameter         Estimate       s.e.
rat.week                                     Sigma2               63.50      12.74
                 rat           Identity      -                        -          -
                 week          Uniform       theta1              0.1896     0.1768
```

- variance $= \hat{\sigma}^2(\hat{\gamma}_s + 1) = 2\hat{\sigma}^2 = 127.0$
- correlation $= (\hat{\gamma}_s + \hat{\theta})/(\hat{\gamma}_s + 1) = (1 + \hat{\theta})/2 = 0.5948$
- so answer is correct but in a slightly unusual form
- better to resolve cause of aliasing, than to untangle results
- in general case, aliasing may cause algorithm to fail

Check residuals

Now we have a choice of residuals - from model 1 or model 2

- in general, residuals from correlated structures may be pre-whitened, ie for

$$\mathrm{var}(\boldsymbol{e}) = \sigma^2 \boldsymbol{R}$$

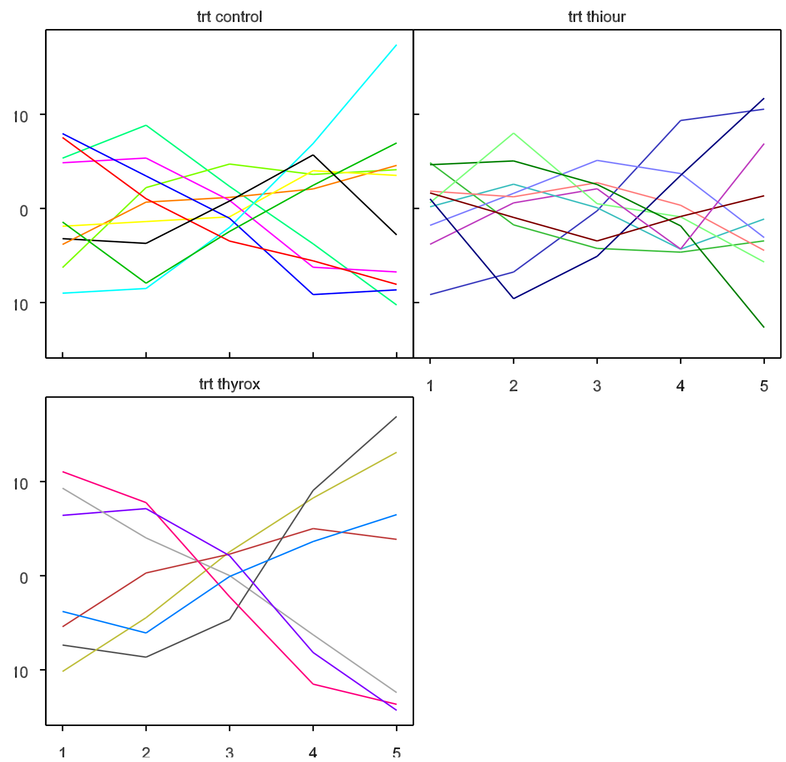
residuals are whitened as

$$\tilde{\boldsymbol{e}}_w = \boldsymbol{L}^{-1}\tilde{\boldsymbol{e}}$$

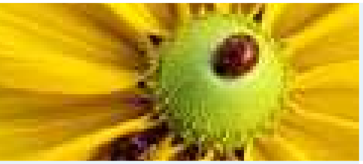
where $\boldsymbol{L}\boldsymbol{L}' = \hat{\boldsymbol{R}}$ and $\tilde{\boldsymbol{e}} = \hat{\boldsymbol{R}}\hat{\boldsymbol{P}}\boldsymbol{y}$.

- the 'best' form of residuals for diagnostics in the linear mixed model is an unresolved issue, see eg Haslett & Dillane (1999)
- in GenStat, residuals are not standardized, not whitened
- then we expect residuals to reflect correlation pattern of fitted matrix
    1. residuals expected to be independent
    2. residuals should reflect uniform correlation structure (??)
- in this case it makes sense to examine model 1 residuals, which should be independent with equal variance

Residuals from model 1



- residuals for each subject are joined by lines
- for independent residuals we expect no pattern
- here there is evidence of temporal correlation, as might be expected
- more sophisticated correlation model required to capture variance pattern



Auto-regressive model

Most common model for temporal correlation for equally-spaced data is auto-regressive model of order 1 (AR1):

$$y_{ij} = \mu + T_v + \beta_j + (T\beta)_{vj} + e_{ij}$$

with AR1 correlation structure applied directly to $\boldsymbol{e}$:

$$\operatorname{cov}(e_{ij}, e_{il}) = \sigma^2 \phi^{|j-l|}$$

This can be written symbolically as subject.AR1(time).

It is easily adapted to unequally-spaced data by using its continuous time analogue:

$$\operatorname{cov}(e_{ij}, e_{il}) = \sigma^2 \phi^{|t_j - t_l|}$$

where t_j is the time at which measurement j was taken.

This is often called the exponential correlation function (the power model in GenStat).

Auto-regressive model

For equally spaced data, the covariance matrix across times within subjects then takes the form:

$$\boldsymbol{C} = \begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 & \dots & \phi^{r-2} & \phi^{r-1} \\ \phi & 1 & \phi & \phi^2 & \dots & \phi^{r-3} & \phi^{r-2} \\ \phi^2 & \phi & 1 & \phi & \dots & \phi^{r-4} & \phi^{r-3} \\ & & & & \ddots & & \\ \phi^{r-3} & \phi^{r-4} & \phi^{r-5} & \phi^{r-6} & \dots & \phi & \phi^2 \\ \phi^{r-2} & \phi^{r-3} & \phi^{r-4} & \phi^{r-5} & \dots & 1 & \phi \\ \phi^{r-1} & \phi^{r-2} & \phi^{r-3} & \phi^{r-4} & \dots & \phi & 1 \end{bmatrix}$$

with inverse

$$\boldsymbol{C}^{-1} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & -\phi & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & \dots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & -\phi & 0 \\ 0 & 0 & 0 & 0 & \dots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & 0 & \dots & -\phi & 1 \end{bmatrix}$$

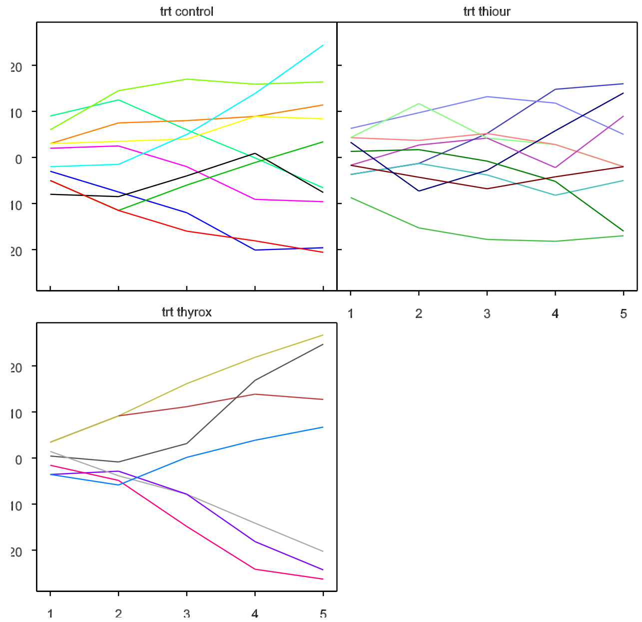
The inverse is sparse (tri-diagonal) and easier to work with.

Fitted AR1 model & residual plots


```
vcomp [fix=trt*week] rat.week
vstruc [rat.week] factor=week; model=ar

Residual variance model
-----------------------

Term       Factor    Model(order)  Parameter     Estimate      s.e.
rat.week                           Sigma2           137.5      36.2
           rat       Identity      -                    -         -
           week      AR(1)         phi_1           0.8821    0.0362
```



High serial correlation + suggestion of variance increasing with time . . .

Composite models

Might consider a composite model:

$$y_{ij} = \mu + T_v + \beta_j + (T\beta)_{vj} + s_i + e_{ij} + a_{ij}$$

with AR1 correlation structure applied directly to $\boldsymbol{e}$:

$$\mathrm{cov}\,(e_{ij}, e_{kl}) = \sigma^2 \phi^{|j-l|}$$

- and $\boldsymbol{s} \sim N(0, \sigma_s^2 \boldsymbol{I})$ to add uniform correlation across time
- this might arise from intrinsic subject differences which stay constant over time
- and $\boldsymbol{a} \sim N(0, \sigma_s^2 \boldsymbol{I})$ to add additional independent error
- this might arise from measurement error on top of the correlated process
- although plausible mechanisms for these extra terms exist, there is rarely sufficient data to fit them all successfully
- variograms can be useful in indicating where extra terms required

Composite models

- For this data, full composite model fails
- Model with additional measurement error fails
- Model with additional subject effects successfully fits the model, but AR1 and subject term are clearly competing for correlation:

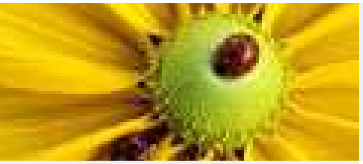
```
Estimated variance components
-----------------------------
Random term                 component       s.e.
rat                            -824.4      112.2

Residual variance model
-----------------------
Term      Factor     Model(order)  Parameter    Estimate       s.e.
rat.week                           Sigma2          909.8      123.8
          rat        Identity      -                   -          -
          week       AR(1)         phi_1          0.9771     0.0033
```

- Unexpected estimates may be trying to tell you something about the model...

Composite models (2)

- This is very much a variance modelling process
- Appears to clash with model determination process discussed earlier?
- May be a problem retaining terms from randomization process with some variance models, *e.g.* subject+subject.ar(week)
- However, this often occurs because terms compete for similar elements of covariation
- Pragmatic approach best - retain randomization terms if no other term is equivalent (or close) or if can sensibly fitted within good variance model



Heterogeneous AR model

We can account for variance heterogeneity apparent in residual plots either indirectly via transformation, or directly by modelling the heterogeneity.

Heterogeneous AR1 correlation structure applied directly to $\boldsymbol{e}$:

$$\operatorname{cov}(e_{ij}, e_{il}) = \sigma_j \sigma_l \phi^{|j-l|}$$

In matrix terms

$$\operatorname{var}(\boldsymbol{e}) = \boldsymbol{I}_N \otimes (\ \boldsymbol{D}^{0.5}\boldsymbol{C}\boldsymbol{D}^{0.5}\ )$$

where

- $\boldsymbol{D}$ is a $r \times r$ diagonal matrix with entries σ_i^2, $i = 1 \ldots r$
- $\boldsymbol{C}$ is a $r \times r$ correlation matrix of form AR1
- pre- and post-multiplication means this is termed outside heterogeneity in GenStat

Ante-dependence model

- An alternative generalization of the AR model is the antedependence model (AD).
- The AR1 model for $\boldsymbol{e}$ can be construed as

$$e_t = \phi e_{t-1} + a_t \quad \text{for } t \text{ large}$$

- where $\boldsymbol{a} \sim N(0, \boldsymbol{D})$, for $\boldsymbol{D} = \boldsymbol{I}$
- $|\phi| < 1$

Note that t is taken to be large so that system is in a steady state.

It follows that

$$\boldsymbol{U}'\boldsymbol{e} = \boldsymbol{a}$$

where $\boldsymbol{U}$ is an upper triangular matrix with value 1 on the diagonal, $-\phi$ on the first off-diagonal and zero elsewhere.

Ante-dependence model

It follows that

$$\begin{aligned} \boldsymbol{U}' \operatorname{var}(e) \boldsymbol{U} &= \operatorname{var}(a) \\ \operatorname{var}(e) &= (\boldsymbol{U}')^{-1} \boldsymbol{D} \boldsymbol{U}^{-1} \\ \operatorname{var}(e) &= (\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}')^{-1} \end{aligned}$$

Generalization of the AR1 model to

$$e_t = \phi_t e_{t-1} + a_t \quad \text{for} \quad t > 1$$

where $\boldsymbol{a} \sim N(0, \boldsymbol{D})$, for $\boldsymbol{D} = \text{diag}\{d_i; d_i > 0\}$

This is the ante-dependence model of order 1

- with covariance matrix $\boldsymbol{C} = (\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}')^{-1}$
- where off-diagonal elements of $\boldsymbol{U}$ are now $\{-\phi_t\}$
- Generalization to higher orders follows by adding lags of e_{t-2} etc

Comparison of variance models

- Likelihood ratio tests are valid only for nested models
- Can use information criteria to compare non-nested variance models (with same fixed effects)
- For a mixed model fitted by REML with
  - N_v variance parameters estimated
  - n data values
  - p DF fitted for fixed terms
  - log-likelihood function maximised under model as RL
- AIC $= -2RL + 2N_v$ (Akaike Information Criterion)
- BIC/SBC $= -2RL + N_v \log(n - p)$ (Bayesian/Schwarz IC)
- BIC tends to be more conservative than AIC
- for criterion chosen, variance model with lowest IC value is chosen

Model selection via IC

For rats data

- $n = 135$
- for fixed=trt*week, $p = 15$
- $n - p = 120$

Model	-2RL	N_v	AIC	BIC
Uniform	676.57	2	680.52	686.14
AR(1)	599.09	2	604.09	608.66
Het. AR(1)	548.61	6	560.61	577.33
AD(1)	544.56	9	562.56	587.65
AD(2)	519.58	12	543.58	577.02
US	517.55	15	547.55	589.36

- In this case, agreement between criteria
- Ordering different: BIC favours more parsimonious models

Impact on fixed effect SEDs

Several purposes of modelling covariance structure:

- to understand patterns of variance and correlation
- to get more appropriate SEDs for fixed effects and PEVs for random effects

Compare SEDs from two models for rat data:

Uniform correlation model

	week 1	week 2	week 3	week 4	week 5
week 1	*				
week 2	1.98	*			
week 3	1.98	1.98	*		
week 4	1.98	1.98	1.98	*	
week 5	1.98	1.98	1.98	1.98	*

Unstructured variance model

	week 1	week 2	week 3	week 4	week 5
week 1	*				
week 2	0.96	*			
week 3	1.42	0.98	*		
week 4	2.35	2.15	1.29	*	
week 5	3.03	2.92	2.11	1.16	*

- Unstructured model reflects both changing variance and correlation

Alternative approach

This analysis fits treatment means at each time-point and uses the covariance matrix to describe the nature of individual variation about treatment means.

There is clear linear trend over time in the profiles and the current analysis does not exploit this structure. We might therefore try an alternative model:

$$y_{ij} = \mu + T_v + at_j + b_v t_j + e_{ij}$$

with a suitable correlation structure applied directly to $\boldsymbol{e}$:

$$\mathrm{var}\,(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_N \otimes \boldsymbol{C}$$

This model fits linear trend in t, with a separate intercept $(\mu + T_v)$ and slope $(a + b_v)$ for each treatment group.

However, if the linear trend is a poor fit to the mean profiles, then the covariance structure will describe the lack of fit as well as variation about treatment means.

If we use the modified model

$$y_{ij} = \mu + T_v + at_j + \beta_j + b_v t_j + (T\beta)_{vj} + e_{ij}$$

then the additional term β_j fits the overall mean at each time exactly, and $(T\beta)_{vj}$ fits the treatment mean at each time point exactly.

Alternative approach (2)

If the model terms are fitted in the order given, then these terms can be used to test for lack of fit of the linear trend model.

In symbolic terms, this model could be represented as

- fixed = Trt + t + fac(t) + Trt.t + Trt.fac(t)
- random = Subject.Cov(Time)

where t represents the numeric variate, and Trt represents the treatment factor.

Tests for rat data:

```
Sequentially adding terms to fixed model

Fixed term                    Wald statistic  n.d.f.   F statistic  d.d.f.    F pr
trt                                     4.09       2          2.05    24.2   0.151
time                                 1407.18       1       1407.18    25.5  <0.001
cweek                                  35.71       3         11.11    25.8  <0.001
trt.time                               24.77       2         12.38    25.5  <0.001
trt.cweek                              27.84       6          4.26    32.3   0.003
```

where cweek is a copy of the week factor = fac(t), time=t, trt=Trt.

Alternative approach (3)

Adding a quadratic term (timesqrd) removes the group-specific lack of fit - although still some lack of fit to overall means at each time:

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
trt	4.09	2	2.05	24.2	0.151
time	1407.18	1	1407.18	25.2	<0.001
timesqrd	0.09	1	0.09	25.9	0.769
cweek	35.63	2	17.22	26.2	<0.001
trt.time	24.77	2	12.38	25.2	<0.001
trt.timesqrd	20.02	2	10.01	25.9	<0.001
trt.cweek	7.82	4	1.87	30.0	0.141

Removing the lack of fit terms has some impact on the across time covariance model (AD2):

Covariance matrix with lack of fit terms

1	21.6				
2	33.0	68.7			
3	31.6	69.1	94.8		
4	27.8	64.5	116.4	181.6	
5	24.9	60.1	122.9	207.2	268.4
	1	2	3	4	5

Covariance model omitting lack of fit

1	22.6				
2	35.6	75.7			
3	33.4	74.0	99.2		
4	33.4	77.0	128.9	218.5	
5	32.3	75.4	134.9	242.4	302.0
	1	2	3	4	5

Unbalanced data

In many cases, longitudinal data will not be balanced in either the allocation of treatments to subjects or in terms of number and frequency of measurements.

Each subject ($i = 1 \ldots N$) then has their own vector of n_i measurement times $\boldsymbol{t}_i = (t_{i1} \ldots t_{in_i})$ and the model is usually written in general terms as

$$y_{ij} = \mu + f(t_{ij}) + f_v(t_{ij}) + e_{ij}$$

where $f()$ is a function describing the population mean profile and $f_j()$ describes deviations of treatment group v from the mean profile.

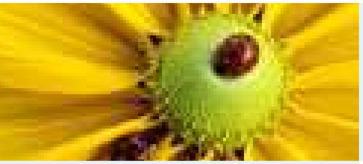
If subjects are measured at different times, then fitting treatment means at each time results in an over-parameterized model that gives little insight into the process.

The variance model for the data (ordered by subjects) is usually written as

$$\mathrm{var}\,(\boldsymbol{e}) = \oplus\{\boldsymbol{C}_i\};\; i = 1 \ldots N$$

where $\boldsymbol{C}_i$ is the across-time covariance matrix for subject i, and the overall variance matrix is block diagonal.

These covariance matrices will be determined by the same underlying model (eg AR1) but will take different numeric values due to the differing sets of measurement times.



Computational issues

The direct product structure is convenient as has a simple inverse

$$(\boldsymbol{I}_N \otimes \boldsymbol{C})^{-1} = \boldsymbol{I}_N \otimes \boldsymbol{C}^{-1}$$

whereas the direct sum structure has inverse

$$\oplus\{\boldsymbol{C}_i^{-1}\}$$

In the former case, $\boldsymbol{C}$ has to be inverted once but in the latter each $\boldsymbol{C}_i$ may have to be inverted separately, and matrix multiplications involving $\boldsymbol{R}^{-1}$ are correspondingly more complex.

If the imbalance is slight eg. an overall set of common measurement times with a few missing values, then it may be computationally more efficient to include the missing measurements to retrieve the balanced structure.

Consider the model

$$\begin{aligned} y_{ij} &= \mu + f(t_j) + f_v(t_j) + e_{ij} && \text{for } t_j \text{present for subject} i \\ 0 &= \theta_{ij} + \mu + f(t_j) + f_v(t_j) + e_{ij} && \text{for } t_j \text{absent for subject i} \end{aligned}$$

Computational issues

Consider the model

$$\begin{aligned} y_{ij} &= \mu + f(t_j) + f_v(t_j) + e_{ij} && \text{for } t_j \text{present for subject} i \\ 0 &= \theta_{ij} + \mu + f(t_j) + f_v(t_j) + e_{ij} && \text{for } t_j \text{absent for subject i} \end{aligned}$$

The parameters θ_{ij} are known as 'missing value covariates' with $\hat{\theta}_{ij} = -\tilde{y}_{ij}$ at times j with no data on subject i and all other estimates are unchanged (see eg. Gilmour *et al*, 2004).

On rearranging these equations back into time order for each subject, the direct product structure of the variance model is retrieved.

Whether this is an efficient strategy depends on the number of missing values.

GenStat technical notes

GenStat deals poorly with direct sum structures - its efficient algorithm is built upon direct product structures. Together with its automatic determination of the residual term, this assumption can lead to surprising results.

Consider a balanced repeated measures structure with several missing values and correlation across time within subject $\boldsymbol{I}_N \otimes \boldsymbol{C}$.

- Then the size of covariance matrix $\boldsymbol{I}_N \otimes \boldsymbol{C} \neq$ the number of data values present, so this term cannot be used as residual matrix $\boldsymbol{R}$
- But, a model must have a residual term
- So an identity residual term is added (and specified in model summary)
- Two ways to deal with this
    - If lack of balance due to missing combinations: put these combinations into data set and use option [mvinclude=yvar] - good for few missing values
    - Explicitly specify extra residual term and fix component to value small enough to not affect rest of model, but not so small it destabilizes fitting process - better for many missing combinations

Random coefficient regression

Random coefficient regression (RCR) uses an individual model for each subject:

- Population mean regression model with subject variation allowed in regression coefficients
- Simplest version uses linear regression, may use higher order polynomials or other functions
- Does not require any form of balance in data
- Simple linear RCR model

$$y_{ij} = \mu + T_r + at_{ij} + b_r t_{ij} + u_i + v_i t_{ij} + e_{ij}$$

  - y_{ij} is jth measurement on subject i at time t_{ij}
  - $r = r(i)$ is the treatment combination allocated to subject i
  - $\mu + T_r$ fixed intercept for treatment r
  - $a + b_r$ fixed slope for treatment r
  - u_i, v_i random deviation in intercept, slope for subject i
  - e_{ij} residual = random variation about subject i linear trend

Random coefficient regression (2)

Matrix version of model for subject i:

$$\boldsymbol{y}_i = [\mathbf{1}\ \boldsymbol{t}_i] \begin{bmatrix} \mu + T_r \\ a + b_r \end{bmatrix} + [\mathbf{1}\ \boldsymbol{t}_i] \begin{bmatrix} u_i \\ v_i \end{bmatrix} + \boldsymbol{e}_i$$

where $\boldsymbol{y}_i$ is set of measurements for subject i taken at times $\boldsymbol{t}_i = (t_{i1} \ldots t_{in_i})'$.

Note: design matrix for fixed and random effects are identical at subject level (if treatment groups/covariates do not change over time).

To finish model, need to specify variance structure:

$$\text{var} \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \boldsymbol{C} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \ ; \quad \text{cov} \left(\begin{bmatrix} u_i \\ v_i \end{bmatrix}, \begin{bmatrix} u_j \\ v_j \end{bmatrix} \right) = \mathbf{0}$$

$$\text{var} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \boldsymbol{C} \otimes \boldsymbol{I}_N$$

$$\text{var}\,(\boldsymbol{e}_i) = \sigma^2 \boldsymbol{I}$$

Then

$$\text{var}\,(y_{ij}) = \sigma_{11} + 2t_{ij}\sigma_{12} + t_{ij}^2\sigma_{22} + \sigma^2.$$

RCR variance model

The covariance σ_{12} is an essential part of the model, required to make it invariant to translations in t.

For example consider the model with the time covariate centered about its mean t_μ:

$$\begin{aligned} y_{ij} &= \mu + T_r + a(t_{ij} - t_\mu) + b_r(t_{ij} - t_\mu) + u_i + v_i(t_{ij} - t_\mu) + e_{ij} \\ &= \mu^* + T_r^* + a t_{ij} + b_r t_{ij} + u_i + v_i(t_{ij} - t_\mu) + e_{ij} \end{aligned}$$

where $\mu^* = \mu - a t_\mu$, $T_r^* = T_r - b t_\mu$.

If the variance matrix of the random effects is written as

$$\text{var}\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \boldsymbol{C} = \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{12}^* & \sigma_{22}^* \end{pmatrix}$$

then

$$\begin{aligned} \text{var}\,(y_{ij}) &= \sigma_{11}^* + 2(t_{ij} - t_\mu)\sigma_{12}^* + (t_{ij} - t_\mu)^2\sigma_{22}^* + \sigma^2 \\ &= \sigma_{11}^* - 2t_\mu\sigma_{12}^* + t_\mu^2\sigma_{22}^* + 2(\sigma_{12}^* - \sigma_{22}^* t_\mu)t_{ij} + t_{ij}^2\sigma_{22}^* + \sigma^2 \end{aligned}$$

RCR variance model (2)

$$\begin{aligned} \operatorname{var}(y_{ij}) &= \sigma_{11}^* + 2(t_{ij} - t_\mu)\sigma_{12}^* + (t_{ij} - t_\mu)^2\sigma_{22}^* + \sigma^2 \\ &= \sigma_{11}^* - 2t_\mu\sigma_{12}^* + t_\mu^2\sigma_{22}^* + 2(\sigma_{12}^* - \sigma_{22}^* t_\mu)t_{ij} + t_{ij}^2\sigma_{22}^* + \sigma^2 \end{aligned}$$

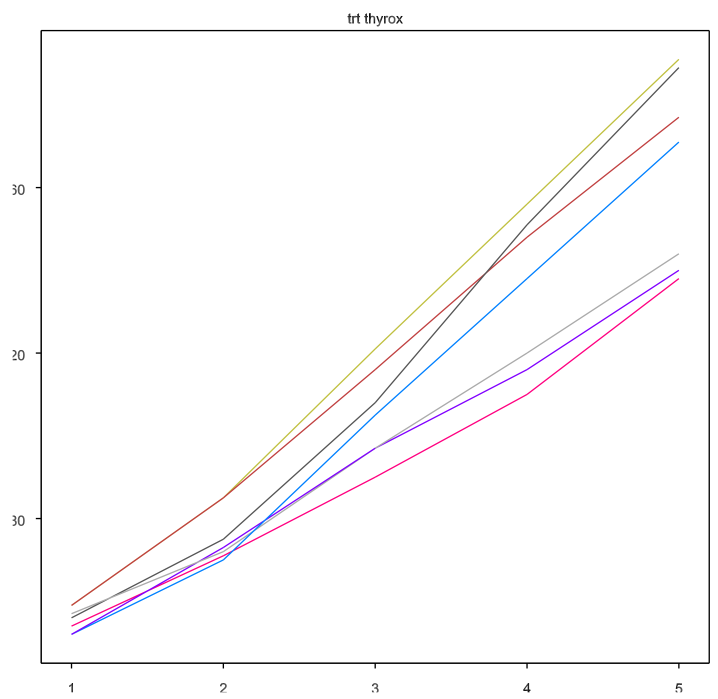
This is equivalent to the original model if

$$\sigma_{11} = \sigma_{11}^* - 2t_\mu\sigma_{12}^* + t_\mu^2\sigma_{22}^*$$
$$\sigma_{12} = \sigma_{12}^* - \sigma_{22}^* t_\mu$$
$$\sigma_{22} = \sigma_{22}^*$$

So a translation in t changes the form of the variance matrix - covariance parameter σ_{12} is required to keep the same model.

Example

- For rat data: little variation in intercept at $t = 1$, much variation at mean(t)
- Clear positive association between intercept and slope at mean(t), much less at $t = 1$



Example (2)

For rat data, fitted RCR with time covariate centered or not:

```
Estimated parameters for covariance models
------------------------------------------

Centered covariate: intercept at t=3 (correlation=0.78)

Random term(s)    Factor          Model(order)  Parameter  Estimate       s.e.
rat + rat.time    Across terms    Unstructured  v_11          4.352      1.496
                                                v_21          1.470      0.561
                                                v_22          0.8018    0.2964
                  Within terms    Identity      -                 -          -

No centering: intercept at t=0 (correlation=-0.11)

Random term(s)    Factor          Model(order)  Parameter  Estimate       s.e.
rat + rat.time    Across terms    Unstructured  v_11          1.678      0.749
                                                v_21        -0.1331     0.3049
                                                v_22         0.8018     0.2964
                  Within terms    Identity      -                 -          -
```

- big change in intercept variance and covariances
- with covariance, can transform between different scales $t - c$
- without covariance, model may be inadequate & interpretation may be wrong

Implicit variance model

RCR has implicit variance model

$$\mathrm{var}\left(\boldsymbol{y}_i\right) = \boldsymbol{X}_i \boldsymbol{C} \boldsymbol{X}_i' + \sigma^2 \boldsymbol{I}$$

where $\boldsymbol{X}_i = [\mathbf{1}\ \boldsymbol{t}_i]$, or

$$\mathrm{var}\left(y_{ij}\right) = \sigma_{11} + 2t_j\sigma_{12} + t_j^2\sigma_{22} + \sigma^2$$

- this is a parsimonious quadratic variance function in terms of t
- presence of covariance makes variance model more flexible
- but - it may not match variance pattern of data - validation of model important
- sometimes more appropriate to fit fixed polynomial regression plus general covariance model

Rat data

- quadratic + AD(2): -2RL=567.99, AIC=591.99, BIC=625.44
- quadratic RCR: -2RL=587.34, AIC=601.34, BIC=620.85

RCR and direct products

One appeal of the RCR is that it has a direct product variance structure

$$\text{var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix} = \boldsymbol{C} \otimes \boldsymbol{I}_N$$

so can be fitted efficiently.

Also, it is conceptually simple - a separate line/curve for each subject of the same form as the population profile.

Two ways of fitting RCR in GenStat:

1. Imposing correlation across terms directly

```
vcomp [fix=trt*time] rat+rat.time
vstruc [terms=rat+rat.time; corr=pos]
reml weight
```

2. Making composite term and imposing direct product structure

```
matrix [r=nval(weight); c=2] X
calc X$[*;1,2]=1,time
vcomp [fix=trt*time] rat.X
vstruc [term=rat.X] model=unstr; factor=X
reml weight
```

Computational issues

Covariance parameters (σ_{12}) in RCR can be difficult to estimate:

- especially where the number of subjects is small ($<$20)
- estimation tends to be more stable with centered covariates
- one strategy to get initial values
  - estimate subject intercept & slope parameters assuming no correlation
  - estimate correlation between intercepts & slopes directly, ie corr($\boldsymbol{u}$,$\boldsymbol{v}$)
  - use these estimates as starting values
  - works well for centered covariates

Exercise

Grizzle & Allen dog data

- Coronary sinus potassium concentrations measured on 36 dogs, divided between 4 different treatment groups
- Seven measurements were made on each dog, every 2 minutes from 1 to 13 minutes after an event (occlusion)
- Aim of this analysis is to quantify the difference in profiles between treatments: a good model is therefore required for both treatment means and within-subject variation
- Data are in spreadsheet dog.xls
- Investigate different analyses for this data & decide on the most appropriate approach

References

Brien CJ & Bailey RA (2006) Multiple randomizations. *Journal of the Royal Statistical Society, Series B*, **68**, 571-599.

Fitzmaurice G, Davidian, M, Molenberghs G, & Verbeke G. (2008) *Longitudinal Data Analysis* Chapman & Hall, CRC Press.

Gilmour A, Cullis B, Welham SJ, Gogel BJ, Thompson R (2004) An efficient computing strategy for prediction in mixed linear models. *Computational Statistics and Data Analysis*, **44**, 571-586.

Haslett, J. & Dillane, D. (2004) Application of 'delete=replace' to deletion diagnostics for variance component estimation in the linear mixed model. *Journal of the Royal Statistical Society, Series B* **66**, 131–143.

Verbeke G & Molenberghs G (2000) *Linear Mixed Models for Longitudinal Data*, Springer, New York.