Index of Lecture 6–L

— in addition to overheads by Davis (2002):
248–53,260–81,291–305 (+ 199,209,227,246–247 from 5–L)

Practical Information

Today's lecture: Chapter 6 in Davis (2002),

- focus on modelling (choice of covariance structure) and estimation procedure,
- brief review of example from last session,
- two new, extensive examples,
- detailed discussion of 3 advanced topics:
  - random slopes modelling (and some caveats),
  - effect of choice of covariance structure on fixed effects,
  - performance comparison of methods,
- R and SAS code at webpage, but discussion of R programming only in next lecture,
- no review of material in McCulloch & Searle (theoretical).

Estimation in Linear Mixed Models

"Likelihood"-based estimation assuming normal distributions for all random terms:

- REML (restricted/residual maximum likelihood) or ML:
  - theoretical properties differ slightly,[1]
  - REML estimates agree with ANOVA-type[2] estimates for balanced data,
  - in practice often only minor differences, but preference generally in favour of REML,
- iterative, numerically robust algorithms[3] that perform well for both balanced and unbalanced data,
- available in many statistical software packages (SAS, S-Plus/R, SPSS, MLwiN, HLM...), that should give same results up to numerical accuracy, except in "difficult" problems, e.g.,
  - large number of covariance parameters,
  - large number of random effects.

[1] REML estimates of variance parameters less biased than ML estimates (biased towards zero); REML estimates adjusted for degrees of freedom of fixed effects model; REML estimates unaffected by β but not by X.

[2] A classical statistical method for variance components models, relies on the ANOVA-table and constructs estimates and test statistics from the MS–column; performs well only for balanced or almost balanced data.

[3] Newton-Raphson or Fisher scoring optimization for "easy" problems, and EM (expectation-maximization) algorithm for "difficult" problems.

Model comparison, AIC and BIC

Primary statistical tool for model comparison
= statistical test of "reduced" model against "full" model;
in linear mixed models,

- Wald test (fixed effects only) or likelihood-ratio test,
- when using REML estimation, likelihood-ratio tests only meaningful when models have same fixed effects.

Secondary tools: AIC/BIC
— for situations where tests don't work (well):

- models to be compared not nested,
- vast amount of data "make everything significant",
  - AIC/BIC → more parsimonious models than tests,
  - BIC more restrictive than AIC.

AIC vs. BIC:

> "I prefer to use AIC with the slight variation that models that are within two units of the lowest AIC are considered to be competitive models for the best. From the competitive models, the one with the fewest parameters is usually selected." (Davis/Jones)

Henrik's view: for non-Bayesian analyses, the AIC seems the most logical choice.

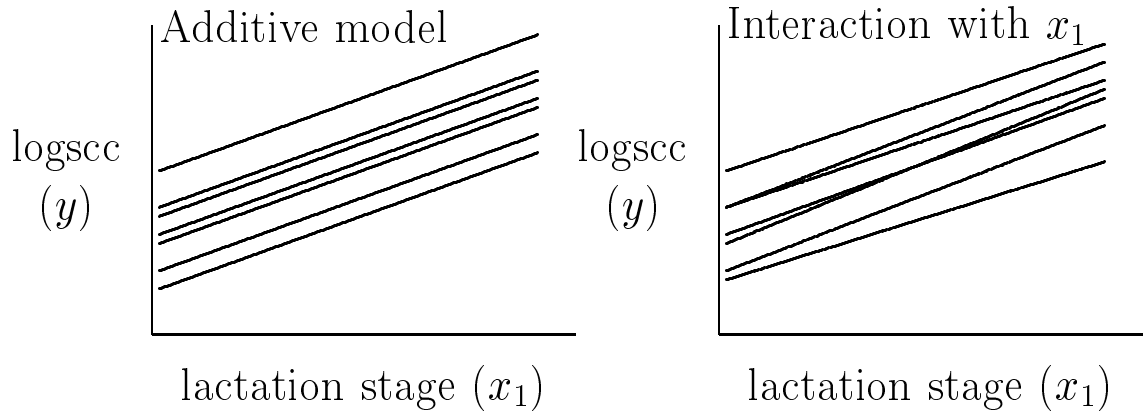
Additivity in Linear Mixed Models

Example: repeated measurements of somatic cell counts on cows (in herds) $\sim$ 3-level *random intercepts* model,

$$y_{hij} = \beta_0 + \beta_1 x_{1hij} + \ldots + \beta_k x_{khij} + u_h + v_i + \varepsilon_{hij}, \quad (1)$$

where $h \sim$ herd, $i \sim$ cow, $j \sim$ measurement, and the predictor x_1 is quantitative, e.g. lactation stage (days in milk).

Additivity (no interactions) implies parallel regression lines (left figure) where different lines correspond to different levels of other variables, e.g. breeds, parity or herds.



Model assumption (additivity):

same impact of x_1 for all groups of cows (breeds. . .)

— not always biologically evident!

Alternative assumption (right figure) = interaction[4]:

different slopes for some different groups of animals, like breeds or parities

— maybe more plausible, but requires good data.

[4] Interaction between x_1 and a categorical predictor.

Random Slope Model

In the example of the previous page, what if x_1-slopes differ (also/instead) between herds, that is, the impacts of different lactation stages is not constant across herds?
$\Rightarrow$ *random slopes model*,

$y_{hij} = \beta_0 + \beta_1 x_{1hij} + \ldots + \beta_k x_{khij} + u_h + b_{1h} x_{1hij} + v_i + \varepsilon_{hij}$,
where $b_{1h} \sim N(0, \sigma_1^2)$, and other variables unchanged from (1).

First details:

- σ_1 interpretable as amount of random fluctuation in x_1-slopes between herds; roughly slopes lie (with prob. 95%) within $\beta_1 \pm 1.96\sigma_1 \approx \beta_1 \pm 2\sigma_1$,
- random slopes usually only at levels above x_1's level.

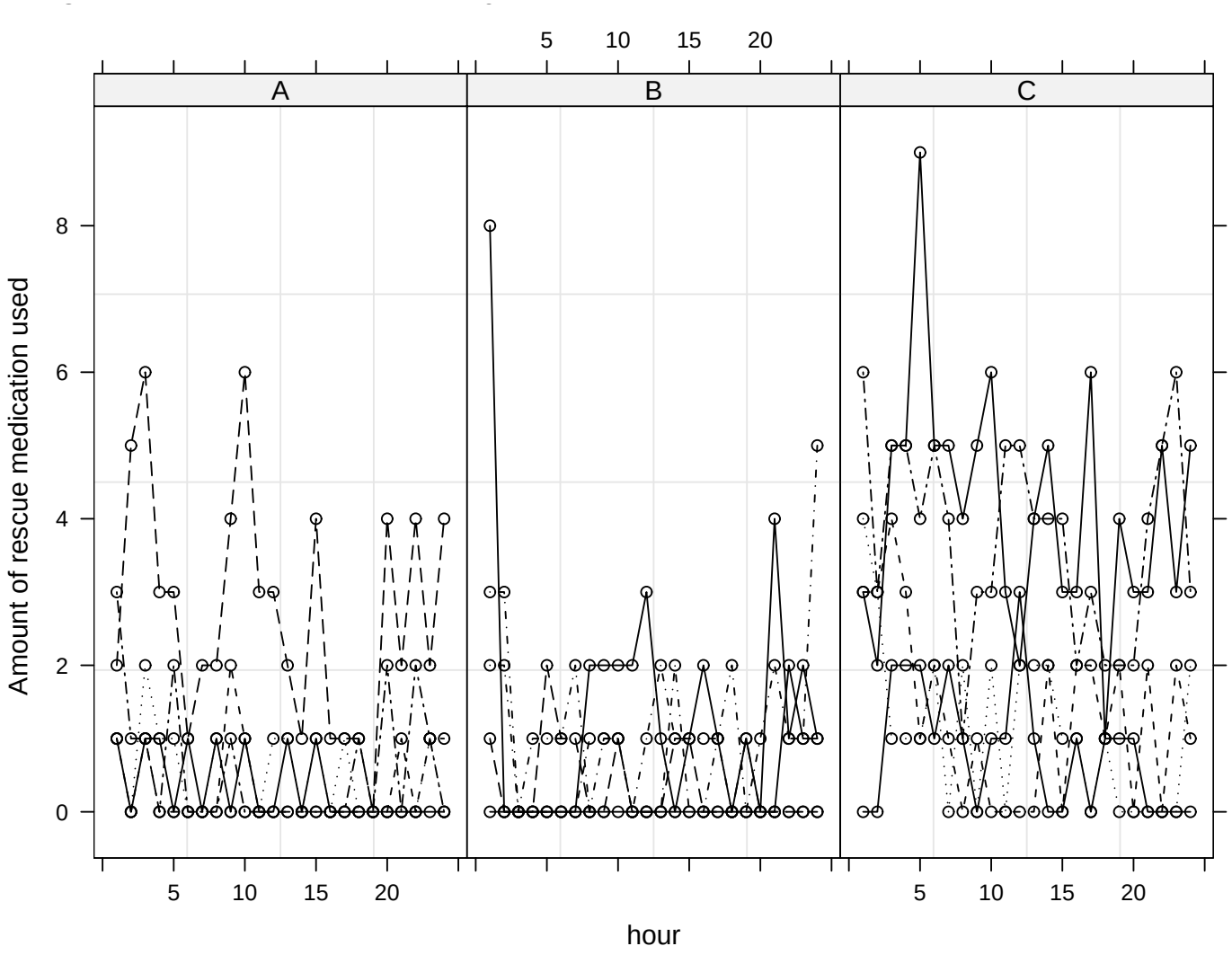
More (technical) details:

- b_{1h}'s enter into the random variation of the model:
  $\mathrm{Var}(y_{hij}) = \sigma_u^2 + x_1^2\sigma_1^2 + 2x_1\sigma_{01} + \sigma_v^2 + \sigma^2, \quad x_1 = x_{1hij}$,
  where b_{1h} are assumed independent of the v_i's and ε_{hij}'s, but possibly dependent on u_h with $\mathrm{Cov}(b_{1h}, u_h) = \sigma_{01}$,[5]
- variances (and covariances) no longer constant (depend on value of x_1) $\Rightarrow$
  - variance components more difficult to interpret,
  - model implicitly models variance heterogeneity.

[5] The b_{1h} and u_h are often negatively correlated, like slope and intercept in a simple regression.

Graph for Rescue Medication Data

Profile plot for 5 subjects per group:



Summary of Rescue Medication Example

Conclusions (textbook):

- autoregressive structure, ar(1), does not fit data well: model correlations die down too quickly,
- significance of treatment effect affected by choice of correlation structure,
- random slopes model very flexible model for variances and correlations,
- random intercepts plus autoregressive errors best structure in terms of AIC and BIC.

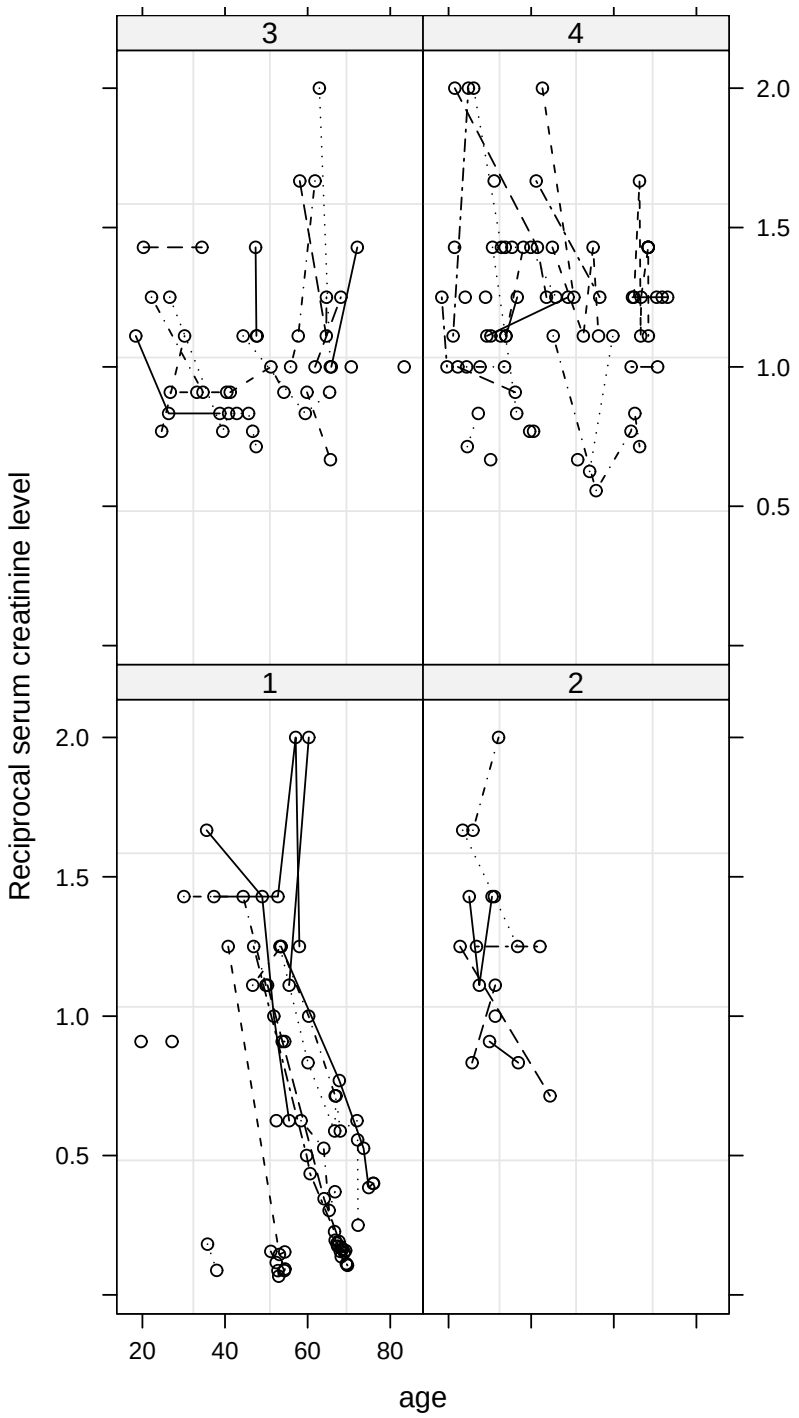
Comparison of models with difference covariance structures (REML estimation, likelihood-ratio statistics G^2):

				test against larger model			
No.	Model	$-2 \log L$	#param.	No.	G^2	df	P
1	ar(1)	9880.35	7	1	336.3	1	<.0001
2	ar(1)+random int.	9544.09	8	5	70.9	1	<.0001
3	random slopes	9589.82	9	7	102.1	1	<.0001
4	arma(1,1)	9494.57	8	5	21.4	1	<.0001
5	arma(1,1)+random int.	9473.19	9	8	14.2	2	.0004
6	ar(1)+random int.+het.var.	9483.44	31	–	–	–	–
7	ar(1)+random slopes	9487.71	10	8	28.8	1	<.0001
8	arma(1,1)+random slopes	9459.04	11	–	–	–	–

- evidence for more complicated correlation structures,
- model 8 is best (also most complex),
- significance for treatments essentially unaffected.

Graph for Serum Creatinine Data

Profile plot for 73 (18,8,18,29) subjects:



Summary of Serum Creatinine Example

Modelling variance/correlation structure:

- difficult, due to long and inhomogeneous series ($\Rightarrow$ correlation matrix structures little useful),
- continuous time structures (e.g., autoregressive (exponential[6]), power, Gaussian):
  - explicit variance functions with few parameters,
  - difficult to check, and at best approximate modelling,
  - developed for spatial models but apply also for time,
  - "observational error" $\sim$ "nugget effect",
- random slopes may be used to model variance/correlation inhomogeneity over time.

Results:

- differences in (some) fits between R and SAS software,
- strong fluctuations in variance/covariance parameters between different covariance models $\Rightarrow$ parameters difficult to interpret,
- considerable fluctuations in estimates and P-values between different covariance models $\Rightarrow$ conclusions partly rely on appropriately chosen covariance structure,
- references given for further discussion/diagnostics.

[6] The relation between autoregressive $\rho(\tau) = \rho^{\tau}$ and exponential $\rho(\tau) = \exp(-\phi\tau)$ parameters is $\phi = -\ln(\rho)$ or $\rho = \exp(-\phi)$.