

Linear Mixed Models

Workshop: Analysis of Longitudinal Data

12th Nov 2024

Jaroslav Harezlak

Armando Teixeira-Pinto



Linear Mixed Effects Model

A mixed model contains both fixed and random effects.

Mixed models for longitudinal data explicitly identify

- individual (random effects)
- population characteristics (fixed effects).

Mixed models are very flexible since they can accommodate any degree of imbalance in the data. That is, we do not necessarily require the same number of observations on each subject or that the measurements be taken at the same times.

Also, the use of random effects allows us to model the covariance structure as a continuous function of time.

Compound Symmetry Model

Correlation between repeated measurements arises because each subject has an underlying level of response that persists over time.

This **subject effect** is treated as random and the mixed model is

$$Y_{ij} = X_{ij}\beta + b_i + e_{ij}$$

$$Y_{ij} = (\beta_0 + b_i) + \beta_1 X_{ij1} + \dots \beta_p X_{ijp} + e_{ij}$$

The response for the i^{th} subject is assumed to differ from the population mean, $X_{ij}\beta$ by
a subject effect, b_i ,
and
a within-subject measurement error, e_{ij} .

Random Intercepts Model

If $\text{var}(b_i) = \sigma_b^2$ and $\text{var}(e_{ij}) = \sigma^2$, the covariance matrix of the repeated measurements has the compound symmetry form:

$$\begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \sigma_b^2 & \dots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma^2 & \dots & \sigma_b^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \dots & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

Note: The introduction of a random subject effect, b_i , induces correlation among the repeated measurements.

The compound symmetry model is the simplest possible example of a mixed model.

However, we can easily generalize these ideas.

Random Intercepts and Slopes Model

Consider the following model with intercepts and slopes that vary randomly among subjects.

For the i^{th} subject at the j^{th} measurement occasion

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + e_{ij}$$

(Note: we are using double subscripting here)

Linear Mixed Model

Notation: Suppose we have n individuals on which we have collected p_i repeated observations at times t_{ij} .

Consider the mixed model

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

where $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of fixed effects;

\mathbf{b}_i is a $(q \times 1)$ vector of random effects and $\mathbf{b}_i \sim N(0, G)$;

\mathbf{X}_i is a $(p_i \times k)$ matrix of covariates;

\mathbf{Z}_i is a $(p_i \times q)$ matrix of covariates (usually the columns of \mathbf{Z}_i are a subset of the columns of \mathbf{X}_i and $q < k$);

\mathbf{e}_i is a $(p_i \times 1)$ vector of errors and $\mathbf{e}_i \sim N(0, R_i)$.

Example

Consider again the random intercepts and slopes model:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + e_{ij}$$

In matrix form this can be represented as

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

where

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & t_{ip_i} \end{bmatrix}$$

Example (cont'd)

Let $\text{var}(b_{i0}) = g_{11}$, $\text{var}(b_{i1}) = g_{22}$, and $\text{cov}(b_{i0}, b_{i1}) = g_{12}$.

These are the three unique elements of the (2×2) covariance matrix, G .

We assume that $\text{var}(e_{ij}) = \sigma^2$. Thus, $\mathbf{R}_i = \sigma^2 \mathbf{I}$.

Then it can be shown that

$$\text{var}(Y_{ij}) = g_{11} + 2t_{ij}g_{12} + g_{22}t_{ij}^2 + \sigma^2$$

And

$$\text{cov}(Y_{ij}, Y_{ik}) = g_{11} + (t_{ij} + t_{ik})g_{12} + g_{22}t_{ij}t_{ik}$$

Note: The covariance is expressed as a function of time.

Covariate effects (e.g. due to treatment) can be expressed by allowing the mean values of the intercept and slope to depend upon the covariates (e.g. by allowing them to differ across the treatment groups).

In the mixed model

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

$\mathbf{R}_i = \text{var}(\mathbf{e}_i)$ describes the covariance among observations when we focus on the response profile of a specific individual.

That is, it is the covariance of the i th subject's deviations from his/her mean profile, $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$.

Usually, it is assumed that $\mathbf{R}_i = \sigma^2\mathbf{I}$, where \mathbf{I} is a $(p_i \times p_i)$ identity matrix

‘conditional independence assumption’

Alternatively, a structured model for \mathbf{R}_i could be assumed, e.g. AR(1).

In the mixed model

$$Y_i = X_i\beta + Z_ib_i + e_i$$

The vector of regression parameters β contains the *fixed effects*, which are assumed to be the same for all individuals.

These regression parameters have population-averaged interpretation (e.g. in terms of changes in the mean response, averaged over individuals).

Although the *conditional* mean of Y_i , given b_i , is

$$E(Y_i | b_i) = X_i\beta + Z_ib_i$$

note that the *marginal* or population-averaged mean of Y_i is

$$E(Y_i) = X_i\beta$$

Interpretation of the coefficients

In contrast to β , the vector \mathbf{b}_i is comprised of subject-specific regression coefficients.

These are the random effects and the \mathbf{b}_i have a distribution (usually, but not necessarily, assumed to be normal).

Combined with the fixed effects, these describe the mean response profile of a specific individual.

That is, the mean response profile for the i^{th} individual is

$$\mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i$$

Technical details

In the mixed model

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

recall that

$$E(\mathbf{Y}_{ij} | \mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

and

$$E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

Similarly,

$$\text{var}(\mathbf{Y}_{ij} | \mathbf{b}_i) = \text{var}(\mathbf{e}_i) = \mathbf{R}_i$$

and

$$\begin{aligned}\text{var}(\mathbf{Y}_i) &= \text{var}(\mathbf{Z}_i\mathbf{b}_i) + \text{var}(\mathbf{e}_i) \\ &= \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i\end{aligned}$$

Of note, even if $\mathbf{R}_i = \sigma^2\mathbf{I}$,

$$\text{var}(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma^2\mathbf{I}$$

is not a diagonal matrix.

Technical details: correlation

Thus, the introduction of random effects, \mathbf{b}_i , induces correlation (marginally) among the \mathbf{Y}_i .

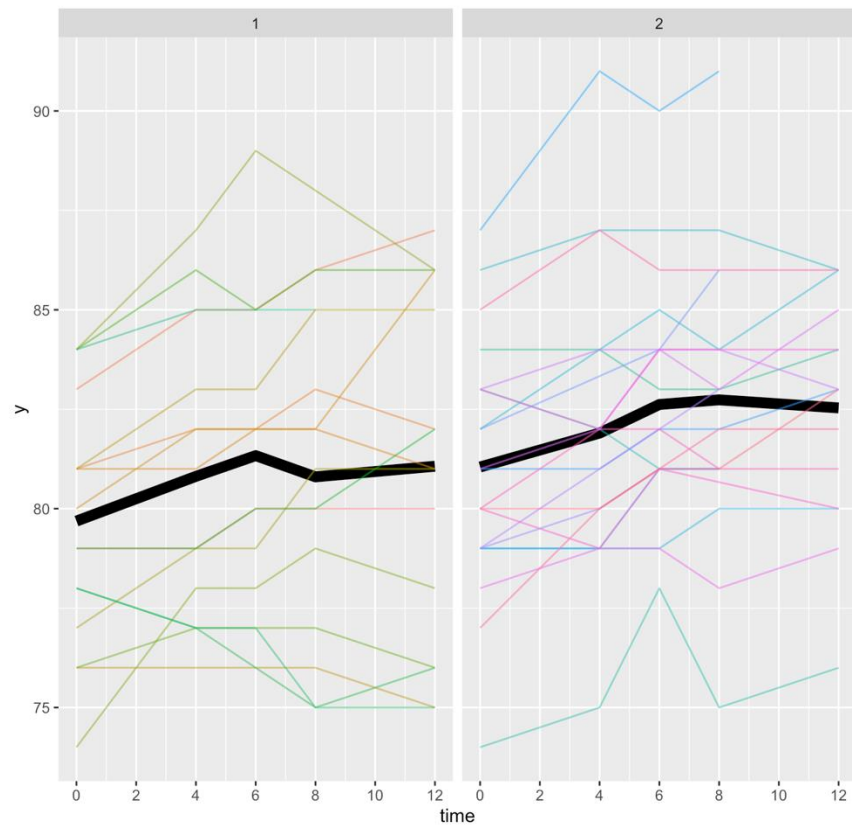
That is,

$$\text{var}(\mathbf{Y}_i) = \Sigma_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$$

which, in general, has non-zero off-diagonal elements.

Finally, note that $\text{var}(\mathbf{Y}_i)$ is described in terms of a set of covariance parameters, some defining \mathbf{G} and some defining \mathbf{R}_i .

Example: strength study



Example: strength study

Consider a model with intercepts and slopes that vary randomly among subjects, and which allows the mean values of the intercept and slope to differ in the two treatment groups.

To fit this model, use the following R code:

```
stren.lme.int.slp <- lme(y~time*trt,  
                        random=~1+time|factor(id),  
                        data=stren.uni.sel)  
summary(stren.lme.int.slp)
```

Random Intercepts and Slopes Model

Linear mixed-effects model fit by REML

Data: stren.uni.sel

	AIC	BIC	logLik
	647.98	673.0192	-315.99

Random effects:

Formula: ~1 + time | factor(id)

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	3.0898360	(Intr)
time	0.1632683	0.106
Residual	0.8283737	

Fixed effects: y ~ time * trt

	Value	Std.Error	DF	t-value	p-value
(Intercept)	78.76993	1.7256244	134	45.64721	0.0000
time	0.09743	0.1037779	134	0.93880	0.3495
trt	1.23486	1.0500206	35	1.17603	0.2475
time:trt	0.03774	0.0636998	134	0.59254	0.5545

Recall:

$$\text{var}(\mathbf{Y}_i) = \text{var}(\mathbf{Z}_i \mathbf{b}_i) + \text{var}(\mathbf{e}_i)$$

$$= \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$$

Given estimates of \mathbf{G} :

$$\begin{bmatrix} 9.547 & 0.053 \\ 0.053 & 0.027 \end{bmatrix}$$

and of $\mathbf{R}_i = \sigma^2 \mathbf{I} = (0.6862) \mathbf{I}$, and with

$$\mathbf{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 12 \end{bmatrix}$$

We can obtain the following estimate of $\text{var}(Y_i)$:

$$\begin{bmatrix} 10.23 & 9.76 & 9.87 & 9.97 & 10.19 \\ 9.76 & 11.09 & 10.72 & 11.04 & 11.68 \\ 9.87 & 10.72 & 11.83 & 11.57 & 12.43 \\ 9.97 & 11.04 & 11.57 & 12.79 & 13.17 \\ 10.19 & 11.68 & 12.43 & 13.17 & 15.35 \end{bmatrix}$$

The corresponding correlation matrix is:

$$\begin{bmatrix} 1.00 & 0.92 & 0.90 & 0.87 & 0.81 \\ 0.92 & 1.00 & 0.94 & 0.93 & 0.90 \\ 0.90 & 0.94 & 1.00 & 0.94 & 0.92 \\ 0.87 & 0.93 & 0.94 & 1.00 & 0.94 \\ 0.81 & 0.90 & 0.92 & 0.94 & 1.00 \end{bmatrix}$$

Next, consider the model with random intercepts only (equivalent to compound symmetry).

To fit this model, use the following R code:

```
stren.lme.int <- lme(y~time*trt,  
                    random=~1|factor(id) ,  
                    data=stren.uni.sel)  
summary(stren.lme.int)
```

Alternatively, we could fit this model by specifying a compound symmetry model for \mathbf{R}_i and assume no random effects:

```
stren.cs <- gls(y~time*trt,  
               correlation=corCompSymm(form= ~time | factor(id)),  
               data=stren.uni.sel)
```

Random Intercepts Model

Linear mixed-effects model fit by REML

Data: stren.uni.sel

	AIC	BIC	logLik
	672.4098	691.1892	-330.2049

Random effects:

Formula: ~1 | factor(id)

(Intercept) Residual

StdDev: 3.294046 1.076052

Fixed effects: y ~ time * trt

	Value	Std.Error	DF	t-value	p-value
(Intercept)	78.72862	1.8594504	134	42.33972	0.0000
time	0.10709	0.0678085	134	1.57935	0.1166
trt	1.28046	1.1314239	35	1.13172	0.2654
time:trt	0.02669	0.0421237	134	0.63366	0.5274

Number of Observations: 173

Number of Groups: 37

Model comparison

However, in the current setting,

	-2 Res Log L
Random Intercepts	660.4
Random Intercepts & Slopes	632.0

$\Rightarrow 2 \times \text{Res log likelihood ratio} = 28.4, 2 \text{ d.f. } (p < 0.0001)$

So, in this case, there is no doubt that the random Intercepts (or compound symmetry) model is not defensible.

For technical reasons to be explained later in this lecture, the chi-square distribution only provides an approximation to the distribution of the likelihood ratio statistic in this instance. We will revisit this test later when we discuss AIC.

Prediction of Random Effects

In most applications, inference is focused on the fixed effects, β .

However, in some studies we may want to predict (or “estimate”) subject-specific response profiles.

Technically, because the \mathbf{b}_i are random, we customarily talk of “predicting” the random effects rather than “estimating” them.

Using maximum likelihood, the prediction of \mathbf{b}_i is given by :

$$\hat{b}_i = \mathbf{GZ}_i' \Sigma_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta})$$

where $\Sigma_i = \text{var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i' + \mathbf{R}_i$.

This is known as the Best Linear Unbiased Predictor (or BLUP).

Prediction of Random Effects

When the unknown covariance parameters are replaced by their ML or REML estimates, the resulting predictor,

$$\hat{b}_i = \hat{G}Z_i'\hat{\Sigma}_i^{-1}(Y_i - X_i\hat{\beta})$$

is often referred to as the “Empirical BLUP” or the “Empirical Bayes” (EB) estimator.

Furthermore, an expression for the variance of this estimate can be derived as a function of the covariates and the estimates of the covariance parameters. Hence, we can estimate the regression coefficients for each individual, as well as their standard errors, confidence intervals.

Prediction of Random Effects

Finally, the i^{th} subject's predicted response profile is estimated by,

$$\hat{Y}_i = X_i \hat{\beta} + Z_i \hat{b}_i$$

$$\hat{Y}_i = X_i \hat{\beta} + Z_i \hat{G} Z_i' \hat{\Sigma}_i^{-1} (Y_i - X_i \hat{\beta})$$

$$\hat{Y}_i = (\hat{R}_i \hat{\Sigma}_i^{-1}) X_i \hat{\beta} + (I - \hat{R}_i \hat{\Sigma}_i^{-1}) Y_i$$

That is, the i^{th} subject's predicted response profile is a weighted combination of the population-averaged mean response profile, $X_i \beta$, and the i^{th} subject's observed response profile, Y_i .

Prediction of Random Effects

Note that the subject's predicted response profile is “shrunk” towards the population-averaged mean response profile.

The amount of “shrinkage” depends on the relative magnitude of \mathbf{R}_i and Σ_i .

\mathbf{R}_i characterizes the within-subject variability, while Σ_i incorporates both within-subject and between-subject sources of variability.

As a result, when \mathbf{R}_i is “large”, and the within-subject variability is greater than the between-subject variability, more weight is given to $\mathbf{X}_i\beta$, the population-averaged mean response profile.

When the between-subject variability is greater than the within-subject variability, more weight is given to the i^{th} subject's observed data \mathbf{Y}_i .

R Code

The Empirical Bayes (EB) estimates, \hat{b}_i , can be obtained by using the following extractor from the `lme()` object:

```
stren.lme.int.slp$coef$random
```

Alternatively, a subject's predicted response profile,

$$\hat{Y}_i = X_i \hat{\beta} + Z_i \hat{b}_i$$

can be obtained by using the following code

```
pred.val <- predict(stren.lme.int.slp)
```

Example: Exercise Therapy Study

Consider a model with intercepts and slopes that vary randomly among subjects, and that allows the mean values of the intercept and slope to differ in the two treatment groups.

To fit this model, use the following R code:

```
stren.lme.int.slp <- lme(y~time*trt,  
                        random=~1+time|factor(id),  
                        data=stren.uni.sel)  
  
summary(stren.lme.int.slp)
```

Empirical Bayes Estimates of b_i

Solution for Random Effects

```
stren.lme.int.slp$coef$random  
$`factor(id)`
```

	(Intercept)	time
1	-1.01112331	-0.0318196571
2	3.37720597	0.1604247683
3	1.28109894	-0.0193501569
4	1.02810647	-0.0768505069
5	0.16950818	0.2365153095

.....

Example: Exercise Therapy Study

Next, we consider how to obtain a subject's predicted response profile.

```
predict(stren.lme.int.slp)

stren.lme.int.slp.pred <- cbind(stren.uni.sel,
predict(stren.lme.int.slp))
names(stren.lme.int.slp.pred)[6] <- "pred.int.slp"

stren.lme.int.slp.pred <- cbind(stren.uni.sel,
                                predict(stren.lme.int.slp),
                                predict(stren.lme.int.slp, level = 0))
names(stren.lme.int.slp.pred)[6:7] <-
  c("pred.int.slp", "pred.mean")
```

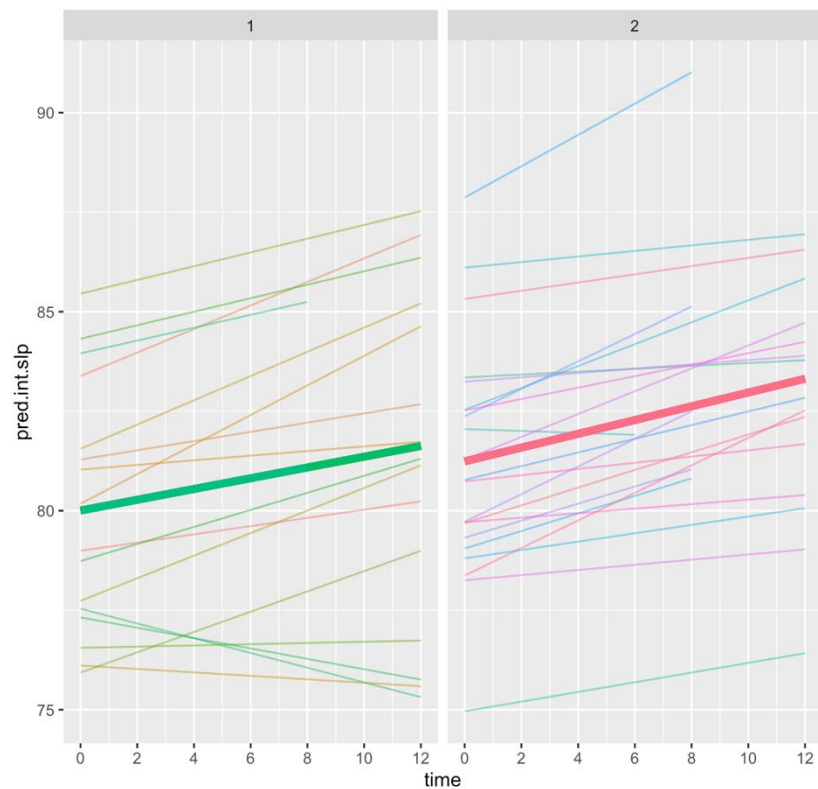
Predicted Response Profiles

```
head(stren.lme.int.slp.pred)
  id trt  y time time.cat pred.int.slp pred.mean
1  1   1 79    0         1    78.99367   80.00479
3  1   1 79    4         3    79.40707   80.54547
4  1   1 80    6         4    79.61377   80.81582
5  1   1 80    8         5    79.82048   81.08616
7  1   1 80   12         7    80.23388   81.62684
8  2   1 83    0         1    83.38199   80.00479
9  ...
```

Predicted subject-specific lines

```
ggplot(data = stren.lme.int.slp.pred,  
       aes(x=time, y=pred.int.slp, group=id,color=factor(id))) +  
  scale_x_continuous(breaks=2*c(0:6)) + #x-axis 0,2,4,...,12  
  geom_line(show.legend = F, alpha=.5) +  
  facet_grid(cols=vars(trt)) +  
  geom_line(aes(y=pred.mean), lwd=2, show.legend=F)
```

Predicted subject-specific lines



Selecting a Model for Covariance

For a given linear model, how can we decide which covariance model to use in the “final analysis?”

There are two general approaches for comparing models for the covariance matrix:

1. Likelihood ratio tests base on the REML likelihood when the models are nested.
2. Information criteria when they are not nested:

Akaike's Information Criterion (AIC)

Schwarz's Bayesian Information Criterion (BIC)

Comparing Models for The Covariance

- The REML likelihood provides a measure of the goodness of fit of an assumed model for the covariance.
- We have described the likelihood-ratio test based on the REML likelihood as the standard approach for comparing two nested models for the covariance matrix.
- Take twice the difference in maximized log likelihoods and compare this value to the chi-squared distribution (with df equal to the difference in number of covariance parameters).
- However, even when two models are nested, the likelihood ratio test is not valid when the null hypothesis implies that a variance must equal 0. The problem arises because the distribution theory justifying the likelihood ratio test is only valid when the null hypothesis is a point on the interior of the parameter space.

Illustration

Suppose the model is

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{i0} + \varepsilon_{ij}; \quad \text{var}(b_{i0}) = g_{11}.$$

In that case, the likelihood ratio test comparing the model with $g_{11} = 0$ (independence) to the model with $g_{11} > 0$ (compound symmetry), does not have a chi-squared distribution with 1 df.

Instead, it is an equally-weighted mixture of chi-squared distributions with 0 and 1 df.

Similarly, to test

H_0 : random intercepts model

H_0 : random intercepts and slopes model,

The asymptotic null distribution of the standard LRT is not a chi-squared distribution with 2 df.

Instead, it is an equally-weighted mixture of chi-squared distributions with 1 and 2 df.

If the classical null distribution is used instead, the resulting p-value will be too large \Rightarrow The test will tend to select the simpler model even when it is not valid.

What Should We Do

Though it is possible in some instances to derive the distribution of the LR statistic in this situation, a simple solution is proposed when this problem arises, namely, use a larger P value as the standard for rejecting the null hypothesis. Specifically, test the null hypothesis at a significance level of 0.10. This will guard against the tendency to select an overly simple model.

Here, we will recommend a single unified approach, the Akaike Information Criterion, for the comparison of a set of models, including non-nested models and models with parameters on the boundary. We shall see model selection based on AIC, combined with the use of empirical variances, provide robust methods for choosing regression models and making inferences about their coefficients.

Comparing Non-Nested, or Non-standard Nested, Models for the Covariance

When models (for the mean or covariance matrix) are not nested, or when hypotheses are on the boundary, we can use *Information Criteria* to compare them. These IC's effectively extract a penalty for the estimation of each additional parameter.

The two most widely used criteria are Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).

Akaike Information Criterion (AIC) is defined as

$$AIC = -2 * \log L + 2 * c$$

where $\log L$ is the maximized REML log likelihood and c is the number of covariance parameters.

AIC can be used to compare any pair of models with the same fixed effects but different covariance structures. The model with the smallest AIC is deemed best.

Comparing Non-Nested, or Non-standard Nested, Models for the Covariance

Schwarz's Bayesian Information Criterion (BIC) is defined as

$$\text{BIC} = -2 * \text{Log } L - c * \ln(n^{\#})$$

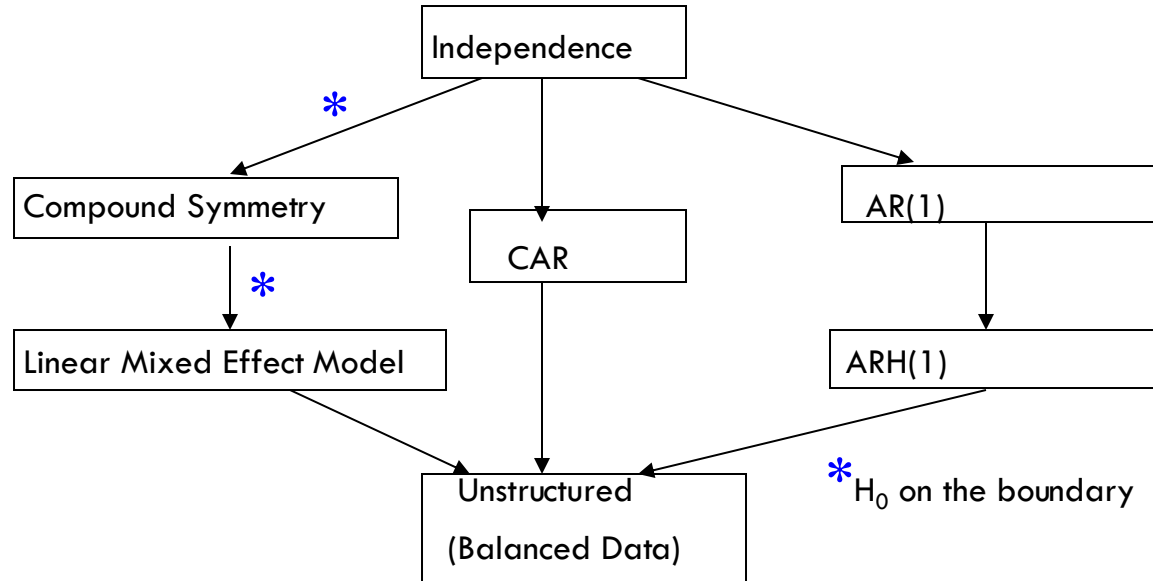
where $\log L$ is either the maximized ML or REML log likelihood, c is the number of covariance parameters, and $n^{\#}$ is the number of effective subjects, n , in the case of ML and $n - k$ in the case of REML estimation (where k is the number of regression coefficients).

When $n^{\#}$ is relatively large, BIC extracts a substantial penalty for the estimation of each additional parameter.

In general, comparing non-nested models using BIC entails a high risk of selecting a model that is too simple for the data at hand. Hence, we recommend consistent use of AIC in the analysis of longitudinal data.

Comparing Models for the Covariance

Note that the models for the covariance structure have the following hierarchical relationship:



Comparing Models for the Covariance

Thus, AR(1) and CAR have hierarchical relationships to independence and unstructured models, but not to the CS or mixed effects models.

Since the exponential correlation model and the random intercepts and slopes model do not have a hierarchical relationship (i.e. they are not nested models), they cannot be compared by a likelihood ratio test.

Instead, we can compare the models in terms of their AIC.

Exercise Therapy Study

Consider a saturated model for the mean response, i.e., time (categorical), treatment, and time*treatment. We can fit a variety of models for the covariance. R gives the following AIC's (smaller is better):

Model	AIC
CS (Random Intercepts)	672.41
AR(1)	648.36
CARH(1)	637.39
CAR(1)	642.55
Random Intercepts and Slopes	647.98
Unstructured	649.96

On the basis of AIC, CARH(1) is the preferred covariance structure, though several other models are close competitors. The random effects models are not competitive.