The Multilevel Change Model

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

GCM, 2010

The Multilevel Change Model

Introduction

In this lecture, we introduce the general multilevel model for repeated measurements, and illustrate it with a simple example.

The General Polynomial Growth Model – Level 1

Raudenbush and Bryk (2002, Chapter 6) describe a general polynomial model for analyzing growth data. An individual i's score at time t is a polynomial (of order P) function of a_{ti}, the age at time t. We will modify the Raudenbush-Bryk notation slightly to agree more closely with the notation in Singer and Willett. Here is the level-1 model.

$$Y_{ti} = \pi_{0i} + \pi_{1i}\, a_{ti} + \pi_{2i}\, a_{ti}^2 + \ldots + \pi_{Pi}\, a_{ti}^P + \epsilon_{ti} \qquad (1)$$

Each person is observed on T_i occasions. (Note that the number and spacing of measurements may vary across persons.) The multivariate distribution of the ϵ_{ti} may be modeled in various ways, to allow for correlation between the measurements across time.

The General Polynomial Growth Model – Level 2

The growth parameters in Equation 1 are free to vary across individuals. The P+1 parameters are modeled at level 2 as

$$\pi_{pi} = \gamma_{p0} + \sum_{q=1}^{Q_p} \gamma_{pq} X_{qi} + \zeta_{pi} \tag{2}$$

where X_{qi} is either a measured characteristic of the individual or a treatment, and ζ_{pi} is a random effect with mean 0. The set of $P + 1$ random effects is assumed to have a multivariate normal distribution with covariance matrix T.

A Linear Growth Model

When the number of observations per individual is small, we find it both convenient and necessary to employ a linear model. In that case, the level-1 equation 1 simplifies to

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{ti} + \epsilon_{ti} \tag{3}$$

and the level-2 equation 2 simplifies to

$$\begin{aligned} \pi_{0i} &= \gamma_{00} + \sum_{q=1}^{Q_0} \gamma_{0q} X_{qi} + \zeta_{0i} \\ \pi_{1i} &= \gamma_{10} + \sum_{q=1}^{Q_1} \gamma_{1q} X_{qi} + \zeta_{1i} \end{aligned} \tag{4}$$

An Example — Alcohol Use among Teenagers

Curran, Stice, and Chassin (1997, *Journal of Consulting and Clinical Psychology*, p. 130) studied longitudinal progression of alcohol use in 82 adolescents. . .

- Three waves of data were gathered, which included a 4-item questionnaire measuring extent of alcohol use
- There were two level-2 predictors, *COA* (child of an alcoholic) and *PEER* (a measure of peer group alcohol use)
- As described in the text, a square root transformation was applied to the data to generate the *PEER* and *ALCUSE* data to enhance linearity.

An Example — Alcohol Use among Teenagers

Curran, Stice, and Chassin (1997, *Journal of Consulting and Clinical Psychology*, p. 130) studied longitudinal progression of alcohol use in 82 adolescents. . .

- Three waves of data were gathered, which included a 4-item questionnaire measuring extent of alcohol use
- There were two level-2 predictors, *COA* (child of an alcoholic) and *PEER* (a measure of peer group alcohol use)
- As described in the text, a square root transformation was applied to the data to generate the *PEER* and *ALCUSE* data to enhance linearity.

An Example — Alcohol Use among Teenagers

Curran, Stice, and Chassin (1997, *Journal of Consulting and Clinical Psychology*, p. 130) studied longitudinal progression of alcohol use in 82 adolescents. . .

- Three waves of data were gathered, which included a 4-item questionnaire measuring extent of alcohol use
- There were two level-2 predictors, *COA* (child of an alcoholic) and *PEER* (a measure of peer group alcohol use)
- As described in the text, a square root transformation was applied to the data to generate the *PEER* and *ALCUSE* data to enhance linearity.

Preliminary Analysis

We would like to get a preliminary feel for the data with some exploratory analyses. We begin by loading the data.

```
> alcohol1 <- read.table("alcohol1_pp.txt", header=T, sep=",")
> attach(alcohol1)
```

The data are in person-period format, as we can see by looking at the first few lines:

```
> alcohol1[1:9,]
```

	id	age	coa	male	age_14	alcuse	peer	cpeer	ccoa
1	1	14	1	0	0	1.732	1.2649	0.2469	0.549
2	1	15	1	0	1	2.000	1.2649	0.2469	0.549
3	1	16	1	0	2	2.000	1.2649	0.2469	0.549
4	2	14	1	1	0	0.000	0.8944	-0.1236	0.549
5	2	15	1	1	1	0.000	0.8944	-0.1236	0.549
6	2	16	1	1	2	1.000	0.8944	-0.1236	0.549
7	3	14	1	1	0	1.000	0.8944	-0.1236	0.549
8	3	15	1	1	1	2.000	0.8944	-0.1236	0.549
9	3	16	1	1	2	3.317	0.8944	-0.1236	0.549

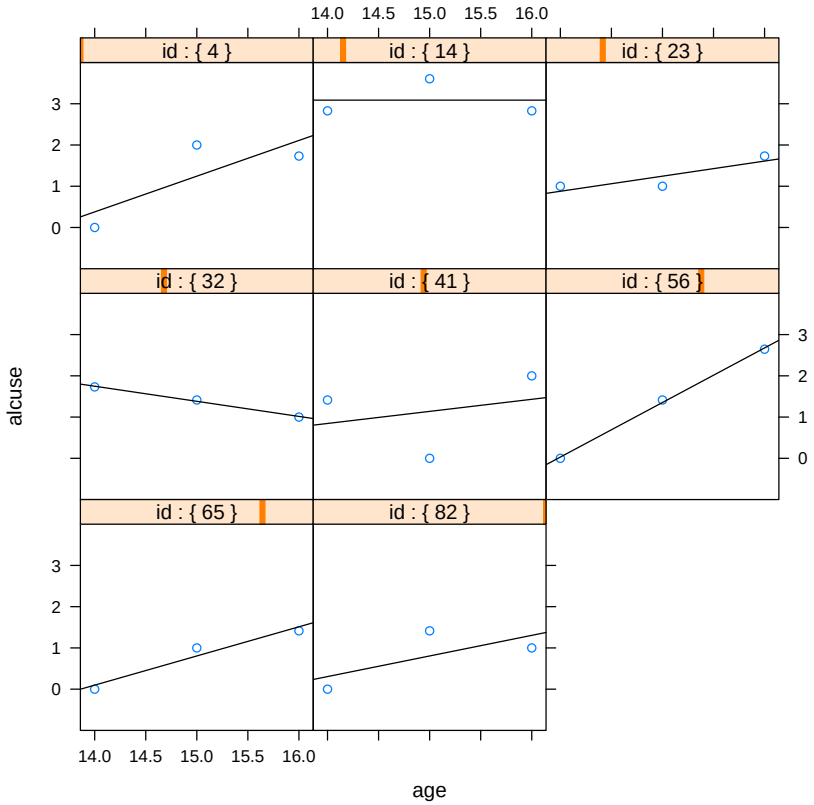
Preliminary Analysis

A good place to start is by examining individual growth curves for a random subset of 8 of the participants in the study.

```
> library(lattice)
```

```
> xyplot(alcuse~age | id,
+    data=alcohol1[alcohol1$id %in%
+      c(4, 14, 23, 32, 41, 56, 65, 82), ],
+    panel=function(x,y){
+         panel.xyplot(x, y)
+         panel.lmline(x,y)
+ }, ylim=c(-1, 4), as.table=T)
> update(trellis.last.object(),
+  strip = strip.custom(strip.names = TRUE,
+  strip.levels = TRUE))
```

Trellis Plot

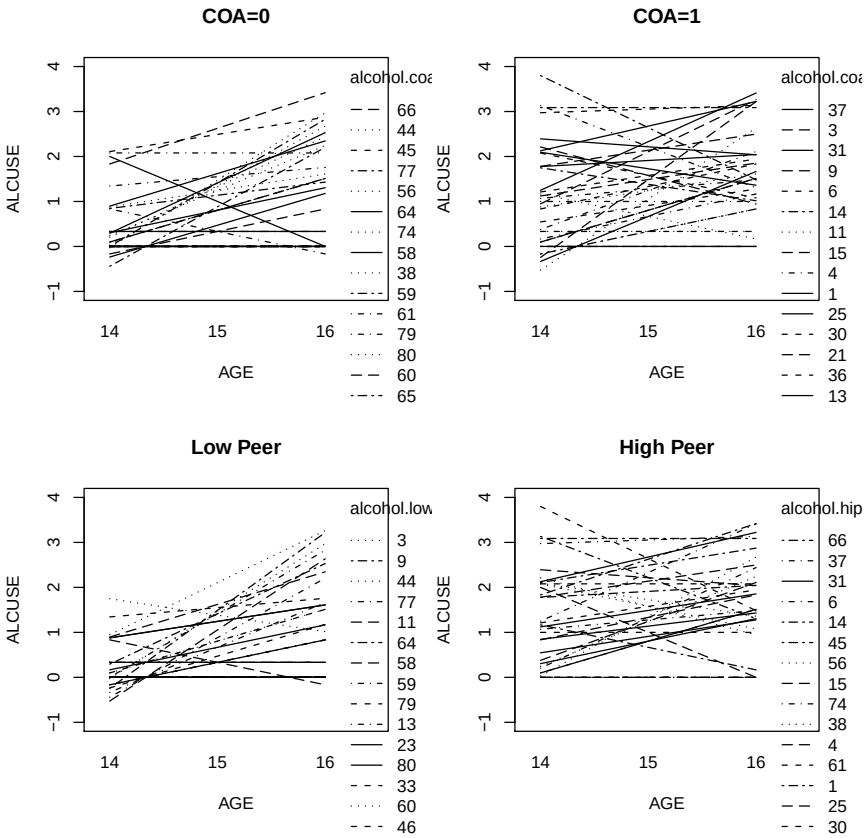


Potential Predictors

```
> #set up a 2x2 panel
> par(mfrow=c(2,2))
> alcohol.coa0 <- alcohol1[alcohol1$coa==0, ]
> #fitting the linear model by id
> f.coa0 <- by(alcohol.coa0, alcohol.coa0$id,
+        function(data) fitted(lm(alcuse~age, data=data)))
> #transforming f.coa from a list to a vector and
> #stripping of the names of the elements in the vector
> f.coa0 <- unlist(f.coa0)
> names(f.coa0) <- NULL
> #plotting the linear fit by id
> interaction.plot(alcohol.coa0$age, alcohol.coa0$id, f.coa0,
+                  xlab="AGE", ylab="ALCUSE", ylim=c(-1, 4), lwd=1)
> title("COA=0")
> alcohol.coa1 <- alcohol1[alcohol1$coa==1, ]
> #fitting the linear model by id
> f.coa1 <- by(alcohol.coa1, alcohol.coa1$id,
+              function(data) fitted(lm(alcuse~age, data=data)))
> #transforming f.coa1 from a list to a vector and
> #stripping of the names of the elements in the vector
> f.coa1 <- unlist(f.coa1)
> names(f.coa1) <- NULL
> #plotting the linear fit by id
> interaction.plot(alcohol.coa1$age, alcohol.coa1$id, f.coa1,
+                  xlab="AGE", ylab="ALCUSE", ylim=c(-1, 4), lwd=1)
> title("COA=1")
> cutoff<-mean(alcohol1$peer)
> alcohol.lowpeer <- alcohol1[alcohol1$peer<=cutoff, ]
> #fitting the linear model by id
> f.lowpeer <- by(alcohol.lowpeer, alcohol.lowpeer$id,
+              function(data) fitted(lm(alcuse~age, data=data)))
> #transforming f.lowpeer from a list to a vector and
> #stripping of the names of the elements in the vector
> f.lowpeer <- unlist(f.lowpeer)
> names(f.lowpeer) <- NULL
> #plotting the linear fit by id
> interaction.plot(alcohol.lowpeer$age, alcohol.lowpeer$id, f.lowpeer,
+                  xlab="AGE", ylab="ALCUSE", ylim=c(-1, 4), lwd=1)
> title("Low Peer")
> #######Lower right panel, peer>1.01756.
> alcohol.hipeer <- alcohol1[alcohol1$peer>cutoff, ]
> #fitting the linear model by id
> f.hipeer <- by(alcohol.hipeer, alcohol.hipeer$id,
+                function(data) fitted(lm(alcuse~age, data=data)))
> #transforming f.hipeer from a list to a vector and
> #stripping of the names of the elements in the vector
> f.hipeer <- unlist(f.hipeer)
> names(f.hipeer) <- NULL
> #plotting the linear fit by id
> interaction.plot(alcohol.hipeer$age, alcohol.hipeer$id, f.hipeer,
+                  xlab="AGE", ylab="ALCUSE", ylim=c(-1, 4), lwd=1)
> title("High Peer")


null device
          1
```

Potential Predictor Display



Evaluation of Potential Predictors

- In the top part of the panel, we see that children of alcoholics have generally higher intercepts than children of nonalcoholics
- In the bottom part of the panel, we see a tendency for adolescents in the higher peer group have higher intercepts but somewhat lower slopes
- These trends suggest that both *COA* and *PEER* may be important predictors of an individual's developmental trajectory

Evaluation of Potential Predictors

- In the top part of the panel, we see that children of alcoholics have generally higher intercepts than children of nonalcoholics
- In the bottom part of the panel, we see a tendency for adolescents in the higher peer group have higher intercepts but somewhat lower slopes
- These trends suggest that both *COA* and *PEER* may be important predictors of an individual's developmental trajectory

Evaluation of Potential Predictors

- In the top part of the panel, we see that children of alcoholics have generally higher intercepts than children of nonalcoholics
- In the bottom part of the panel, we see a tendency for adolescents in the higher peer group have higher intercepts but somewhat lower slopes
- These trends suggest that both *COA* and *PEER* may be important predictors of an individual's developmental trajectory

Introduction

In this section, we present the R code for generating the models discussed in Singer and Willett, Chapter 4.

The models are presented algebraically in their Table 4.2.

The output from an analysis with MLwiN (full IGLS) is presented in their Table 4.1.

We shall present the R code and output corresponding to each model.

Model A – The Unconditional Means Model

This model, corresponding to one-way random effects ANOVA, states in effect that all individual trajectories are flat, but that intercepts vary in a normal distribution around a population mean γ_{00}. Be sure to load the `lme4` library.

```
> library(lme4)
```

Fitting Model A

```
> model.a <- lmer(alcuse~ 1 + (1|id),REML=FALSE)
> summary(model.a)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ 1 + (1 | id)
 AIC BIC logLik deviance REMLdev
 676 687   -335      670     673
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 0.564    0.751
 Residual             0.562    0.749
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.9220     0.0957    9.63
```

The Intraclass Correlation Revisited

The intraclass correlation is computed on page 96 of Willett and Singer (2003). This is

$$\rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2} \tag{5}$$

which we estimate in this case from our R output as .57313/(.57313+.56175) = .505.

The Intraclass Correlation Revisited

The authors make the point that the composite model demonstrates, i.e., that the "residuals" in the composite model are the sum of two terms, one of which remains constant across time. So the intraclass correlation also represents the autocorrelation between measurements at two times the ith individual. For example, consider the outcome scores for individual i at times 1 and 2. These are, from the composite model,

$$\begin{aligned} Y_{i1} &= \gamma_{00} + \zeta_{0i} + \epsilon_{i1} \\ Y_{i2} &= \gamma_{00} + \zeta_{0i} + \epsilon_{i2} \end{aligned} \tag{6}$$

(C.P.) Using the heuristic rules for linear combinations, prove that the correlation between Y_{i1} and Y_{i2} is the intraclass correlation ρ.

Model B — The Unconditional Growth Model

This model allows a non-flat trajectory by including *TIME* as the predictor in the level-1 model.

It also allows the slopes and intercepts to correlate across individuals.

The data file contains a variable called `age14` that represents time from the beginning of the study, which is a reasonable metric to use in this case. However, I prefer the name *TIME* and have effectively renamed the variable in the code below.

Fitting Model B

```
> time <- age_14
> model.b <- lmer(alcuse ~ time +(time | id),REML=FALSE)
> summary(model.b)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ time + (time | id)
 AIC BIC logLik deviance REMLdev
 649 670   -318      637     643
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.624    0.790
          time        0.151    0.389    -0.223
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.6513     0.1051    6.20
time          0.2707     0.0625    4.33

Correlation of Fixed Effects:
     (Intr)
time -0.441
```

Interpreting Model B Output

Note that the residual variance dripped sharply from .562 to .337. Since .337/.562 = .600, Singer and Willett conclude that the 40% of the within-person variation alcohol use is systematically associated with linear *TIME*.

Note also that the correlation between the two random effects is negative, −.227, and weak.

Model C – *COA* as a Level-2 Predictor

In this model, we use *COA* at level 2 to predict slopes and intercepts.

Fitting Model C

```
> model.c <- lmer(alcuse ~ coa + time + coa:time + (time | id),REML=FALSE)
> summary(model.c)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ coa + time + coa:time + (time | id)
 AIC BIC logLik deviance REMLdev
 637 665   -311      621     632
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.488    0.698
          time        0.151    0.388    -0.219
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.3160     0.1307    2.42
coa           0.7432     0.1946    3.82
time          0.2930     0.0842    3.48
coa:time     -0.0494     0.1254   -0.39

Correlation of Fixed Effects:
         (Intr) coa    time
coa      -0.672
time     -0.460  0.309
coa:time  0.309 -0.460 -0.672
```

Pseudo-R^2 Statistics

On pages 102–104, Singer and Willett discuss three "pseudo-R^2" statistics for quantifying performance of the various models. The first statistic, $R^2_{y,\hat{y}}$ is the squared correlation, across all participants, between predicted scores (using model estimates in the composite model formula) and actual outcome scores. In this case, $R^2_{y,\hat{y}} = .043$, as computed below.

```
> cor(alcuse,.6513 +.2707*time)^2
```

```
[1] 0.04339
```

Pseudo-R^2 Statistics

Residual variation—that portion of the outcome variation *unexplained* by a model's level-1 predictors—provides another criterion for comparing two models. For models A and B, we have

$$R^2_\epsilon = \frac{\hat{\sigma}^2_{\epsilon_A} - \hat{\sigma}^2_{\epsilon_B}}{\hat{\sigma}^2_{\epsilon_A}} \tag{7}$$

In this case, we get $(.562 - .337)/.562 = .400$.

Pseudo-R^2 Statistics

We can use an approach similar to that taken in the previous slide to compute pseudo-R^2 statistics for the proportional reduction in level-2 variance attributable to the addition of level-2 predictors. We have, for example

$$R^2_C = \frac{\hat{\sigma}^2_{\epsilon_B} - \hat{\sigma}^2_{\epsilon_C}}{\hat{\sigma}^2_{\epsilon_B}} \tag{8}$$

One well-known problem with these statistics is that unlike more familiar R^2 indices, they can be negative.

Model D – COA and PEER as Level-2 Predictors

```
> model.d <- lmer(alcuse ~ coa + time + coa:time+ peer + peer:time +(time | id),REML=FALSE)
> summary(model.d)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ coa + time + coa:time + peer + peer:time + (time | id)
 AIC BIC logLik deviance REMLdev
 609 644   -294      589     606
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.241    0.491
          time        0.139    0.373    -0.033
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -0.3165     0.1481   -2.14
coa           0.5792     0.1625    3.56
time          0.4294     0.1137    3.78
peer          0.6943     0.1115    6.23
coa:time     -0.0140     0.1248   -0.11
time:peer    -0.1498     0.0856   -1.75

Correlation of Fixed Effects:
          (Intr) coa    time   peer   coa:tm
coa       -0.371
time      -0.436  0.162
peer      -0.686 -0.162  0.299
coa:time   0.162 -0.436 -0.371  0.071
time:peer  0.299  0.071 -0.686 -0.436 -0.162
```

Recentering Predictors to Improve Interpretation

Time-invariant predictors can be centered, either to a group average or a particularly meaningful or interesting value, in order to facilitate interpretation. In general, such recentering will affect intercepts but not slopes. $\hat{\gamma}_{10}$ and $\hat{\gamma}_{11}$ represent values of their respective growth parameters when all other predictors in the associated level-2 model are zero.

This can make interpretation problematic if 0 is an impossible value.

Model E

In model E, neither `PEER` nor `COA` are centered. The intercepts therefore represent a child of non-alcoholic parents whose peers at age 14 are totally abstinent (`PEER` $= 0$ and `COA` $= 0$).

Model E

```
> model.e <- lmer(alcuse ~ coa + peer + time + peer:time +(time | id),REML=FALSE)
> summary(model.e)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ coa + peer + time + peer:time + (time | id)
 AIC BIC logLik deviance REMLdev
 607 638   -294      589     604
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.241    0.491
          time        0.139    0.373    -0.034
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -0.3138     0.1461   -2.15
coa           0.5712     0.1462    3.91
peer          0.6952     0.1113    6.25
time          0.4247     0.1056    4.02
peer:time    -0.1514     0.0845   -1.79

Correlation of Fixed Effects:
          (Intr) coa    peer   time
coa       -0.338
peer      -0.709 -0.146
time      -0.410  0.000  0.351
peer:time  0.334  0.000 -0.431 -0.814
```

Model F

In model F, `PEER` is centered while `COA` is not. The intercepts therefore represent a child of non-alcoholic parents whose peers at age 14 are average consumers (`PEER` $= 1.018$ and `COA` $= 0$).

Model F

```
> model.f <- lmer(alcuse ~ coa + cpeer + time + cpeer:time + (time | id),REML=FALSE)
> summary(model.f)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ coa + cpeer + time + cpeer:time + (time | id)
 AIC BIC logLik deviance REMLdev
 607 638   -294      589     604
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.241    0.491
          time        0.139    0.373    -0.034
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.3939     0.1035    3.80
coa           0.5712     0.1462    3.91
cpeer         0.6952     0.1113    6.25
time          0.2706     0.0613    4.42
cpeer:time   -0.1514     0.0845   -1.79

Correlation of Fixed Effects:
           (Intr) coa    cpeer  time
coa        -0.637
cpeer       0.094 -0.146
time       -0.336  0.000  0.000
cpeer:time  0.000  0.000 -0.431  0.001
```

Model F

In model G, `PEER` and `COA` are centered. The intercepts therefore represent an average study participant(`PEER` $= 1.018$ and `COA` $= 0.451$).

Model G

```
> model.g <- lmer(alcuse ~ ccoa+ cpeer + time + cpeer:time + (time | id),REML=FALSE)
> summary(model.g)

Linear mixed model fit by maximum likelihood
Formula: alcuse ~ ccoa + cpeer + time + cpeer:time + (time | id)
 AIC BIC logLik deviance REMLdev
 607 638   -294      589     604
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.241    0.491
          time        0.139    0.373    -0.034
 Residual             0.337    0.581
Number of obs: 246, groups: id, 82

Fixed effects:
            Estimate Std. Error t value
(Intercept)   0.6515     0.0798    8.17
ccoa          0.5712     0.1462    3.91
cpeer         0.6952     0.1113    6.25
time          0.2706     0.0613    4.42
cpeer:time   -0.1514     0.0845   -1.79

Correlation of Fixed Effects:
           (Intr) ccoa   cpeer  time
ccoa        0.000
cpeer       0.001 -0.146
time       -0.436  0.000  0.000
cpeer:time  0.000  0.000 -0.431  0.001
```

Deviance Statistics for Comparing Nested Models

As explained in the text, with ML estimation $-2 \times LL$ is a chi-square "deviance" or "badness of fit" statistic. If models are nested, the difference in deviance statistics has a chi-square distribution with degrees of freedom equal to the difference in the number of estimated parameters for the two models.

However, if REML estimation is used, then deviance-based comparisons can only be made for models with identical fixed effects but varying random effects.

Comparing Model A to Model B

```
> options(digits=4)
> anova(model.a,model.b)

Data:
Models:
model.a: alcuse ~ 1 + (1 | id)
model.b: alcuse ~ time + (time | id)
        Df AIC BIC logLik Chisq Chi Df Pr(>Chisq)
model.a  3 676 687   -335
model.b  6 649 670   -318  33.5      3    2.5e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing Model B to Model C

```
> anova(model.b,model.c)

Data:
Models:
model.b: alcuse ~ time + (time | id)
model.c: alcuse ~ coa + time + coa:time + (time | id)
        Df AIC BIC logLik Chisq Chi Df Pr(>Chisq)
model.b  6 649 670   -318
model.c  8 637 665   -311  15.4      2    0.00045 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wald Statistics for Testing Composite Hypotheses

If a set of parameters is collected in a vector $\boldsymbol{\gamma}$, for example, then composite linear hypotheses can be written in the form

$$H_0: \quad \boldsymbol{C}'\boldsymbol{\gamma} = \mathbf{0}$$

Example (A Composite Hypothesis)

Suppose we want to test whether γ_1 and γ_2 are both zero. Then, in matrix formulation, we have

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Wald Statistics for Testing Composite Hypotheses

Suppose we estimate $\boldsymbol{\gamma}$ with the ML estimates $\hat{\boldsymbol{\gamma}}$ having estimated covariance matrix $\hat{\boldsymbol{T}}$. Then, if the null hypothesis is true, it can be shown rather easily that

$$\hat{\boldsymbol{\gamma}}'\boldsymbol{C}'(\boldsymbol{C}'\hat{\boldsymbol{T}}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}'\hat{\boldsymbol{\gamma}} \tag{9}$$

has an asymptotic χ^2 distribution with the number of degrees of freedom equal to the number of rows in the $\boldsymbol{C}$ matrix. This is the multivariate equivalent of our general linear combination hypothesis procedure discussed in Psychology 310. It should be noted that this is based on asymptotic theory that, in turn, depends on the assumption of multinormality, *and* tends to exhibit slow convergence. Only at very large samples would such statistics be accurate if used to test random effects. This parallels the reason why, in Psychology 310, we avoided Z tests on variances.

The simplified treatment of the Wald Statistic in Willett and Singer is incomplete and, in its attempt to maintain simplicity, not really correct.

Information-Based Criteria for Comparing Non-Nested Models

We have a variety of null hypothesis tests to compare models for exact equivalence. Frankly, the modern view is that these are useful but should not be taken too seriously, for the same reasons that hypothesis tests in general should not be taken too seriously.

Deviance-based hypothesis tests aren't available for comparing non-nested models. Moreover, when models are nested, the more complex model always fits better (except in artificial examples) because models almost never fit perfectly. Added complexity in a nested model framework always improves fit. For example, in multiple regression, adding predictors always improves R^2.

Information-Based Criteria for Comparing Non-Nested Models

Deviance-based hypothesis tests aren't available for comparing non-nested models. Moreover, when models are nested, the more complex model always fits better (except in artificial examples) because models almost never fit perfectly. Added complexity in a nested model framework always improves fit. For example, in multiple regression, adding predictors always improves R^2. Adding factors in factor analysis always improves fit.

So there is an inevitable tradeoff between complexity and the quality of a model's fit. What we seek is a model that has good fit and good parsimony.

The Akaike (AIC) and Schwarz (BIC) Criteria

A variety of model-fitting statistics have been developed to help us select a model in a good region of the complexity-fit tradeoff. Suppose we fit a set of models to the same data.

The Akaike Information Criterion (AIC) is applied to all the models, and the model with the lowest value of the AIC criterion is selected.

The Schwarz Bayesian Information Criterion (BIC) is used in a similar manner.

The Akaike (AIC) and Schwarz (BIC) Criteria

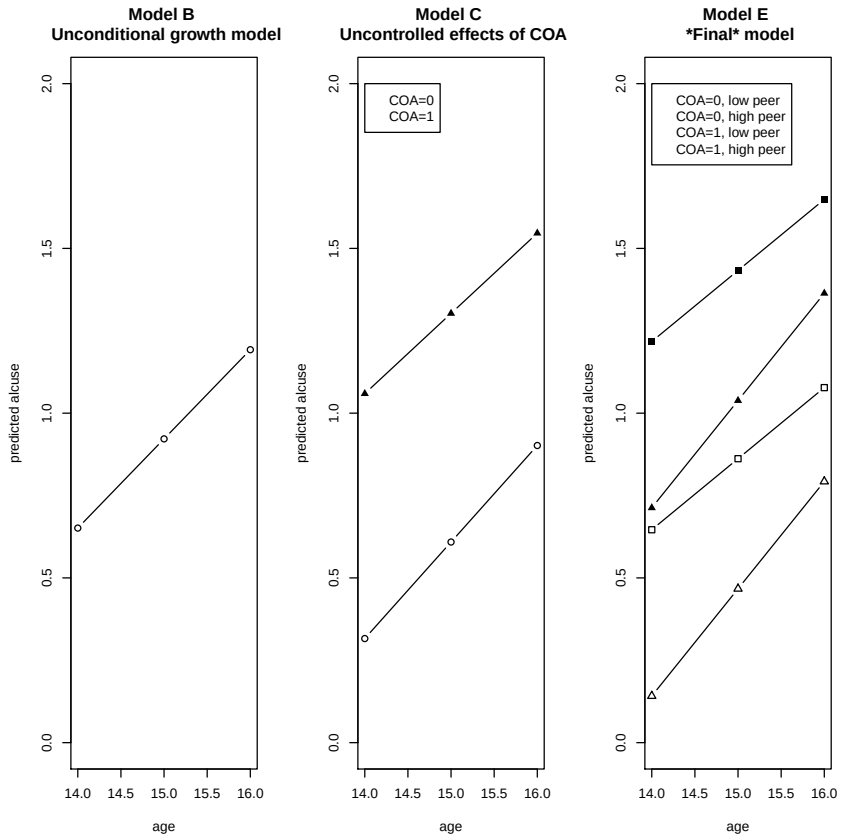
The AIC and BIC are only useful in a *relative* sense, and must be applied to the same data, for models explaining the same variables. Because these statistics are only used in a relative sense, they may be rescaled monotonically in any way you find convenient. Consequently, you will see different versions of the criteria.

$$\begin{aligned} \text{AIC} &= \chi^2 + 2k \\ \text{BIC} &= \chi^2 + \ln nk \end{aligned}$$

Plotting Model Trends

```
> pdf("ModelFitPanel.pdf")
> par(mfrow = c(1,3))
> #Plots
> #Model B
> fixef.b <- fixef(model.b)
> fit.b <- fixef.b[[1]] + time[1:3]*fixef.b[[2]]
> plot(alcohol1$age[1:3], fit.b, ylim=c(0, 2), type="b",
+      ylab="predicted alcuse", xlab="age")
> title("Model B \n Unconditional growth model")
> #Model C
> fixef.c <- fixef(model.c)
> fit.c0 <- fixef.c[[1]] + time[1:3]*fixef.c[[3]]
> fit.c1 <- fixef.c[[1]] + fixef.c[[2]] +
+           time[1:3]*fixef.c[[3]] +
+           time[1:3]*fixef.c[[4]]
> plot(alcohol1$age[1:3], fit.c0, ylim=c(0, 2), type="b",
+      ylab="predicted alcuse", xlab="age")
> lines(alcohol1$age[1:3], fit.c1, type="b", pch=17)
> title("Model C \n Uncontrolled effects of COA")
> legend(14, 2, c("COA=0", "COA=1"))
> #Model E
> fixef.e <- fixef(model.e)
> fit.ec0p0 <- fixef.e[[1]] + .655*fixef.e[[3]] +
+              time[1:3]*fixef.e[[4]] +
+              .655*time[1:3]*fixef.e[[5]]
> fit.ec0p1 <- fixef.e[[1]] + 1.381*fixef.e[[3]] +
+              time[1:3]*fixef.e[[4]] +
+              1.381*time[1:3]*fixef.e[[5]]
> fit.ec1p0 <- fixef.e[[1]] + fixef.e[[2]] + .655*fixef.e[[3]] +
+              time[1:3]*fixef.e[[4]] +
+              .655*time[1:3]*fixef.e[[5]]
> fit.ec1p1 <- fixef.e[[1]] + fixef.e[[2]] + 1.381*fixef.e[[3]] +
+              time[1:3]*fixef.e[[4]] +
+              1.381*time[1:3]*fixef.e[[5]]
> plot(alcohol1$age[1:3], fit.ec0p0, ylim=c(0, 2), type="b",
+      ylab="predicted alcuse", xlab="age", pch=2)
> lines(alcohol1$age[1:3], fit.ec0p1, type="b", pch=0)
> lines(alcohol1$age[1:3], fit.ec1p0, type="b", pch=17)
> lines(alcohol1$age[1:3], fit.ec1p1, type="b", pch=15)
> title("Model E \n *Final* model")
> legend(14, 2, c("COA=0, low peer", "COA=0, high peer",
+        "COA=1, low peer", "COA=1, high peer"))
> dev.off()
```

Plotting Model Trends



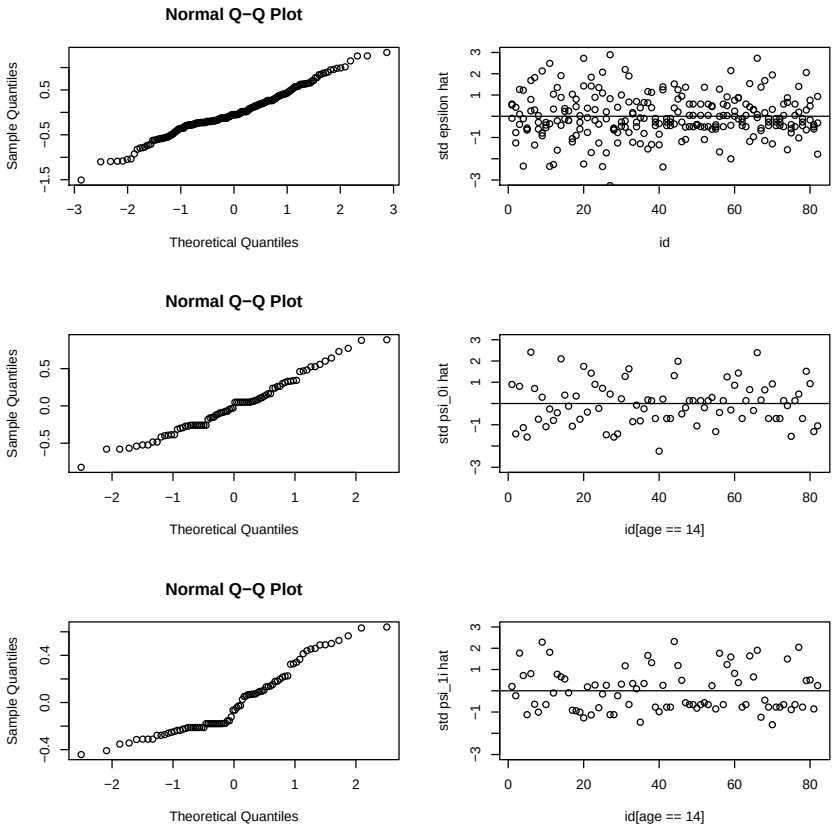
Empirical Bayes Estimates of Individual Trajectories

In our preliminary exploratory analyses, we plotted individual trajectories based on OLS estimation from an individual's data. Singer and Willett (pp. 132–137) explain how to calculate improved estimates of an individual's trajectory. In a lab exercise, we will examine how to compute these with R.

Displaying Residual Plots

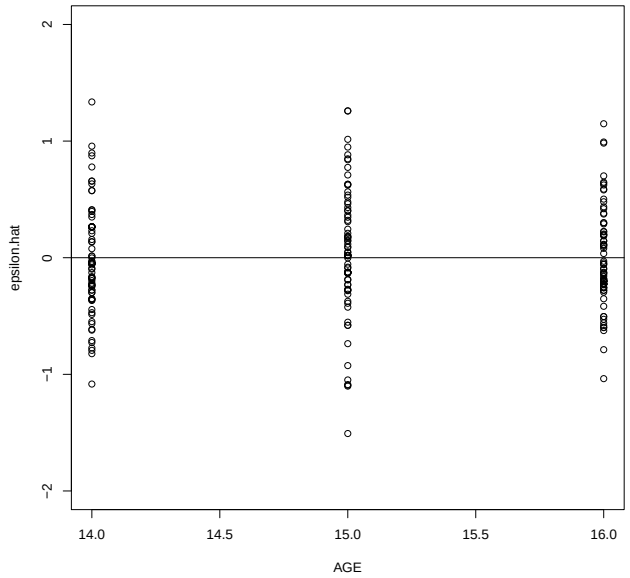
```
> pdf("NormalityPanel.pdf")
> par(mfrow = c(3,2))
> resid <- residuals(model.f)
> qqnorm(resid)
> #creating the standardized residual (std epsilon.hat)
> resid.std <- resid/sd(resid)
> plot(id, resid.std, ylim=c(-3, 3), ylab="std epsilon hat")
> abline(h=0)
> #Middle left panel
>
> #extracting the random effects of model f
> ran <- attr(model.f,"ranef")[1:82]
> qqnorm(ran)
> #Middle right panel
>
> #standardizing the ksi0i.hat
> ran1.std <- ran/sd(ran)
> plot(id[age==14], ran1.std, ylim=c(-3, 3), ylab="std psi_0i hat")
> abline(h=0)
> #Lower left panel
> ran2 <- attr(model.f,"ranef")[83:164]
> qqnorm(ran2)
> #Lower right panel
>
> #standardizing the ksi1i.hat
> ran2.std <- ran2/sd(ran2)
> plot(id[age==14], ran2.std, ylim=c(-3, 3), ylab="std psi_1i hat")
> abline(h=0)
> dev.off()
```

Displaying Residual Plots



Examining Residual Variance

```
> plot(age, resid, ylim=c(-2, 2), ylab="epsilon.hat",
+      xlab="AGE")
> abline(h=0)
```



Examining Residual Variance

```
> pdf("ResidPanel.pdf")
> par(mfrow=c(2,2))
> #Upper left panel
> plot(coa[age==14], ran, ylim=c(-1, 1),
+      ylab="ksi0i.hat", xlab="COA")
> abline(h=0)
> #Upper right panel
> plot(peer[age==14], ran, ylim=c(-1, 1),
+      xlim=c(0, 3), ylab="ksi0i.hat", xlab="PEER")
> abline(h=0)
> #Lower left panel
> plot(coa[age==14], ran2, ylim=c(-1, 1),
+      ylab="ksi1i.hat", xlab="COA")
> abline(h=0)
> #Lower right panel
> plot(peer[age==14], ran2, ylim=c(-1, 1),
+      xlim=c(0, 3), ylab="ksi1i.hat", xlab="PEER")
> abline(h=0)
> dev.off()
```

Examining Residual Variance