Bayesian Hierarchical Modeling for the Social Sciences

An Introduction to Hierarchical Modeling in Quantitative Research

JEFF GILL
Department of Government
Department of Mathematics and Statistics
Center for Behavioral Neurosciences
American University

Advantages of Multilevel Models

- Removes the restriction that the estimated coefficients are constant across individual cases by specifying levels of additional effects.
- Provides a notationally efficient way to organize groups in the model.
- Accounting for individual versus group-level variation.
- Modeling variation among individual-level regression coefficients.
- Estimating regression coefficients for groups of interest.
- Gets the standard errors right.

Features of Multilevel Models

- Each level of the model is its *own* regression, with its own assumptions about: functional form, linearity, independence, variance, distribution of errors, etc.
- Models are usually "mixed," meaning some coefficients are *modeled* and some are *unmodeld*.
- Multilevel models are highly symbiotic with Bayesian specifications because the focus in both cases is on making reasonable distributional assumptions.
- These approaches are generally more demanding of statistical estimation process (software) to produce results.

Linear Model Illustration

- Start with a standard linear model specification indexed by subjects and a first level of grouping, the *context* level.
- Now use a single explanatory variable that has the form:

$$y_i = \beta_{j0[i]} + \beta_{j1[i]} X_i + \epsilon_i.$$

- Suppose we have group-level explanatory variables, $Z_{j\cdot}$, in that their effect is measured at the aggregated rather than at the individual level.
- Now add a second level to the model that explicitly nests effects within groups and index these groups $j = 1$ to J:

$$\begin{aligned} \beta_{j0[i]} &= \gamma_{00} + \gamma_{10} Z_{j0} + u_{j0} \\ \beta_{j1[i]} &= \gamma_{01} + \gamma_{11} Z_{j1} + u_{j1}, \end{aligned}$$

Linear Model Illustration

- The two-level model is produced by inserting the context level specifications into the original linear expression for the outcome variable of interest:

$$y_i = \gamma_{00} + \gamma_{01}X_i + \gamma_{10}Z_{j0} + \gamma_{11}X_iZ_{j1} + u_{j1}X_i + u_{j0} + \epsilon_i.$$

- This equation shows that the composite error structure, $u_{j1}X_i + u_{j0} + \epsilon_i$, is now clearly heteroscedastic since it is conditioned on levels of the explanatory variable, causing additional estimation complexity.
- Notice that there is an "automatic" interaction component: $\gamma_{11}X_iZ_{j1}$.
- Now we are going model *distributions* for y, β_{j0}, and β_{j1}.
- This means we will make distributional regression statements:

$$\beta_{j0} \sim f(\gamma_{00} + \gamma_{10}Z_{j0}, \sigma_{\beta_0}).$$

Vocabulary Fun

- The varying coefficients are sometimes called random effects, since they are associated with distributional statements like $\alpha_j = N(\mathbf{Z}_j\gamma, \sigma^2_\alpha)$.
- The term fixed effects is more nebulous with different meanings from different authors:
  - coefficients that are constant across individuals (most common definition)
  - factor contrasts
  - nuisance coefficients that are uninteresting but included
  - coefficients in population models
  - realized random variables
  - MLE values assuming infinite group-level variance

  (see the discussion in G&H, p.245).
- Sometimes these models labeled as mixed effects models.
- Prescription: use multilevel models or hierarchical models with appropriate descriptor or detailed specification.

Vocabulary Overview

- For the data matrices, X_i for individual i in cluster j, and $\mathbf{Z}_j$ for cluster j, there are five canonical models that we will look at:

"Completely Pooled"	$y_i = \beta_0 + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z} + e_i$
"Fixed Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + e_i$
"Random Effect"	$y_i = \beta_{j0[i]} + \beta_1 \mathbf{X}_i + \gamma \mathbf{Z}_j + e_i$
"Random Intercept and Random Slope"	$y_i = \beta_{j0[i]} + \beta_{j1[i]} \mathbf{X}_i + e_i$
"Completely Unpooled'	$y_{j[i]} = \beta_{j0} + (\beta_{j1} \mathbf{X}_{j[i]} + \gamma \mathbf{Z}_j) + e_{j[i]}$

- This is produced by replacing the previous γ coefficient names with common regression-style language.

- "Fixed" and "random" can differ in definition by literature (Kreft and De Leeuw 1988, Section 1.3.3, Gelman 2005), and better notation is "random intercepts" for "fixed effect," and "varying-intercept, varying-slope" for "random intercept and random slope."

- Best to conceptualize these specifications as members of a larger multilevel family where indices are *turned-on* or *turned-off* systematically depending on the hierarchical purpose.

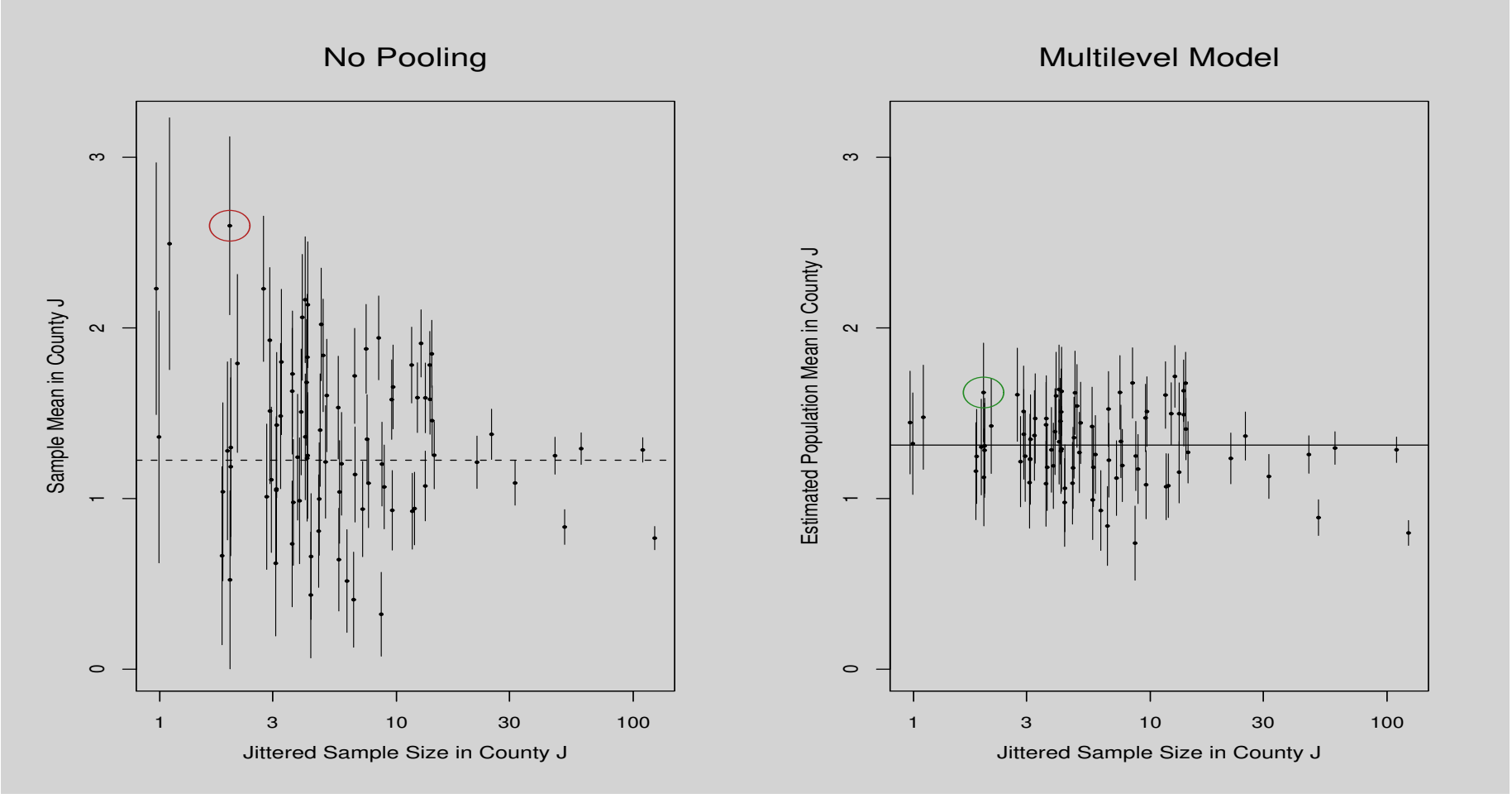
Comparison with Variable Contrasts

- The most commonly used scheme for dealing with purely categorical explanatory variables is called a **treatment contrast** where one category is selected as a baseline and for k categories of the explanatory variable there are $k - 1$ estimated difference parameters.
- Such schema are necessary because including all categorical information creates an uninvertible $\mathbf{X}'\mathbf{X}$ matrix (perfect collinearity).
- This can be awkward for large k or where there is not a logical baseline category.
- Multilevel models allow inclusion of all categorical values through specification in a hierarchy: they become part of the model specification rather than just additional $\mathbf{X}$ columns.

Back to Pooling

- Complete-Pooling: excluding categorical predictors completely (no hierarchy).
- This ignores (possibly important) variation between categories.
- No-Pooling: estimating separate models for each level of the categorical predictors.
- This overstates variation between categories, making them look more different than they really are (unless the categories are not meaningful).
- Multilevel Models: a compromise between these two extremes that captures within category uniqueness and between category similarities.
- Running example from Gelman & Hill: Radon gas by county ($J = 85$) in Minnesota.

Figure 12.1 from Gelman & Hill



Partial Pooling Estimates with No Explanatory Variables

- Panel 2 of Figure 12.1 gives the average log radon level α_j by county $j = 1, \ldots, J$, plotted by n_j which varies by county.

- These were produced from the multilevel model estimated with `JAGS` , but are approximated by:

$$\alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

where:

$\bar{y}_j$	unpooled estimate for county j
$\bar{y}_{\text{all}}$	completely pooled estimate
σ_y^2	within-county variance (assumed equal for now)
σ_α^2	variance among the mean estimates

Partial Pooling Estimates with No Explanatory Variables

- County Sample Size Consequences from:

$$\alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{\text{all}}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

- Averages from counties with smaller sample sizes contribute less and are closer to the state average, and county estimate equal to the state average for $n_j = 0$.
- Averages from counties with larger sample sizes contribute more and pull the state average towards them, and state average equal to the county estimate for $n_j = \infty$.
- Output from `JAGS` that generated the second panel of Figure 12.1:

```
          Mean      SD   NaiveSE TimeseriesSE
sigma.y 0.7991 0.01956 0.0001956    0.0002304
a1      1.0602 0.25501 0.0025501    0.0031325
mu.a    1.3136 0.05011 0.0005011    0.0009859
sigma.a 0.3171 0.04888 0.0004888    0.0014624
```

where `NaiveSE` $= \sqrt{\text{sample variance}}/\sqrt{n}$ and: `TIMEseriesSE` $= \sqrt{\text{spectral density var}}/\sqrt{n} =$ asymptotic SE.

Partial Pooling Estimates with No Explanatory Variables

- We can also re-express this as a weighted average of the no-pooling estimate and the pooled estimate mean:

$$\alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}}(\bar{y}_j - \beta\bar{X}_j) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}}\mu_\alpha$$

  where the first term is a weighted no-pooled regression of the jth case and the second term is a weighted completely pooled mean.

- This shows the *shrinkage* that occurs when some n_j gets large relative to the others (sometimes called the "Manhattan Effect").

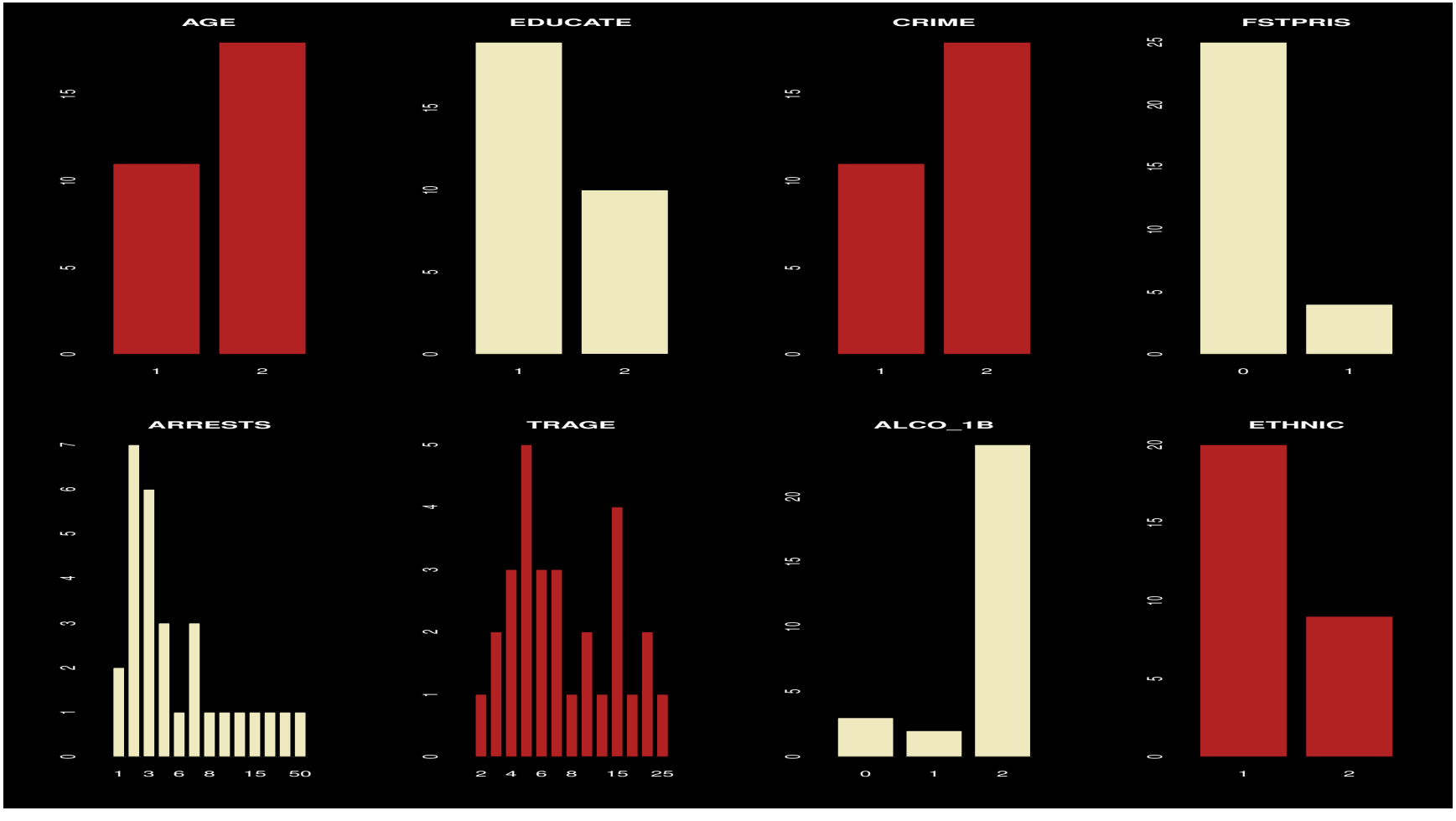
- When $\sigma_\alpha^2 \to 0$, then the results move toward the complete-pooling model.

- When $\sigma_\alpha^2 \to \infty$, then the results move toward the no-pooling model.

Sample Analysis, Data

- 29 Incarcerated Women with Substance Use Disorder and Post-traumatic Stress Disorder in Providence, Rhode Island, 1999-2001.
- Outcome variable: `PTSD diagnosis` (13 negative, 16 positive).
- Explanatory variables
  - `AGE`, 1 for 20-29, 2 for 30+
  - `ETHNIC`, 1 for white (nonhispanic), 2 for nonwhite
  - `EDUCATE`, 1 for no HS diploma, 2 for HS diploma
  - `FSTPRIS`, 0 for in prison before, 1 for first time
  - `CRIME`, 0 for misdomeanor, 1 for felony
  - `ARRESTS`, the number of arrests with convictions
  - `TRAGE`, age of first trauma (robbery/mugging, sexual abuse, physical abuse)
  - `ALCO1B`, alchohol issue: 0=None, 1=Abuse, 2=Dependent

Sample Analysis, Data



Sample Analysis, Model

$$p(y_i = 1) = \text{logit}^{-1}\Big(\beta^0 + \beta^{\text{AGE}} \cdot \texttt{AGE}_i + \beta^{\text{EDUCATE}} \cdot \texttt{EDUCATE}_i + \beta^{\text{AGE·EDUCATE}} \cdot (\texttt{AGE}_i \cdot \texttt{EDUCATE}_i) + \beta^{\text{CRIME}} \cdot \texttt{CRIME}_i + \beta^{\text{FSTPRIS}} \cdot \texttt{FSTPRIS}_i + \beta^{\text{ARRESTS}} \cdot \log(\texttt{ARRESTS}_i) + \beta^{\text{TRAGE}} \cdot \exp(\texttt{TRAGE}_i) + \beta^{\text{ALCO1B}} \cdot \texttt{ALCO1B}_i + \alpha^{\text{ETHNIC}}_{j[i]}\Big)$$

$$\alpha^{\text{ETHNIC}}_j \sim N\left(\alpha_0 + \alpha^{\text{AGE}}_{m[j]} \cdot \texttt{AGE} + \alpha^{\text{EDUCATE}}_{m[j]} \cdot \texttt{EDUCATE},\ \sigma^2_{\text{ETHNIC}}\right)$$

Fixed effects:

	Estimate	Std. Error	z value
(Intercept)	-5.02e+00	4.14e+00	-1.214
AGE	5.12e+00	2.78e+00	1.841
EDUCATE	7.37e+00	3.76e+00	1.959
CRIME	-2.00e+00	9.01e-01	-2.221
FSTPRIS	2.48e-01	1.11e+00	0.222
log(ARRESTS)	-1.19e+00	6.17e-01	-1.927
exp(TRAGE)	5.63e-08	3.48e-08	1.617
ALCO_1B	1.10e-01	4.18e-01	0.264
AGE:EDUCATE	-3.98e+00	1.99e+00	-2.003

Random effects:

Groups	Name	Variance	Std.Dev.
ETHNIC	(Intercept)	0.1839	0.429
	AGE	0.0643	0.253
	EDUCATE	0.0904	0.301
	Residual	0.7299	0.854

AIC	BIC	logLik	deviance
57.2	77.7	-13.6	27.2

Data Exercise 7: Trauma Data

```
trauma.short.complete <-
    read.table("http://jeffgill.org/files/jeffgill/files/trauma.short_.dat_.txt",
    header=TRUE)
library(nlme); library(arm)
trauma.out <- glmer(PTSD2 ~ "some individual explanatory variables"
                + (1 + "some group-level variables" | "grouping"),
                family=binomial(link="probit"),
                data=trauma.short.complete)
summary(trauma.out)
```

Data Exercise 7: Trauma Data

- Run your version of the Trauma/Prison model replacing the quoted amounts with your definitions.
- Consider putting different categorical variables in the hierarchy (parenthetical part).
- Consider interactions.
- Look at different mixes of covariates.
- Now run a completely pooled version of your model (no hierarchy) and compare the coefficient esimates at the individual level.

A Bayesian Take On Hierarchical Models

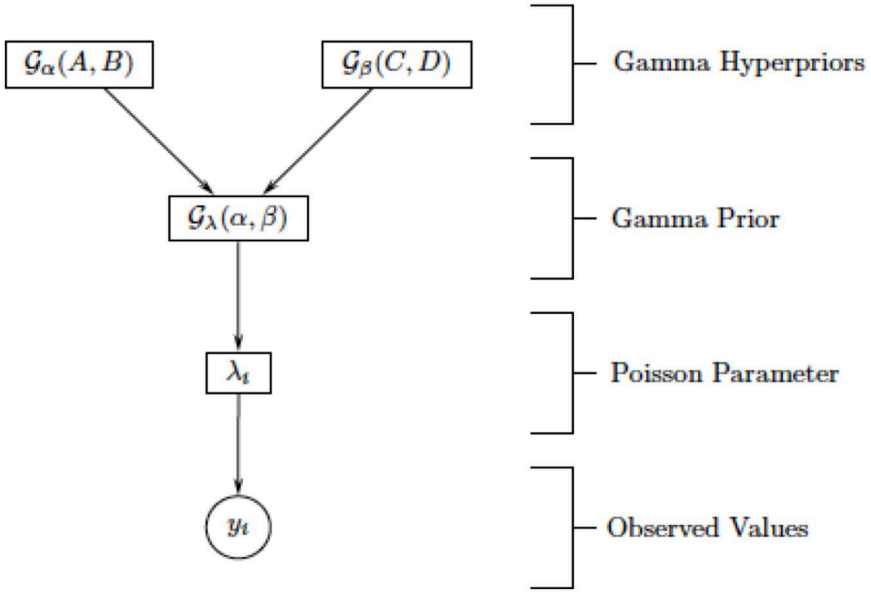


FIGURE 12.1: Poisson-Gamma Hierarchical Model

the example model, this representation is:

$$\begin{aligned} y_i &\sim \mathcal{P}(\lambda_i) \\ \lambda_i &\sim \mathcal{G}(\alpha,\beta) \\ \alpha &\sim \mathcal{G}(A,B) \\ \beta &\sim \mathcal{G}(C,D), \end{aligned} \tag{12.24}$$

A Nonlinear Multilevel Approach

- Suppose we have the same binary outcome at the individual-level for $i = 1, \ldots, n$:

$$p(y_i = 1) = \text{logit}^{-1}(\beta_{j0[i]} + X_i\beta_{j1}),$$

  where $\mathbf{X}$ contains individual-level explanatory variables only.

- Now make the following assumption for the $j = 1, \ldots, J$ intercepts:

$$\beta_j = N(Z_j\gamma, \sigma^2_\beta),$$

  where Z_J is the group-level matrix analogous to $\mathbf{X}$, the coefficients to be estimated are γ, σ^2_y and σ^2_β.

- So the full model in "stacked notation" is:

$$p(y_i = 1) = \text{logit}^{-1}(\beta_{j0[i]} + X_i\beta_{j1}),$$

$$\beta_j = N(Z_j\gamma, \sigma^2_\beta).$$

- Notice that we are being more efficient and more explicit about distributions in our notation.

Panel Data as Group Membership

- 2,000 Australian adolescents with smoking measured every 6 months for 3 years.
- So observations are nested (grouped) with persons.
- Specified model for case j at wave t:

  $$p(y_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{psmoke}_{j[i]} + \beta_2 \text{female}_{j[i]} + \beta_3(1 - \text{female}_{j[i]})t[i] + \beta_4(\text{female}_{j[i]})t[i] + \alpha_j)$$

  for a representation that has a matrix for individual effects that don't change ($[i]$) and another for group effects that do not change ($j[i]$).
- So the jth person has multiple waves on the jth row of the $\mathbf{X}$ matrix.
- Software note: for logit/probit link functions `lmer` sets $\sigma_y = 1$ for identifiability.

Data Load

```
lapply(c("lme4","arm"),library, character.only=TRUE)
smoking <- read.table("http://JeffGill.org/data/smoke_pub.dat",header=TRUE)
smoking[c(1:8,(nrow(smoking)-7):nrow(smoking)),]
1         1         1      0    1      0
2         1         1      0    2      0
3         1         1      0    4      0
4         1         1      0    5      0
5         1         1      0    6      0
6         2         0      0    1      0
7         2         0      0    2      0
8         2         0      0    3      0
8723   1757         1      0    6      0
8724   1758         1      1    4      0
8725   1758         1      1    5      0
8726   1758         1      1    6      0
8727   1759         1      1    4      0
8728   1759         1      1    5      0
8729   1759         1      1    6      0
8730   1760         0      0    5      0
```

Data Exploration

```
summary(smoking)
     newid          sex.1.F.         parsmk          wave          smkreg
 Min.   :   1   Min.   :0.000   Min.   :0.00   Min.   :1.00   Min.   :0.000
 1st Qu.: 417   1st Qu.:0.000   1st Qu.:0.00   1st Qu.:2.00   1st Qu.:0.000
 Median : 830   Median :1.000   Median :0.00   Median :4.00   Median :0.000
 Mean   : 847   Mean   :0.542   Mean   :0.35   Mean   :3.68   Mean   :0.125
 3rd Qu.:1280   3rd Qu.:1.000   3rd Qu.:1.00   3rd Qu.:5.00   3rd Qu.:0.000
 Max.   :1760   Max.   :1.000   Max.   :1.00   Max.   :6.00   Max.   :1.000
table(smoking$wave)
  1    2    3    4    5    6
876 1571 1601 1587 1570 1525
cor(smoking)
            newid sex.1.F.     parsmk       wave   smkreg
newid    1.000000 0.085630  0.0229614  0.1660314 0.040525
sex.1.F. 0.085630 1.000000  0.0149581  0.0143264 0.045383
parsmk   0.022961 0.014958  1.0000000 -0.0072693 0.153724
wave     0.166031 0.014326 -0.0072693  1.0000000 0.076927
smkreg   0.040525 0.045383  0.1537244  0.0769268 1.000000
```

Varying Intercept Logit Multilevel Model, *No Group-Level Explanatory Variables*

```
lmer.out <- lmer(smkreg ~ wave + (1|newid), data=smoking,
                    family=binomial(link=logit))
display(lmer.out)
glmer(formula = smkreg ~ wave + (1 | newid), data = smoking,
    family = binomial(link = logit))
            coef.est coef.se
(Intercept) -6.41     0.27
wave         0.21     0.05

Error terms:
 Groups   Name        Std.Dev.
 newid    (Intercept) 4.17
 Residual             1.00
number of obs: 8730, groups: newid, 1760
AIC = 17774.7, DIC = 17769
deviance = 17768.7
```

Varying Intercept Logit Multilevel Model, *Adding Group-Level Explanatory Variables*

```
lmer.out <- lmer(smkreg ~ wave + sex.1.F. + parsmk + (1|newid), data=smoking,
                    family=binomial(link=logit))
display(lmer.out)
glmer(formula = smkreg ~ wave + sex.1.F. + parsmk + (1 | newid),
    data = smoking, family = binomial(link = logit))
             coef.est coef.se
(Intercept) -6.60     0.30
wave         0.25     0.04
sex.1.F.     0.05     0.32
parsmk       1.17     0.31

Error terms:
 Groups   Name        Std.Dev.
 newid    (Intercept) 4.18
 Residual             1.00
number of obs: 8730, groups: newid, 1760
AIC = 3935.8, DIC = 3925.8
deviance = 3925.8
```

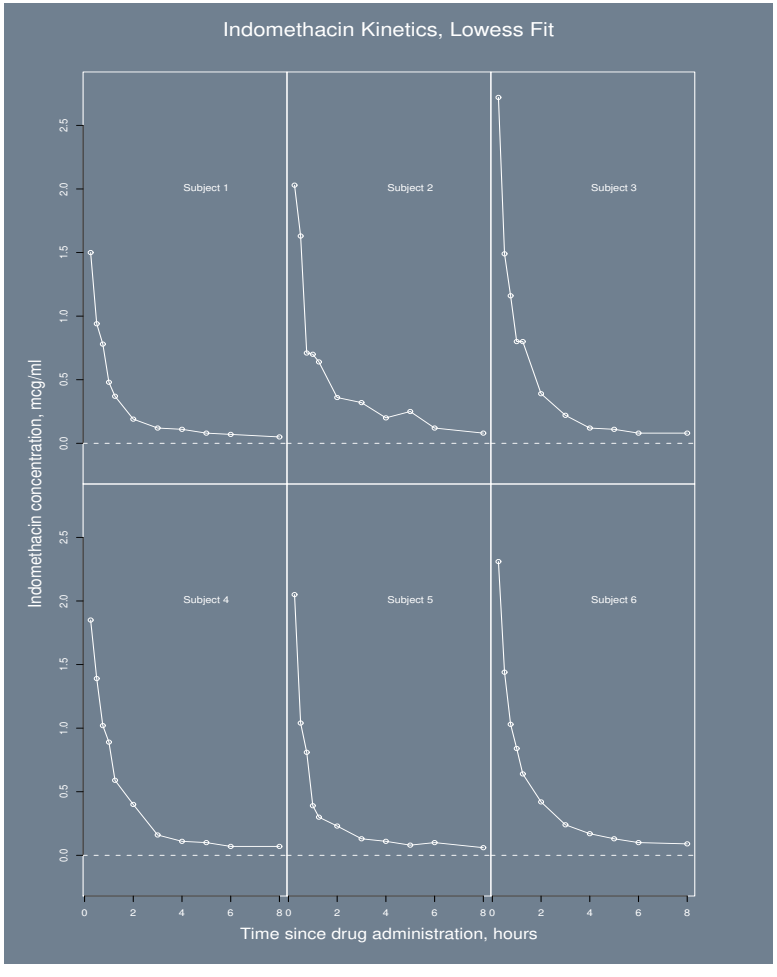
Varying-Intercept, Varying-Slope Logit Multilevel Model, *Drop Parents Smoking, Add an Interaction*

```
lmer.out <- lmer(smkreg ~ wave + sex.1.F. + wave:sex.1.F. + (1|newid), data=smoking,
                    family=binomial(link=logit))
display(lmer.out)
glmer(formula = smkreg ~ wave + sex.1.F. + wave:sex.1.F. + (1 |
    newid), data = smoking, family = binomial(link = logit))
             coef.est coef.se
(Intercept)   -4.85     0.26
wave           0.10     0.05
sex.1.F.      -0.88     0.36
wave:sex.1.F.  0.21     0.07

Error terms:
 Groups   Name        Std.Dev.
 newid    (Intercept) 3.47
 Residual             1.00
number of obs: 8730, groups: newid, 1760
AIC = 4086.4, DIC = 4076.4
deviance = 4076.4
```

Nonlinear Random Effects Example for Indomethacin Trials

- A pharmacokinetic analysis of the nonsteroidal anti-inflammatory drug Indomethacin from Kwan, Breault, Umbenhauer, McMahon, and Duggan, (1976) *Journal of Pharmacokinetics and Biopharmaceutics*, 4, 255-280.
- 6 volunteers received bolus intravenous injections of the same dose of Indomethacin.
- Plasma concentration (in mcg/ml) is measured 11 times between 15 minutes and 8 hours postinjection.
- Motivation for multilevel analysis: all subjects show a characteristic decay curve, but rates differ.



Nonlinear Random Effects Example for Indomethacin Trials

```
"Indometh" <-
  structure(list(
   Subject = structure(ordered(c(1, 1, 1, 1, 1, 1,
     1, 1, 1, 1, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 6, 6, 6, 6, 6,
     6, 6, 6, 6, 6, 6, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 4,
     4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5), levels=1:6),
     class = c("ordered", "factor"),
     .Label = c("1", "4", "2", "5", "6", "3")),
   time = c(0.25, 0.5, 0.75, 1, 1.25, 2, 3, 4, 5, 6, 8, 0.25, 0.5,
     0.75, 1, 1.25, 2, 3, 4, 5, 6, 8, 0.25, 0.5, 0.75, 1, 1.25, 2, 3,
     4, 5, 6, 8, 0.25, 0.5, 0.75, 1, 1.25, 2, 3, 4, 5, 6, 8, 0.25,
     0.5, 0.75, 1, 1.25, 2, 3, 4, 5, 6, 8, 0.25, 0.5, 0.75, 1, 1.25,
     2, 3, 4, 5, 6, 8),
   conc = c(1.5, 0.94, 0.78, 0.48, 0.37, 0.19, 0.12, 0.11,
     0.08, 0.07, 0.05, 2.03, 1.63, 0.71, 0.7, 0.64, 0.36, 0.32, 0.2,
     0.25, 0.12, 0.08, 2.72, 1.49, 1.16, 0.8, 0.8, 0.39, 0.22, 0.12,
     0.11, 0.08, 0.08, 1.85, 1.39, 1.02, 0.89, 0.59, 0.4, 0.16, 0.11,
     0.1, 0.07, 0.07, 2.05, 1.04, 0.81, 0.39, 0.3, 0.23, 0.13, 0.11,
     0.08, 0.1, 0.06, 2.31, 1.44, 1.03, 0.84, 0.64, 0.42, 0.24, 0.17,
     0.13, 0.1, 0.09)),
row.names = 1:66,
class = c("nfnGroupedData", "nfGroupedData", "groupedData", "data.frame"),
formula = conc ~ time | Subject,
labels = list(X = "Time since drug administration",
  y = "Indomethacin concentration"),
units = list(X = "(hr)", y = "(mcg/ml)"))
```

Nonlinear Random Effects Example for Indomethacin Trials

```
Indometh

Grouped Data: conc ~ time | Subject
   Subject time conc
1        1 0.25 1.50
2        1 0.50 0.94
:
65       6 6.00 0.10
66       6 8.00 0.09

names(Indometh)
"Subject" "time"    "conc"

class(Indometh)
"nfnGroupedData" "nfGroupedData" "groupedData" "data.frame"

formula(Indometh)
Subject ~ time + conc
```

Nonlinear Random Effects Example for Indomethacin Trials

- Biexponential model:

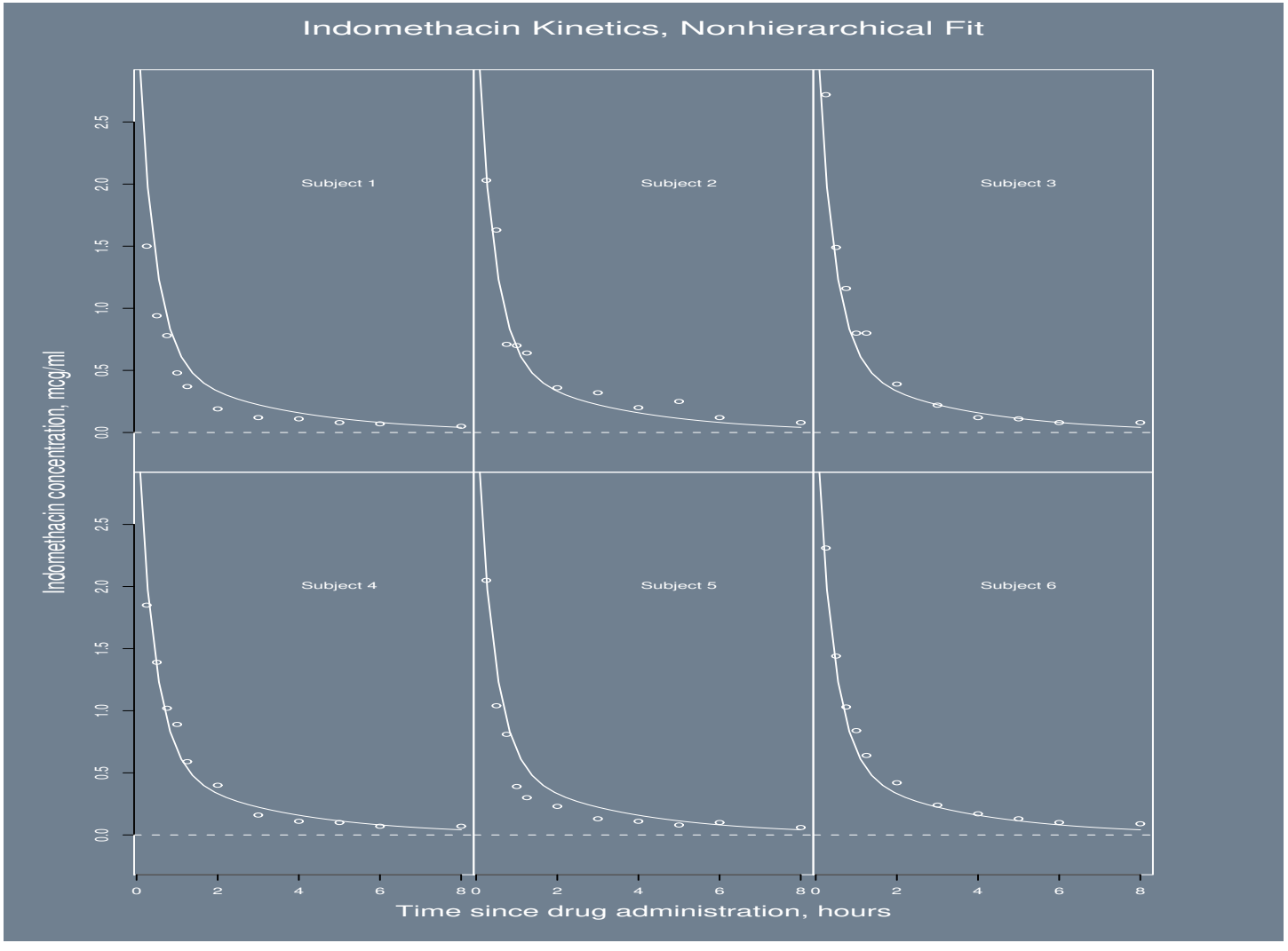
$$y_{ij} = \phi_1 \exp(-\exp(\phi_2)t_j) + \phi_3 \exp(-\exp(\phi_4)t_j) + \epsilon_{ij},$$

  with: individual i at time t_j, $e_{ij} \sim N(0, \sigma^2)$, $\phi_2 > \phi_4$ for identifiability.

- First fit a model ignoring hierarchy, getting fixed effect estimates (*full pooling*):

```
library(nlme)
indo.pop.nls <- nls(conc ~ SSbiexp(time, A1, lrc1, A2, lrc2), data = Indometh)
summary(indo.pop.nls)
Parameters:
     Estimate Std. Error t value Pr(>|t|)
A1      2.773      0.253   10.95    4e-16
lrc1    0.886      0.222    3.99  0.00018
A2      0.607      0.267    2.27  0.02660
lrc2   -1.092      0.409   -2.67  0.00966
---
Residual standard error: 0.1745 on 62 degrees of freedom
```

Nonlinear Random Effects Example for Indomethacin Trials



Nonlinear Random Effects Example for Indomethacin Trials

- Now fit a separate model to each individual subject ignoring population effects (*no pooling*).

```
indo.term.lis <- nlsList(conc ~ SSbiexp(time, A1, lrc1, A2, lrc2), data = Indometh)
summary(indo.term.lis)

Coefficients:
   A1
  Estimate Std. Error   t value     Pr(>|t|)
1 2.029277  0.2023875 10.026695 3.388285e-07
2 2.827673  0.2311604 12.232518 3.536510e-04
3 5.468312  1.8759966  2.914884 1.087070e-02
4 2.198132  0.3155032  6.967066 7.942588e-06
5 3.566103  0.3245732 10.987053 5.630248e-06
6 3.002250  0.3503106  8.570251 3.069467e-07
   lrc1
   Estimate Std. Error  t value     Pr(>|t|)
1 0.5793887  0.2295508 2.524011 2.347391e-03
2 0.8013195  0.1803742 4.442540 5.193756e-02
3 1.7497936  0.3108862 5.628406 2.937754e-04
4 0.2423124  0.2427792 0.998077 1.402737e-01
```

```
5 1.0407660  0.1636874 6.358253 1.986106e-04
6 1.0882119  0.2564197 4.243870 3.504364e-05
   A2
   Estimate Std. Error   t value     Pr(>|t|)
1 0.1915474  0.2037201 0.9402482 0.1269760002
2 0.4989175  0.1822390 2.7377104 0.1927034117
3 1.6757521  0.2814723 5.9535238 0.0002075745
4 0.2545223  0.3716832 0.6847828 0.2914158655
5 0.2914970  0.1592207 1.8307727 0.0811792830
6 0.9685230  0.2905245 3.3337056 0.0001646898

   lrc2
    Estimate Std. Error   t value     Pr(>|t|)
1 -1.7877849  1.4495070 -1.233374 0.0573694854
2 -1.6353512  0.4779239 -3.421781 0.1146482023
3 -0.4122004  0.1680153 -2.453351 0.0232031740
4 -1.6026860  1.4786607 -1.083877 0.1138959965
5 -1.5068522  0.7133811 -2.112268 0.0511506022
6 -0.8731358  0.2715939 -3.214858 0.0002066297

Residual standard error: 0.0755502 on 42 degrees of freedom # 66-24
```

Nonlinear Random Effects Example for Indomethacin Trials

- Finally we get a fully mixed effects biexponential model:

$$y_{ij} = (\beta_1 - b_{1i}\exp(-\exp(\beta_2 - b_{2i})t_j) + (\beta_3 - b_{3i}\exp(-\exp(\beta_4 - b_{4i})t_j + \epsilon_{ij}$$

  where the β's give the estimated mean population effects, and the b_i's give the individual deviations, with assumed mean zero.

- Also we assume no covariances by using only the diagonal for computation.

```
indo.nlme <- nlme( indo.term.lis, random = pdDiag(A1 + lrc1 + A2 + lrc2 ~ 1) )

summary(indo.nlme)

Data: Indometh
        AIC       BIC   logLik
  -91.18562 -71.47873 54.59281
```

```
Random effects:
 Formula: list(A1 ~ 1, lrc1 ~ 1, A2 ~ 1, lrc2 ~ 1)
 Level: Subject
 Structure: Diagonal
             A1    lrc1      A2       lrc2 Residual
StdDev: 0.57141 0.15808 0.11160 8.2778e-06 0.081493

Fixed effects: list(A1 ~ 1, lrc1 ~ 1, A2 ~ 1, lrc2 ~ 1)
        Value Std.Error DF t-value p-value
A1    2.82754   0.26401 57 10.7099   0e+00
lrc1  0.77362   0.11003 57  7.0313   0e+00
A2    0.46147   0.11281 57  4.0908   1e-04
lrc2 -1.34410   0.23108 57 -5.8167   0e+00

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max
-3.17338 -0.35627 -0.12853  0.34232  3.00251

Number of Observations: 66
Number of Groups: 6
```

Nonlinear Random Effects Example for Indomethacin Trials

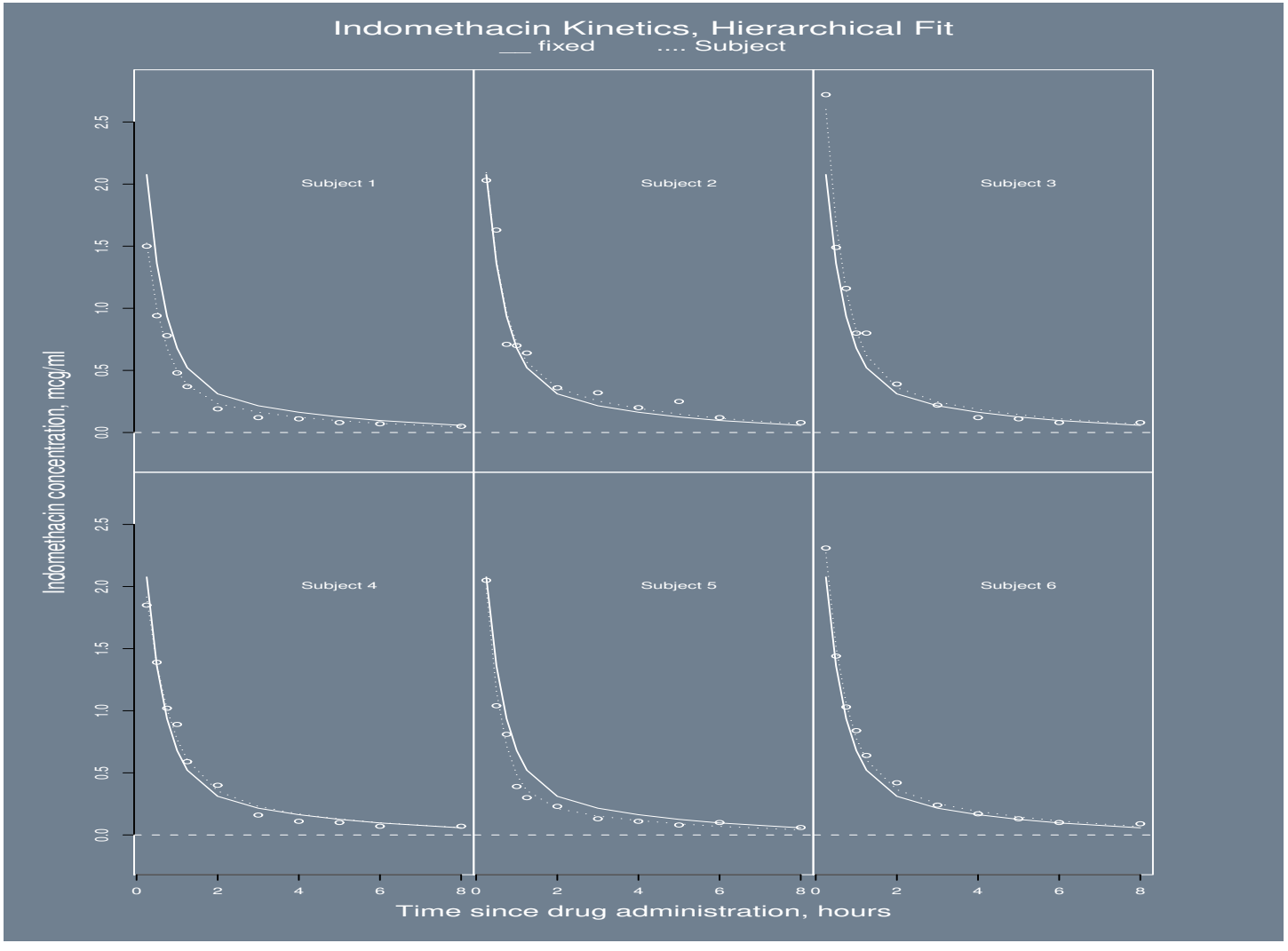
```
( be.fixed.terms <- as.vector(indo.nlme$coefficients$fixed) )

[1]  2.8275982  0.7732937  0.4610679 -1.3449124

( be.random.terms <- as.matrix(indo.nlme$coefficients$random$Subject) )

           A1          lrc1          A2          lrc2
1 -0.73966229  0.027813190 -0.11124141  6.553013e-17
2 -0.07054535  0.028667640  0.08711800 -1.855721e-16
3  0.80060558  0.004134323  0.06711258  4.906331e-17
4 -0.56554287 -0.231256957  0.01169103  7.287592e-17
5  0.41273288  0.198153725 -0.13174119  3.397749e-17
6  0.16241204 -0.027511921  0.07706099 -3.587477e-17
```

Nonlinear Random Effects Example for Indomethacin Trials



Data Exercise 8: Functional Data Analysis

- Rerun the Indomethacin hierarchical functional data analysis model.
- Get the data from above but cut-and-paste or get from:

`http://jeffgill.org/files/jeffgill/files/indometh.dat_.txt`

- Retype the simple commands from the slides making your changes.
- Graph the results of the final model against the non-hierarchical model.
- Be creative about graphing coefficients, predictions, etc.

Nested Classification Factors

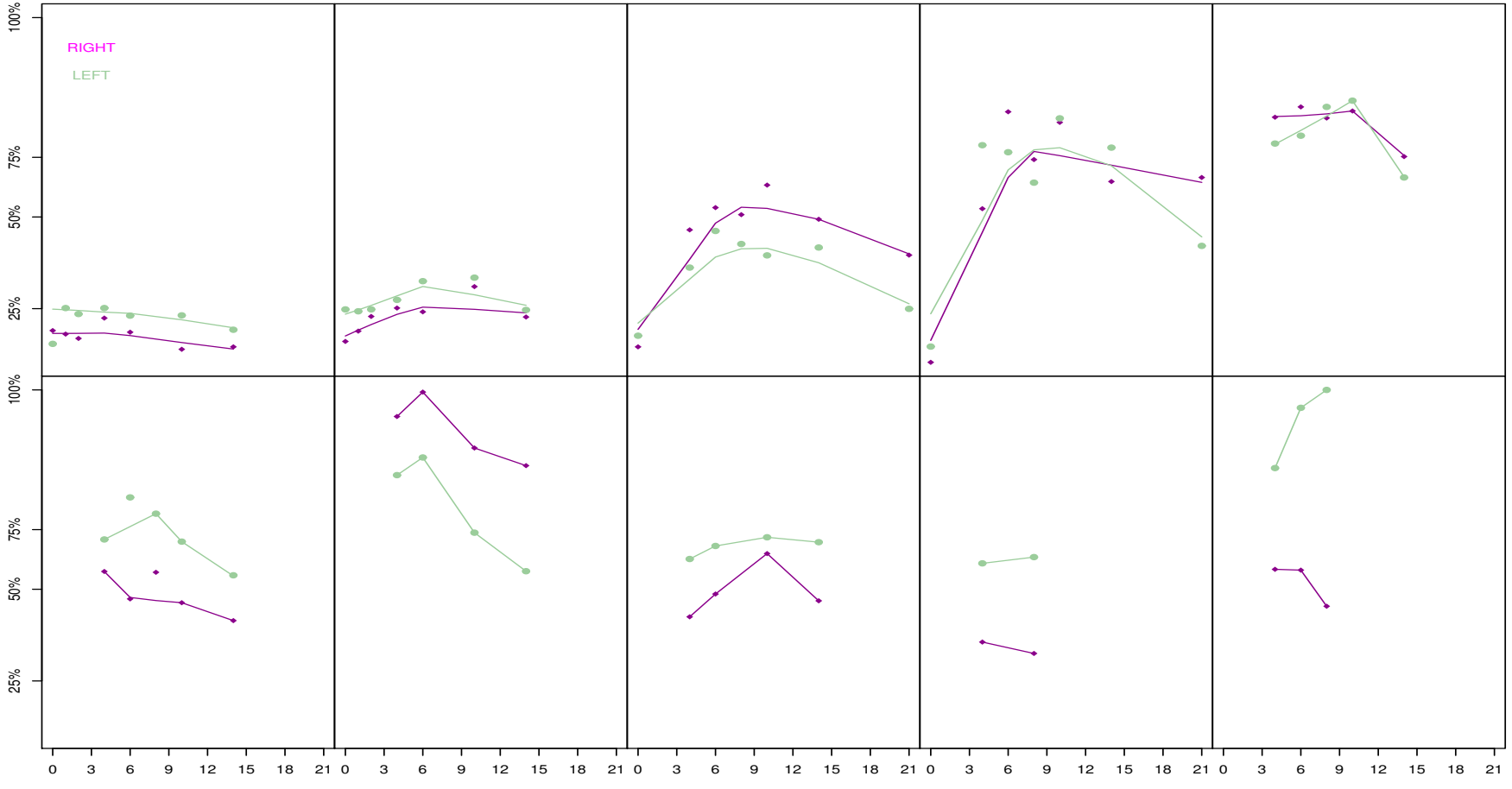
- Experiment recording mean pixel intensity of the R and L lymph-nodes in the auxiliary region from (tomography) CT scans of 10 dogs over 14 days after intravenous application of a dye contrast.
- Data: go to

  `http://jeffgill.org/classes/applied-bayesian-data-analysis?admin_panel=1`

  and download the zip file Pixel.zip, unzip it into your directory, and use `load(Pixel.rda` to load the file.

```
Pixel[,"Side"] <- as.numeric( Pixel[,"Side"] )
Pixel[1:10,]
   Dog Side day  pixel
1    1    2   0 1045.8
2    1    2   1 1044.5
3    1    2   2 1042.9
4    1    2   4 1050.4
5    1    2   6 1045.2
6    1    2  10 1038.9
```

Nested Classification Factors



Nested Classification Factors

```
postscript("Class.Multilevel/dogs1.ps")
J = max(as.numeric(Pixel$Dog))
day.range <- c(min(as.numeric(Pixel$day)), max(as.numeric(Pixel$day)))
pixel.range <- c(min(as.numeric(Pixel$pixel)), max(as.numeric(Pixel$pixel)))
par(oma=c(5,5,1,1),mar=c(0,0,0,0),mfrow=c(J/(J/2),J/2),bg="white")
for (i in 1:J)  {
    dog.i <- Pixel[Pixel["Dog"]==i,]
    plot(dog.i[dog.i["Side"]==2,][,3:4], xlim=day.range, ylim=pixel.range,
        pch=18,col="darkmagenta",yaxt="n",xaxt="n")
    lines(lowess(dog.i[dog.i["Side"]==2,][,3:4], f=0.9), col="darkmagenta")
    points(dog.i[dog.i["Side"]==1,][,3:4], pch=19,col="darkseagreen3")
    lines(lowess(dog.i[dog.i["Side"]==1,][,3:4], f=0.9), col="darkseagreen3")
    if (i > 5) axis(side=1,labels=seq(0,21,by=3),at=seq(0,21,by=3))
    if (i == 1 | i==6) axis(side=2,labels=names(quantile(Pixel$pixel)[-1]),
        at=quantile(Pixel$pixel)[-1])
    if (i == 1)  { text(3,1150,"RIGHT",col="magenta")
                   text(3,1140,"LEFT",col="darkseagreen3") }
}
dev.off()
```

Nested Classification Factors

- We want to account for nesting of sides with dogs.
- We also need to think carefully about where to put intercepts in the various levels (choices).
- Note that with nesting levels we can qualitatively different results for standard error terms.
- There is an obvious nonlinear effect to account for in days with a peak at about 10.
- Most general model:

$$
\begin{aligned}
y_{ijt} &= \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 S_{ij} + \beta_{4,j[i]} + \beta_{5,j[i]} + \epsilon_{ijt} \\
\beta_4 &\sim N(\gamma_0 + \gamma_1 S_{ij} + \gamma_2 T_i, \sigma^2_{\beta_4,}) \\
\beta_5 &\sim N(\delta_0 + \delta_1 T_i, \sigma^2_{\beta_5,})
\end{aligned}
$$

where:

y_{ijt}	pixels for ith dog, side j, at time t
T_{ij}	ith dog's time t (not all measured at regular intervals!)
S_{ij}	ith dog's side: 1 (left) or 2 (right).

Nested Classification Factors

- Random intercepts model:

```
fm1Pixel <- lmer(pixel ~ day + I(day^2) + (1 | Dog), data = Pixel)
summary(fm1Pixel)
   AIC   BIC logLik deviance REMLdev
 895.2 908.4 -442.6    886.9   885.2
Random effects:
 Groups   Name        Variance Std.Dev.
 Dog      (Intercept) 633.89   25.177
 Residual             263.95   16.247
Number of obs: 102, groups: Dog, 10

Fixed effects:
             Estimate Std. Error t value          Correlation of Fixed Effects:
(Intercept) 1074.16253    9.12496  117.72                    (Intr) day
day            4.94201    1.03686    4.77          day       -0.426
I(day^2)      -0.25047    0.05303   -4.72          I(day^2)   0.355 -0.945
```

Nested Classification Factors

- Checking for a non-grouped difference between the left and right sides across dogs:

```
fm2Pixel <- lmer(pixel ~ day + I(day^2) + Side + (1 | Dog), data = Pixel)
summary(fm2Pixel)
   AIC BIC logLik deviance REMLdev
 890.2 906 -439.1      884   878.2
Random effects:
 Groups   Name        Variance Std.Dev.
 Dog      (Intercept) 634.54   25.190
 Residual             258.56   16.080
Number of obs: 102, groups: Dog, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept) 1082.28413   10.28304  105.25
day            4.93811    1.02628    4.81
I(day^2)      -0.25029    0.05249   -4.77
Side          -5.40196    3.18425   -1.70

Correlation of Fixed Effects:
            (Intr) day    I(d^2)
day         -0.374
I(day^2)     0.311 -0.945
Side        -0.464  0.000  0.000
```

Nested Classification Factors

```
anova(fm2Pixel,fm1Pixel)

fm1Pixel: pixel ~ day + I(day^2) + (1 | Dog)
fm2Pixel: pixel ~ day + I(day^2) + Side + (1 | Dog)
          Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm1Pixel  5  896.87  910.00 -443.44
fm2Pixel  6  895.94  911.69 -441.97 2.9352      1    0.08667
```

Nested Classification Factors

- A model that nests the side in dogs as well as the days in dogs, days has intercept:

```
fm3Pixel <- lmer(pixel ~ day + I(day^2) + (Side | Dog) + (-1 + day | Dog),
                 data = Pixel); summary(fm3Pixel)

   AIC   BIC logLik deviance REMLdev
 843.4 864.4 -413.7    829.7   827.4
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 Dog      (Intercept) 1937.2268 44.014
          Side         565.3087 23.776   -0.746
 Dog      day            3.1791  1.783
 Residual                81.3617  9.020
Number of obs: 102, groups: Dog, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept) 1073.6530     9.7331  110.31
day            6.2335     0.8655    7.20
I(day^2)      -0.3685     0.0340  -10.84

Correlation of Fixed Effects:
            (Intr) day
day         -0.202
I(day^2)     0.195 -0.679
```

Nested Classification Factors

```
anova( fm3Pixel, fm2Pixel )

fm2Pixel: pixel ~ day + I(day^2) + Side + (1 | Dog)
fm3Pixel: pixel ~ day + I(day^2) + (Side | Dog) + (-1 + day | Dog)
          Df     AIC     BIC  logLik Chisq Chi Df Pr(>Chisq)
fm2Pixel  6  895.94  911.69 -441.97
fm3Pixel  8  845.69  866.69 -414.84 54.25      2  1.658e-12
```

Nested Classification Factors

- Is it worth having **Side** as a fixed effect too?

```
fm4Pixel <- lmer(pixel ~ day + I(day^2) + Side + (Side | Dog) + (-1 + day | Dog),
                 data = Pixel); summary(fm4Pixel)
 AIC   BIC logLik deviance REMLdev
 838 861.7   -410    828.2     820
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 Dog      (Intercept) 1908.1242 43.6821
          Side         544.6817 23.3384  -0.741
 Dog      day            3.1794  1.7831
 Residual               81.2434  9.0135
Number of obs: 102, groups: Dog, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept) 1086.41495   14.41456   75.37
day            6.23301    0.86512    7.20
I(day^2)      -0.36852    0.03398  -10.85
Side          -9.19399    7.62640   -1.21

Correlation of Fixed Effects:
            (Intr) day    I(d^2)
day         -0.136
I(day^2)     0.131 -0.679
Side        -0.737  0.000  0.000
```

Nested Classification Factors

```
anova( fm4Pixel, fm3Pixel )

fm3Pixel: pixel ~ day + I(day^2) + (Side | Dog) + (-1 + day | Dog)
fm4Pixel: pixel ~ day + I(day^2) + Side + (Side | Dog) + (-1 + day | Dog)
          Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm3Pixel  8  845.69  866.69 -414.84
fm4Pixel  9  846.20  869.83 -414.10 1.4834      1     0.2232
```