Treating Time More Flexibly

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

GCM, 2010

Treating Time More Flexibly

Introduction

Our introductory examples have shared some simplifying features. Each is:

1. *Balanced.* Each individual is assessed an equal number of times.
2. *Time-Structured.* Each set of occasions is identical across individuals.

Moreover, we have used only:

1. *Time-Invariant Predictors.*
2. *A Standard Time Representation* which led to an easy interpretation of parameters.

Introduction

The multilevel change model can handle more ambitious examples, where the data are not necessarily either balanced or time-structured. Moreover, we can include time-varying predictors.

Singer and Willett begin their Chapter 5 with a discussion of the difficulties of obtaining time-structured and balanced data in the real world.

Psychological Consequences of Unemployment

Example (Psychological Consequences of Unemployment)

- Ginexi, Howe, and Caplan (2000) designed a time-structured study with interviews *scheduled* a 1, 5, and 11 months after job loss.
- Once in the field, however, the interview times varied considerably around these targets, with increasing variability as the study proceeded
- First interview (2–61 days), Second interview (111–220 days), Third interview (319–458 days)
- Ginexi et al. argued that number of days rather than target time should be used.
- As a result, data were not time-structured

Accelerated Cohort Design

Example (Accelerated Cohort Design)

- Age-heterogeneous group is followed for a constant period of time
- Age is the appropriate time measure
- Different people are interviewed at different ages, for example
  - $14.2 \rightarrow 15.2 \rightarrow 16.2$
  - $15.7 \rightarrow 16.7 \rightarrow 17.7$

The CNLSY Study

Singer and Willett illustrate the structure of variably spaced data with an example from the Children of the National Longitudinal Study of Youth (CNLSY).

- The study assessed 3 waves of data on 89 African-American kids
- Ages 6.5,8.5,10.5.
- Outcome variable was the reading subtest of the Peabody Individual Achievement Test (PIAT)
- Actual times of measurement were unstructured.

We'll jump to their slide set for a discussion of the example, then return for an analysis in R.

The CNLSY Study – AGE Model

```
> data <- read.table("reading_pp.txt",header=T,sep=",")
> attach(data)
> library(lme4)
> age_c <- age - 6.5
> agegrp_c <- agegrp - 6.5
> fit.age <- lmer(piat ~ age_c + (1+age_c|id),REML=FALSE)
> fit.age

Linear mixed model fit by maximum likelihood
Formula: piat ~ age_c + (1 + age_c | id)
  AIC  BIC logLik deviance REMLdev
 1816 1837   -902     1804    1804
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept)  5.11    2.26
          age_c        3.30    1.82     0.576
 Residual             27.45    5.24
Number of obs: 267, groups: id, 89

Fixed effects:
            Estimate Std. Error t value
(Intercept)   21.061      0.559    37.7
age_c          4.540      0.261    17.4

Correlation of Fixed Effects:
      (Intr)
age_c -0.287
```

The CNLSY Study – AGEGRP Model

```
> fit.agegrp <- lmer(piat ~ agegrp_c + (1+agegrp_c|id),REML=FALSE)
> fit.agegrp

Linear mixed model fit by maximum likelihood
Formula: piat ~ agegrp_c + (1 + agegrp_c | id)
  AIC  BIC logLik deviance REMLdev
 1832 1853   -910     1820    1820
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 11.0     3.32
          agegrp_c     4.4     2.10     0.236
 Residual             27.0     5.20
Number of obs: 267, groups: id, 89

Fixed effects:
            Estimate Std. Error t value
(Intercept)   21.163      0.614    34.5
agegrp_c       5.031      0.296    17.0

Correlation of Fixed Effects:
         (Intr)
agegrp_c -0.316
```

The NLSY Wages Study – Model A

This is an unconditional growth model.

```
> detach(data)
> data <- read.table("wages_pp.txt",header=T,sep=",")
> attach(data)
> hgc_9 <- hgc - 9
> fit.A <- lmer(lnw ~ exper + (1 + exper | id), REML=FALSE)
> fit.A

Linear mixed model fit by maximum likelihood
Formula: lnw ~ exper + (1 + exper | id)
  AIC  BIC logLik deviance REMLdev
 4933 4974  -2461     4921    4939
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.05427  0.2330
          exper       0.00173  0.0415   -0.301
 Residual             0.09510  0.3084
Number of obs: 6402, groups: id, 888

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.71560    0.01080   158.9
exper        0.04568    0.00234    19.5

Correlation of Fixed Effects:
      (Intr)
exper -0.565
```

The NLSY Wages Study – Model B

This model uses `black` and `hgc_9` to predict slopes and intercepts of the individual's trajectory.

```
> fit.B <- lmer(lnw~exper+black+hgc_9+black:exper +hgc_9:exper + (1+exper|id),REML=FALSE)
> fit.B

Linear mixed model fit by maximum likelihood
Formula: lnw ~ exper + black + hgc_9 + black:exper + hgc_9:exper + (1 +      exper | id)
  AIC  BIC logLik deviance REMLdev
 4894 4961  -2437     4874    4925
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.05175  0.2275
          exper       0.00164  0.0404   -0.310
 Residual             0.09519  0.3085
Number of obs: 6402, groups: id, 888

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.71714    0.01254   136.9
exper        0.04934    0.00263    18.7
black        0.01540    0.02393     0.6
hgc_9        0.03492    0.00788     4.4
exper:black -0.01821    0.00550    -3.3
exper:hgc_9  0.00128    0.00172     0.7

Correlation of Fixed Effects:
            (Intr) exper  black  hgc_9  expr:b
exper       -0.575
black       -0.523  0.301
hgc_9        0.071 -0.020 -0.020
exper:black  0.275 -0.478 -0.573  0.011
exper:hgc_9 -0.019 -0.003  0.011 -0.578 -0.023
```

The NLSY Wages Study – Model C

This “pared-back” model uses `black` to predict only the intercepts and `hgc_9` to predict only the slopes of the individual’s trajectory.

```
> fit.C <- lmer(lnw~exper+hgc_9+black:exper + (1+exper|id),REML=FALSE)
> fit.C

Linear mixed model fit by maximum likelihood
Formula: lnw ~ exper + hgc_9 + black:exper + (1 + exper | id)
  AIC  BIC logLik deviance REMLdev
 4891 4945  -2437     4875    4910
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.05183  0.2277
          exper       0.00165  0.0406   -0.312
 Residual             0.09517  0.3085
Number of obs: 6402, groups: id, 888

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.72147    0.01070   160.9
exper        0.04885    0.00251    19.4
hgc_9        0.03836    0.00643     6.0
exper:black -0.01612    0.00451    -3.6

Correlation of Fixed Effects:
            (Intr) exper  hgc_9
exper       -0.515
hgc_9        0.077 -0.023
exper:black -0.036 -0.391 -0.015
```

The NLSY Wages Study – Model C – Reduced Data

To demonstrate convergence problems, Model C was also fit to a reduced data set.

```
> detach(data)
> data <- read.table("wages_small_pp.txt",header=T,sep=",")
> attach(data)
> fit.C.small <- lmer(lnw~exper+hcg.9+black:exper + (1+exper|id),REML=FALSE)
> fit.C.small

Linear mixed model fit by maximum likelihood
Formula: lnw ~ exper + hcg.9 + black:exper + (1 + exper | id)
 AIC BIC logLik deviance REMLdev
 300 328   -142      284     305
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 8.22e-02 0.28662
          exper       3.52e-06 0.00188  1.000
 Residual             1.15e-01 0.33907
Number of obs: 257, groups: id, 124

Fixed effects:
            Estimate Std. Error t value
(Intercept)   1.7373     0.0476    36.5
exper         0.0516     0.0211     2.4
hcg.9         0.0461     0.0245     1.9
exper:black  -0.0597     0.0348    -1.7

Correlation of Fixed Effects:
            (Intr) exper  hcg.9
exper       -0.612
hcg.9        0.051 -0.133
exper:black -0.129 -0.297  0.023
```

Models for Missing Data

Certain kinds of missing data can be handled effectively by special methods. Some of the key *Random Component Selection Models* models for missing data include:

1. Missing Completely at Random (MCAR)
2. Covariate Dependent Dropout (CDD)
3. Missing at Random (MAR)

Missing Completely at Random

Suppose we denote the potential outcome variable by $\boldsymbol{y}_i$, the random effect coefficients by $\boldsymbol{b}_i$, and the covariates as $\boldsymbol{X}_i$. The missingness mechanism is modeled as a random process R_i. When data are *missing completely at random* (MCAR), then

$$[R_i|\boldsymbol{X}_i, \boldsymbol{y}_i, \boldsymbol{b}_i] = [R_i] \tag{1}$$

That is, the missingness mechanism is independent of the covariates, the outcome, and the random coefficients or, in other words, completely random.

Covariate Dependent Dropout

When data show covariate dependent dropout (CDD), we have

$$[R_i|\boldsymbol{X}_i, \boldsymbol{y}_i, \boldsymbol{b}_i] = [R_i|\boldsymbol{X}_i] \tag{2}$$

That is, the missingness mechanism is independent of the outcome and the random coefficients given the covariates. This model allows dependence of drop-out on both between-subject and within-subject covariates that can be treated as fixed in the model.

Missing at Random

Data are *Missing at Random* (MAR) if the distribution of the dropout mechanism depends on $\boldsymbol{y}_i$ only through its observed components $\boldsymbol{y}_{obs,i}$. That is

$$[R_i|\boldsymbol{X}_i, \boldsymbol{y}_{obs,i}, \boldsymbol{y}_{mis,i}\boldsymbol{b}_i] = [R_i|\boldsymbol{X}_i, \boldsymbol{y}_{obs,i}] \tag{3}$$

What to Do?

If a reasonable case can be made that the missing data mechanism is MCAR, CDD, or MAR, then ML methods applied to all the data will work well. However, if missingness depends on the random coefficients themselves or on the unobserved values in a way that cannot be predicted from covariates, then special approaches may be necessary.

This is a complex topic, probably worthy of a course in itself. The books by Joe Shafer and Little and Rubin, and the 1995 JASA article (vol 90, pp. 1112–1121, available online) are primary references.

What to Do?

A MCAR test is available, and rejecting the null hypothesis rejects the MCAR assumption. However, since the goal is *not* to reject, the standard caveats about Accept-Support testing apply.

If missingness is clearly non-ignorable, you need to either model the mechanism or use a pattern mixture model.

Time-Varying Predictors

Time-varying predictors can change values at any recording instance.

Fortunately, the person-period data format handles such data effortlessly.

The Ginexi et al. Unemployment Study

This study examined the relationship over time between unemployment and depression.

```
> detach(data)
> data <- read.table("unemployment_pp.txt",
+ header=T,sep=",")
> attach(data)
```

(Jump to Singer-Willett Chapter 5 slide set.)

Model A – An Unconditional Growth Model

$$Y_{ij} = \pi_{0i} + \pi_{1i}\,TIME_{ij} + \epsilon_{ij}$$

with

$$\begin{aligned}\pi_{0i} &= \gamma_{00} + \zeta_{0i}\\ \pi_{1i} &= \gamma_{10} + \zeta_{1i}\end{aligned}$$

and the standard assumption. Substituting, we get the model

$$Y_{ij} = \gamma_{00} + \gamma_{10}\,TIME_{ij} + \zeta_{0i} + \zeta_{1i}\,TIME_{ij} + \epsilon_{ij}$$

Fitting Model A

```
> fit.A <- lmer(cesd ~ 1 + months +
+ (1+months|id),REML=FALSE)
> fit.A

Linear mixed model fit by maximum likelihood
Formula: cesd ~ 1 + months + (1 + months | id)
  AIC  BIC logLik deviance REMLdev
 5145 5172  -2567     5133    5135
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 86.848   9.319
          months       0.355   0.596    -0.551
 Residual             68.850   8.298
Number of obs: 674, groups: id, 254

Fixed effects:
            Estimate Std. Error t value
(Intercept)   17.669      0.776   22.78
months        -0.422      0.083   -5.09

Correlation of Fixed Effects:
       (Intr)
months -0.632
```

Model B – Adding Unemployment as a Time-Varying Predictor

Next, unemployment is added as a direct level-1 predictor, yielding the composite model

$$Y_{ij} = \gamma_{00} + \gamma_{10}\,TIME_{ij} + \gamma_{20}\,UNEMP_{ij} + \zeta_{0i} + \zeta_{1i}\,TIME_{ij} + \epsilon_{ij}$$

Fitting Model B

```
> fit.B <- lmer(cesd ~ 1 + months +
+ unemp + (1+months|id),REML=FALSE)
> fit.B

Linear mixed model fit by maximum likelihood
Formula: cesd ~ 1 + months + unemp + (1 + months | id)
  AIC  BIC logLik deviance REMLdev
 5122 5153  -2554     5108    5108
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 93.519   9.671
          months       0.465   0.682    -0.591
 Residual             62.388   7.899
Number of obs: 674, groups: id, 254

Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6656     1.2421   10.20
months       -0.2020     0.0933   -2.16
unemp         5.1113     0.9888    5.17

Correlation of Fixed Effects:
       (Intr) months
months -0.715
unemp  -0.780  0.459
```

Model C – Allowing the Effect of Unemployment to Vary over Time

Next, the effect of unemployment is allowed to change over time via the addition of an interaction term.

$$
\begin{aligned}
Y_{ij} \quad = \quad & \gamma_{00} + \gamma_{10}\, TIME_{ij} + \gamma_{20}\, UNEMP_{ij} + \\
& \gamma_{30}\, UNEMP_{ij} \times TIME_{ij} + \zeta_{0i} + \zeta_{1i}\, TIME_{ij} + \epsilon_{ij}
\end{aligned}
$$

Fitting Model C

```
> fit.C <- lmer(cesd ~ 1 + months +
+ unemp + months:unemp + (1+months|id),REML=FALSE)
> fit.C

Linear mixed model fit by maximum likelihood
Formula: cesd ~ 1 + months + unemp + months:unemp + (1 + months | id)
  AIC  BIC logLik deviance REMLdev
 5119 5155  -2552     5103    5105
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 93.713   9.681
          months       0.451   0.672    -0.596
 Residual             62.031   7.876
Number of obs: 674, groups: id, 254

Fixed effects:
             Estimate Std. Error t value
(Intercept)     9.617      1.889    5.09
months          0.162      0.194    0.84
unemp           8.529      1.878    4.54
months:unemp   -0.465      0.217   -2.14

Correlation of Fixed Effects:
            (Intr) months unemp
months      -0.888
unemp       -0.911  0.863
months:unmp  0.755 -0.878 -0.852
```

Model D – Constraining the Trajectory of the Employed

In this model, the trajectory is constrained to have a zero slope when individuals are employed.

This is done by including both a main effect for unemployment and an interaction effect between unemployment and time at both the fixed and random levels, and removing the fixed and random effects for time.

Since unemployment is a binary variable, the net effect of this is that when unemployment is 1, the interaction effect solely determines the slope of the relationship between Y and time. When unemployment is zero, there is no slope term, and so the slope effectively becomes zero.

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{20} UNEMP_{ij} + \gamma_{30} UNEMP_{ij} \times TIME_{ij} \\ &\quad + \zeta_{0i} + \zeta_{2i} UNEMP_{ij} + \zeta_{3i} UNEMP_{ij} \times TIME_{ij} + \epsilon_{ij} \end{aligned}$$

Fitting Model C

```
> fit.D <- lmer(cesd ~ 1 + unemp +
+ months:unemp + (1+unemp + months:unemp|id),REML=FALSE)
> fit.D

Linear mixed model fit by maximum likelihood
Formula: cesd ~ 1 + unemp + months:unemp + (1 + unemp + months:unemp |      id)
  AIC  BIC logLik deviance REMLdev
 5115 5160  -2548     5095    5096
Random effects:
 Groups   Name         Variance Std.Dev. Corr
 id       (Intercept)  45.254   6.727
          unemp        44.968   6.706     0.145
          unemp:months  0.753   0.868     0.112 -0.967
 Residual              59.018   7.682
Number of obs: 674, groups: id, 254

Fixed effects:
             Estimate Std. Error t value
(Intercept)    11.195      0.790   14.17
unemp           6.927      0.930    7.45
unemp:months   -0.303      0.112   -2.70

Correlation of Fixed Effects:
            (Intr) unemp
unemp       -0.563
unemp:mnths -0.074 -0.443
```

Recentering the Effects of Time

So far, time has been centered on the initial status point.

However, other alternatives are possible, and any meaningful constant can be used.

Singer and Willett discuss some options in the context of a study by Tomarken, et al. (1997).

The Effect of Treatment on Mood over Time

The composite model is

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01} TREAT_i + \gamma_{10} TIME_{ij} \\ &\quad + \gamma_{11} TREAT_i \times TIME_{ij} + \epsilon_{ij} + (\zeta_{1i} TIME_{ij} + \zeta_{0i}) \end{aligned}$$

Fitting the Model

```
> detach(data)
> data <- read.table("medication_pp.txt",header=T,sep=",")
> attach(data)
> fit.initial <- lmer(pos ~ treat + time + treat:time + (1 + time | id),REML=FALSE)
> fit.initial

Linear mixed model fit by maximum likelihood
Formula: pos ~ treat + time + treat:time + (1 + time | id)
   AIC   BIC logLik deviance REMLdev
 12696 12737  -6340    12680   12663
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 2111.4   45.95
          time          63.7    7.98    -0.332
 Residual             1229.9   35.07
Number of obs: 1242, groups: id, 64

Fixed effects:
            Estimate Std. Error t value
(Intercept)   167.46       9.33   17.96
treat          -3.11      12.33   -0.25
time           -2.42       1.73   -1.40
treat:time      5.54       2.28    2.43

Correlation of Fixed Effects:
           (Intr) treat  time
treat      -0.756
time       -0.404  0.305
treat:time  0.307 -0.408 -0.760
```

Fitting the Model Centered at Midpoint

```
> fit.midpoint <- lmer(pos ~ treat + time333 + treat:time333 + (1 + time333 | id),REML=FALSE)
> fit.midpoint

Linear mixed model fit by maximum likelihood
Formula: pos ~ treat + time333 + treat:time333 + (1 + time333 | id)
   AIC   BIC logLik deviance REMLdev
 12696 12737  -6340    12680   12663
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 2008.8   44.82
          time333       63.7    7.98    0.254
 Residual             1229.9   35.07
Number of obs: 1242, groups: id, 64

Fixed effects:
              Estimate Std. Error t value
(Intercept)     159.40       8.76   18.19
treat            15.35      11.54    1.33
time333          -2.42       1.73   -1.40
treat:time333     5.54       2.28    2.43

Correlation of Fixed Effects:
            (Intr) treat  tim333
treat       -0.759
time333      0.229 -0.173
treat:tm333 -0.174  0.221 -0.760
```

Fitting the Model Centered at Endpoint

```
> fit.endpoint <- lmer(pos ~ treat + time667 + treat:time667 + (1 + time667 | id),REML=FALSE)
> fit.endpoint

Linear mixed model fit by maximum likelihood
Formula: pos ~ treat + time667 + treat:time667 + (1 + time667 | id)
   AIC   BIC logLik deviance REMLdev
 12696 12737  -6340    12680   12663
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 3322.5   57.64
          time667       63.7    7.98    0.659
 Residual             1229.9   35.07
Number of obs: 1242, groups: id, 64

Fixed effects:
              Estimate Std. Error t value
(Intercept)     151.34      11.54   13.11
treat            33.80      15.16    2.23
time667          -2.42       1.73   -1.40
treat:time667     5.54       2.28    2.43

Correlation of Fixed Effects:
            (Intr) treat  tim667
treat       -0.761
time667      0.673 -0.513
treat:tm667 -0.512  0.670 -0.760
```