Modeling Discontinuous and Nonlinear Change

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

GCM, 2010

Modeling Discontinuous and Nonlinear CHange

Introduction

So far we have been investigating individual growth using a linear model that, in many applied situations, is an unacceptable oversimplification.

Individual change can be nonlinear. Depending on the process, a linear model can be predicted to be wrong on the basis of simple logical considerations.

Individual change can also be discontinuous. For example, an intervention can produce a sudden, permanent change in behavior.

In this module, we investigate models for fitting discontinuous and nonlinear change processes.

Wage Trajectories and the GED – Shift in Elevation Only

Singer and Willett investigate the modeling of discontinuous change in the context of a study of the effect of attaining a GED on log wages (Mournane, Bourdett, & Willett, 1999).

The first notion they explore is a discontinuity in slope only. Modeling this is straightforward — one simply adds GED, coded as a binary 0–1 variable, as a time-varying predictor at level-1.

When GED is 0, it essentially disappears from the equation.

When GED is 1, it adds a fixed component to the intercept, thus creating a discontinuous shift in elevation of the trajectory. (SW6 Slides 3–5).

Shift in Slope Alone

In order to model a shift in slope that is *not* accompanied by a corresponding shift in elevation, Singer and Willett rely on a neat trick. They add an additional temporal predictor that is actually time recentered in terms of the shift point.

In-Class Group Exercise

In slide 6 of their Chapter 6 powerpoints, there is an algebraic error that is fairly obvious. Although their meaning is clear and their interpretation correct, the equation is technically incorrect. Singer and Willett themselves point out in the recorded lecture that there is an error, but do not say what it is. Put your heads together, detail the error, and explain it geometrically in terms of the red lines in slide 6.

Discontinuities in Slope *and* Intercept

We combine the two previous approaches to obtain discontinuities in both slope and intercept. [SW Slide 7]. The level-1 model is

$$\begin{aligned} Y_{ij} &= \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{2i} GED_{ij} + \\ &\quad \pi_{3i} POSTEXP_{ij} + \epsilon_{ij} \end{aligned} \tag{1}$$

[GROUP EXERCISE. Re-express $POSTEXP$ in terms of $EXPER$ and re-write the above model when $GED = 0$ and when $GED = 1$.]

An Alternate Model

An alternate approach to modeling changes in both slope and intercept is describe by S&W on page 199. The level-1 model is

$$\begin{aligned} Y_{ij} &= \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{2i} GED_{ij} + \\ &\quad \pi_{3i} GED_{ij} \times EXPER_{ij} + \epsilon_{ij} \end{aligned} \tag{2}$$

[GROUP EXERCISE. Re-write the above model when $GED = 0$ and when $GED = 1$. Then compare the model of Equation 1 with that of Equation 2.]

Introduction

In their section 6.1.2, S&W present an extensive analysis of fitting their wage data to a wide variety of models. In homework assignment 3, you will set up the two-level *and* composite versions of these models and reproduce the results in their Tables 6.2 and 6.3. We'll take a look at the baseline model and one of the followup models for clues on how to proceed.

A Baseline — Model A

Here is Model A, the baseline model described in detail on S&W page 201:

1. Level 1.

$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{2i}(UERATE_{ij} - 7) + \epsilon_{ij}$$

2. Level 2.

$$\begin{aligned} \pi_{0i} &= \gamma_{00} + \gamma_{01}(HGC_i - 9) + \zeta_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{12} BLACK_i + \zeta_{1i} \\ \pi_{2i} &= \gamma_{20} \end{aligned}$$

3. Composite.

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01}(HGC_i - 9) + \gamma_{10} EXPER_{ij} + \\ &\quad \gamma_{12} BLACK_i \times EXPER_{ij} + \gamma_{20}(UERATE_{ij} - 7) \\ &\quad + [\zeta_{0i} + \zeta_{1i} EXPER_{ij}] + \epsilon_{ij} \end{aligned}$$

Fitting Model A

Using `lmer()` to fit the model is straightforward.

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01}(HGC_i - 9) + \gamma_{10} EXPER_{ij} + \\ &\quad \gamma_{12} BLACK_i \times EXPER_{ij} + \gamma_{20}(UERATE_{ij} - 7) \\ &\quad + [\zeta_{0i} + \zeta_{1i} EXPER_{ij}] + \epsilon_{ij} \end{aligned}$$

```
model.a <-
lmer(lnw ~ 1 + hgc.9 + exper + black:exper +
 ue.7 + (1 + exper | id), REML=FALSE)
```

Fitting Model A

```
> data <- read.table("wages_pp.txt",header=T, sep=",")
> attach(data)
> model.a <- lmer(lnw ~ 1 + hgc.9 + exper + black:exper +
+   ue.7 + (1 + exper | id), REML=FALSE)
> model.a

Linear mixed model fit by maximum likelihood
Formula: lnw ~ 1 + hgc.9 + exper + black:exper + ue.7 + (1 + exper | id)
  AIC  BIC logLik deviance REMLdev
 4849 4909  -2415     4831    4877
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.05064  0.2250
          exper       0.00163  0.0404   -0.320
 Residual             0.09480  0.3079
Number of obs: 6402, groups: id, 888

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.74899    0.01140   153.4
hgc.9        0.04001    0.00636     6.3
exper        0.04405    0.00260    16.9
ue.7        -0.01195    0.00179    -6.7
exper:black -0.01818    0.00448    -4.1

Correlation of Fixed Effects:
            (Intr) hgc.9  exper  ue.7
hgc.9        0.086
exper       -0.566 -0.033
ue.7        -0.363 -0.039  0.277
exper:black -0.059 -0.018 -0.354  0.070
```

Model B

Model B adds *GED* as both a fixed and random effect. This occurs when the *GED* term is added to the level-1 model, and, at level 2, the coefficient for *GED* has a fixed and random term.

Model B

1. Level 1.

$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{2i}(UERATE_{ij} - 7) + \pi_{3i} GED_{ij} + \epsilon_{ij}$$

2. Level 2.

$$\begin{aligned} \pi_{0i} &= \gamma_{00} + \gamma_{01}(HGC_i - 9) + \zeta_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{12} BLACK_i + \zeta_{1i} \\ \pi_{2i} &= \gamma_{20} \\ \pi_{3i} &= \gamma_{30} + \zeta_{3i} \end{aligned}$$

3. Composite.

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01}(HGC_i - 9) + \gamma_{10} EXPER_{ij} + \\ & \gamma_{12} BLACK_i \times EXPER_{ij} + \gamma_{20}(UERATE_{ij} - 7) \\ & + \gamma_{30} GED_{ij} + [\zeta_{0i} + \zeta_{1i} EXPER_{ij} + \zeta_{3i} GED_{ij}] + \epsilon_{ij} \end{aligned}$$

Model B

```
> model.b <- lmer(lnw ~ 1 + hgc.9 + exper + black:exper + ue.7 +
+ ged + (1 + exper + ged| id), REML=FALSE)
> model.b

Linear mixed model fit by maximum likelihood
Formula: lnw ~ 1 + hgc.9 + exper + black:exper + ue.7 + ged + (1 + exper +      ged | id)
  AIC  BIC logLik deviance REMLdev
 4832 4919  -2403     4806    4858
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.04361  0.2088
          exper       0.00166  0.0407   -0.308
          ged         0.02824  0.1680    0.067 -0.318
 Residual             0.09416  0.3069
Number of obs: 6402, groups: id, 888

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.73421    0.01180   147.0
hgc.9        0.03833    0.00627     6.1
exper        0.04322    0.00262    16.5
ue.7        -0.01161    0.00179    -6.5
ged          0.06132    0.01845     3.3
exper:black -0.01820    0.00447    -4.1

Correlation of Fixed Effects:
            (Intr) hgc.9  exper  ue.7   ged
hgc.9        0.098
exper       -0.508 -0.029
ue.7        -0.370 -0.045  0.268
ged         -0.284 -0.045 -0.119  0.055
exper:black -0.047 -0.018 -0.350  0.069 -0.020
```

Comparing Model A and Model B

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

Data:
Models:
model.a: lnw ~ 1 + hgc.9 + exper + black:exper + ue.7 + (1 + exper | id)
model.b: lnw ~ 1 + hgc.9 + exper + black:exper + ue.7 + ged + (1 + exper +
model.b:     ged | id)
        Df  AIC  BIC logLik Chisq Chi Df Pr(>Chisq)
model.a  9 4849 4909  -2415
model.b 13 4832 4919  -2403    25      4      5e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Modeling Nonlinear Change via Transformations

A major advantage of a linear regression system with a normal outcome variable is that some classic mathematical results hold.

When the relationship between X and Y is nonlinear, a first option is to consider transformation to linearity. This approach is discussed in many fundamental regression texts, because of benefits like these:

1. Transformation to linearity often simultaneously normalizes the dependent variable
2. Since outcome variables are often delivered in a scale that is essentially arbitrary, transformations, especially when simple and clearly specified, aren't doing any serious harm
3. Parameters in a linear regression have an especially simple interpretation

The Mosteller-Tukey Transformation Ladder

Mosteller and Tukey introduced their “rule of the bulge” and “transformation ladder” for transforming to linearity.

The Mosteller-Tukey Transformation Ladder

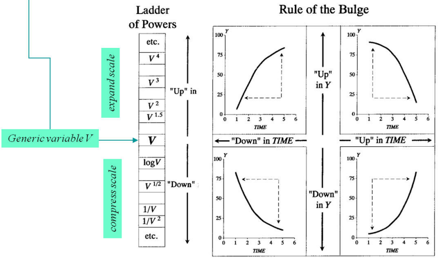
The "Rule of the Bulge" and the "Ladder of Transformations"

Mosteller & Tukey (1977): EDA techniques for straightening lines

Step 2: How do we know when to use which transformation?

1. *Plot many empirical growth trajectories*
2. *You find linearizing transformations by moving "up" or "down" in the direction of the "bulge"*

Step 1: What kinds of transformations do we consider?



(ALDA, Section 6.2.1, pp. 210-212)

Illustration — Using the Bulge Rule

Singer and Willett discuss how to employ the rule of the bulge in the context of some data for a single participant in the Berkeley Growth Study.

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

Illustration — Using the Bulge Rule

Example (Using the Bulge Rule)

Here is the plot of the original data:

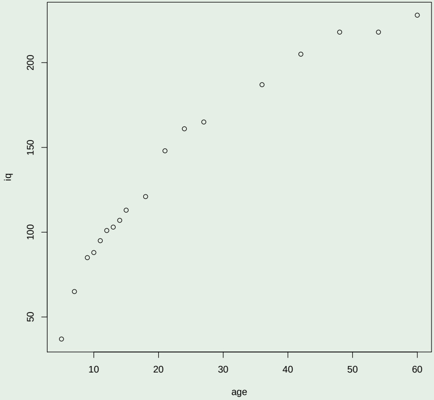


Illustration — Using the Bulge Rule

Example (Using the Bulge Rule)

Here is the plot of the data with iq raised to the 2nd power:

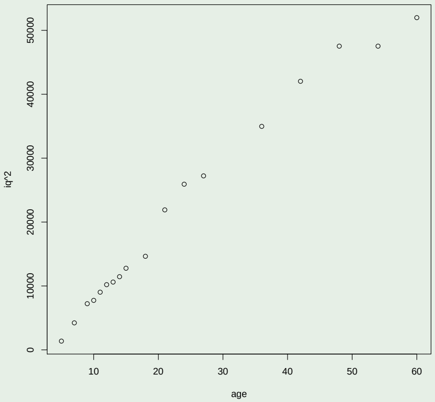


Illustration — Using the Bulge Rule

Example (Using the Bulge Rule)

Here is the plot of the data with iq raised to the 2.5th power:

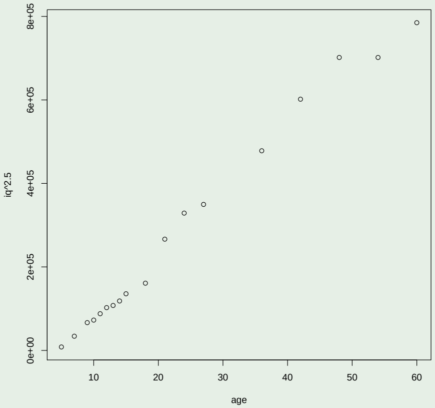


Illustration — Using the Bulge Rule

Example (Using the Bulge Rule)

Singer and Willett settle on raising iq to the 2.3 power:

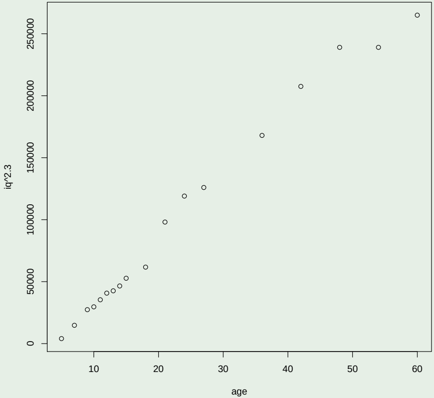
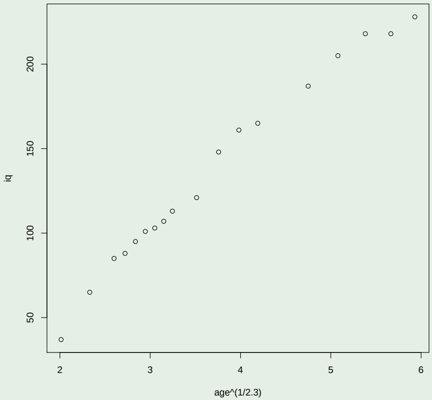


Illustration — Using the Bulge Rule

Example (Using the Bulge Rule)

A similar result is obtained by raising age to the 1/2.3 power:



Polynomial Regression Models

Many nonlinear functions can be approximated very well with polynomials of a reasonably low order.

Interpreting a Polynomial

Polynomials in X have an "order" given by the highest power of X in the equation. So, for example, $Y = 4$ is a 0th order polynomial, $Y = 2X + 3$ is a first order polynomial (linear), $Y = 4X^2 + 2X + 7$ is a second order (quadratic) polynomial, etc.

We are all familiar with the fact that in a first order polynomial of the form $Y = b_1X + b_0$, b_1 is the slope and b_0 is the intercept.

A fact we learned in calculus is that the first derivative of the function gives its slope. So, of course, if $Y = b_1X + b_0$, the first derivative of Y with respect to X, i.e., b_1, is the slope of the function, and this slope is constant with respect to X (hence the line is straight).

Interpreting a Polynomial

Consider the second-order polynomial $Y = b_0 + b_1 X + b_2 X^2$.

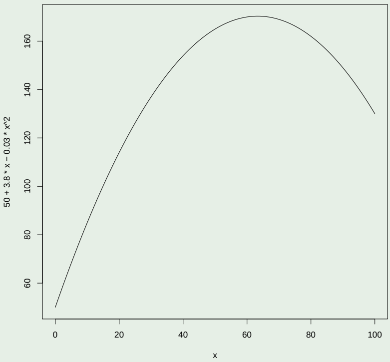
Of course the intercept, the value of Y when $X = 0$, is still b_0.

The slope of the function is $2b_2 X + b_1$. The slope of the function is changing as a function of X, so now the plot is curved. The slope starts out positive, but becomes zero when $X = -b_1/2b_2$.

Interpreting a Polynomial

Here is a graph of the function $Y = 50 + 3.8X - .03X^2$. Note the slope becomes 0 at $X = -3.8/2(-.03) = 63.33$.

Example (Quadratic Function)



The Externalizing Behavior Study

Keily, Bates, Dodge, and Pettit (2000) examined changes over time in externalizing behaviors, using Achenbach's (1991) Child Behavior Checklist.

Investigation of the individual trajectories shows that the degree polynomial required to obtain adequate fit varies widely.

In order to apply multilevel modeling to the data, we choose as our level-1 model the highest order polynomial required to fit any child.

We might settle on a quartic model, but this runs the danger of overfitting.

The Externalizing Behavior Study

We begin by loading in the data and creating some ancillary variables.

```
> detach(data)
> data <- read.table("external_pp.txt",header=T,sep=",")
> attach(data)
> TIME <- GRADE - 1
> TIME2 <- TIME^2
> TIME3 <- TIME^3
```

Choosing a Polynomial Model

Next we fit a sequence of models. We begin with the basic “no change” model.

```
> model.a <- lmer(EXTERNAL ~ 1 + (1|ID),REML=FALSE)
> model.a
```

```
Linear mixed model fit by maximum likelihood
Formula: EXTERNAL ~ 1 + (1 | ID)
  AIC  BIC logLik deviance REMLdev
 2016 2027  -1005     2010    2008
Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 87.4     9.35
 Residual             70.2     8.38
Number of obs: 270, groups: ID, 45

Fixed effects:
            Estimate Std. Error t value
(Intercept)    12.96       1.48    8.74
```

Choosing a Polynomial Model

Next comes the linear change model:

```
> model.b <- lmer(EXTERNAL ~ 1 + TIME +
+ (1+TIME|ID),REML=FALSE)
> model.b

Linear mixed model fit by maximum likelihood
Formula: EXTERNAL ~ 1 + TIME + (1 + TIME | ID)
  AIC  BIC logLik deviance REMLdev
 2004 2025   -996     1992    1989
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ID       (Intercept) 123.52   11.11
          TIME          4.69    2.17    -0.521
 Residual              53.72    7.33
Number of obs: 270, groups: ID, 45

Fixed effects:
            Estimate Std. Error t value
(Intercept)   13.290      1.836    7.24
TIME          -0.131      0.415   -0.31

Correlation of Fixed Effects:
     (Intr)
TIME -0.589
```

Choosing a Polynomial Model

Here is the quadratic:

```
> model.c <- lmer(EXTERNAL ~ 1 + TIME + TIME2 +
+ (1+TIME + TIME2|ID),REML=FALSE)
> model.c

Linear mixed model fit by maximum likelihood
Formula: EXTERNAL ~ 1 + TIME + TIME2 + (1 + TIME + TIME2 | ID)
  AIC  BIC logLik deviance REMLdev
 1996 2032   -988     1976    1974
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ID       (Intercept) 107.08   10.35
          TIME         24.61    4.96    -0.072
          TIME2         1.22    1.10    -0.119 -0.908
 Residual              41.98    6.48
Number of obs: 270, groups: ID, 45

Fixed effects:
            Estimate Std. Error t value
(Intercept)   13.970      1.774    7.88
TIME          -1.151      1.107   -1.04
TIME2          0.204      0.228    0.89

Correlation of Fixed Effects:
      (Intr) TIME
TIME  -0.322
TIME2  0.131 -0.932
```

Choosing a Polynomial Model

Here is the cubic:

```
> model.d <- lmer(EXTERNAL ~ 1 + TIME + TIME2 + TIME3 +
+ (1+TIME + TIME2 + TIME3 |ID),REML=FALSE)
> model.d

Linear mixed model fit by maximum likelihood
Formula: EXTERNAL ~ 1 + TIME + TIME2 + TIME3 + (1 + TIME + TIME2 + TIME3 |      ID)
  AIC  BIC logLik deviance REMLdev
 1997 2051   -984     1967    1968
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ID       (Intercept) 128.861  11.352
          TIME        106.819  10.335   -0.479
          TIME2        16.651   4.081    0.531 -0.975
          TIME3         0.177   0.421   -0.682  0.939 -0.981
 Residual              37.824   6.150
Number of obs: 270, groups: ID, 45

Fixed effects:
            Estimate Std. Error t value
(Intercept)  13.7945     1.9159    7.20
TIME         -0.3501     2.3279   -0.15
TIME2        -0.2343     1.0593   -0.22
TIME3         0.0584     0.1300    0.45

Correlation of Fixed Effects:
      (Intr) TIME   TIME2
TIME  -0.511
TIME2  0.450 -0.957
TIME3 -0.447  0.887 -0.978
```

Comparing Models

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

Data:
Models:
model.a: EXTERNAL ~ 1 + (1 | ID)
model.b: EXTERNAL ~ 1 + TIME + (1 + TIME | ID)
model.c: EXTERNAL ~ 1 + TIME + TIME2 + (1 + TIME + TIME2 | ID)
model.d: EXTERNAL ~ 1 + TIME + TIME2 + TIME3 + (1 + TIME + TIME2 + TIME3 |
model.d:     ID)
        Df  AIC  BIC logLik Chisq Chi Df Pr(>Chisq)
model.a  3 2016 2027  -1005
model.b  6 2004 2025   -996 18.51      3    0.00035 ***
model.c 10 1996 2032   -988 15.91      4    0.00315 **
model.d 15 1997 2051   -984  8.48      5    0.13170
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What do you think? (C.P.)