

[Page One](#)[Campus
Computing
News](#)[Mindset 1946](#)[New Computer
Based Training
Courses](#)[Today's Cartoon](#)[RSS Matters](#)[SAS Corner](#)[The Network
Connection](#)[Link of the
Month](#)[WWW@UNT.EDU](#)[Short Courses](#)[IRC News](#)[Staff Activities](#)[Subscribe to
Benchmarks
Online](#)

Research and Statistical Support

University of North Texas

RSS Matters

The previous issue in this series can be found in the July, 2002 issue of *Benchmarks Online*: [Using Robust Mean and Robust Variance Estimates to Calculate Robust Effect Size](#).

Please make sure and read the farewell message from Dr. Karl Ho in this month's [SAS Corner](#). - Ed.

An Introduction to Multilevel Models (Part I): Exploratory Growth Curve Modeling

By [Dr. Rich Herrington](#), Research and Statistical Support Consultant

This month we discuss exploratory growth curve modeling. This is Part I in a series on Multilevel Modeling. The GNU S language, "R" is used to implement this procedure. R is a statistical programming environment that is a clone of the S and S-Plus language developed at Lucent Technologies. In the following document we illustrate the use of a GNU Web interface to the R engine on the "rss" server (<http://rss.acs.unt.edu/cgi-bin/R/Rprog>). This GNU Web interface is a derivative of the "Rcgi" Perl scripts available for download from the CRAN Website (<http://www.cran.r-project.org>), the main "R" Website. Scripts can be submitted interactively, edited, and then be re-submitted with changed parameters by selecting the hypertext link buttons that appear below the figures. For example, clicking the "Run Program" button below creates a vector of 100 random normal deviates; sorts and displays the results; then creates a histogram of the random numbers. To view any text output, scroll to the bottom of the browser window. To view any graphical output, select the "Display Graphic" link. The script can be edited and resubmitted by changing the script in the form window and then selecting "Run the R Program". Selecting the browser "back page" button will return the reader to this document.

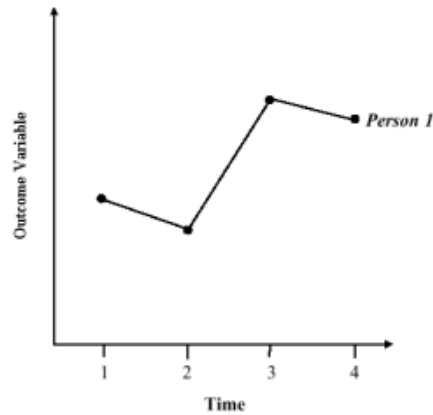
Introduction to Repeated Measure Multilevel Models

In this article we will discuss an exploratory approach to fitting a repeated measures multilevel model. This will serve as a warm up to the more formal estimation methods in multilevel modeling, which we will cover in coming issues of *RSS Matters*. We will motivate the model with a simulated example that uses Iteratively Reweighted Least Squares (IRLS) regression to estimate parameters. Our goal here is to keep the example simple to elucidate the main insights that growth curve modeling (multilevel modeling) offers. Additionally, we will learn S language commands that allow us to graphically explore and statistically estimate parameters of a simple multilevel model.

The following graph represents the time course of what we will call the **level-one observed change**, within an individual. Examining such a graph allows the data modeler to attempt to characterize the change process as a simple mathematical function (i.e. linear, quadratic, or cubic) for the individual entity. An assumption of growth curve modeling approaches and multilevel approaches in general, is that a family of functions (e.g. linear) can characterize the change process in different individuals, by allowing the parameters of the linear function to vary

randomly across individuals. Individuals express their unique variability through the parameters of the mathematical function. These randomly varying estimates of the change process (e.g. intercept and slope) can be related to background conditions of the individuals. It is here that we wish to get efficient and unbiased estimates of the predictors of the change process. Do background covariates predict rates of change (slope) or initial status of the outcome variable?

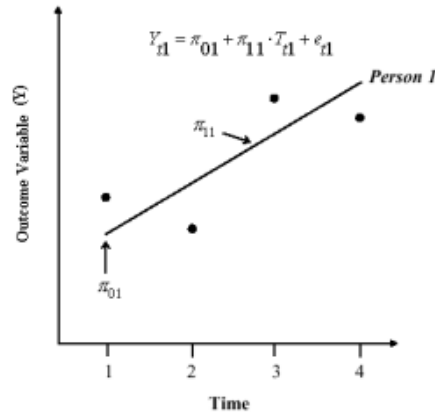
Four Repeated Measures for a Single Individual



Summary Measures of the Level-One Model Change Process

Since the observed growth record for an individual is subject to measurement error and sampling variability, it is useful to characterize the growth process as a mathematical function. We will call this the **level-one model** of change. This mathematical function (e.g. linear), if representative, should capture the essential features of the growth profile. The parameters of the function should summarize the characteristics of the change process for the individual. In the following graph, a linear function is considered representative of the change process. An intercept is calculated; the intercept gives a best estimate of the initial status of the outcome variable at the first observation period. Additionally a slope is calculated; a slope gives an estimate of the rate of change in the outcome variable per unit change in the time indicator variable. In our simulated example we have electrophysiological assessments from the scalp of an individual who is engaged in EEG biofeedback (also referred to as neurofeedback - NF), gathered over four observation periods. Individuals use a process of trial and error and an electrical feedback signal from the scalp to change their own scalp electrical potentials. An initial question is posed for our example: Does the rate of change in cortical electrical activity (amplitude in millivolts) relate to background conditions of depression in the individual?

Level One Linear Regression Model – Fitted “True” Growth Trajectory for a Single Individual



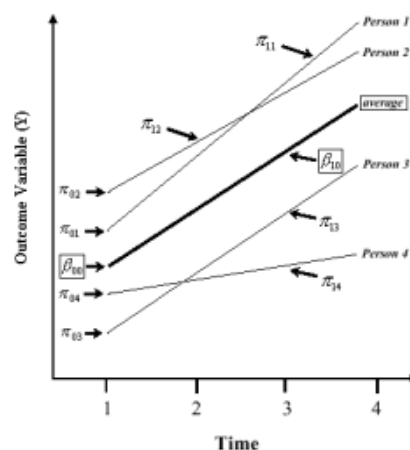
$$Y_{ti} = \pi_{0i} + \pi_{1i} \cdot T_{ti} + e_{ti}$$

where Y_{ti} is the response variable of person i measured at time point $t=1,2,3,4$; T is the time variable that indicates the time point. For example Y could represent NF or QEEG assessments at four periods of time.

Individual Variation (Level-One) and Group Level Variation (Level-Two)

The following graph depicts four linear growth curves fit for four different individuals. Additionally, an average curve is estimated which represents the change process of the group as a whole, collapsing over individual variation in the growth process. Each person has an intercept estimate and a slope estimate; these individual estimates can vary from the group intercept and group slope (level-two). The source of this variation may be related to other predictor variables, such as background characteristics, or may represent variation that is not related to a substantial variable of interest. We will call this the **level-two model**. In our example, it might be that the rate of change in electrical activity on the scalp (as measured in millivolts), is related to initial status of depression, as measured behaviorally by a paper and pencil inventory - the Beck's Depression Inventory (BDI).

Fitted Growth Trajectories for Four Individuals



Predicting Outcome as a Function of Time - Allowing Individual Variation in Slopes and Intercepts

We can represent both level-one and level-two models of change as a set of regression equations. The outcome variable (Y) is predicted by the estimates of individuals' slopes and intercepts. Each individuals' slope and

intercept is composed of a constant (fixed) portion and a residual (random) portion. The fixed portion can be thought of as the mean of the group estimate (i.e. group slope and group intercept), and the residual variation can be thought of as unique variation that is not accounted for at the group level. We can substitute level two regression equations into level one regression equations to obtain a single regression equation.

Unconditional Random Effects Growth Model

Level 1 model:

$$Y_{ti} = \pi_{0i} + \pi_{1i} \cdot T_{ti} + e_{ti}$$

Level 2 model:

$$\pi_{0i} = \beta_{00} + u_{0i}$$

$$\pi_{1i} = \beta_{10} + u_{1i}$$

Combining Level 1 and Level 2 by substitution with simple rearrangement of terms give:

$$Y_{ti} = \underbrace{(\beta_{00} + T_{ti} \cdot \beta_{10})}_{\text{1st term}} + \underbrace{(u_{0i} + T_{ti} \cdot u_{1i} + e_{ti})}_{\text{2nd term}}$$

The first term represents the **fixed** component of growth; this represents the **group** level characteristics of growth. The second term represents the **random** component of growth; this represents the **individual** level characteristics of growth. This model is referred to as an unconditional growth model because we are NOT attempting to predict growth using additional measures. To predict growth we must use a **conditional growth model**. Both time variant and time invariant predictors of growth can be used in a conditional growth model

It is important to note that this single equation has "heteroscedastic" components. The second term in this regression equation is heteroscedastic because the time variable (T), which is increasing, multiplies by a variance component (u). Hence, the variance increases as a function of observation period, thus violating the assumption of homoscedasticity in classical OLS estimates of the regression equation. It is for this reason that a non-OLS estimate is needed for the parameters in the model; heteroscedasticity can lead to biased and/or inefficient estimates of slope and intercept coefficients

Predicting the Outcome Variable as a Function of Level-One Covariates and Level-Two Covariates

In the system of regression equations below, both time varying level-one covariates, and time-invariant level-two covariates are combined. X is an individual level, time varying covariate, and Z is a time invariant, individual level characteristic, or group level predictor; the six regression parameters (beta coefficients) relate the two predictors to variability in initial status and rate of change over time.

Conditional Random Effects Growth Model

Level 1 model:

$$Y_{ti} = \pi_{0i} + \pi_{1i} \cdot T_{ti} + \pi_{2i} \cdot X_{ti} + e_{ti}$$

Level 2 model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \cdot Z_i + u_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \cdot Z_i + u_{1i}$$

$$\pi_{2i} = \beta_{20} + \beta_{21} \cdot Z_i + u_{2i}$$

Y = Neurofeedback Amplitudes
T = Time
X = BDI Score
Z = Gender (coded 0,1)

Combining Level 1 and Level 2 by substitution gives:

$$Y_{ti} = \beta_{00} + \beta_{10} \cdot T_{ti} + \beta_{20} \cdot X_{ti} + \beta_{01} \cdot Z_i + \beta_{11} \cdot Z_i \cdot T_{ti} + \beta_{21} \cdot Z_i \cdot X_{ti} + u_{1i} \cdot T_{ti} + u_{2i} \cdot X_{ti} + u_{0i} + e_{ti}$$

Using Variable Labels: NFA (Neurofeedback Amplitudes), BDI (Beck's Depression Inventory) and Gender.

$$NFA_{ti} = \beta_{00} + \beta_{10} \cdot Time_{ti} + \beta_{20} \cdot BDI_{ti} + \beta_{01} \cdot Gender_i + \beta_{11} \cdot Gender_i \cdot Time_{ti} + \beta_{21} \cdot Gender_i \cdot BDI_{ti} + u_{1i} \cdot Time_{ti} + u_{2i} \cdot BDI_{ti} + u_{0i} + e_{ti}$$

This model is a conditional growth model because the individual variability in initial status and change over time are conditioned upon exogenous predictor variables. The conditional random effects model is sometimes referred to as a: random coefficient model, mixed-effects model, hierarchical linear model (Bryk & Raudenbush, 1992), empirical Bayes model, "slopes as outcomes" model - or, more generally, a multilevel model.

In longitudinal research, we sometimes have repeated measures of individuals who are all measured together on a small number of fixed occasions. This is typically the case with experimental designs involving repeated measures and panel designs. For our example, our group level predictor can be an indicator variable indicating either control or experimental group membership.

For our simulation example, we will use a much simpler model (displayed below), and will only focus on one coefficient of interest. In the following example Y will be electrical voltages from the scalp (amplitude - measured in millivolts) measured over observation period (T). Individuals change their own scalp electrical potential through trial and error, relying on a biofeedback signal coming from their scalp (neurofeedback - NF). Additionally, we have measured levels of depression using a behavioral (self-report) assessment device (Beck Depression Inventory - BDI). Several questions can be posed: Is there evidence for systematic change and individual variability in NF amplitudes over time? Is the post BDI assessment related to the initial levels or rates of change in NF amplitudes? What is the relationship between the initial levels of NF amplitudes and the rates of change in NF amplitudes over time? Is a linear relationship a good description of within-person change?

Our Example Model

Level 1 model:

$$Y_{ti} = \pi_{0i} + \pi_{1i} \cdot T_{ti} + e_{ti}$$

Level 2 model:

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01} \cdot Z_i + u_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} \cdot Z_i + u_{1i} \end{aligned}$$

We are interested in estimating an unbiased estimate of the correlation between the level one slopes and the background covariate Z

Parameter Estimation Using Iteratively Re-weighted Least Squares (IRLS)

We are interested in unbiased and efficient estimates of the beta coefficients which relate the background covariate Z to the individual level estimates of rates of change and initial status (pi coefficients). Willet (1988) outlines a fairly straightforward approach to estimating a growth curve model for a single, level-two coefficient (relating individual slope coefficients to individual background covariates). This method relies on weighted least squares (WLS) and can be improved by iterating until the weights or residuals converge (IRLS). This method is appealing because it provides insight into a central point of multilevel modeling: Level-two coefficient estimates are weighted by the precision of the level-one coefficients. Individuals whose growth coefficients are more reliably estimated, provide more input (information) in the estimation of level-two coefficients. Individuals with large residual variability in their growth record, and hence more unreliable estimates of level-one growth coefficients, are down weighted in the second level of analysis of the level-two coefficients. In the table below, a modified form of Willet's algorithm is outlined.

A Simple Algorithm for a Two Stage Analysis of Growth Rates with a Selected Covariate (See Willet, 1988)

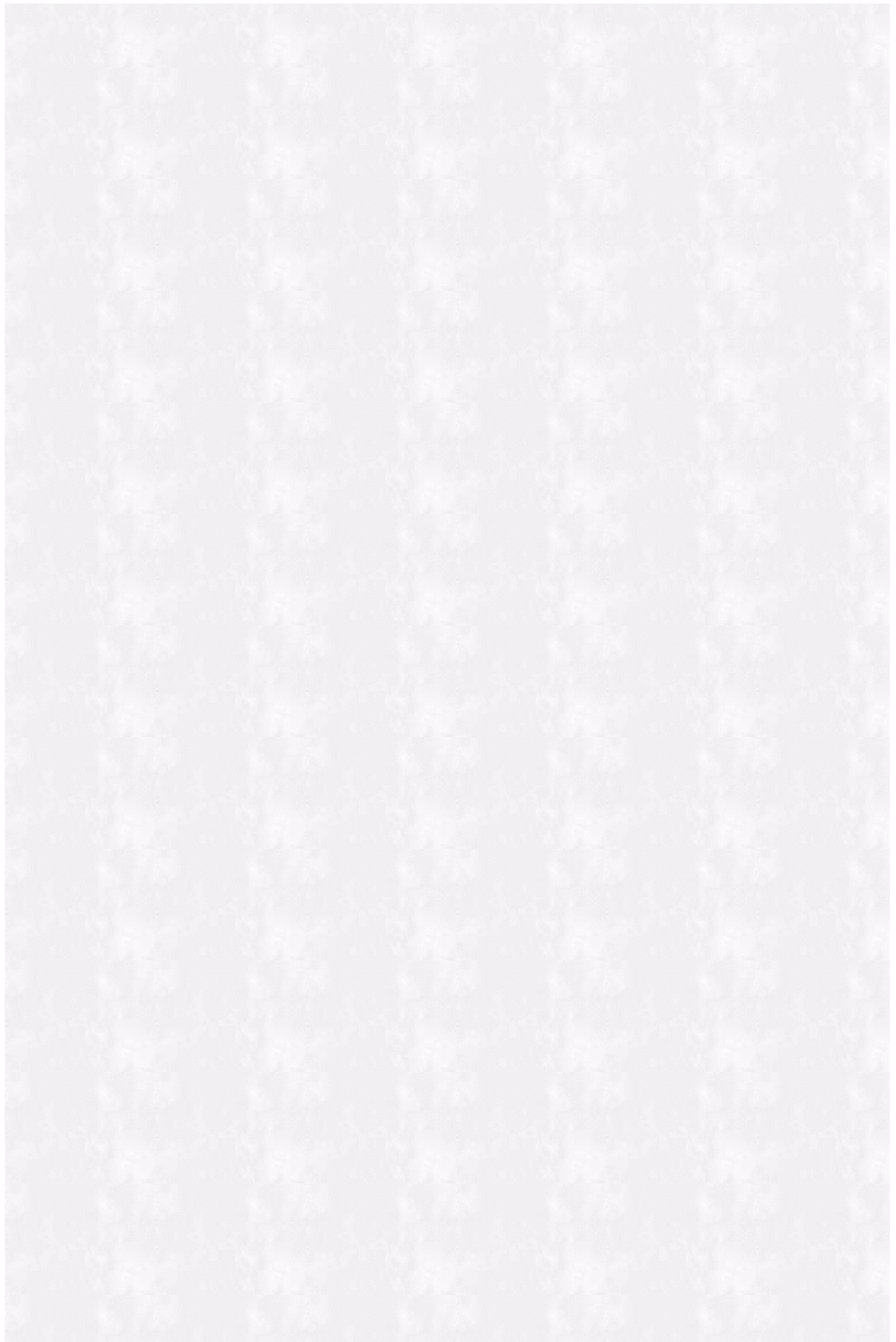
- 1) Fit the within-individual growth models, retaining the estimated individual growth rates $\hat{\pi}_i$ and their standard errors, $s.e.(\hat{\pi}_i)$, for subsequent between person analysis
- 2) Regress the estimated individual growth rates $\hat{\pi}_i$ on the selected covariate Z_i by ordinary least squares (OLS) analysis, retaining the estimated residual variance s^2 as a global measure of the "scatter" present in the second phase of analysis
- 3) Compute a weight matrix using weighted least squares (WLS) regression:

$$W_1 = \frac{s^2}{s^2 + s.e.(\hat{\pi}_i)^2}$$

- 4) Iterate steps 2 and 3 until W_1 or s^2 converges within some tolerance (Iteratively Re-weighted Least Squares - IRLS)

An Example Using GNU-S ("R")

We will create a simulated data set that is modeled after real neurofeedback data. The observations represent average neurofeedback session amplitudes (10 averaged sessions per observation period). The covariate represents either a pre or post BDI assessment (we will use a post BDI assessment). We will go through most sections of the S code to illustrate how S can be used to both simulate data according to an assumed model, and then estimate the model, to try and recapture the population parameters. It is recommended that the reader change the population parameters and experiment with the model. In this way, intuition can be gained about how the observed data varies according to changes in the model. A "Live" script is presented below (this script can be run, examined, and re-edited for re-submission and re-examination); below this live script is annotated program output.



Population parameters are set according to our previous model with one background covariate (Z, the BDI). A small number of growth curves are generated to examine the effect of small sample size in getting unbiased estimates using OLS versus IRLS parameter estimation. Both the fixed portion and the residual portion (beta and u) for the level-two regressions are set in the population. These will be used with the background covariate to generate the slopes and intercepts for individuals (pi's).

```
> ##### Number of Growth Curves
>
> nobs.curves<-20
>
> ##### Level Two Parameters
>
> beta.00<-5.5
> beta.01<-0.005
>
> beta.10<-(-.20)
> beta.11<-0.20
>
> u.0.std<-0.03
> u.1.std<-0.15
>
> ##### Parameter for Level One Residuals
>
> error.std<-3
>
> ##### Parameters for covariate
>
> covar.mean<-10
> covar.std<-2.5
>
```

A matrix of observation periods must be generated for each individual.

```
> ##### Generate Time vector
>
> time<-matrix(rep(c(0,1,2,3), length=4*nobs.curves),
+             nrow=nobs.curves, ncol=4, byrow=T)
>
> time.list<-vector('list', nobs.curves)
> for (i in 1:nobs.curves){time.list[[i]]<-time[i,]}
>
```

Slopes and intercepts of individuals can be correlated across individuals (p0 - intercept, p1 - slope). We create slopes and intercepts whose correlation is -.30: high intercepts lead to rapid rates of change, and low intercepts lead to small rates of change.

```
> ##### Generate Correlated Level One Parameters
>
> Sigma.param <- matrix(c(1,-.3,-.3,1),2,2)
> pi.indiv<-mvrnorm(n=nobs.curves, rep(0, 2), Sigma.param,
+                  empirical=TRUE)
> cor(pi.indiv)
      [,1] [,2]
[1,]  1.0 -0.3
[2,] -0.3  1.0
```

Residuals at level-one (within an individual) are likely to be correlated. We assume a correlation structure that is equal across time periods (compound symmetry).

```

> ##### Generate Correlated Residuals
>
> Sigma.error <- matrix(c(1,.2,.2,.2,.2,1,.2,.2,.2,.2,1,.2,.2,.2,.2,1),
+                       4,4)
> z.error<-mvrnorm(n=nobs.curves, rep(0, 4), Sigma.error,
+                 empirical=TRUE)
>
> indiv.error<-z.error*error.std
> indiv.error
      [,1]      [,2]      [,3]      [,4]
[1,]  5.1871378  3.92186871  4.7762044  1.5447049
[2,] -2.7297285 -1.58349909  1.1641684  1.4861057
[3,] -0.1038213 -2.00845366  2.3072686  1.6031268
[4,]  2.4789522 -3.73737318 -5.1474588 -6.3880011
[5,]  1.2983139 -0.09995374 -3.1199230  2.9683871
[6,] -1.6679028  3.66215983  1.0135153  2.0962094
[7,] -2.1744504  3.40328453 -2.6881090 -3.5218952
[8,]  0.5073552  3.84594669  2.3063680 -2.8651158
[9,]  0.4993936 -0.28892692  0.2376112 -0.1347145
[10,]  4.4785767 -1.30473322  1.7369103 -1.1471286
[11,]  3.1462809 -0.70600709  0.4854152 -1.4005639
[12,] -6.0050461 -7.08117031 -4.6103369 -2.8585091
[13,]  0.6299895  2.26310343  2.7280101  1.1231027
[14,] -2.9222697  2.10340089 -3.0063778 -1.2331652
[15,] -1.0199459 -0.60136666 -0.9304446  4.2259268
[16,] -2.9397596  1.48798761 -0.4206570 -4.3487363
[17,]  3.8387073 -3.28142206  1.5184872  1.7138605
[18,] -3.9686504 -1.87733734  5.5083218 -0.5025358
[19,] -0.7807658 -1.76999505  0.2529361  2.1127997
[20,]  2.2476334  3.65248662 -4.1119095  5.5261419
~

```

Next, a covariate is generated for each individual:

```

> ##### Generate Covariate Scores
>
> covar<-matrix(rnorm(nobs.curves, covar.mean,
+                   covar.std), nrow=nobs.curves)
> covar
      [,1]
[1,] 10.106585
[2,] 12.353063
[3,] 10.182889
[4,]  6.040697
[5,] 11.045452
[6,] 12.187145
[7,]  8.272381
[8,] 14.054151
[9,] 11.690640
[10,] 13.048148
[11,]  9.764001
[12,]  6.764276
[13,] 11.518807
[14,] 10.221173
[15,] 13.825145
[16,] 10.972019
[17,]  8.968430
[18,] 14.283469
[19,] 11.429488
[20,]  6.278810

```

With the group betas (b_0 , b_1), individual background covariate (covar), and level-two residual terms (u_0 , u_1), we can combine the level-two parameters together to obtain the level-one parameters for each individual (p_0 , p_1). We also estimate the average population slope and intercept estimates, and the correlation between the two, once the betas are combined with the covariate and the residual terms. Later, in estimating a statistical model for this simulated data, we might want to use "centered" estimates of the covariate (covar) to aid in interpretation and estimation (Hox, 2002).

Once the pi coefficients (p_0, p_1) are calculated for each individual, we can combine them with the time index and add the level-one individual residual term to obtain the "observed score" for each individual, indexed by T - each individual gets four observations since T ranges from 1 to 4.

```
> ##### Construct Level One Parameters from level two parameters
>
> pi.0<-matrix(beta.00+beta.01*(covar)+pi.indiv[,1]*u.0.std, ncol=1)
> pi.1<-matrix(beta.10+beta.11*(covar)+pi.indiv[,2]*u.1.std, ncol=1)
>
> mean(pi.0)
[1] 5.553252
> mean(pi.1)
[1] 1.930068
> sd(pi.0)
[1] 0.03260155
> sd(pi.1)
[1] 0.5144264
> cor(pi.0, pi.1)
      [,1]
[1,] 0.2941906
>
> ##### Construct observed data set
>
> y.obs<-matrix(0, nrow=nobs.curves, ncol=4)
> for(i in 1:nobs.curves){
+   for(j in 1:4){
+     y.obs[i,j]<-pi.0[i]+pi.1[i]*time[i,j]+indiv.error[i,j]
+   }
+ }
>
```

Some attention needs to be paid to the fact that some observations can fall out of range of reasonable limits (dictated by physical logistics and limitations). We can replace the min and max observation with the mean observation, for each observation period. This process "draws" the tails of the distribution of scores, for an observation period, toward the mean of the distribution. This can be iterated to shape the growth curves is so desired. Here, we only iterate one time, taking the one smallest and one largest observation and replacing with the mean for that observation period.

```
> ##### Eliminate unlikely observations in the simulation
> #       by drawing tails of distribution inward toward mean
>
> y.obs[y.obs[,1]==min(y.obs[,1]),1]<-mean(y.obs[,1])
> y.obs[y.obs[,1]==max(y.obs[,1]),1]<-mean(y.obs[,1])
>
> y.obs[y.obs[,2]==min(y.obs[,2]),2]<-mean(y.obs[,2])
> y.obs[y.obs[,2]==max(y.obs[,2]),2]<-mean(y.obs[,2])
>
> y.obs[y.obs[,3]==min(y.obs[,3]),3]<-mean(y.obs[,3])
> y.obs[y.obs[,3]==max(y.obs[,3]),3]<-mean(y.obs[,3])
>
> y.obs[y.obs[,4]==min(y.obs[,4]),4]<-mean(y.obs[,4])
> y.obs[y.obs[,4]==max(y.obs[,4]),4]<-mean(y.obs[,4])
>
```

Finally, we list out the observed scores as simulated from a "known" model.

```
> y.obs.list
[[1]]
[1] 5.856097 11.391735 14.167513 12.857456

[[2]]
[1] 2.862320 6.348366 11.435850 14.097604

[[3]]
[1] 5.404281 5.435123 11.686319 12.917652
```

.....


```

.....
[[17]]
[1]  9.384307  3.693321  9.922374 11.546891

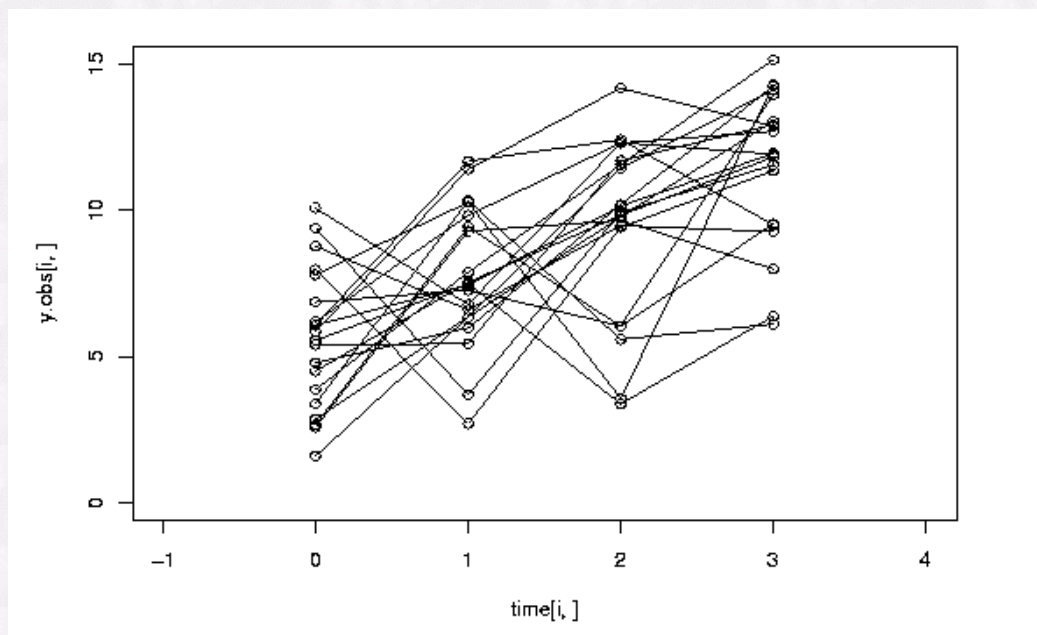
[[18]]
[1]  1.600006  6.347898  9.771714 13.035857

[[19]]
[1]  4.778620  5.975849 10.185238 14.231559

[[20]]
[1]  7.792503 10.265817  3.569881 14.276393

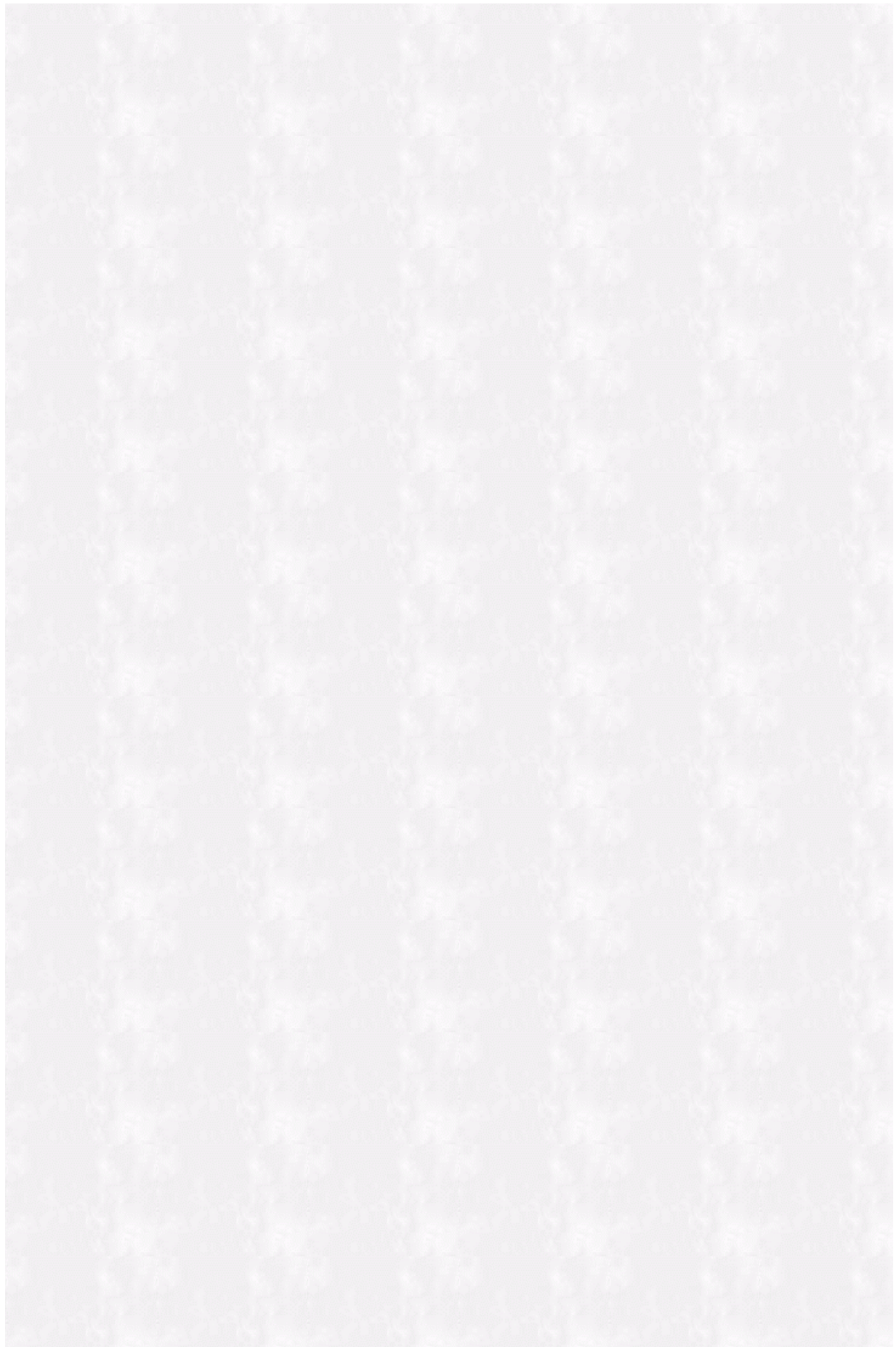
```

A "not-so-useful" depiction of all the observed growth profiles is a "Profile" plot. A profile plot allows one to discern some random effects structures in the data - we can see that the slopes and intercepts are potentially correlated; we see that there is heteroscedasticity prevalent in the data; and we can examine potential "outlying" growth records.



Parametric Estimation of Growth Curve Models Using OLS and IRLS

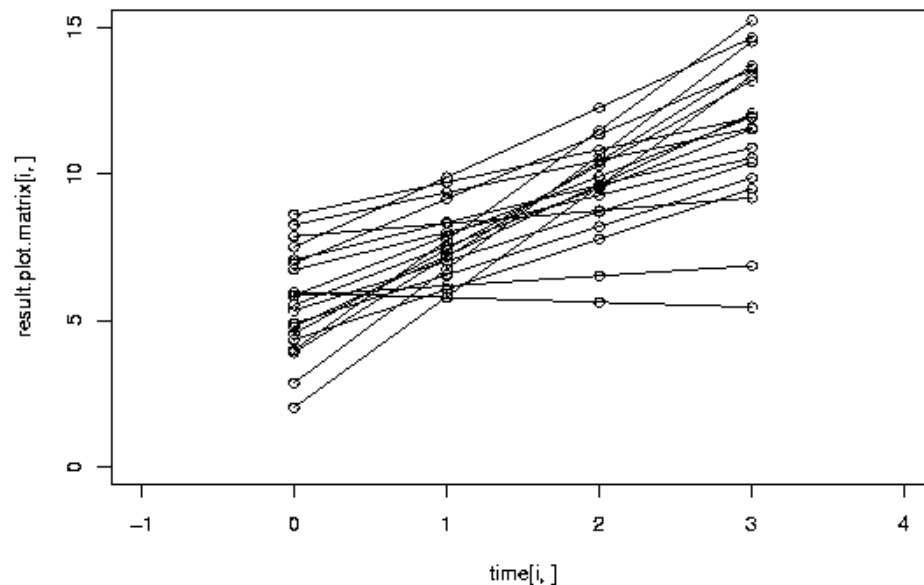
Ordinary Least Squares (OLS) is used to estimate the level-one linear regressions for all individuals. Standard errors for each individual regression are used to create weights for the level-two regression of the level-one slopes and background covariates. Additionally, at the level-two regression, iteratively re-weighted least squares (IRLS) is used to converge on the best estimates given the initial level-one weights.



All 20 regressions are performed; the beta coefficients, standard errors (used for weights later), covariates, and other interesting information are combined into a single data frame; informative names are then assigned to the columns.

```
> #####
> ### Perform OLS Regression on all Curves
>
> results<-lapply(1:nobs.curves, function(i, x, y) lm(y[[i]]~x[[i]]),
+             x = time.list, y = y.obs.list)
>
> ##### Extract Results into a Dataframe
>
> result.matrix<-matrix(0, ncol=11, nrow=nobs.curves)
> for(i in 1:nobs.curves){
+ result.matrix[i,]<-cbind(summary(results[[i]])$coefficients[1],
+                          summary(results[[i]])$coefficients[3],
+                          summary(results[[i]])$coefficients[2],
+                          summary(results[[i]])$coefficients[4],
+                          summary(results[[i]])$sigma,
+                          summary(results[[i]])$r.squared,
+                          covar[i],
+                          results[[i]]$fitted.values[1],
+                          results[[i]]$fitted.values[2],
+                          results[[i]]$fitted.values[3],
+                          results[[i]]$fitted.values[4])
+ }
>
> ##### Label Columns of DataFrame
>
> result.matrix<-data.frame(result.matrix)
> names(result.matrix)<-c('intercept', 'int.std.err', 'slope',
+                          'slope.std.err', 'resid.std.err', 'R.squared',
+                          'covar', 't0.pred', 't1.pred', 't2.pred',
+                          't3.pred')
>
```

A nice alternative to a profile plot of the observed data, is a profile plot of the level-one predicted values for each observation period. This plot better summarizes the correlation structure of the parameters (slopes and intercepts), and indicates more clearly the nature of the heteroscedasticity in the data.



A comparison of the OLS and IRLS solution indicates that the slope estimates are close for both the IRLS and OLS procedures (.25 for OLS versus .36 for IRLS). However, the IRLS procedure gives a substantially smaller residual standard error (1.001 for OLS versus .5483 for IRLS), thereby giving a more efficient estimate for the IRLS procedure. Furthermore, the residuals for the OLS indicate fairly high values (-1.7009, 1.5025), whereas the IRLS residuals are relatively smaller (-.84154, 1.00915) - using 2.0 as a cutoff.

```
> ##### OLS Results
>
> result.level2

Call:
lm(formula = result.matrix[, 3] ~ result.matrix[, 7])

Residuals:
    Min       1Q   Median       3Q      Max
-1.7009 -0.5330 -0.0875  0.8958  1.5025

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.85485    1.02297   -2.791  0.0121 *
result.matrix[, 7]  0.25108    0.09372    2.679  0.0153 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.001 on 18 degrees of freedom
Multiple R-Squared:  0.2851,    Adjusted R-squared:  0.2453
F-statistic: 7.177 on 1 and 18 DF,  p-value: 0.01532
```

```

> ##### WLS Results
>
> summary(result.weighted)

Call:
lm(formula = result.matrix[, 3] ~ result.matrix[, 7], weights = weight)

Residuals:
    Min       1Q   Median       3Q      Max
-0.84154 -0.41529 -0.09472  0.26591  1.00915

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -3.8360     1.4213  -2.699  0.01469 *
result.matrix[, 7]  0.3636     0.1178   3.086  0.00638 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5483 on 18 degrees of freedom
Multiple R-Squared: 0.346,    Adjusted R-squared: 0.3096
F-statistic: 9.522 on 1 and 18 DF,  p-value: 0.006375

```

OLS gives a good fit to the observed data even though growth curve parameters are not weighted by the parameter's precision at level-one. This indicates a relatively efficient solution, however IRLS gives a much better fit to the data. Consequently, the estimates of R-squared are significantly larger for IRLS indicating that the OLS is not as efficient in the presence of heterogeneous slopes having differing measurement precisions.

Descriptive statistics for the observed data and the level-one regressions are generated.

```

> ##### Descriptives on Observed Scores
>
> cor(y.obs)
           [,1]      [,2]      [,3]      [,4]
[1,]  1.00000000 -0.28130365  0.06113171  0.08572414
[2,] -0.28130365  1.00000000 -0.04777873 -0.22536795
[3,]  0.06113171 -0.04777873  1.00000000  0.38395990
[4,]  0.08572414 -0.22536795  0.38395990  1.00000000
> apply(y.obs, 2, mean)
[1] 5.558872 5.499548 5.127406 5.080371
> apply(y.obs, 2, sd)
[1] 2.385257 2.384481 3.006104 2.612348
> mean(covar)
[1] 10.65034
>
> ##### Level Two Means for Intercept and Slope
>
> mean(result.matrix[,1])
[1] 5.587696
> mean(result.matrix[,3])
[1] -0.1807644
> cor(result.matrix[,1], result.matrix[,3])
[1] -0.709922

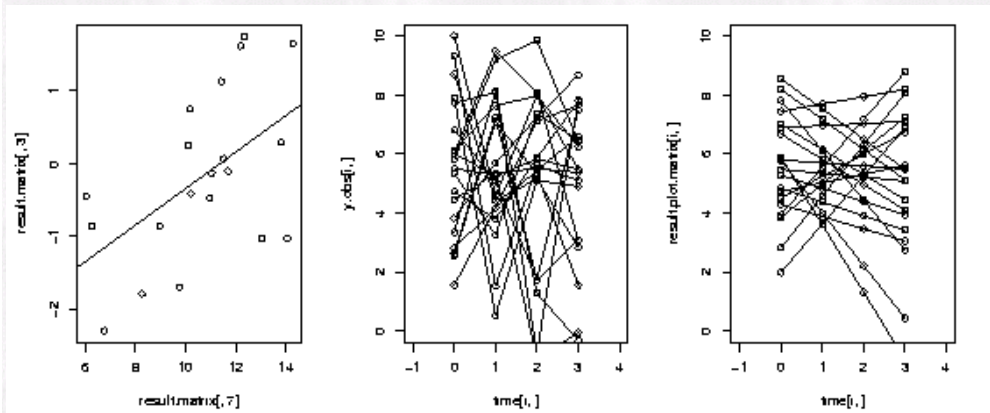
```

A Graphical Depiction of the Relationship Between the Observed Background Covariate and the Estimate Growth Rates

Another useful graphical depiction is a scatter plot (with an OLS best fit line) of the estimated growth rates (slopes) with the background covariate. One can graphically characterize the impact of the heteroscedasticity present in the data on the level-two regression of slopes and covariates. The heteroscedasticity in the plot appears to be quite large. While significant "outliers" are present, these outliers seem to "counterbalance" one another across the mean of the Y-axis, and counterbalance one another across the mean of the X-axis. This could account for the closeness of the IRLS and OLS parameter estimates. If the size of the residuals are: 1) symmetrical and counterbalanced across both Y and X axes (as is in this case), and 2) the outlying values on the X-axis are close to the mean of the X-axis, this would lead to less bias in the parameter estimates. An interesting experiment would be to generate different data sets with differing patterns of heteroscedasticity, and

observe the difference in the parameter estimates of the OLS and IRLS procedures.

The graph below (panel 1) depicts the relationship between the covariate (Y-axis) and the slopes for individuals (X-axis). The second graph (panel 2) plots the observed data as a function of observation period, and the third graph (panel 3) is the predicted values from the level-one regressions plotted as a function of observation period.



Conclusions

A Two Stage IRLS model of the Post BDI covariate and the individual growth rates indicates that there is reliable differences in growth rates in individuals and is correlated with the post BDI assessment (for the simulated data set). In general, larger negative slopes are correlated with smaller values on the BDI assessment, and lesser negative slopes are correlated with larger values on the post BDI assessment.

In general, larger negative slopes are correlated with larger NF amplitudes at initial status, and smaller negative slopes are correlated with smaller NF amplitudes at initial status. The IRLS and OLS beta estimates (slopes) were fairly close in value - this may be due to the symmetrical pattern of heteroscedasticity in the data. However, IRLS estimation gave significantly smaller level-two residual standard error than OLS estimation, for the level-two regression - IRLS is more efficient than OLS in this example. Consequently, the percentage variance accounted for in the outcome variable (NF amplitudes), by knowledge of the background covariate (BDI), was substantially larger for the IRLS solution.

Next Time

Next time we will explore the use of the S-Plus and R library NLME (linear and nonlinear mixed effects) with the simulated data set used in this article. Additionally, we will look at other parameter estimation algorithms (e.g. Restricted Maximum Likelihood - REML), and other model diagnostic approaches (e.g. AIC).

References

Bryk, A.& Raudenbush, S.W. (1992). *Hierarchical linear Models*. Newbury Park, CA: Sage.

Hox, J. (2002). *Multilevel analysis: Techniques and Applications*. Lawrence Erlbaum, New Jersey.

Willet, J. (1988). Questions and answers in the measurement of change. *Review of Research Education*, Vol. 15, AERA Research Associates, Washington, DC.