Quantitative Analytic Techniques (Class #12)

Gregory M. Eirich
QMSS

1. Growth curve analysis

1. Growth curve analysis

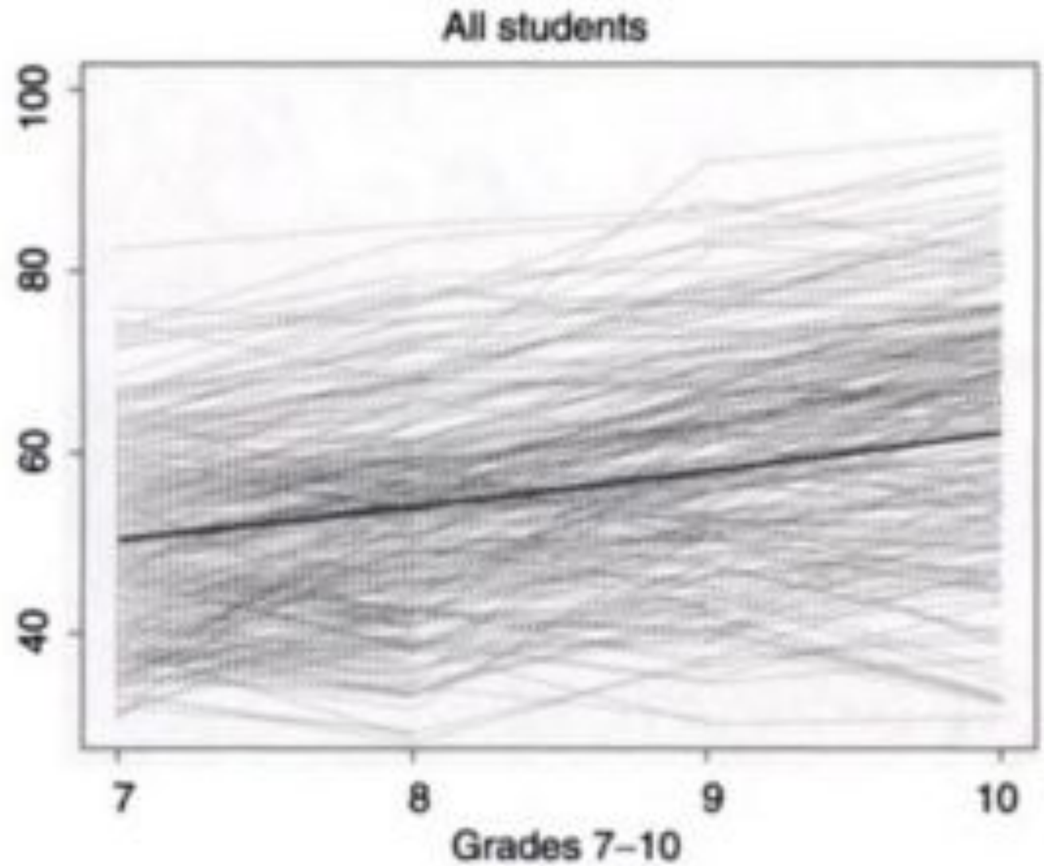
An example of growth curves

- Muthén, Bengt. "Latent variable analysis." *The Sage handbook of quantitative methodology for the social sciences. Thousand Oaks, CA: Sage Publications* (2004): 345-68.

Time often is necessary to see how people sort out

- Many processes take a long time to have effects
- Trajectories can converge or diverge
- We want to study groups of people as they change over time in similar and different ways
- Do certain events change people's trajectories up or down?

The substantive issue

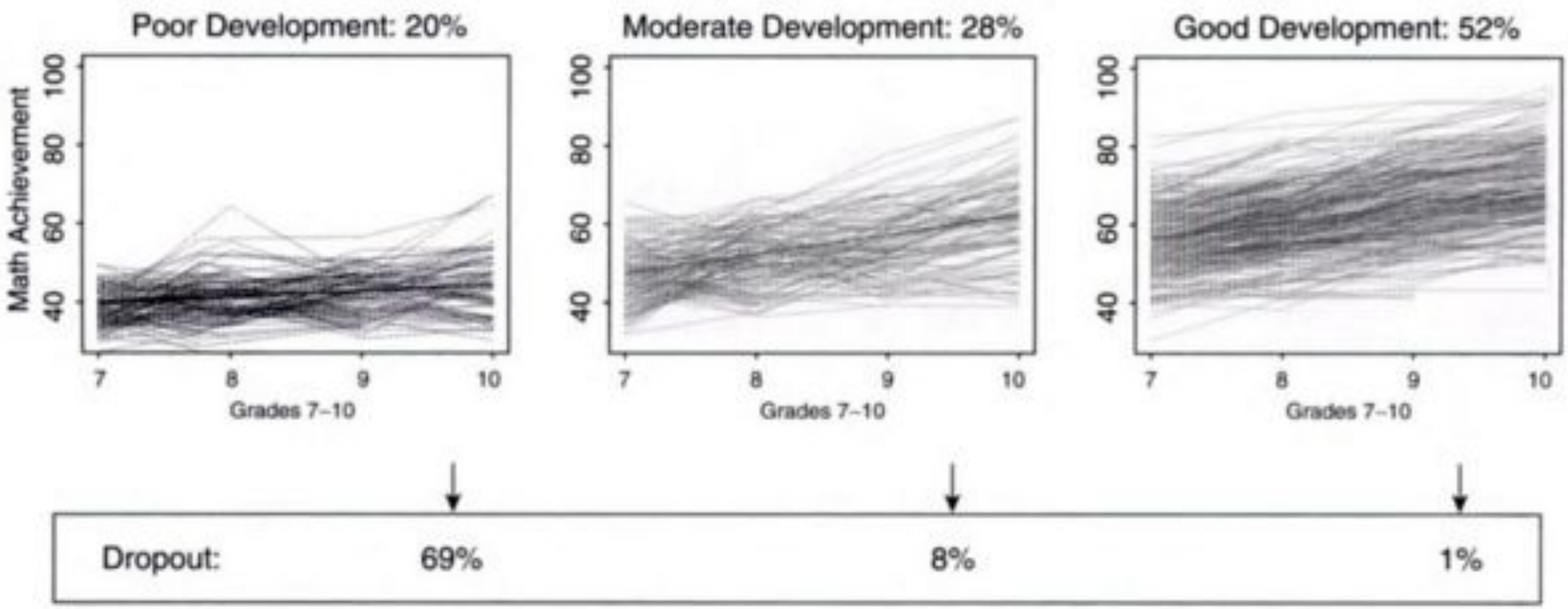


- Growth in math scores for Longitudinal Study of Youth
- Lots of variability! Lots

The question

Are there typical ways that math scores move? Can we see them?

Three basic patterns emerge



- Some covariates predict which path students are on and then that path corresponds with some outcomes, too, like drop-out

Why do growth curve analysis?

Why do growth curves I?

- We want to model the shape of growth over time
- One step more complicated than merely plotting the values over time
- How much does everyone follow the same path or do they deviate from it?
- Similar to trend analysis over pooled cross-sectional data, but now, we have the same people over time (i.e., panel data)

Why do growth curves II?

- We want to compare the starting points and trajectories (and final destinations) of groups over time
- We are usually looking for diverging trajectories
- Or converging ones
- Or ones with stable differences over time

Why do growth curves III?

- We want to see if there are inflection points, turning points, tipping points, etc. when something happens in the course of someone's life:
- How to criminal trajectories change when someone gets married?
- How do children's depression trajectories change after their parents get divorced?
- Etc.

What makes growth curves unique?

What makes growth curves unique I?

- We are modeling a full trajectory (a whole series of Y values over time) for each person as our dependent variable.
- Compare that to first differences, where we broke everything down to year-to-year changes only
- Growth curves take a longer-term perspective

What makes growth curves unique II?

- Tricky to think about how to incorporate time-varying characteristics, since we are modeling the whole growth curve
- Most unchanging variables can be entered "at baseline" – no problem
- Unchanging variables can change in importance over time; i.e., an interaction of that variable with time

What makes growth curves unique III?

- Must think hard about cumulative processes that take place over time
- How does one time period affect the next?

What makes growth curves unique IV?

- How much do we deal with individual heterogeneity and how much do we deal with overall patterns?

About growth curves

Growth curves can be modeled in a few ways

- Structural equation modeling (SEM)
- Hierarchical linear modeling (HLM)
- MANOVA

2. How do I do growth curve analysis?

How to do growth curves I

1. Look at the overall trend of Y over time
2. Plot the growth curves for each person
3. You will include time, but do you need a quadratic too?

How to do growth curves II

4. Do you need separate intercepts for each person? [*Almost always*.]
5. Do you need separate slopes (on time) for each person as well? [*Usually*.]
6. Do you need other Xs in the model? [*Sure*.]
7. Do you need interactions of some Xs with time? [*Hmmmm*?]

How to do growth curves I

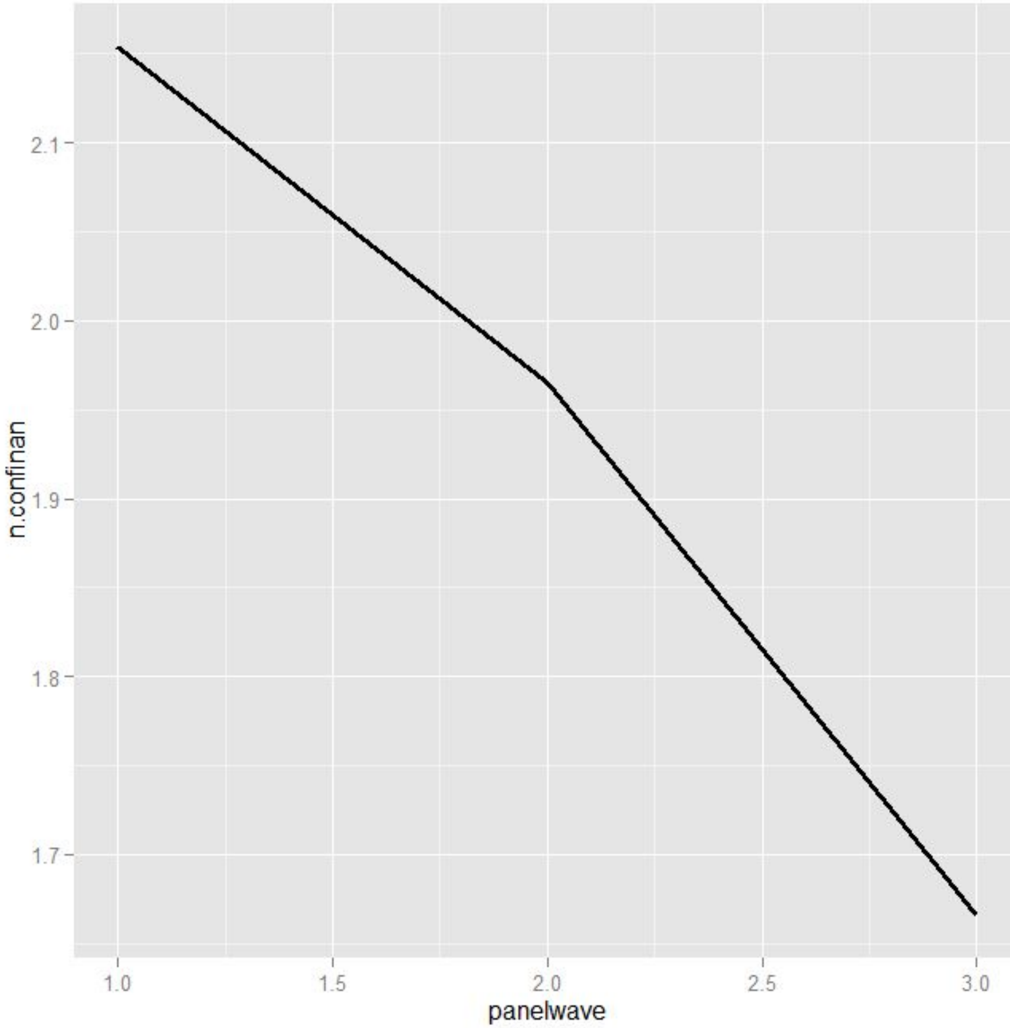
1. Look at the overall trend of Y over time

Setting up the data

```
library(QMSS)
library(ggplot2)

pd=read.csv(file.choose()) ### choose the GSS panel ###
```

Overall trend in confidence in banks

This is from Wave 1 (2006) to Wave 3 (2010)



How did I do this?

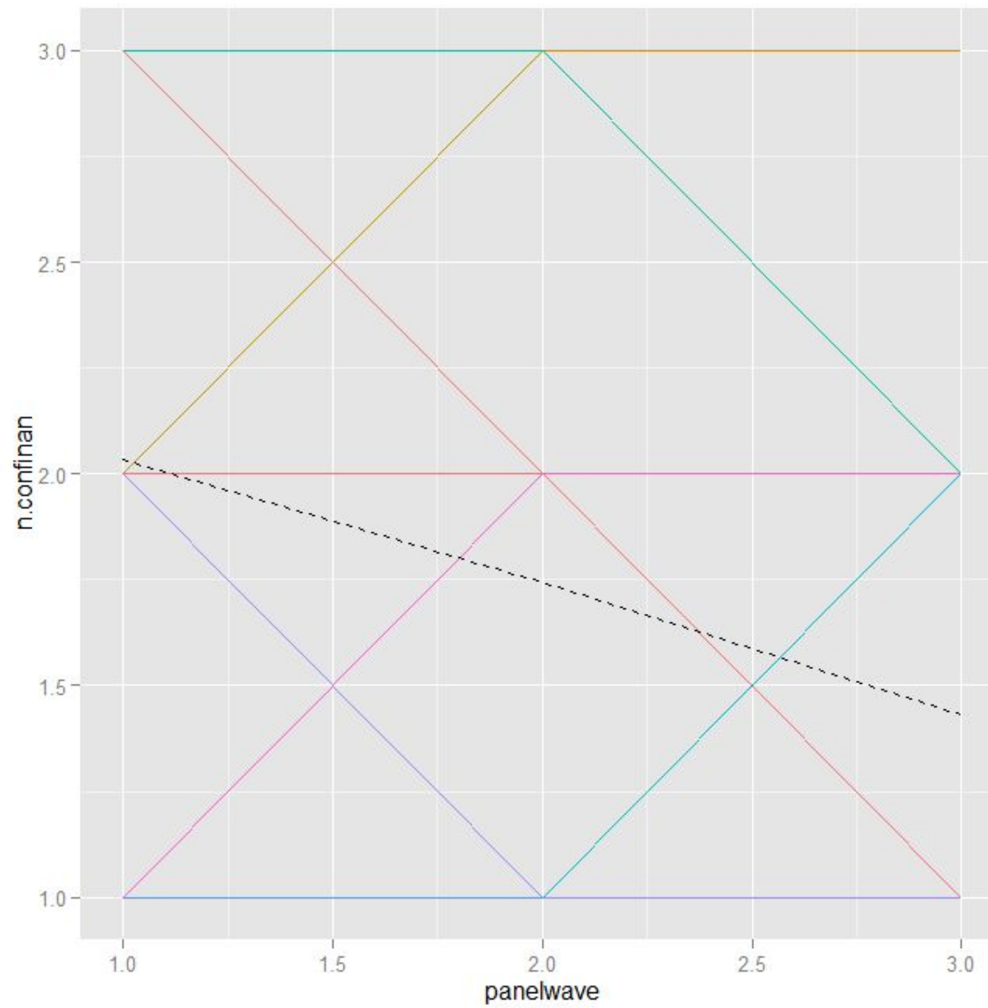
```
vars <- c("idnum", "panelwave", "sex", "age", "educ", "race", "polviews", "confinan")
sub <- pd[, vars]
sub$n.confinan <- ReverseThis(sub$confinan)

# Overall trend in confidence in banks
g_trend <- ggplot(sub, aes(x = panelwave, y = n.confinan))
(g_trend <- g_trend + stat_summary(fun.y=mean, geom="line", lwd = 1.25))
```

How to do growth curves I

2. Plot the growth curves for each person

Empirical growth curves

This is from Wave 1 (2006) to Wave 3 (2010); for the first hundred+ cases

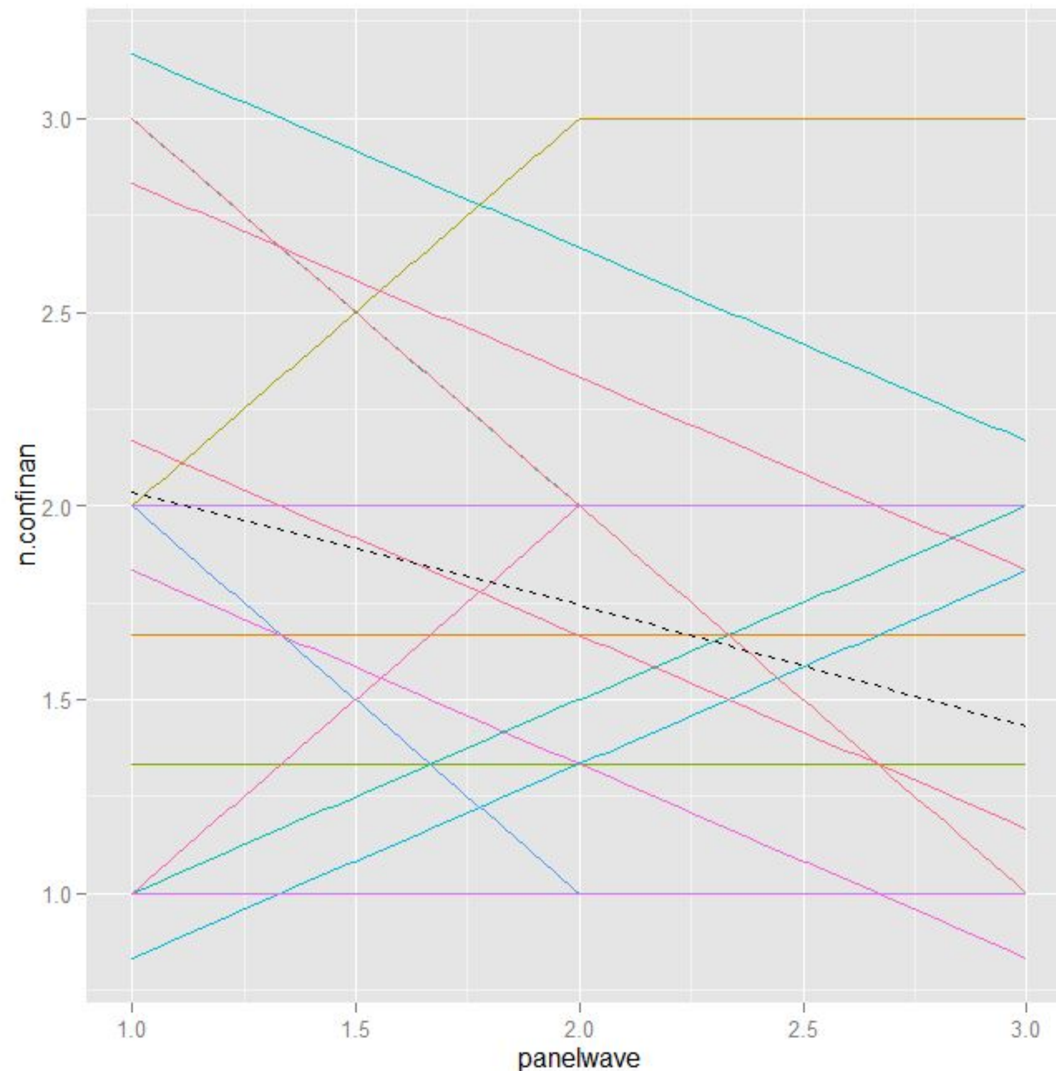


How did I do this?

```
# Empirical growth curves for idnum < 200 (& overall)
g_growth <- ggplot(subset(sub, idnum<200),
                   aes(x = panelwave, y = n.confinan, group = idnum, color =
factor(idnum)))
no_legend <- theme(legend.position="none")

g_id <- g_growth + geom_line() + no_legend
g_id + stat_summary(fun.y=mean, geom="line", aes(group=1), lty = 2, color="black")
```

Individual regressions for each person

This is from Wave 1 (2006) to Wave 3 (2010); for the first hundred+ cases



How did I do this?

```
# individual regression lines for idnum < 200 (& overall)
g_reg <- g_growth + stat_smooth(method = lm, se = F) + no_legend
g_reg + stat_summary(fun.y=mean, geom="smooth", aes(group=1), lty = 2, color = "black")
```

What do we see here?

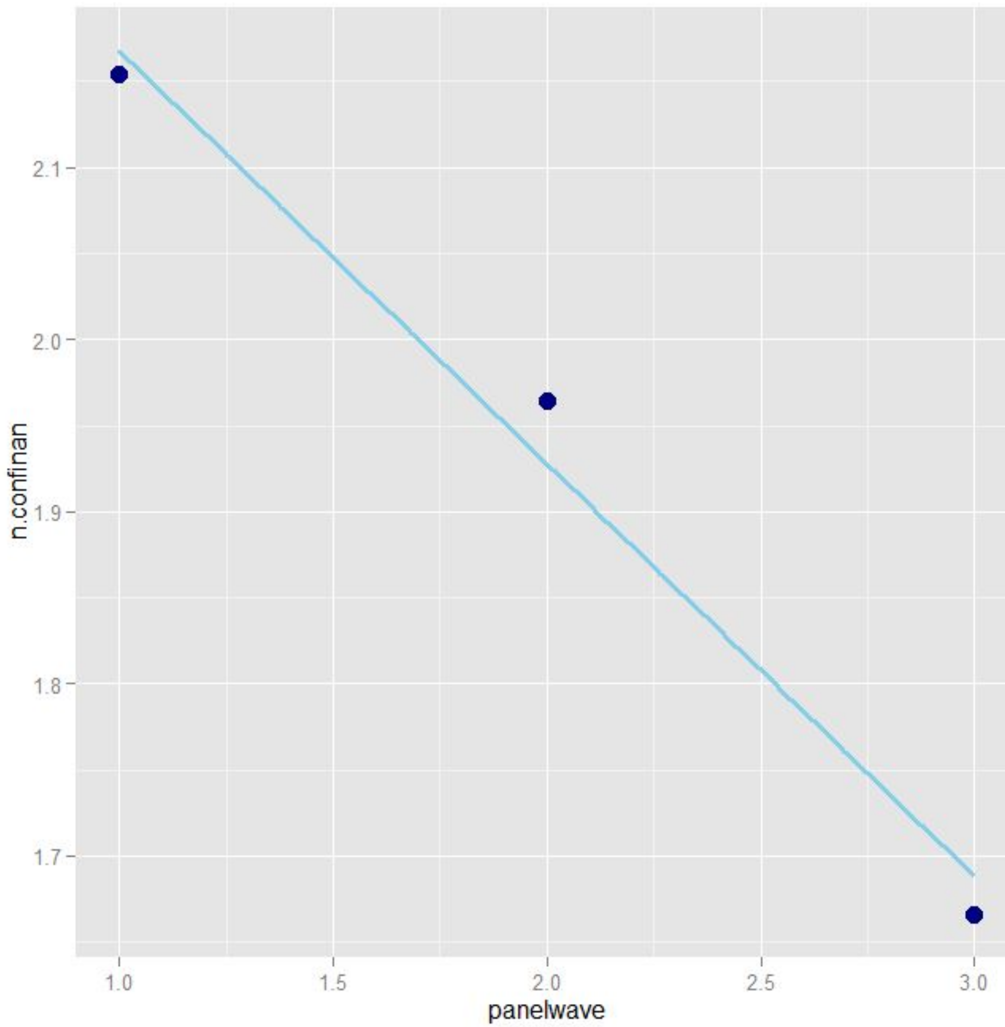
There also looks like a lot of individual variation over time regarding their opinions on banks

How to do growth curves I

3. You will include time, but do you need a quadratic too?

Overall linear prediction is this

Not a bad fit here

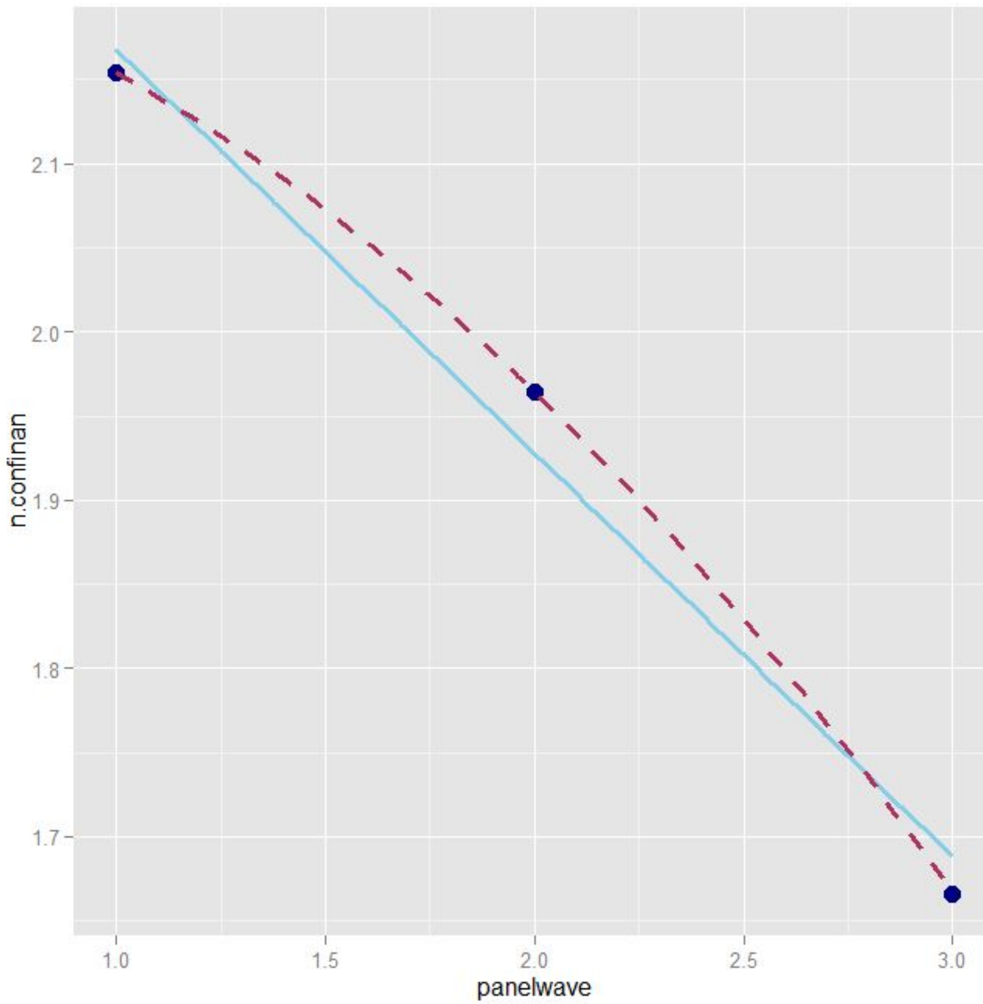


How did I do this?

```
# overall linear prediction
g_lm <- ggplot(sub, aes(x = panelwave, y = n.confinan))
g_lm <- g_lm + stat_summary(fun.y=mean, geom="point", aes(group=1), size=4, color =
"navyblue")
g_lm <- g_lm + stat_smooth(method = lm, se = F, color = "skyblue", lwd = 1.25)
g_lm
```

Overall quadratic prediction is this

This is an amazing fit, but still pretty close to linear …



How did I do this?

```
# add quadratic prediction curve
g_lm <- g_lm + stat_smooth(formula = y ~ poly(x,2), method = lm, se = F,
                           color = "maroon", lty = 2, lwd = 1.25)
g_lm
```

How to do growth curves II

4. Do you need separate intercepts for each person? [*Almost always*.]

Model with no separate intercepts …

Just an OLS with clustered and robust standard errors; at Wave 2, confidence in banks has dropped -.189, relative to Wave 1 – and at Wave 3, it has dropped -.487

```
library(plyr)
library(psych)
library(multcomp)
library(rms)
library(lme4)

# ols with clustered & robust SEs
robcov(ols(n.confinan ~ factor(panelwave), x = T, y = T, data = sub),
       cluster = sub$idnum)

                                        Model Likelihood     Discrimination
                                           Ratio Test           Indexes
Obs                              3163    LR chi2    280.89    R2       0.085
sigma                          0.6433    d.f.            2    R2 adj   0.084
d.f.                             3160    Pr(> chi2) 0.0000    g        0.208
Clusters                         1324

            Coef    S.E.   t      Pr(>|t|)
Intercept    2.1538 0.0180 119.83 <0.0001
panelwave=2 -0.1893 0.0229  -8.26 <0.0001
panelwave=3 -0.4876 0.0252 -19.34 <0.0001
```

Model with no separate intercepts …

Just an OLS with clustered and robust standard errors; with a continuous measure of time, this means that confidence in banks drops -.239 points per Wave

```
robcov(ols(n.confinan ~ panelwave, x = T, y = T, data = sub), cluster = sub$idnum)

                                               Model Likelihood     Discrimination
                                                  Ratio Test           Indexes
Obs                                    3163    LR chi2    276.04    R2        0.084
sigma                                0.6436    d.f.            1    R2 adj    0.083
d.f.                                   3161    Pr(> chi2) 0.0000    g         0.210
Clusters                               1324


          Coef    S.E.   t      Pr(>|t|)
Intercept  2.4077 0.0269  89.62 <0.0001
panelwave -0.2398 0.0126 -19.03 <0.0001
```

Model with squared time …

Just an OLS with clustered and robust standard errors; confidence in banks starts to drop only a little and then drops much further too

```
sub$panelwavesq = sub$panelwave^2

robcov(ols(n.confinan ~ panelwave + panelwavesq, x = T, y = T, data = sub), cluster =
sub$idnum)
                                            Model Likelihood     Discrimination
                                               Ratio Test           Indexes
Obs                                3163     LR chi2    280.89    R2        0.085
sigma                            0.6433     d.f.            2    R2 adj    0.084
d.f.                               3160     Pr(> chi2) 0.0000    g         0.208
Clusters                           1324


            Coef    S.E.   t      Pr(>|t|)
Intercept    2.2343 0.0690 32.40 <0.0001
panelwave   -0.0260 0.0801 -0.32 0.7453
panelwavesq -0.0544 0.0200 -2.72 0.0065
```

Now, the random intercept model

The random intercept model is this:

$y_i = \alpha_{i[j]} + \beta x_i + e_i$

which provides a unique intercept for each person *j*

Our random intercept model

The random intercept model is this:

$$nconfin_i = \alpha_{i[j]} + \beta_1 time_i + \beta_1 time_i^2 + e_i$$

which provides a unique intercept for each person *j*

A model with random intercepts …

By typing "1 | idnum" this is like saying, for each unique idnum, give me a random parameter for each variable listed. If nothing is specified, it just gives for each idnum, a unique intercept

```
lmer.confinan <- lmer(n.confinan ~ panelwave + (1|idnum), data = sub, REML = F)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
     AIC      BIC   logLik deviance df.resid
  5959.8   5984.1  -2975.9   5951.8     3159

Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1287   0.3588
 Residual             0.2852   0.5341
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41046    0.02582   93.34
panelwave   -0.24151    0.01210  -19.95

Correlation of Fixed Effects:
          (Intr)
panelwave -0.841
```

A model with random intercepts …

Give me a random parameter for each variable listed. If nothing is specified, it just gives for each idnum, a unique intercept

```
> lmer.confinan <- lmer(n.confinan ~ panelwave + (1|idnum), data = sub, REML = F)
> summary(lmer.confinan)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: n.confinan ~ panelwave + (1 | idnum)
   Data: sub

     AIC      BIC   logLik deviance df.resid
  5959.8   5984.1  -2975.9   5951.8     3159

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.26693 -0.64452 -0.03549  0.50991  2.74135

Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1287   0.3588
 Residual             0.2852   0.5341
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41046    0.02582   93.34
```

A model with random intercepts …

Confidence in banks drops -.242 points per Wave (slightly different from OLS earlier)

```
lmer.confinan <- lmer(n.confinan ~ panelwave + (1|idnum), data = sub, REML = F)

Linear mixed model fit by maximum likelihood  ['lmerMod']
     AIC      BIC   logLik deviance df.resid
  5959.8   5984.1  -2975.9   5951.8     3159

Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1287   0.3588
 Residual             0.2852   0.5341
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41046    0.02582   93.34
panelwave   -0.24151    0.01210  -19.95

Correlation of Fixed Effects:
          (Intr)
panelwave -0.841
```

What is all this “Random-effects” at bottom?

- *Std.Dev.(idnum)* is how much individual variation there is in the constant, once everyone gets their own intercept
- I.e., the constant has a mean of 2.41 and now a sd of 0.359

```
lmer.confinan <- lmer(n.confinan ~ panelwave + (1|idnum), data = sub, REML = F)

Linear mixed model fit by maximum likelihood  ['lmerMod']
     AIC       BIC    logLik deviance df.resid
  5959.8    5984.1   -2975.9   5951.8     3159

Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1287   0.3588
 Residual             0.2852   0.5341
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41046    0.02582   93.34
panelwave   -0.24151    0.01210  -19.95

Correlation of Fixed Effects:
          (Intr)
panelwave -0.841
```

What is all this “Random-effects” at bottom?

- Is giving everyone their own intercept better than OLS? Let’s compare AICs between the OLS model (run as a GLS) and the random-intercept one.
- Random-intercept AIC (i.e., 5959.8) < OLS (i.e., 6192.9), so random-intercept is an improved fit.

```
> glm.confinan = glm(n.confinan ~ panelwave, data=sub)
> summary(glm.confinan)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.40768    0.02851   84.44   <2e-16 ***
panelwave   -0.23978    0.01412  -16.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.414281)

    Null deviance: 1429.0  on 3162  degrees of freedom
Residual deviance: 1309.5  on 3161  degrees of freedom
  (2837 observations deleted due to missingness)

AIC: 6192.9
```

What is all this “Random-effects” at bottom?

This is just the calculation of Rho again.

$$\text{Rho} = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$$
$$= (.36\text{^}2)/(.36\text{^}2 + .53\text{^}2)$$
$$= 0.31$$

```
-- from earlier --
Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1287   0.3588
 Residual             0.2852   0.5341
```

```
> rho(lmer.confinan)
[1] 0.3109777
```

We have already estimated this model

Using *plm*, with option="random": that gives same coefficients and same Rho as random intercepts model

Random intercepts but quadratic on time

This may have improved the fit marginally (lower AIC); confidence in banks goes down and down even faster

```
> lmer.confinan2 <- update(lmer.confinan, ~ . + I(panelwave^2))
> summary(lmer.confinan2)

     AIC      BIC   logLik deviance df.resid
  5954.3   5984.6  -2972.1   5944.3     3158

Random effects:
 Groups   Name        Variance Std.Dev.
 idnum    (Intercept) 0.1293   0.3596
 Residual             0.2840   0.5329
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
               Estimate Std. Error t value
(Intercept)     2.22899    0.07077  31.496
panelwave      -0.01744    0.08226  -0.212
I(panelwave^2) -0.05700    0.02070  -2.754

Correlation of Fixed Effects:
            (Intr) panlwv
panelwave   -0.966
I(panlwv^2)  0.931 -0.989

> rho(lmer.confinan2)
[1] 0.3129095
```

How to do growth curves II

5. Do you need separate slopes (on time) for each person as well? [*Usually*.]

The equations -

The random intercept and random coefficient model is this:

$$y_i = \alpha_{i[j]} + \beta_{i[j]}x_i + u_{\alpha i} + u_{\beta i[j]} + e_i$$

Random intercepts + *random slopes*

A lot going on here. 1) By typing *panelwave* after the "1 +" this is like saying, allow *panelwave* to have a random coefficient (slope) for each idnum

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML = F)
> summary(lmer.confinan3)

     AIC      BIC   logLik deviance df.resid
  5956.7   5993.1  -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Random intercepts + *random slopes*

The covariance between each idnum's intercept and slope can be unstructured; we don't want the default to be covariance of zero

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML
= F)
> summary(lmer.confinan3)

     AIC      BIC   logLik deviance df.resid
  5956.7   5993.1  -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Random intercepts + *random slopes*

Interpretation of constant. At Wave 0, on average, people have a 2.41 score, but there is substantial variation around that mean, with a st dev of .499

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML
= F)
> summary(lmer.confinan3)

     AIC      BIC   logLik deviance df.resid
  5956.7   5993.1  -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Random intercepts + *random slopes*

Interpretation of slope. On average, people's scores drop by -.24 per Wave, but there is substantial variation around that average slope, with a st dev of .149

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML
= F)
> summary(lmer.confinan3)

     AIC      BIC   logLik deviance df.resid
  5956.7   5993.1  -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Random intercepts + *random slopes*

Interpretation of corr(panel~ve,_cons) I. There is a high negative correlation between the constant and the slope for each idnum (ρ=-.69).

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML
= F)
> summary(lmer.confinan3)

     AIC       BIC    logLik deviance df.resid
  5956.7    5993.1   -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Random intercepts + *random slopes*

Interpretation of corr(panel~ve,_cons) II. When people started out with a high nconfin score, they fell more rapidly over time; conversely, those with low initial nconfin scores *increased* them rapidly

```
> lmer.confinan3 <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub, REML
= F)

     AIC       BIC   logLik deviance df.resid
  5956.7    5993.1  -2972.3   5944.7     3157

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.24955  0.4995
          panelwave   0.02236  0.1495   -0.69
 Residual             0.26280  0.5126
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.41022    0.02680   89.94
panelwave   -0.24128    0.01245  -19.38

Correlation of Fixed Effects:
          (Intr)
panelwave -0.854
```

Are random slopes necessary?

We can determine if random slopes better fit our data by comparing the likelihood ratio from the random intercept model (LL=-2976) with the LR from the random intercept + random slopes model (LL=-2972)

```
> anova(lmer.confinan, lmer.confinan3)

Data: sub
Models:
lmer.confinan: n.confinan ~ panelwave + (1 | idnum)
lmer.confinan3: n.confinan ~ panelwave + (1 + panelwave | idnum)

               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer.confinan   4 5959.8 5984.1 -2975.9   5951.8
lmer.confinan3  6 5956.7 5993.1 -2972.3   5944.7 7.1503      2    0.02801 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Are random slopes necessary?

The random intercept + random slopes model provides a superior fit to the data than the random intercepts model (p<.05). But the AIC change is pretty minimal (=3 points)

```
> anova(lmer.confinan, lmer.confinan3)

Data: sub
Models:
lmer.confinan: n.confinan ~ panelwave + (1 | idnum)
lmer.confinan3: n.confinan ~ panelwave + (1 + panelwave | idnum)

               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer.confinan   4 5959.8 5984.1 -2975.9   5951.8
lmer.confinan3  6 5956.7 5993.1 -2972.3   5944.7 7.1503      2    0.02801 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How to do growth curves II

6. Do you need other Xs in the model? [Sure.]

Add in time-invariant characteristics

Net of time, men are -.11 lower on the nconfin scale. Easy enough.

```
> sub$male <- ifelse(sub$sex==1, "male", "female")
> lmer.confinan4 <- update(lmer.confinan3, ~ . + male)
> summary(lmer.confinan4)

     AIC      BIC   logLik deviance df.resid
  5943.3   5985.7  -2964.6   5929.3     3156

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.25402  0.5040
          panelwave   0.02217  0.1489   -0.71
 Residual             0.26298  0.5128
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.45701    0.02934   83.74
panelwave   -0.24174    0.01242  -19.46
malemale    -0.11033    0.02784   -3.96

Correlation of Fixed Effects:
          (Intr) panlwv
panelwave -0.788
malemale  -0.403  0.010
```

Add in a time-varying characteristics too

Net of time and gender, each move up the political conservativism scale increases one's nconfin score by 0.016 (p<.1)

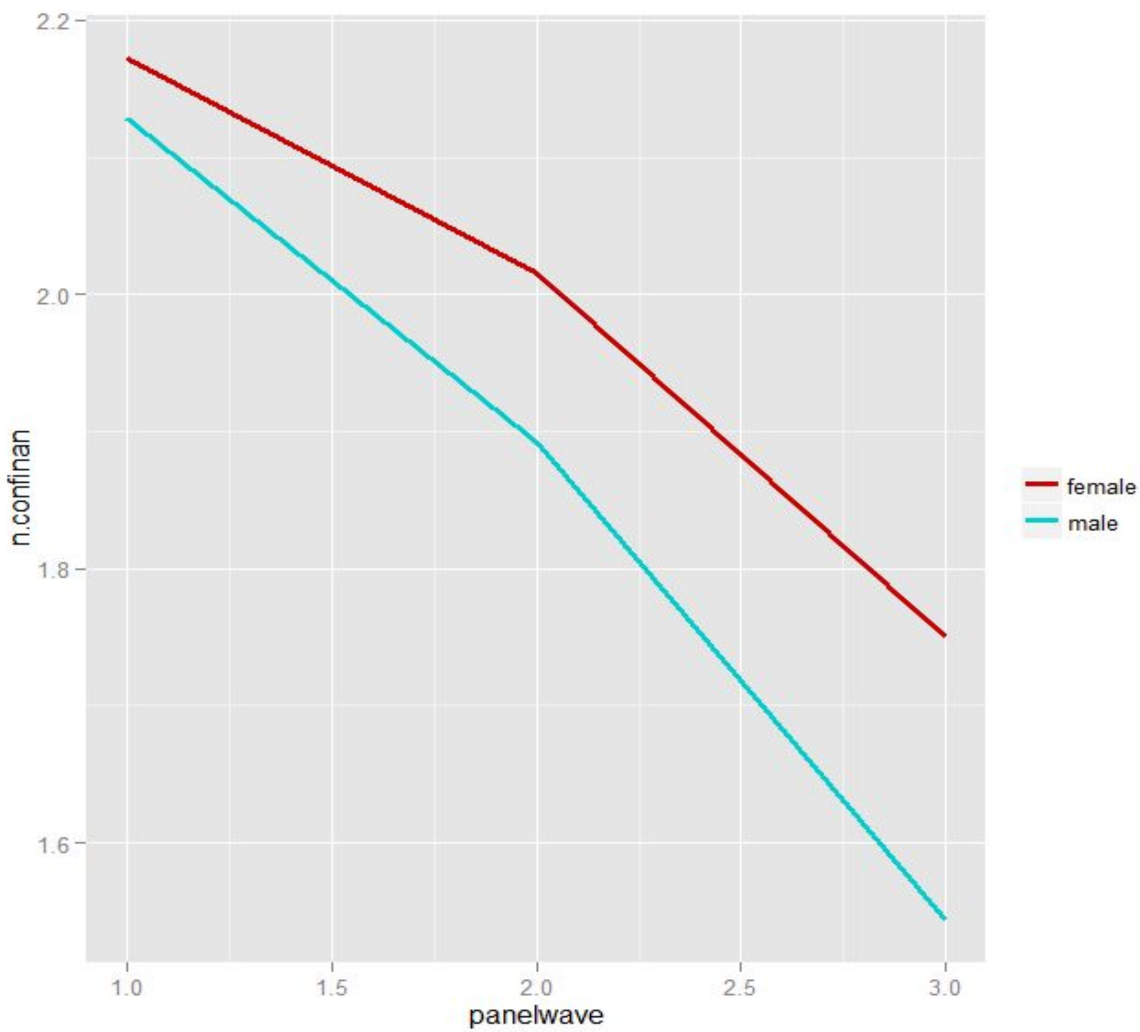
```
> lmer.confinan5 <- update(lmer.confinan4, ~ . + polviews)

     AIC      BIC   logLik deviance df.resid
  5759.6   5807.9  -2871.8   5743.6     3061

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.2612   0.5110
          panelwave   0.0240   0.1549   -0.72
 Residual             0.2610   0.5109
Number of obs: 3069, groups:  idnum, 1301

Fixed effects:
             Estimate Std. Error t value
(Intercept)  2.395228   0.046015   52.05
panelwave   -0.242785   0.012653  -19.19
malemale    -0.103955   0.028089   -3.70
polviews     0.015831   0.008601    1.84

Correlation of Fixed Effects:
          (Intr) panlwv maleml
panelwave -0.503
malemale  -0.268  0.008
polviews  -0.762 -0.012  0.013
```

How to do growth curves II

7. Do you need interactions of some Xs with time?

Changes in confidence in banks, by sex

This is from Wave 1 (2006) to Wave 3 (2010)



How did I do this?

```
colors_and_labels <- scale_color_manual(values = c("red3", "cyan3"), name = "")
g_sex <- ggplot(sub, aes(x = panelwave, y = n.confinan, color = male))
(g_sex <- (g_sex + stat_summary(fun.y=mean, geom="line", lwd = 1.25)))
g_sex + colors_and_labels
```

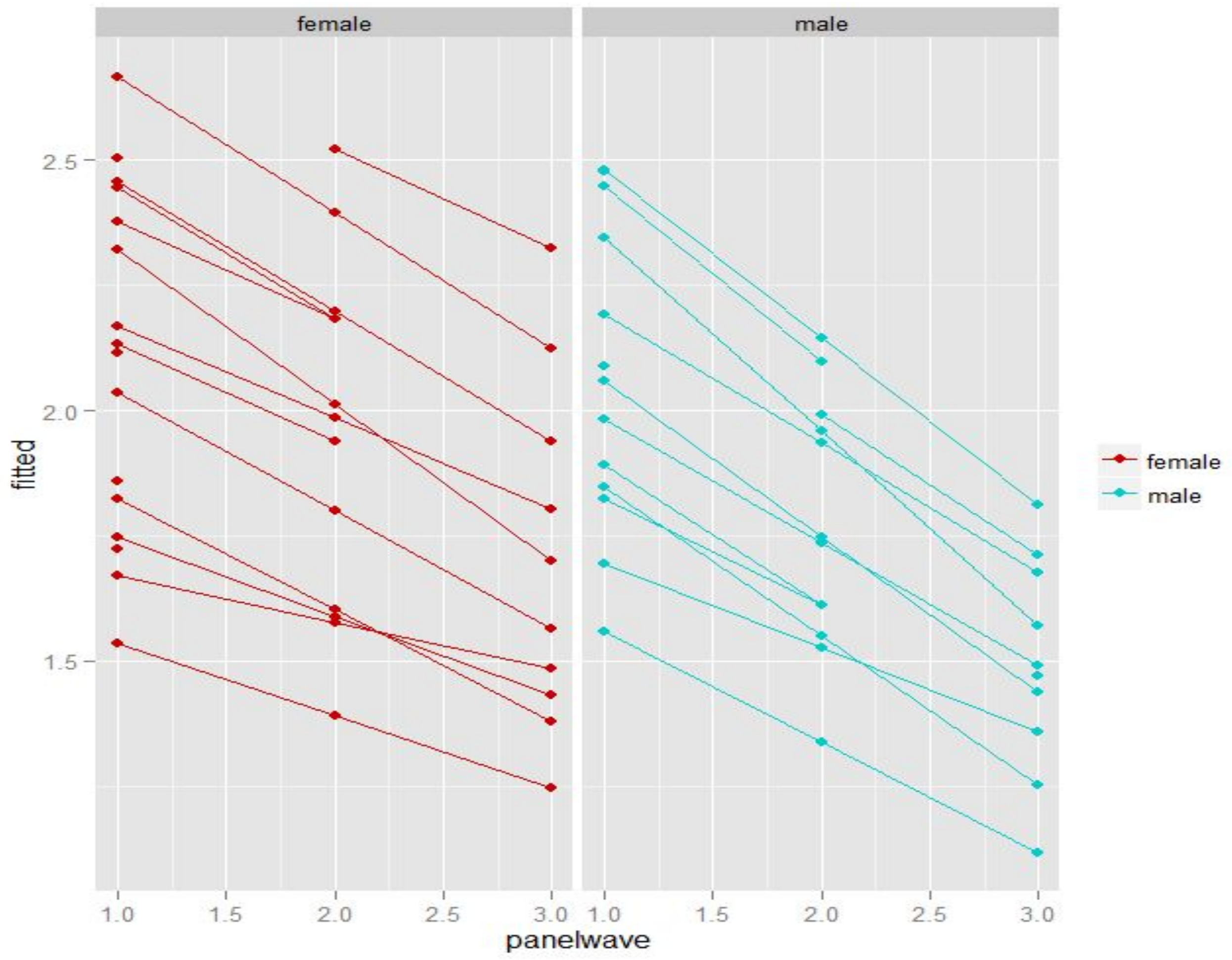
Add in a time-invariant characteristic

Males are losing confidence in banks faster than females (B on the interaction = -.081, p<.01)

```
> lmer.confinan6 <- update(lmer.confinan5, ~ . + male:panelwave - polviews)

     AIC      BIC   logLik deviance df.resid
  5934.9   5983.3  -2959.4   5918.9     3155

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.2466   0.4966
          panelwave   0.0200   0.1414   -0.70
 Residual             0.2635   0.5133
Number of obs: 3163, groups:  idnum, 1324

Fixed effects:
                   Estimate Std. Error t value
(Intercept)         2.39458    0.03505   68.32
panelwave          -0.20828    0.01611  -12.93
malemale            0.04022    0.05424    0.74
panelwave:malemale -0.08145    0.02516   -3.24

Correlation of Fixed Effects:
            (Intr) panlwv maleml
panelwave   -0.857
malemale    -0.646  0.555
panlwv:mlml  0.551 -0.642 -0.858
```

The predicted curves for males and females

What do you see?



How did I do this?

```
model.dat <- cbind(model.frame(lmer.confinan6), fitted = fitted(lmer.confinan6))
model.dat <- subset(model.dat, idnum < 200)
g_sex_fit <- ggplot(model.dat, aes(x = panelwave, y = fitted, group = idnum, color =
male))
(g_sex_fit <- g_sex_fit + geom_line() + geom_point() + facet_grid( . ~ male))
g_sex_fit + colors_and_labels
```

Maybe I should enter these interactions as random effects?

Why? Because perhaps the intercepts and/or slopes are more variable for one gender over the other.

Random intercepts/random slopes for men

Males are losing confidence in banks at -.292 per Wave, with a sd on that slope of .105; the constant has a mean of 2.43 and a sd of .509

```
> lmer.confinanM <- lmer(n.confinan ~ panelwave + (1 + panelwave | idnum), data = sub,
subset = sex == 1)
> summary(lmer.confinanM)

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.26126  0.5111
          panelwave   0.01148  0.1072   -0.78
 Residual             0.27684  0.5262
Number of obs: 1309, groups:  idnum, 568

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.43758    0.04240   57.49
panelwave   -0.29204    0.01914  -15.26

Correlation of Fixed Effects:
          (Intr)
panelwave -0.856
```

Random intercepts/random slopes for women

Females are losing confidence in banks at -.209 per Wave, with a sd on that slope of .171; the constant has a mean of 2.397 and a sd of .501

```
> lmer.confinanW <- update(lmer.confinanM, subset = sex == 2)
> summary(lmer.confinanW)

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 idnum    (Intercept) 0.25308  0.5031
          panelwave   0.02961  0.1721   -0.70
 Residual             0.24831  0.4983
Number of obs: 1854, groups:  idnum, 777

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.39741    0.03455    69.4
panelwave   -0.20966    0.01625   -12.9

Correlation of Fixed Effects:
          (Intr)
panelwave -0.856
```

Are the intercepts and/or slopes more variable for one gender over the other?

Maybe a little bit. The standard deviation of the slope for males is only .10 but for females it is .17. This means that the growth curves for males are bunched a bit more together than for females. How much should we worry about this?

Not unlike sequence analysis too

Be creative about “types” of experiences, i.e.:

- “Early delinquents” vs. “late delinquents”
- “Young divorce” vs. “old divorce”
- Etc.