Example 2 Part 1: Alternative Metrics of Time in Accelerated Longitudinal Designs
(complete syntax, data, and output available for STATA and R electronically)

These data come from Hoffman (2015) chapter 10, which examined prediction of a memory outcome (prose recall) from three metrics of time: years since birth (centered at 85 years) and years in study (centered at the first occasion) in a sample of 557 observations from 207 older adults. This example first estimates empty models for each time-varying variable (recall, age, time), saturated means for recall by age and time, and then a series of unconditional and conditional models for change evaluating age convergence (i.e., age cohort effects). I am using maximum likelihood estimation to facilitate all model comparisons (and Part 2's translation into SEM).

STATA Syntax for Importing and Preparing Data for Analysis:

```
// Defining global variable for file location to be replaced in code below
// \\Client\ precedes path in Virtual Desktop outside H drive
     global filesave "C:\Dropbox\23_PSQF7375_AdvLong\PSQF7375_AdvLong_Example2"

// Import chapter 10a stacked data and create centered predictors for analysis
   use "$filesave\STATA_Chapter10a.dta", clear
// Time in study
   gen time = tvage-aget0
   gen timesq = time*time
// Fixing 1 case rounded to 9
   replace occasion=8 if (occasion==9)
// Age (years since birth) variables
   gen roundage = round(tvage,1)
// Fixing 2 cases above 95
   replace roundage=95 if (roundage==97)
   replace roundage=95 if (roundage==100)
   gen tvage84 = tvage-84
   gen tvage84sq = tvage84*tvage84
   gen aget084 = aget0-84
   gen aget084sq = aget084*aget084
   label variable time         "time: Years since Time 0"
   label variable timesq       "timesq: Squared Years since Time 0"
   label variable roundage     "roundage: Age Rounded to Nearest Year"
   label variable tvage84      "tvage84: Time-Varying Age (0=84 years)"
   label variable tvage84sq    "tvage84sq: Squared Time-Varying Age (0=84 years)"
   label variable aget084      "aget084: Age at Time 0 (0=84 years)"
   label variable aget084sq    "aget084: Squared Age at Time 0 (0=84 years)"

// Subset sample to complete cases for all predictors
   egen nummiss = rowmiss(tvage aget0 recall)
   drop if nummiss>0
```

Time is current age – baseline age, which was created in order to represent the purely longitudinal variance of time-varying age.

Also, I am building squared variables to simplify the model syntax below (but these versions won't work with the predicted outcomes code used in the chapter 10 syntax at the book's website).

R Syntax for Importing and Preparing Data for Analysis (after loading packages haven, expss, TeachingDemos, psych, lmerTest, performance, and prediction as shown online):

```
# Define variables for working directory and data name
filesave = "C:\\Dropbox/23_PSQF7375_AdvLong/PSQF7375_AdvLong_Example2/"
filename = "SAS_Chapter10a.sas7bdat"
setwd(dir=filesave)

# Import chapter 10a stacked data with labels
Example2 = read_sas(data_file=paste0(filesave,filename))
# Convert to data frame as data frame without labels to use for analysis
Example2 = as.data.frame(Example2)

# Create centered predictors for analysis
# Time in study
Example2$time=Example2$tvage-Example2$ageT0
Example2$timesq=Example2$time*Example2$time
# Fixing 1 case rounded to 9
Example2$occasion[which(Example2$occasion==9)]=8
# Age (years since birth) variables
Example2$roundage=round(Example2$tvage,digits=0)
```

```
# Fixing 2 cases above 95
Example2$roundage[which(Example2$roundage>95)]=95
Example2$tvage84=Example2$tvage-84
Example2$tvage84sq=Example2$tvage84*Example2$tvage84
Example2$ageT084=Example2$ageT0-84
Example2$ageT084sq=Example2$ageT084*Example2$ageT084

# Subset sample to complete cases for all predictors
Example2 = Example2[complete.cases(Example2[ , c("tvage","ageT0","recall")]),]
```

<u>STATA</u> Syntax and Output for Data Description:

```
display "Descriptive Statistics"
summarize aget0 tvage time recall
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       aget0 |        557    82.97113    2.688563   79.41918   97.77808
       tvage |        557    85.64534     3.55728   79.41918   99.89863
        time |        557    2.674216    2.603101          0   8.502732
      recall |        557     10.1939    3.826512          0         16
```

```
pwcorr aget0 tvage time recall, sig
             |    aget0    tvage     time   recall
-------------+------------------------------------
       aget0 |   1.0000
             |
       tvage |   0.6852   1.0000
             |   0.0000
             |
        time |  -0.0965   0.6589   1.0000
             |   0.0228   0.0000
             |
      recall |  -0.1230  -0.0630   0.0409   1.0000
             |   0.0037   0.1375   0.3352
```

<u>R</u> Syntax for Data Description:

```
print("Descriptive Statistics")
describe(x=Example2[ , c("ageT0","tvage","time","recall")])
       vars   n  mean   sd median trimmed  mad   min   max range  skew kurtosis   se
ageT0     1 557 82.97 2.69  82.25   82.64 2.24 79.42 97.78 18.36  1.23     1.84 0.11
tvage     2 557 85.65 3.56  85.26   85.45 3.92 79.42 99.90 20.48  0.50    -0.16 0.15
time      3 557  2.67 2.60   2.09    2.37 3.09  0.00  8.50  8.50  0.64    -0.72 0.11
recall    4 557 10.19 3.83  11.00   10.49 2.97  0.00 16.00 16.00 -0.66    -0.32 0.16
```

```
corr.test(x=Example2[ , c("ageT0","tvage","time","recall")])

Correlation matrix
       ageT0 tvage  time recall
ageT0   1.00  0.69 -0.10  -0.12
tvage   0.69  1.00  0.66  -0.06
time   -0.10  0.66  1.00   0.04
recall -0.12 -0.06  0.04   1.00
Sample Size
[1] 557

Probability values (Entries above the diagonal are adjusted for multiple tests.)
       ageT0 tvage time recall
ageT0   0.00  0.00 0.07   0.01
tvage   0.00  0.00 0.00   0.27
time    0.02  0.00 0.00   0.34
recall  0.00  0.14 0.34   0.00
```

Because baseline age and time do have some (negative) correlation, they do not provide a complete separation of the between-person and within-person variance in age (i.e., as would have been obtained by centering using person mean age instead of baseline age to create time, analogous to chapter 8). Instead, we will think of **baseline age** as representing **cross-sectional** age variance and **time** as representing **longitudinal** age variance.

STATA and R Syntax and Partial Output for Empty Means, Random Intercept Model for Age:

```
display "Empty Means, Random Intercept Model for Age"
mixed tvage , || personid: , mle nolog
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
personid: Identity           |
                  var(_cons) |   5.091145   .8861275      3.619608    7.160929
-----------------------------+------------------------------------------------
               var(Residual) |    7.77283   .5869496      6.703513    9.012721
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 72.93         Prob >= chibar2 = 0.0000

estat icc  // Intraclass correlation
------------------------------------------------------------------------------
                       Level |        ICC   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                    personid |   .3957676   .0500664      .3029102    .4968037
------------------------------------------------------------------------------

print("Empty Means, Random Intercept Model for Age")
Age = lmer(data=Example2, REML=FALSE, formula=tvage~1+(1|PersonID))
summary(Age); icc(Age)

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 5.0911   2.2564
 Residual             7.7728   2.7880

# Intraclass Correlation Coefficient
    Adjusted ICC: 0.396
  Unadjusted ICC: 0.396
```

$$ICC = \frac{5.0911}{5.0911 + 7.7728} = .396$$

So 40% of the variance in age is actually cross-sectional—due to age mean differences! This means that age can have both cross-sectional (~BP) and longitudinal (~WP) effects simultaneously.

STATA and R Syntax and Partial Output for Empty Means, Random Intercept Model for Time:

```
display "Empty Means, Random Intercept Model for Time"
mixed time , || personid: , mle nolog

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
personid: Identity           |
                  var(_cons) |   3.74e-20   1.03e-19      1.70e-22    8.23e-18
-----------------------------+------------------------------------------------
               var(Residual) |   6.763969   .4053182      6.014436    7.606909
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 0.00          Prob >= chibar2 = 1.0000

estat icc  // Intraclass correlation
------------------------------------------------------------------------------
                       Level |        ICC   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                    personid |   5.53e-21          0      5.53e-21    5.53e-21
------------------------------------------------------------------------------

print("Empty Means, Random Intercept Model for Time")
Time = lmer(data=Example2, REML=FALSE, formula=time~1+(1|PersonID))
summary(Time); icc(Time)

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 0.000    0.0000
 Residual             6.764    2.6008

> icc(Time)
[1] NA
```

$$ICC = \frac{0}{0 + 6.764} = 0$$

All of the variance in time is within persons—this means time can only have a longitudinal (~WP) effect.

Syntax and Partial Output for Empty Means, Random Intercept Model for Recall:

Level-1: $\boldsymbol{recall_{ti} = \beta_{0i} + e_{ti}}$

Level-2: $\boldsymbol{\beta_{0i} = \gamma_{00} + U_{0i}}$

STATA:

```
display "Model 0: Empty Means, Random Intercept Model for Recall Outcome"
mixed recall , || personid: , mle nolog

Log likelihood = -1428.6775
------------------------------------------------------------------------------
      recall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   9.734908   .2505783    38.85   0.000     9.243784    10.22603
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
personid: Identity           |
                  var(_cons) |   10.45792   1.309559      8.181943    13.36701
-----------------------------+------------------------------------------------
               var(Residual) |   5.164586   .3930054      4.448999     5.99527
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 217.28        Prob >= chibar2 = 0.0000

display "-2LL = " e(ll)*-2  // Print -2LL for model
-2LL = 2857.355

estat icc  // Intraclass correlation
------------------------------------------------------------------------------
                       Level |        ICC   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                    personid |   .6694138   .0347124       .598231    .7336009
------------------------------------------------------------------------------
```

R:

```
print("Model 0: Empty Means, Random Intercept Model for Recall Outcome")
Empty = lmer(data=Example2, REML=FALSE, formula=recall~1+(1|PersonID))
llikAIC(Empty); summary(Empty); icc(Empty)

 $AICtab
      AIC        BIC     logLik   deviance   df.resid
 2863.3550  2876.3227 -1428.6775  2857.3550   554.0000   deviance = -2LL (for homework)

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 10.4579  3.2339   Level-2 variance of U0i
 Residual              5.1646  2.2726   Level-1 variance of eti
Number of obs: 557, groups:  PersonID, 207

Fixed effects:
             Estimate Std. Error        df t value  Pr(>|t|)
(Intercept)   9.73491    0.25058 197.01209   38.85 < 2.2e-16

# Intraclass Correlation Coefficient
    Adjusted ICC: 0.669
  Unadjusted ICC: 0.669
```

$$ICC = \frac{10.4579}{10.4579 + 5.1646} = .669$$

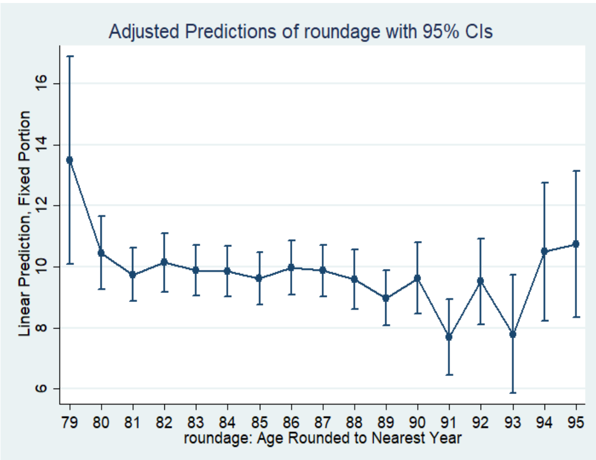
So 67% of the variance in recall is initially due to person mean differences.

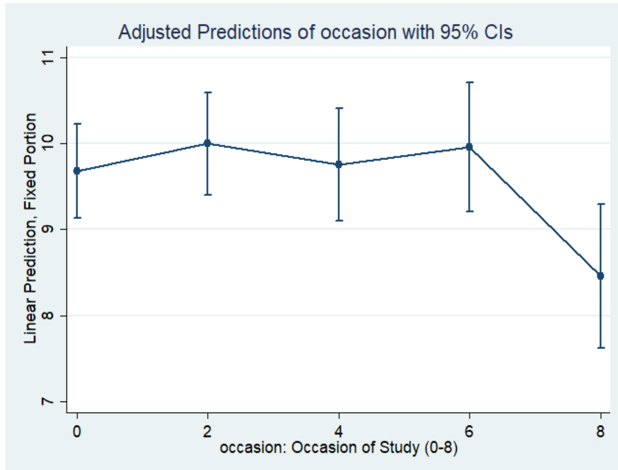
Next, we will see what the mean trajectory for recall looks like over age and time…

STATA Syntax and Plots for Saturated Means for Recall by Age and Time

```
display "Saturated Means by Rounded Age, Random Intercept Model"
mixed recall i.roundage, || personid: , mle nolog
margins  i.roundage  // get saturated means per age and plot them
marginsplot, xdimension(roundage) name(by_age, replace)
graph export "$filesave\STATA plots\STATA Recall by Age.png", replace

display "Saturated Means by Rounded Time, Random Intercept Model"
mixed recall i.occasion, || personid: , mle nolog
margins  i.occasion  // get saturated means per occasion and plot them
marginsplot, xdimension(occasion) name(by_time, replace)
graph export "$filesave\STATA plots\STATA Recall by Time.png", replace
```





R Syntax for Saturated Means for Recall by Age and Time (see syntax online for plots)

```
print("Saturated Means by Rounded Age, Random Intercept Model")
SatAge = lmer(data=Example2, REML=FALSE, formula=recall~0+as.factor(roundage)+(1|PersonID))
summary(SatAge)

print("Saturated Means by Rounded Occasion, Random Intercept Model")
SatTim = lmer(data=Example2, REML=FALSE, formula=recall~0+as.factor(occasion)+(1|PersonID))
summary(SatTim)
```

Model 1a. Syntax and Partial Output for Fixed Quadratic Age, Random Intercept for Recall:

Level-1 Age: $recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - 84) + \beta_{2i}(age_{ti} - 84)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + U_{0i},\ \ \beta_{1i} = \gamma_{10},\ \ \beta_{2i} = \gamma_{20}$

Model 1a STATA:

```
display "Model 1a Age: Fixed Quadratic, Random Intercept Model"
mixed recall c.tvage84 c.tvage84sq, || personid: , mle nolog
display "-2LL = " e(ll)*-2  // Print -2LL for model
estat ic, n(207)            // Information criteria at N=# persons
estimates store FitRIAge    // Save for LRT
predict PredAge, xb         // Save fixed-pred outcomes for total-R2
corr recall PredAge         // Get total r to make R2
     display r(rho)^2       // Print total R2 relative to empty model
```

Model 1a R:

```
print("Model 1a Age: Fixed Quadratic, Random Intercept Model")
RIAge = lmer(data=Example2, REML=FALSE, formula=recall~1+tvage84+tvage84sq+(1|PersonID))
llikAIC(RIAge); summary(RIAge)
```

```
$AICtab
       AIC        BIC     logLik   deviance   df.resid
 2860.6264  2882.2393 -1425.3132  2850.6264   552.0000   deviance = −2LL

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 10.4804  3.2373  Level-2 variance of U0i
 Residual              5.0716  2.2520  Level-1 variance of eti
Number of obs: 557, groups:  PersonID, 207

Fixed effects:
               Estimate  Std. Error          df t value Pr(>|t|)
(Intercept)   9.8196505   0.2634293 230.9528321 37.2762  < 2e-16   gamma00
tvage84      -0.1189889   0.0516498 465.3551305 -2.3038  0.02168   gamma10
tvage84sq     0.0047917   0.0075791 474.9664062  0.6322  0.52755   gamma20
```

Interpret these fixed effects:

Intercept γ_{00} =

Slope for age_{ti} γ_{10} =

Slope for age_{ti}^2 γ_{20} =

```
print("Total R2 for fixed age slopes")
Example2$PredRIAge = predict(RIAge, re.form=NA)
rRIAge = cor.test(Example2$PredRIAge, Example2$recall, method="pearson")
rRIAge$estimate^2
         cor
0.0046803019
```

Model 1b. Syntax and Partial Output for Fixed Quadratic Time, Random Intercept for Recall:

Level-1 Time: $recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - ageT0_i) + \beta_{2i}(age_{ti} - ageT0_i)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + U_{0i},\ \ \beta_{1i} = \gamma_{10},\ \ \beta_{2i} = \gamma_{20}$

Where $age_{ti} - ageT0_i = time_{ti}$ (as years-in-study rather than years-since-birth)

Model 1b STATA:

```
display "Model 1b Time: Fixed Quadratic, Random Intercept Model"
mixed recall c.time c.timesq, || personid: , mle nolog
display "-2LL = " e(ll)*-2  // Print -2LL for model
estat ic, n(207)            // Information criteria at N=# persons
estimates store FitRITim    // Save for LRT
predict PredTim, xb         // Save fixed-pred outcomes for total-R2
corr recall PredTim         // Get total r to make R2
     display r(rho)^2       // Print total R2 relative to empty model
```

Model 1b R:

```
print("Model 1b Time: Fixed Quadratic, Random Intercept Model")
RITim = lmer(data=Example2, REML=FALSE, formula=recall~1+time+timesq+(1|PersonID))
llikAIC(RITim); summary(RITim)

$AICtab
       AIC        BIC     logLik   deviance   df.resid
 2856.0088  2877.6216 -1423.0044  2846.0088   552.0000   deviance = −2LL

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 10.622   3.2591  Level-2 variance of U0i
 Residual              4.983   2.2323  Level-1 variance of eti
```

```
Fixed effects:
              Estimate Std. Error         df t value  Pr(>|t|)
(Intercept)   9.660987   0.274986 282.359753 35.1326 < 2.2e-16  gamma00
time          0.261331   0.119243 377.995056  2.1916  0.029019  gamma10
timesq       -0.046907   0.015826 366.791758 -2.9640  0.003235  gamma20
```

Interpret these fixed effects:

Intercept γ_{00} =

Slope for $time_{ti}$ γ_{10} =

Slope for $time_{ti}^2$ γ_{20} =

```
print("Total R2 for fixed time slopes")
Example2$PredRITim = predict(RITim, re.form=NA)
rRITim = cor.test(Example2$PredRITim, Example2$recall, method="pearson")

rRITim$estimate^2
         cor
0.0027296325
```

Pseudo-R^2 Table—level-1 time explains more level-1 residual variance than level-1 age:

Model	Random Intercept Variance	Residual Variance	100*% Random Intercept Reduced	100*% Residual Variance Reduced
Model 0: Empty Means, Random Intercept	10.4578	5.1646		
Model 1a Age: Fixed Quadratic, Random Intercept Model	10.4803	5.0716		
R2 change from level-1 fixed effects			-0.22	1.80
Model 0: Empty Means, Random Intercept	10.4578	5.1646		
Model 1b Time: Fixed Quadratic, Random Intercept Model	10.6213	4.9831		
R2 change from level-1 fixed effects			-1.56	3.51

From here, I am adding terms in a deliberately different order than I typically recommend so that I can make some pedagogical points. First, I add effects of **baseline age** ($ageT0_i - 84$) to each fixed quadratic, random intercept model to see how its fixed slopes differ in interpretation when using *age* versus *time* in the level-1 model. Then I add **random slopes** for each level-1 predictor to see how their level-2 random effect variances differ between models. Finally, I show syntax and partial output for novel models (relative to what was covered in chapter 10) that try to repair the mis-specification introduced into the random slope age-as-time model.

Model 2a. Syntax and Partial Output for Fixed Quadratic Age, Random Intercept for Recall, adding Age at Baseline to Add Contextual Birth Cohort Effects (that Test Age Convergence):

Level-1 Age: $\boldsymbol{recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - 84) + \beta_{2i}(age_{ti} - 84)^2 + e_{ti}}$

Level-2: $\boldsymbol{\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i}}$

$\boldsymbol{\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i),\ \ \beta_{2i} = \gamma_{20}}$

Model 2a STATA:

```
display "Model 2a Age: Fixed Quadratic, Random Intercept Model"
display "Controlling for Birth Cohort as Contextual Effects"
mixed recall c.tvage84 c.tvage84sq c.aget084 c.aget084sq c.tvage84#c.aget084, ///
        || personid: , mle nolog
display "-2LL = " e(ll)*-2  // Print -2LL for model
estat ic, n(207)   // Information criteria at N=# persons
```

```
// Total Linear Birth Cohort on Intercept
   lincom c.aget084*1 + c.tvage84*1
// Total Quadratic Birth Cohort on Intercept
   lincom c.aget084sq*1 + c.tvage84#c.aget084*1 + c.tvage84sq*1
// Total Linear Birth Cohort on Linear Slope
   lincom c.tvage84#c.aget084*1 + c.tvage84sq*2
estimates store FitRICohAge     // Save for LRT
lrtest FitRICohAge FitRIAge     // LRT for birth cohort contextual fixed slopes
predict PredCohAge, xb          // Save fixed-pred outcomes for total-R2
corr recall PredCohAge          // Get total r to make R2
     display r(rho)^2           // Print total R2 relative to empty model
```

Model 2a R:

```
print("Model 2a Age: Fixed Quadratic, Random Intercept Model")
print("Controlling for Birth Cohort as Contextual Effects")
RICohAge = lmer(data=Example2, REML=FALSE, formula=recall~1+tvage84+tvage84sq
                +ageT084+ageT084sq+tvage84:ageT084+(1|PersonID))
llikAIC(RICohAge); summary(RICohAge)

$AICtab
       AIC         BIC      logLik   deviance    df.resid
 2851.3846   2885.9651 -1417.6923   2835.3846   549.0000

Random effects:
 Groups    Name        Variance Std.Dev.
 PersonID (Intercept) 10.2325   3.1988 → reduced by another 2.37%
 Residual               4.9275   2.2198 → reduced by another 1.12%

Fixed effects:
                  Estimate Std. Error         df t value  Pr(>|t|)
(Intercept)        9.388079   0.341412 263.136000 27.4978 < 2.2e-16  gamma00
tvage84            0.287618   0.119256 380.694774  2.4118 0.0163477  gamma10
tvage84sq         -0.043547   0.015815 366.603154 -2.7535 0.0061904  gamma20
ageT084           -0.575545   0.157213 556.003996 -3.6609 0.0002753  gamma01
ageT084sq         -0.075633   0.028853 549.018161 -2.6213 0.0090012  gamma02
tvage84:ageT084    0.126026   0.034676 377.633390  3.6344 0.0003172  gamma11
```

Interpret these fixed effects:

Intercept γ_{00} =

Slope for age_{ti} γ_{10} =

Slope for age_{ti}^2 γ_{20} =

Slope for $ageT0_i$ γ_{01} =

Slope for $ageT0_i^2$ γ_{02} =

Slope for $age_{ti} * ageT0_i^2$ γ_{11} =

Above, the slopes of baseline age represent **contextual** birth cohort effects (or age non-convergence). The **total** birth cohort effects of baseline age are then linear combinations as shown below (see chapter 10 for the math).

```
print("Total Linear    Birth Cohort on Intercept"); contest1D(RICohAge, L=c(0,1,0,1,0,0))
lincom c.aget084*1 + c.tvage84*1  // STATA Code
     Estimate Std. Error        df     t value     Pr(>|t|)
  -0.28792735 0.10003879 250.76393 -2.8781571 0.0043449637 == gamma10 + gamma01

print("Total Quadratic Birth Cohort on Intercept"); contest1D(RICohAge, L=c(0,0,1,0,1,1))
lincom c.aget084sq*1 + c.tvage84#c.aget084*1 + c.tvage84sq*1  // STATA Code
      Estimate  Std. Error        df     t value    Pr(>|t|)
  0.0068458763 0.018496716 235.73005 0.37011307 0.71163056 == gamma20 + gamma02 + gamma11
```

```
print("Total Linear Birth Cohort on Linear Slope"); contest1D(RICohAge, L=c(0,0,2,0,0,1))
lincom c.tvage84#c.aget084*1 + c.tvage84sq*2  // STATA Code
     Estimate  Std. Error        df    t value    Pr(>|t|)
  0.038932654 0.017779819 400.38829 2.1897104 0.029121678 == 2*gamma20 + gamma11

print("LRT for birth cohort contextual fixed slopes"); anova(RICohAge,RIAge)
         npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
RIAge       5 2860.63 2882.24 -1425.31  2850.63
RICohAge    8 2851.39 2885.97 -1417.69  2835.39 15.2418  3  0.0016212

print("Total R2 for all fixed slopes and change in total R2 for birth cohort effects")
Example2$PredRICohAge = predict(RICohAge, re.form=NA)
rRICohAge = cor.test(Example2$PredRICohAge, Example2$recall, method="pearson")

rRICohAge$estimate^2
        cor
0.025145884
rRICohAge$estimate^2-rRIAge$estimate^2
        cor
0.020465582
```

Model 2b. Syntax and Partial Output for Fixed Quadratic Time, Random Intercept for Recall, adding Age at Baseline to Introduce Total Cross-Sectional Birth Cohort Effects:

Level-1 Time: $recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - ageT0_i) + \beta_{2i}(age_{ti} - ageT0_i)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i),\ \ \beta_{2i} = \gamma_{20}$

Where $age_{ti} - ageT0_i = time_{ti}$ (as years-in-study rather than years-since-birth)

Although it may not appear so, this is an equivalent model to the previous Model 2a using age as the level-2 predictor (centered at age 84) instead…. Even though the difference is in the level-1 predictor, it's going to be the level-2 slopes and cross-level interactions slopes for baseline age that change their values and interpretation!

Model 2b STATA:

```
display "Model 2b Time: Fixed Quadratic, Random Intercept Model"
display "Controlling for Birth Cohort as Total Effects"
mixed recall time c.timesq c.aget084 c.aget084sq c.time#c.aget084, ///
              || personid: , mle nolog
display "-2LL = " e(ll)*-2  // Print -2LL for model
estat ic, n(207)            // Information criteria at N=# persons
// Contextual Linear Birth Cohort on Intercept
   lincom c.aget084*1 + time*-1
// Contextual Quadratic Birth Cohort on Intercept
   lincom c.aget084sq*1 + c.time#c.aget084*-1 + c.timesq*1
// Contextual Linear Birth Cohort on Linear Slope
   lincom c.time#c.aget084*1 + c.timesq*-2
estimates store FitRICohTim     // Save for LRT
lrtest FitRICohTim FitRITim     // LRT for birth cohort total fixed slopes
predict PredCohTim, xb          // Save fixed-pred outcomes for total R2
corr recall PredCohTim          // Get total r to make R2
     display r(rho)^2           // Print total R2 relative to empty model
```

Model 2b R:

```
print("Model 2b Time: Fixed Quadratic, Random Intercept Model")
print("Controlling for Birth Cohort as Total Effects")
RICohTim = lmer(data=Example2, REML=FALSE, formula=recall~1+time+timesq
                +ageT084+ageT084sq+time:ageT084+(1|PersonID))
llikAIC(RICohTim); summary(RICohTim);
```

```
$AICtab
      AIC        BIC     logLik   deviance   df.resid
 2851.3846  2885.9651 -1417.6923  2835.3846   549.0000

Random effects:
 Groups   Name        Variance Std.Dev.
 PersonID (Intercept) 10.2325  3.1988   → reduced by another 3.66%
 Residual              4.9275  2.2198   → reduced by another 1.12%

Fixed effects:
                Estimate  Std. Error          df t value  Pr(>|t|)
(Intercept)    9.3880788   0.3414122 263.1359999 27.4978 < 2.2e-16  gamma00
time           0.2876175   0.1192557 380.6947769  2.4118  0.016348  gamma10
timesq        -0.0435466   0.0158152 366.6031559 -2.7535  0.006190  gamma20
ageT084       -0.2879274   0.1000388 250.7639296 -2.8782  0.004345  gamma01
ageT084sq      0.0068459   0.0184967 235.7300477  0.3701  0.711631  gamma02
time:ageT084   0.0389327   0.0177798 400.3882883  2.1897  0.029122  gamma11
```

Interpret these fixed effects:

Intercept γ_{00} =

Slope for $time_{ti}$ γ_{10} =

Slope for $time_{ti}^2$ γ_{20} =

Slope for $ageT0_i$ γ_{01} =

Slope for $ageT0_i^2$ γ_{02} =

Slope for $time_{ti} * ageT0_i^2$ γ_{11} =

Above, the slopes of baseline age represent **total** birth cohort age effects. The **contextual** birth cohort effects (that test age convergence) are then linear combinations as shown below (see chapter 10 for the math).

```
print("Contextual Linear    Birth Cohort on Intercept"); contest1D(RICohTim, L=c(0,-1, 0,1,0, 0))
lincom c.aget084*1 + time*-1  // STATA Code
     Estimate Std. Error      df     t value       Pr(>|t|)
  -0.57554488 0.15721289 556.004 -3.6609268 0.00027528003  == gamma01 - gamma10

print("Contextual Quadratic Birth Cohort on Intercept"); contest1D(RICohTim, L=c(0, 0, 1,0,1,-1))
lincom c.aget084sq*1 + c.time#c.aget084*-1 + c.timesq*1  // STATA Code
      Estimate   Std. Error         df     t value     Pr(>|t|)
  -0.075633385 0.028852904 549.01816 -2.6213439 0.0090012011  == gamma20 + gamma02 - gamma11

print("Contextual Linear Birth Cohort on Linear Slope"); contest1D(RICohTim, L=c(0, 0,-2,0,0, 1))
lincom c.time#c.aget084*1 + c.timesq*-2  // STATA Code
    Estimate  Std. Error         df    t value      Pr(>|t|)
  0.12602587 0.034675651 377.63339 3.6344197 0.00031718011  == gamma11 - 2*gamma20

print("LRT for birth cohort total fixed slopes"); anova(RICohTim,RITim)
         npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
RITim       5 2856.01 2877.62 -1423.00  2846.01
RICohTim    8 2851.39 2885.97 -1417.69  2835.39 10.6242  3   0.013942

print("Total R2 for all fixed slopes and change in total R2 for age effects")
Example2$PredRICohTim = predict(RICohTim, re.form=NA)
rRICohTim = cor.test(Example2$PredRICohTim, Example2$recall, method="pearson")
rRICohTim$estimate^2
        cor
0.025145884
rRICohTim$estimate^2-rRITim$estimate^2
        cor
0.022416251
```

Comparing the two solutions directly—btw, −2LL and variance components are all the same:

```
Fixed effects:
                  Estimate  Std. Error         df t value  Pr(>|t|)
(Intercept)       9.388079    0.341412 263.136000 27.4978 < 2.2e-16  gamma00
tvage84           0.287618    0.119256 380.694774  2.4118 0.0163477  gamma10
tvage84sq        -0.043547    0.015815 366.603154 -2.7535 0.0061904  gamma20
ageT084          -0.575545    0.157213 556.003996 -3.6609 0.0002753  gamma01
ageT084sq        -0.075633    0.028853 549.018161 -2.6213 0.0090012  gamma02
tvage84:ageT084   0.126026    0.034676 377.633390  3.6344 0.0003172  gamma11

Fixed effects:
                  Estimate  Std. Error          df t value  Pr(>|t|)
(Intercept)      9.3880788   0.3414122 263.1359999 27.4978 < 2.2e-16  gamma00
time             0.2876175   0.1192557 380.6947769  2.4118  0.016348  gamma10
timesq          -0.0435466   0.0158152 366.6031559 -2.7535  0.006190  gamma20
ageT084         -0.2879274   0.1000388 250.7639296 -2.8782  0.004345  gamma01
ageT084sq        0.0068459   0.0184967 235.7300477  0.3701  0.711631  gamma02
time:ageT084     0.0389327   0.0177798 400.3882883  2.1897  0.029122  gamma11
```

Model 3a. Syntax and Partial Output to add Random Linear Age to Model 2a:

Level-1 Age: $Recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - 84) + \beta_{2i}(age_{ti} - 84)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i) + U_{1i},\ \ \beta_{2i} = \gamma_{20}$

Model 3a STATA:

```
display "Model 3a Age: Add Random Linear TVage to Model 2a"
mixed recall c.tvage84 c.tvage84sq c.aget084 c.aget084sq c.tvage84#c.aget084, ///
           || personid: tvage84, mle nolog covariance(unstructured)
estat recovariance, relevel(personid) correlation  // GCORR matrix
display "-2LL = " e(ll)*-2       // Print -2LL for model
estat ic, n(207)                 // Information criteria at N=# persons
estimates store FitRLCohAge      // Save for LRT
lrtest FitRLCohAge FitRICohAge   // LRT for random linear TVage slope
```

Model 3a R:

```
print("Model 3a Age: Add Random Linear TVage to Model 2a")
RLCohAge = lmer(data=Example2, REML=FALSE, formula=recall~1+tvage84+tvage84sq
                +ageT084+ageT084sq+tvage84:ageT084+(1+tvage84|PersonID))
llikAIC(RLCohAge); summary(RLCohAge)

$AICtab
       AIC        BIC     logLik   deviance   df.resid
 2843.8455  2887.0711 -1411.9227  2823.8455   547.0000

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 PersonID (Intercept) 11.154725 3.33987
          tvage84      0.090726 0.30121  -0.340  → new random age slope variance
 Residual              4.110764 2.02750

Fixed effects:
                  Estimate Std. Error         df t value  Pr(>|t|)
(Intercept)       9.414130   0.350985 256.483500 26.8220 < 2.2e-16
tvage84           0.295904   0.113458 353.980589  2.6081 0.0094915
tvage84sq        -0.045388   0.015083 333.443196 -3.0092 0.0028184
ageT084          -0.579338   0.154367 522.386360 -3.7530 0.0001943
ageT084sq        -0.077490   0.030650 162.364800 -2.5282 0.0124203
tvage84:ageT084   0.125601   0.034497 356.744786  3.6409 0.0003119

print("LRT for random linear TVage slope"); anova(RLCohAge,RICohAge)
         npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
RICohAge    8 2851.39 2885.97 -1417.69  2835.39
RLCohAge   10 2843.84 2887.07 -1411.92  2823.84 11.5391  2  0.0031211
```

Model 3b. Syntax and Partial Output to add Random Linear Time to Model 2b:

Level-1 Time: $recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - ageT0_i) + \beta_{2i}(age_{ti} - ageT0_i)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i) + U_{1i},\ \beta_{2i} = \gamma_{20}$

Model 3b STATA:

```
display "Model 3b Time: Add Random Linear Time to Model 2b"
mixed recall c.time c.timesq c.aget084 c.aget084sq c.tvage84#c.aget084, ///
            || personid: time, mle nolog covariance(unstructured)
estat recovariance, relevel(personid) correlation  // GCORR matrix
display "-2LL = " e(ll)*-2       // Print -2LL for model
estat ic, n(207)                 // Information criteria at N=# persons
estimates store FitRLCohTim      // Save for LRT
lrtest FitRLCohTim FitRICohTim   // LRT for random linear time slope
```

Model 3b R:

```
print("Model 3b Time: Add Random Linear Time to Model 2b")
RLCohTim = lmer(data=Example2, REML=FALSE, formula=recall~1+time+timesq
                +ageT084+ageT084sq+time:ageT084+(1+time|PersonID))
llikAIC(RLCohTim); summary(RLCohTim)

$AICtab
       AIC        BIC     logLik   deviance   df.resid
 2838.5453  2881.7709 -1409.2726  2818.5453   547.0000
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 PersonID (Intercept) 12.4835  3.53320
          time         0.1272  0.35665  -0.473  → new random time slope variance
 Residual              3.9405  1.98508

Fixed effects:
               Estimate  Std. Error        df t value  Pr(>|t|)
(Intercept)    9.3402104  0.3515637 230.2831758 26.5676 < 2.2e-16
time           0.3132277  0.1123665 366.1427566  2.7876  0.005588
timesq        -0.0455538  0.0149692 340.1904601 -3.0432  0.002523
ageT084       -0.2972341  0.1050941 205.1743603 -2.8283  0.005144
ageT084sq      0.0091296  0.0183207 229.8606122  0.4983  0.618735
time:ageT084   0.0442743  0.0207981 126.2247447  2.1288  0.035217

print("LRT for random linear TVage slope"); anova(RLCohTim,RICohTim)

         npar     AIC     BIC   logLik deviance   Chisq Df Pr(>Chisq)
RICohTim    8 2851.39 2885.97 -1417.69  2835.39
RLCohTim   10 2838.55 2881.77 -1409.27  2818.55 16.8393  2 0.00022049
```

Comparing variance components across random level-1 slope models:

Model	Random Intercept Variance	Random L1 Slope Variance	Residual Variance		Ratio of Intercept Variance	Ratio L1 Slope Variance	Ratio of Residual Variance
Model 3a Age: Add Random Linear TVage to Model 2a	11.1546	0.0907	4.1108				
Model 3b Time: Add Random Linear Time to Model 2b	12.4837	0.1272	3.9406				
Ratio of Variance Components					0.89	0.71	1.04

The models are no longer equivalent because the age-as-time model assumes the same pattern of variance heterogeneity occurs across longitudinal age (as time) and cross-sectional age (as baseline age). This is a testable assumption in theory, but no software I tried (SAS, STATA, or R) would cleanly estimate the model needed to do so below! That's because it requires a random level-2 "slope" of baseline age (a level-2 predictor)!

Model 4a. Syntax and Partial Output to add Random Linear Baseline Age to Model 3a:

Level-1 Age: $Recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - 84) + \beta_{2i}(age_{ti} - 84)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i} + ‽(ageT0_i)$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i) + U_{1i}, \ \beta_{2i} = \gamma_{20}$

Btw, this punctuation mark ‽ is called an interrobang

Model 4a STATA (showed convergence problems):

```
display "Model 4a Age: Add Random Linear AgeCoh to Model 3a -- extra iterations"
mixed recall c.tvage84 c.tvage84sq c.aget084 c.aget084sq c.tvage84#c.aget084, ///
             || personid: tvage84 aget084, mle nolog emiterate(100) ///
                covariance(unstructured)
//estimates store FitRL2CohAge     // Save for LRT
//lrtest FitRL2CohAge FitRLCohAge  // LRT for random linear baseline age slope

------------------------------------------------------------------------------------
              recall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             tvage84 |   .3135771   .1123666     2.79   0.005     .0933427    .5338115
           tvage84sq |  -.0455586   .0149703    -3.04   0.002    -.0748999   -.0162172
             aget084 |  -.6110121   .1582063    -3.86   0.000    -.9210908   -.3009334
           aget084sq |  -.0807496   .0299605    -2.70   0.007     -.139471   -.0220282
c.tvage84#c.aget084  |   .1358292   .0348349     3.90   0.000     .0675541    .2041042
               _cons |   9.336705   .3540701    26.37   0.000      8.64274    10.03067
------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+-------------------------------------------------
personid: Unstructured       |  all estimates now closer to time level-1 model!
                var(tvage84) |   .1269809          .
                var(aget084) |    .144648          .
                  var(_cons) |   12.53306          .
        cov(tvage84,aget084) |  -.1355204          .
          cov(tvage84,_cons) |  -.6000273          .
          cov(aget084,_cons) |   .6519523          .
-----------------------------+-------------------------------------------------
               var(Residual) |   3.941131          .
-------------------------------------------------------------------------------
LR test vs. linear model: chi2(6) = 238.31                Prob > chi2 = 0.0000

estat recovariance, relevel(personid) correlation  // GCORR matrix

Random-effects correlation matrix for level personid

             |   tvage84    aget084      _cons
-------------+---------------------------------
     tvage84 |         1
     aget084 | -.9999523          1
       _cons | -.4756344   .4842071          1

display "-2LL = " e(ll)*-2          // Print -2LL for model
-2LL = 2818.506

estat ic, n(207)                    // Information criteria at N=# persons
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        207         .  -1409.253       6    2830.506   2850.502
-----------------------------------------------------------------------------
```

Model 4a R (won't converge):

```
print("Model 4a Age: Add Random Linear AgeCoh to Model 3a -- won't run")
RL2CohAge = lmer(data=Example2, REML=FALSE, formula=recall~1+tvage84+tvage84sq
                 +ageT084+ageT084sq+tvage84:ageT084+(1+tvage84+ageT084|PersonID))
llikAIC(RL2CohAge); summary(RL2CohAge)
```

Model 4b. Syntax and Partial Output to add Random Linear Baseline Age to Model 3b:

Level-1 Time: $recall_{ti} = \beta_{0i} + \beta_{1i}(age_{ti} - ageT0_i) + \beta_{2i}(age_{ti} - ageT0_i)^2 + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(ageT0_i) + \gamma_{01}(ageT0_i)^2 + U_{0i} + \text{‽}(ageT0_i)$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(ageT0_i) + U_{1i},\ \ \beta_{2i} = \gamma_{20}$

Btw, this punctuation mark ‽ is called an interrobang

Model 4b STATA (did converge):

```
display "Model 4b Time: Add Random Linear AgeCoh to Model 3b -- extra iterations"
mixed recall c.time c.timesq c.aget084 c.aget084sq c.tvage84#c.aget084, ///
             || personid: time aget084, mle nolog emiterate(100) covariance(unstructured)

------------------------------------------------------------------------------------
              recall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
                time |   .3135834   .1123693     2.79   0.005     .0933437    .5338231
              timesq |  -.0455585   .0149707    -3.04   0.002    -.0749006   -.0162165
             aget084 |  -.2974395   .1053803    -2.82   0.005     -.503981    -.090898
           aget084sq |  -.0351879   .0268437    -1.31   0.190    -.0878005    .0174248
c.tvage84#c.aget084 |   .0447104    .020784     2.15   0.031     .0039745    .0854463
               _cons |    9.33669   .3540657    26.37   0.000     8.642734    10.03065
------------------------------------------------------------------------------------
-----------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+-----------------------------------------------
personid: Unstructured       |
                   var(time) |   .1269922    .043199      .0651964    .2473607
                var(aget084) |   .0005876   .0075139      7.65e-15    4.51e+07  → new term!
                  var(_cons) |   12.53216   1.637893      9.700139    16.19101
           cov(time,aget084) |  -.0085308   .0550702     -.1164663    .0994048
             cov(time,_cons) |  -.6000931   .2283879     -1.047725   -.1524611
          cov(aget084,_cons) |   .0520171    .290053     -.5164764    .6205106
-----------------------------+-----------------------------------------------
               var(Residual) |   3.941339   .3597755      3.295671    4.713501
-----------------------------------------------------------------------------
LR test vs. linear model: chi2(6) = 238.31                Prob > chi2 = 0.0000

display "-2LL = " e(ll)*-2       // Print -2LL for model
-2LL = 2818.506

estat recovariance, relevel(personid) correlation  // GCORR matrix
Random-effects correlation matrix for level personid
             |      time    aget084      _cons
-------------+---------------------------------
        time |         1
     aget084 | -.9875708          1
       _cons | -.4756824   .6061784          1

estat ic, n(207)                 // Information criteria at N=# persons
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        207         .  -1409.253      13    2844.506   2887.831
-----------------------------------------------------------------------------

estimates store FitRL2CohTim     // Save for LRT
lrtest FitRL2CohTim FitRLCohTim  // LRT for random linear baseline age slope
Likelihood-ratio test                                 LR chi2(3)  =      0.04
(Assumption: FitRLCohTim nested in FitRL2CohTim)      Prob > chi2 =    0.9980
```

Model 4b R (won't converge):

```
print("Model 4b Time: Add Random Linear Time to Model 3b -- won't run")
RL2CohTim = lmer(data=Example2, REML=FALSE, formula=recall~1+time+timesq
                 +ageT084+ageT084sq+time:ageT084+(1+time+ageT084|PersonID))
llikAIC(RL2CohTim); summary(RL2CohTim)
```

So it appears there is basically 0 variance for the level-2 random “slope” for baseline age, which is why the level-2 random slope variance for level-1 age was too small—it was being dragged down (downwardly biased) by the incorrect assumption of equal heterogeneity of variance across level-2 baseline years-since-birth (as cross-sectional age) and level-1 years-in-study (time as longitudinal age). Stay tuned for more info on how and why this happens, but the bottom line is that one needs to remove level-2 cross-sectional variance in time before estimating a random slope in order to avoid likely model mis-specification!

Chapter 10 contains an example results section using these models, as well as predicted trajectories by age and time, as shown below. The vertical lines show where the intercept is for each model; the dotted continuous of the lines convey impossible extrapolations predicted by the model!

Take-home point: Use within-person time as your level-1 predictor instead of time-varying age (or any accelerated time metric that has both cross-sectional and longitudinal variance)!!!!!