Example 1: Time-Invariant Predictors in Polynomial Models for Change over Time
complete data, syntax, and output available electronically for SAS, SPSS, or STATA (1a) and Mplus (1b)

These simulated data are from Hoffman (2015) chapter 7b (continued in chapter 9), and include 200 girls measured approximately annually from ages 12–18 (time 0 = age 18) on their risky behavior (the outcome, a sum ranging from 10 to 50) and the extent to which their mothers monitored their activities (the time-varying predictor, a mean ranging from 1 to 5, centered at 3). A time-invariant predictor of the conservativeness of mothers' attitudes about the smoking and drinking (a mean ranging from 1 to 5, centered at 4) was also collected at the age 12 occasion. In this example we predict risky behavior from age and mothers' attitudes.

SAS Code for Data Manipulation (used in Example 1a):

```
* Define global variable for file location -- CHANGE THIS TO YOUR DIRECTORY;
%LET example=C:\Dropbox\Workshop_OSU_2018\Download\SAS;
LIBNAME example "&example.";

* Import chapter 7b multivariate data in work library;
DATA work.Example1multiv; SET example.Example1; RUN;

* Stack multivariate data;
DATA work.Example1; SET work.Example1multiv;
occasion=12; age=age12; risky=risky12; monitor=monitor12; OUTPUT;
occasion=13; age=age13; risky=risky13; monitor=monitor13; OUTPUT;
occasion=14; age=age14; risky=risky14; monitor=monitor14; OUTPUT;
occasion=15; age=age15; risky=risky15; monitor=monitor15; OUTPUT;
occasion=16; age=age16; risky=risky16; monitor=monitor16; OUTPUT;
occasion=17; age=age17; risky=risky17; monitor=monitor17; OUTPUT;
occasion=18; age=age18; risky=risky18; monitor=monitor18; OUTPUT;
* Drop old unnecessary multivariate variables;
DROP age12-age18 risky12-risky18 monitor12-monitor18;
LABEL
occasion = "occasion: Occasion of Measurement (12-18)"
age = "age: Exact Age at Occasion"
risky = "risky: Risky Behavior at Occasion"
monitor = "monitor: Monitoring at Occasion";
RUN;

* Center predictors for analysis in stacked data;
DATA work.Example1; SET work.Example1;
agec18 = age - 18;
att4 = attitude12 - 4;
LABEL
agec18 = "agec18: Exact Age (0=18)"
att4 = "att4: Age 12 Attitudes (0=4)";
* Subset sample to complete cases for all eventual predictors in chapter 9;
IF NMISS(agec18, att4, risky, monitor)>0 THEN DELETE;
RUN;
```

SPSS Code for Data Manipulation (used in Example 1a):

```
* Define location of files used in code below.
FILE HANDLE example /NAME = "C:\Dropbox\Workshop_OSU_2018\Download\SPSS".

* Import chapter 7b multivariate data.
GET FILE = " example/Example1.sav".
DATASET NAME Example1multiv WINDOW=FRONT.
EXECUTE.

* Stack multivariate data.
VARSTOCASES
   /MAKE age FROM age12 TO age18
   /MAKE risky FROM risky12 TO risky18
   /MAKE monitor FROM monitor12 TO monitor18
   /INDEX = wave (7)
   /KEEP = ALL.
DATASET NAME Example1 WINDOW=FRONT.
COMPUTE occasion = wave + 11.
VARIABLE LABELS
wave "wave: Wave of Study (1-7)"
occasion "occasion: Occasion of Measurement (12-18)"
```

```
age "age: Exact Age at Occasion"
risky "risky: Risky Behavior at Occasion"
monitor "monitor: Monitoring at Occasion".

* Center predictors for analysis in stacked data.
COMPUTE agec18 = age - 18.
COMPUTE att4 = attitude12 - 4.
VARIABLE LABELS
agec18 "agec18: Exact Age (0=18)"
att4 "att4: Age 12 Attitudes (0=4)".
* Subset sample to complete cases for all eventual predictors in chapter 9.
SELECT IF NVALID(agec18, att4, risky, monitor)=4.
EXECUTE.
```

STATA Code for Data Manipulation

```
* Defining global variable for file location to be replaced in code below
global example "C:\Dropbox\Workshop_OSU_2018\Download\STATA"

* Import and stack chapter7b multivariate data
* List time-varying variables first, i(level2ID) j(newtimeID)
use "$example\Example1.dta", clear
reshape long age risky monitor, i(personid) j(occasion)
label variable occasion "occasion: Occasion of Measurement (12-18)"
label variable age "age: Exact Age at Occasion"
label variable risky "risky: Risky Behavior at Occasion"
label variable monitor "monitor: Monitoring at Occasion"

* Center predictors for analysis in stacked data
gen agec18 = age - 18
gen agec18sq = agec18 * agec18
gen att4 = attitude12 - 4
label variable agec18 "agec18: Exact Age (0=18)"
label variable agec18sq "agec18sq: Exact Quadratic Age (0=18)"
label variable att4 "att4: Age 12 Attitudes (0=4)"

* Subset sample to complete cases for all eventual predictors in chapter 9
egen nummiss = rowmiss(agec18 att4 risky monitor)
drop if nummiss>0
```

Note: I am not using the new "small" options available in STATA v 14+ that allow denominator DF options because Satterthwaite and KR are only available when using REML. (The PilesOfVariance.com website for example syntax on how to use these options.) As a result, STATA's output should exactly match that of Mplus (with no denominator DF).

Data Description via a Model with Saturated Means *by Rounded Occasion*, Unstructured Variances

```
TITLE1 'SAS Model 0: Saturated Means by Rounded Occasion, Unstructured Variance Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID occasion;
     MODEL risky = occasion / SOLUTION DDFM=Satterthwaite;
     REPEATED occasion / R RCORR TYPE=UN SUBJECT=PersonID;
     LSMEANS occasion / DIFF=ALL;
RUN; TITLE1;


ECHO 'SPSS Model 0: Saturated Means by Rounded Occasion,
      Unstructured Variance Model'.
MIXED risky BY PersonID occasion
     /METHOD   = ML
     /PRINT    = SOLUTION TESTCOV R
     /FIXED    = occasion
     /REPEATED = occasion | COVTYPE(UN) SUBJECT(PersonID)
     /EMMEANS  = TABLES(occasion) COMPARE(occasion).
```

When included as a model predictor, placing *occasion* on the CLASS/BY/i. statements treats it as a categorical predictor. So this is an ANOVA saturated means model using rounded age for descriptive purposes. No RANDOM statements mean no random effects are included yet (we have unstructured variances and covariances treating the occasions as multivariate at one level instead).

```
display as result "STATA Model 0: Saturated Means by Rounded Occasion, Unstructured Variance Model"
mixed risky i.occasion, ///
              || personid: , noconstant variance mle covariance(unstructured) ///
              residuals(unstructured,t(occasion)),
      estat ic, n(200),
      estat wcorrelation, covariance,
      estat wcorrelation,
      contrast i.occasion,
      margins  i.occasion,
      margins  i.occasion, pwcompare(pveffects)
```

i. indicates categorical predictor of *occasion* (ref=last to match others)
noconstant = no random intercept (just **R** matrix)

SAS output:

This Unstructured **R matrix** estimates all variances and covariances separately—it best approximates the patterns we will try to capture with random effects in our model for the variance.

Estimated R Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	**20.8919**	11.6035	10.7002	8.8325	9.1931	6.9827	5.8521
2	11.6035	**19.5582**	11.1286	9.3914	11.5883	10.5351	9.9714
3	10.7002	11.1286	**20.2464**	12.5830	12.9565	11.7780	9.2230
4	8.8325	9.3914	12.5830	**20.8101**	14.0117	15.0484	14.3849
5	9.1931	11.5883	12.9565	14.0117	**21.9129**	16.0068	15.2538
6	6.9827	10.5351	11.7780	15.0484	16.0068	**27.1119**	19.0865
7	5.8521	9.9714	9.2230	14.3849	15.2538	19.0865	**29.2478**

Estimated R Correlation Matrix for PersonID 1 (R converted to correlation matrix)

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.5740	0.5203	0.4236	0.4297	0.2934	0.2367
2	0.5740	1.0000	0.5592	0.4655	0.5598	0.4575	0.4169
3	0.5203	0.5592	1.0000	0.6130	0.6151	0.5027	0.3790
4	0.4236	0.4655	0.6130	1.0000	0.6562	0.6335	0.5831
5	0.4297	0.5598	0.6151	0.6562	1.0000	0.6567	0.6025
6	0.2934	0.4575	0.5027	0.6335	0.6567	1.0000	0.6778
7	0.2367	0.4169	0.3790	0.5831	0.6025	0.6778	1.0000

Solution for Fixed Effects

Effect	occasion	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		23.5211	0.3824	200	61.51	<.0001
occasion	12	-6.7988	0.4384	200	-15.51	<.0001
occasion	13	-6.3383	0.3799	200	-16.68	<.0001
occasion	14	-5.6585	0.3940	200	-14.36	<.0001
occasion	15	-4.5393	0.3263	200	-13.91	<.0001
occasion	16	-3.7482	0.3213	200	-11.66	<.0001
occasion	17	-1.8704	0.3016	200	-6.20	<.0001
occasion	18	0	.	.	.	.

Mean diffs relative to occasion 18 (which is the intercept given that it is the highest value)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
occasion	6	200	55.47	<.0001

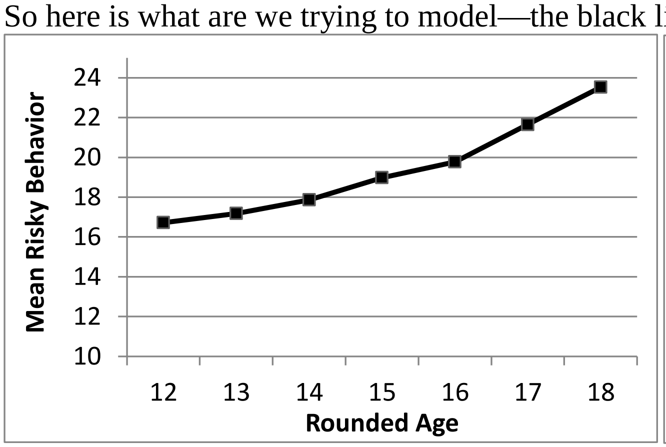
This is the omnibus test of mean differences across the 7 occasions.

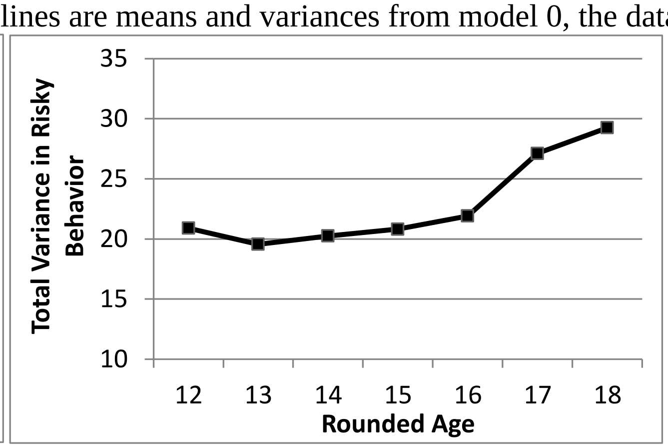
Least Squares Means

Effect	occasion	Estimate	Standard Error	DF	t Value	Pr > |t|
occasion	12	16.7223	0.3232	200	51.74	<.0001
occasion	13	17.1828	0.3127	200	54.95	<.0001
occasion	14	17.8626	0.3182	200	56.14	<.0001
occasion	15	18.9818	0.3226	200	58.85	<.0001
occasion	16	19.7729	0.3310	200	59.74	<.0001
occasion	17	21.6507	0.3682	200	58.80	<.0001
occasion	18	23.5211	0.3824	200	61.51	<.0001

These are the means per (rounded) occasion that the fixed effects of time will try to reproduce.

So here is what are we trying to model—the black lines are means and variances from model 0, the data:





Model 1. Most Conservative Baseline—Empty Means, Random Intercept

Level 1: $y_{ti} = \beta_{0i} + e_{ti}$

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

SAS and SPSS: METHOD = ML or REML (default)
CLASS/BY = categorical predictors, nesting
MODEL dv = fixed effects / print solution
RANDOM = person variances in **G**

```
TITLE1 'SAS Model 1: Empty Means, Random Intercept Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky =  / SOLUTION DDFM=Satterthwaite OUTPM=PredEmpty;
     RANDOM INTERCEPT / V VCORR TYPE=UN SUBJECT=PersonID;
     ODS OUTPUT CovParms=CovEmpty; RUN; TITLE1;

ECHO 'SPSS Model 1: Empty Means, Random Intercept Model'.
MIXED risky BY PersonID
     /METHOD    = ML
     /PRINT     = SOLUTION TESTCOV
     /FIXED     =
     /RANDOM    = INTERCEPT | COVTYPE(UN) SUBJECT(PersonID).
```

STATA: DV = nm3rt, random part after ||
Level 2 ID is PersonID, random intercept by default
Print variances instead of SD, use reml
covariance(unstructured) refers to G matrix
residuals(independent) → type of R matrix by session
estat ic → Print IC given N = 101 persons

```
display as result "STATA Model 1: Empty Means, Random Intercept Model"
mixed risky , || personid: , variance mle covariance(unstructured),
      estat ic, n(200),
      estat icc,
      estat wcorrelation, covariance,
      estat wcorrelation,
```

The **V** matrix is the total variance-covariance matrix after combining the level-2 **G** and level-1 **R** matrices. Right now, everyone has the same predicted **V** matrix (although it prints it for the 1st person by default).

SAS output:

Estimated V Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	**28.0843**	10.8431	10.8431	10.8431	10.8431	10.8431	10.8431
2	10.8431	**28.0843**	10.8431	10.8431	10.8431	10.8431	10.8431
3	10.8431	10.8431	**28.0843**	10.8431	10.8431	10.8431	10.8431
4	10.8431	10.8431	10.8431	**28.0843**	10.8431	10.8431	10.8431
5	10.8431	10.8431	10.8431	10.8431	**28.0843**	10.8431	10.8431
6	10.8431	10.8431	10.8431	10.8431	10.8431	**28.0843**	10.8431
7	10.8431	10.8431	10.8431	10.8431	10.8431	10.8431	**28.0843**

Estimated V Correlation Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.3861	0.3861	0.3861	0.3861	0.3861	0.3861
2	0.3861	1.0000	0.3861	0.3861	0.3861	0.3861	0.3861
3	0.3861	0.3861	1.0000	0.3861	0.3861	0.3861	0.3861
4	0.3861	0.3861	0.3861	1.0000	0.3861	0.3861	0.3861
5	0.3861	0.3861	0.3861	0.3861	1.0000	0.3861	0.3861
6	0.3861	0.3861	0.3861	0.3861	0.3861	1.0000	0.3861
7	0.3861	0.3861	0.3861	0.3861	0.3861	0.3861	1.0000

VCORR ICC = $\frac{10.8431}{10.8431 + 17.2412} = .3861$

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	PersonID	10.8431	1.3344	8.13	<.0001	L2 BP random intercept variance
Residual		17.2412	0.7039	24.49	<.0001	L1 WP residual variance

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	345.71	<.0001

This null model LRT tells us that the random intercept variance is significantly greater than 0, and thus so is the ICC.

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
8296.6	3	8302.6	8302.6	8306.6	8312.5	8315.5

ML only counts **all** parameters in the model (whereas REML does not include fixed effects).

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	19.3849	0.2579	200	75.15	<.0001

This is the fixed intercept, gamma00 (the grand mean of the person means so far).

Model 2a. Fixed Linear Exact Age, Random Intercept

$$\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i}\left(\text{Age}_{ti} - 18\right) + e_{ti}$$
$$\text{Level 2: Intercept: } \beta_{0i} = \gamma_{00} + U_{0i}$$
$$\text{Linear Age: } \beta_{1i} = \gamma_{10}$$

The predictor of *agec18* will be treated as continuous given that it is not on the CLASS statement (SAS), it is on WITH (SPSS), and uses c. (STATA).

```
TITLE1 'SAS Model 2a: Fixed Linear Age, Random Intercept Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky = agec18 / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=PersonID;
    * CovParms will be used for pseudo-R2, InfoCrit for LRT against next model;
     ODS OUTPUT CovParms=CovFixLin InfoCrit=FitFixLin;
RUN; TITLE1;

ECHO 'SPSS Model 2a: Fixed Linear Age, Random Intercept Model'.
MIXED risky BY PersonID WITH agec18
     /METHOD    = ML
     /PRINT     = SOLUTION TESTCOV
     /FIXED     = agec18
     /RANDOM    = INTERCEPT | COVTYPE(UN) SUBJECT(PersonID).
```

DV = risky, c. means continuous fixed slope for *agec18*
Level 2 ID is id, random intercept by default
estimates → save results as “FixLin” for next LRT

```
display as result "STATA Model 2a: Fixed Linear Age, Random Intercept Model"
mixed risky c.agec18 || personid: , variance mle covariance(unstructured),
      estat ic, n(200),
      estimates store FitFixLin,
```

SAS output:

The predicted **V** matrix (not shown) still has a compound symmetry pattern because we have not yet added to the model for the variance (still a random intercept variance only in **G**).

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	PersonID	11.6594	1.3302	8.77	<.0001	L2 BP random intercept variance
Residual		11.3819	0.4647	24.49	<.0001	L1 WP residual variance

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
7798.0	4	7806.0	7806.0	7811.3	7819.2	7823.2

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	22.7421	0.2910	321	78.16	<.0001
agec18	1.1194	0.04503	1200	24.86	<.0001

The fixed linear effect of exact age is significant according to the Wald test (*p*-value for fixed effect).

```
* Call SAS macro to calculate pseudo R2;
  %PseudoR2(NCov=2, CovFewer=CovEmpty, CovMore=CovFixLin);
```

PsuedoR2 (% Reduction) for CovEmpty vs. CovFixLin

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2	
CovEmpty	UN(1,1)	PersonID	10.8431	1.3344	8.13	<.0001	.	
CovEmpty	Residual		17.2412	0.7039	24.49	<.0001	.	
CovFixLin	UN(1,1)	PersonID	11.6594	1.3302	8.77	<.0001	**-0.07528**	for L2 BP int var
CovFixLin	Residual		11.3819	0.4647	24.49	<.0001	**0.33984**	for L1 WP res var

Relative to the empty means, random intercept model 1, the fixed linear effect of *agec18* explained 34% of the residual variance (which made the random intercept variance increase by 7.5% due to its smaller correction factor).

Note that, in total, we have explained 34% of the 61% of the variance that was originally within persons over time (from the empty means, random intercept model 1).

Model 2b. Random Linear Exact Age

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}\left(Age_{ti} - 18\right) + e_{ti}$

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

Linear Age: $\beta_{1i} = \gamma_{10} + U_{1i}$

SAS and SPSS: Now there are 2 random effects: intercept and linear age, given by *agec18* on the RANDOM statements.

STATA: DV = risky, c. means continuous fixed slope for agec18 Level 2 ID is personid, random intercept and agec18 now estimates → save results as "RandLin" for LRT

```
TITLE1 'SAS Model 2b: Random Linear Age Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky = agec18 / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT agec18 / G GCORR V VCORR TYPE=UN SUBJECT=PersonID;
     ODS OUTPUT CovParms=CovRandLin InfoCrit=FitRandLin; RUN; TITLE1;

ECHO 'SPSS Model 2b: Random Linear Age Model'.
MIXED risky BY PersonID WITH agec18
     /METHOD    = ML
     /PRINT     = SOLUTION TESTCOV
     /FIXED     = agec18
     /RANDOM    = INTERCEPT agec18 | COVTYPE(UN) SUBJECT(PersonID).

display as result "STATA Model 2b: Random Linear Age Model"
mixed risky c.agec18, || personid: agec18, variance mle covariance(unstructured),
      estat ic, n(200),
      estimates store FitRandLin,
      lrtest FitRandLin FitFixLin,
```

SAS output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	PersonID	21.5137	2.5634	8.39	<.0001	L2 BP random intercept variance
UN(2,1)	PersonID	2.4328	0.4139	5.88	<.0001	L2 BP random int-linear covariance
UN(2,2)	PersonID	0.5699	0.08919	6.39	<.0001	L2 BP random linear age variance
Residual		8.7179	0.3898	22.36	<.0001	L1 WP residual variance

Estimated G Matrix

Row	Effect	PersonID	Col1	Col2
1	Intercept	1	**21.5137**	2.4328
2	agec18	1	2.4328	**0.5699**

G shows the variance–covariance matrix of random effects, and **GCORR** shows their correlation(s).

Estimated G Correlation Matrix

Row	Effect	PersonID	Col1	Col2
1	Intercept	1	1.0000	0.6948
2	agec18	1	**0.6948**	1.0000

Below: the **V** matrix is the total variance-covariance matrix after combining the level-2 **G** and level-1 **R** matrices. Now the variances and covariances (and the correlations in **VCORR**) are predicted to change by age.

Estimated V Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	**22.2046**	11.7266	10.7657	9.3141	8.6628	7.2868	6.2674
2	11.7266	**19.9982**	11.0367	10.6686	10.5035	10.1546	9.8961
3	10.7657	11.0367	**19.9026**	11.4082	11.5084	11.7203	11.8773
4	9.3141	10.6686	11.4082	**21.2432**	13.0265	14.0854	14.8699
5	8.6628	10.5035	11.5084	13.0265	**22.4255**	15.1466	16.2127
6	7.2868	10.1546	11.7203	14.0854	15.1466	**26.1067**	19.0497
7	6.2674	9.8961	11.8773	14.8699	16.2127	19.0497	**29.8692**

Estimated V Correlation Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.5565	0.5121	0.4289	0.3882	0.3026	0.2434
2	0.5565	1.0000	0.5532	0.5176	0.4960	0.4444	0.4049
3	0.5121	0.5532	1.0000	0.5548	0.5447	0.5142	0.4871
4	0.4289	0.5176	0.5548	1.0000	0.5968	0.5981	0.5903
5	0.3882	0.4960	0.5447	0.5968	1.0000	0.6260	0.6264
6	0.3026	0.4444	0.5142	0.5981	0.6260	1.0000	0.6822
7	0.2434	0.4049	0.4871	0.5903	0.6264	0.6822	1.0000

How to calculate the predicted V matrix variances and covariances in a random linear age model:

$$\mathbf{V}_i \text{ matrix: Variance}\left[y_{time}\right] = \tau^2_{U_0} + \left[\left(Age-18\right)^2 \tau^2_{U_1}\right] + \left[2\left(Age-18\right)\tau_{U_{01}}\right] + \sigma^2_e$$

$$\mathbf{V}_i \text{ matrix: Covariance}\left[y_A, y_B\right] = \tau^2_{U_0} + \left[\left(A+B\right)\tau_{U_{01}}\right] + \left[\left(AB\right)\tau^2_{U_1}\right]$$

```
                              Information Criteria
Neg2LogLike     Parms        AIC        AICC        HQIC         BIC        CAIC
     7677.2         6     7689.2      7689.2      7697.2      7709.0      7715.0

                     Solution for Fixed Effects
                           Standard
Effect        Estimate        Error      DF    t Value     Pr > |t|
Intercept      22.7434       0.3575     200      63.62       <.0001
agec18          1.1199      0.06638     200      16.87       <.0001
```

```
* Call SAS macro to calculate LRT for nested models;
%FitTest(FitFewer=FitFixLin, FitMore=FitRandLin);
```

Is the random linear age model (2b) better than the fixed linear age, random intercept model (2a)?
Yep, −2ΔLL= 120, which is bigger than the critical value of 5.99ish on df = ~2ish

```
Likelihood Ratio Test for FitFixLin vs. FitRandLin
              Neg2Log
   Name        Like      Parms        AIC         BIC      DevDiff     DFdiff     Pvalue
FitFixLin     7798.0       4       7806.0      7819.2         .           .          .
FitRandLin    7677.2       6       7689.2      7709.0      120.785        2          0
```

We will not calculate pseudo-R^2 for this random linear age slope model relative to the previous fixed linear age slope, random intercept model because random effects *do not* explain variance—they partition it instead.

Model 2c. Fixed Quadratic, Random Linear Exact Age

$$\begin{aligned}
&\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i}\left(Age_{ti}-18\right) + \beta_{2i}\left(Age_{ti}-18\right)^2 + e_{ti}\\
&\text{Level 2: Intercept: } \beta_{0i} = \gamma_{00} + U_{0i}\\
&\qquad\text{Linear Age: } \beta_{1i} = \gamma_{10} + U_{1i}\\
&\qquad\text{Quadratic Age: } \beta_{2i} = \gamma_{20}
\end{aligned}$$

Interactions can be defined on the fly in SAS and SPSS using *. Interactions can be defined on the fly in STATA using # for fixed effects, but not for random effects.

The OUTPM in SAS, /SAVE in SPSS, and predict in STATA calculate outcomes predicted by the fixed effects. We can then correlate the predicted and actual outcomes to get total R^2 (actual variance explained).

```
TITLE1 'SAS Model 2c: Fixed Quadratic, Random Linear Age Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky = agec18 agec18*agec18 / SOLUTION DDFM=Satterthwaite OUTPM=PredAge;
     RANDOM INTERCEPT agec18 / TYPE=UN SUBJECT=PersonID;
     ODS OUTPUT CovParms=CovFixQuad InfoCrit=FitFixQuad;
     ESTIMATE 'Intercept at Age=12'           intercept 1 agec18 -6 agec18*agec18 36;
     ESTIMATE 'Intercept at Age=13'           intercept 1 agec18 -5 agec18*agec18 25;
     ESTIMATE 'Intercept at Age=14'           intercept 1 agec18 -4 agec18*agec18 16;
     ESTIMATE 'Intercept at Age=15'           intercept 1 agec18 -3 agec18*agec18 9;
     ESTIMATE 'Intercept at Age=16'           intercept 1 agec18 -2 agec18*agec18 4;
     ESTIMATE 'Intercept at Age=17'           intercept 1 agec18 -1 agec18*agec18 1;
     ESTIMATE 'Intercept at Age=18'           intercept 1 agec18 0  agec18*agec18 0;
     ESTIMATE 'Linear Slope at Age=12'        agec18 1 agec18*agec18 -12;
     ESTIMATE 'Linear Slope at Age=13'        agec18 1 agec18*agec18 -10;
     ESTIMATE 'Linear Slope at Age=14'        agec18 1 agec18*agec18 -8;
     ESTIMATE 'Linear Slope at Age=15'        agec18 1 agec18*agec18 -6;
     ESTIMATE 'Linear Slope at Age=16'        agec18 1 agec18*agec18 -4;
     ESTIMATE 'Linear Slope at Age=17'        agec18 1 agec18*agec18 -2;
     ESTIMATE 'Linear Slope at Age=18'        agec18 1 agec18*agec18  0; RUN; TITLE1;
```

```
ECHO 'SPSS Model 2c: Fixed Quadratic, Random Linear Age Model'.
MIXED risky BY PersonID WITH agec18
     /METHOD   = ML
     /PRINT    = SOLUTION TESTCOV
     /FIXED    = agec18 agec18*agec18
     /RANDOM   = INTERCEPT agec18 | COVTYPE(UN) SUBJECT(PersonID)
```

```
      /TEST     = 'Intercept at Age=12'          intercept 1 agec18 -6 agec18*agec18 36
      /TEST     = 'Intercept at Age=13'          intercept 1 agec18 -5 agec18*agec18 25
      /TEST     = 'Intercept at Age=14'          intercept 1 agec18 -4 agec18*agec18 16
      /TEST     = 'Intercept at Age=15'          intercept 1 agec18 -3 agec18*agec18 9
      /TEST     = 'Intercept at Age=16'          intercept 1 agec18 -2 agec18*agec18 4
      /TEST     = 'Intercept at Age=17'          intercept 1 agec18 -1 agec18*agec18 1
      /TEST     = 'Intercept at Age=18'          intercept 1 agec18 0  agec18*agec18 0
      /TEST     = 'Linear Slope at Age=12'       agec18 1 agec18*agec18 -12
      /TEST     = 'Linear Slope at Age=13'       agec18 1 agec18*agec18 -10
      /TEST     = 'Linear Slope at Age=14'       agec18 1 agec18*agec18 -8
      /TEST     = 'Linear Slope at Age=15'       agec18 1 agec18*agec18 -6
      /TEST     = 'Linear Slope at Age=16'       agec18 1 agec18*agec18 -4
      /TEST     = 'Linear Slope at Age=17'       agec18 1 agec18*agec18 -2
      /TEST     = 'Linear Slope at Age=18'       agec18 1 agec18*agec18  0.
      /SAVE     = FIXPRED(PredAge)
CORRELATIONS risky PredAge.

display as result "STATA Model 2c: Fixed Quadratic, Random Linear Age Model"
mixed risky c.agec18 c.agec18#c.agec18, ///
            || personid: agec18, variance mle covariance(unstructured),
      estat ic, n(200),
      lincom _cons*1 + c.agec18*-6 + c.agec18#c.agec18*36   * Intercept at Age=12
      lincom _cons*1 + c.agec18*-5 + c.agec18#c.agec18*25   * Intercept at Age=13
      lincom _cons*1 + c.agec18*-4 + c.agec18#c.agec18*16   * Intercept at Age=14
      lincom _cons*1 + c.agec18*-3 + c.agec18#c.agec18*9    * Intercept at Age=15
      lincom _cons*1 + c.agec18*-2 + c.agec18#c.agec18*4    * Intercept at Age=16
      lincom _cons*1 + c.agec18*-1 + c.agec18#c.agec18*1    * Intercept at Age=17
      lincom _cons*1 + c.agec18*0  + c.agec18#c.agec18*0    * Intercept at Age=18
      lincom c.agec18*1 + c.agec18#c.agec18*-12      * Linear Slope at Age=12
      lincom c.agec18*1 + c.agec18#c.agec18*-10      * Linear Slope at Age=13
      lincom c.agec18*1 + c.agec18#c.agec18*-8       * Linear Slope at Age=14
      lincom c.agec18*1 + c.agec18#c.agec18*-6       * Linear Slope at Age=15
      lincom c.agec18*1 + c.agec18#c.agec18*-4       * Linear Slope at Age=16
      lincom c.agec18*1 + c.agec18#c.agec18*-2       * Linear Slope at Age=17
      lincom c.agec18*1 + c.agec18#c.agec18*0        * Linear Slope at Age=18
      estimates store FitFixQuad,
      predict PredAge, xb,
corr risky PredAge
```

SAS output:

The **V** matrix (not shown) has the same pattern as in the previous model, because we have not altered the model for the variance (just add a fixed quadratic effect of age).

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	PersonID	21.6917	2.5636	8.46	<.0001	L2 BP random intercept variance
UN(2,1)	PersonID	2.4754	0.4142	5.98	<.0001	L2 BP random int-linear covariance
UN(2,2)	PersonID	0.5846	0.08927	6.55	<.0001	L2 BP random linear age variance
Residual		8.3520	0.3735	22.36	<.0001	L1 WP residual variance

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
7634.8	7	7648.8	7648.9	7658.1	7671.9	7678.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	23.4655	0.3740	239	62.74	<.0001
agec18	1.9877	0.1476	1188	13.47	<.0001
agec18*agec18	0.1446	0.02197	1010	6.58	<.0001

The fixed quadratic effect of *agec18* is significant according to the Wald test (*p*-value for fixed effect). The linear slope changes by twice the quadratic coefficient per unit time (year of age).

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept at Age=12	16.7458	0.3108	260	53.88	<.0001
Intercept at Age=13	17.1426	0.2640	200	64.93	<.0001
Intercept at Age=14	17.8287	0.2609	228	68.35	<.0001
Intercept at Age=15	18.8040	0.2724	249	69.02	<.0001
Intercept at Age=16	20.0686	0.2872	223	69.88	<.0001
Intercept at Age=17	21.6224	0.3139	200	68.87	<.0001
Intercept at Age=18	23.4655	0.3740	239	62.74	<.0001

These are the quadratic-model-predicted means (intercepts) per age.

```
Linear Slope at Age=12      0.2522      0.1476     1188       1.71      0.0878
Linear Slope at Age=13      0.5414      0.1102      932       4.92      <.0001
Linear Slope at Age=14      0.8307     0.07965      396      10.43      <.0001
Linear Slope at Age=15      1.1199     0.06645      200      16.86      <.0001
Linear Slope at Age=16      1.4092     0.07966      397      17.69      <.0001
Linear Slope at Age=17      1.6985      0.1102      932      15.42      <.0001
Linear Slope at Age=18      1.9877      0.1476     1188      13.47      <.0001
```

These are the instantaneous linear slopes at each age. Note how the SEs narrow towards the middle ages.

```
* Call macro to calculate pseudo R2;
%PseudoR2(NCov=4, CovFewer=CovRandLin, CovMore=CovFixQuad);
```

Relative to the random linear time model 2b, fixed quadratic age explained another 4% of the residual (which increased the random intercept variance).

PsuedoR2 (% Reduction) for CovRandLin vs. CovFixQuad

Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2	
CovRandLin	UN(1,1)	PersonID	21.5137	2.5634	8.39	<.0001	.	
CovRandLin	UN(2,2)	PersonID	0.5699	0.08919	6.39	<.0001	.	
CovRandLin	Residual		8.7179	0.3898	22.36	<.0001	.	
CovFixQuad	UN(1,1)	PersonID	21.6917	2.5636	8.46	<.0001	-0.008274	for L2 int var
CovFixQuad	UN(2,2)	PersonID	0.5846	0.08927	6.55	<.0001	-0.025771	for L2 lin var
CovFixQuad	Residual		8.3520	0.3735	22.36	<.0001	0.041969	for L1 res var

```
* Call macro to calculate Total R2 for overall model;
%TotalR2(DV=risky, PredFewer=PredEmpty, PredMore=PredAge);
```

Total R2 (% Reduction) for PredEmpty vs. PredAge

Name	Pred Corr	TotalR2	Total R2Diff
PredEmpty	.	.	.
PredAge	0.43427	0.18859	.

r = .43427 → **TOTAL R^2 = .19**
19% of total risky behavior variance is accounted for by linear and quadratic effects of age

---

Model 2d. Random Quadratic Exact Age

$$\text{Level 1: } y_{ti} = \beta_{0i} + \beta_{1i}\left(\text{Age}_{ti} - 18\right) + \beta_{2i}\left(\text{Age}_{ti} - 18\right)^2 + e_{ti}$$
$$\text{Level 2: Intercept: } \beta_{0i} = \gamma_{00} + U_{0i}$$
$$\text{Linear Age: } \beta_{1i} = \gamma_{10} + U_{1i}$$
$$\text{Quadratic Age: } \beta_{2i} = \gamma_{20} + U_{2i}$$

```
TITLE1 'SAS Model 2d: Random Quadratic Age Model';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky = agec18 agec18*agec18 / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT agec18 agec18*agec18 / G GCORR V VCORR TYPE=UN SUBJECT=PersonID;
     ODS OUTPUT InfoCrit=FitRandQuad; RUN; TITLE1;

ECHO 'SPSS Model 2d: Random Quadratic Age Model'.
MIXED risky BY PersonID WITH agec18
     /METHOD   = ML
     /PRINT    = SOLUTION TESTCOV
     /FIXED    = agec18 agec18*agec18
     /RANDOM   = INTERCEPT agec18 agec18*agec18 | COVTYPE(UN) SUBJECT(PersonID).

display as result "STATA Model 2d: Random Quadratic Age Model"
mixed risky c.agec18 c.agec18#c.agec18, ///
              || personid: agec18 agec18sq, variance mle covariance(unstructured),
      estat ic, n(200),
      estimates store FitRandQuad,
      lrtest FitRandQuad FitFixQuad,
```

The STATA random statement will not accept interaction terms, so we are using *agec18sq*.

SAS output:

Estimated G Matrix

Row	Effect	PersonID	Col1	Col2	Col3
1	Intercept	1	21.2827	2.4761	-0.00542
2	agec18	1	2.4761	1.3561	0.1277
3	agec18*agec18	1	-0.00542	0.1277	0.02152

G shows the variance–covariance matrix of random effects, and **GCORR** shows their correlation(s).

Estimated G Correlation Matrix

Row	Effect	PersonID	Col1	Col2	Col3
1	Intercept	1	1.0000	0.4609	-0.00801
2	agec18	1	**0.4609**	1.0000	0.7475
3	agec18*agec18	1	**-0.00801**	**0.7475**	1.0000

Below: the **V** matrix total variances and covariances (and the correlations in **VCORR**) are predicted to change differently by age than before.

Estimated V Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	**21.5024**	11.2582	10.1076	8.4858	7.8036	6.4548	5.5367
2	11.2582	**19.6663**	11.6825	11.3511	11.0763	10.2386	9.3932
3	10.1076	11.6825	**20.1821**	12.5393	12.5125	12.0949	11.4698
4	8.4858	11.3511	12.5393	**21.8080**	14.2123	14.6183	14.5677
5	7.8036	11.0763	12.5125	14.2123	**22.7692**	15.6407	15.9425
6	6.4548	10.2386	12.0949	14.6183	15.6407	**25.5552**	18.8163
7	5.5367	9.3932	11.4698	14.5677	15.9425	18.8163	**28.8959**

Estimated V Correlation Matrix for PersonID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.5475	0.4852	0.3919	0.3527	0.2754	0.2221
2	0.5475	1.0000	0.5864	0.5481	0.5234	0.4567	0.3940
3	0.4852	0.5864	1.0000	0.5977	0.5837	0.5326	0.4750
4	0.3919	0.5481	0.5977	1.0000	0.6378	0.6192	0.5803
5	0.3527	0.5234	0.5837	0.6378	1.0000	0.6484	0.6215
6	0.2754	0.4567	0.5326	0.6192	0.6484	1.0000	0.6924
7	0.2221	0.3940	0.4750	0.5803	0.6215	0.6924	1.0000

How to calculate the predicted V matrix variances and covariances in a random quadratic age model:

Predicted Variance at Age *T*:

$$\mathrm{Var}(y_T) = \sigma_e^2 + \tau_{U_0}^2 + 2*T*\tau_{U_{01}} + T^2*\tau_{U_1}^2 + 2*T^2*\tau_{U_{02}} + 2*T^3*\tau_{U_{12}} + T^4*\tau_{U_2}^2$$

Predicted Covariance between Time *A* and *B*:

$$\mathrm{Cov}(y_A, y_B) = \tau_{U_0}^2 + (A+B)*\tau_{U_{01}} + (AB)*\tau_{U_1}^2 + (A^2+B^2)*\tau_{U_{02}} + (AB^2)+(A^2B)*\tau_{U_{12}} + (A^2B^2)*\tau_{U_2}^2$$

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	PersonID	21.2827	2.7450	7.75	<.0001	L2 BP random intercept variance
UN(2,1)	PersonID	2.4761	0.9474	2.61	0.0090	L2 BP random int-linear covariance
UN(2,2)	PersonID	1.3561	0.5325	2.55	0.0054	L2 BP random linear age variance
UN(3,1)	PersonID	-0.00542	0.1309	-0.04	0.9670	L2 BP random int-quad age covariance
UN(3,2)	PersonID	0.1277	0.07761	1.65	0.0999	L2 BP random lin-quad age covariance
UN(3,3)	PersonID	0.02152	0.01240	1.74	0.0414	L2 BP random quad age variance
Residual		7.9778	0.3996	19.97	<.0001	L2 WP residual variance

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
7629.5	10	7649.5	7649.7	7662.9	7682.5	7692.5

Solution for Fixed Effects

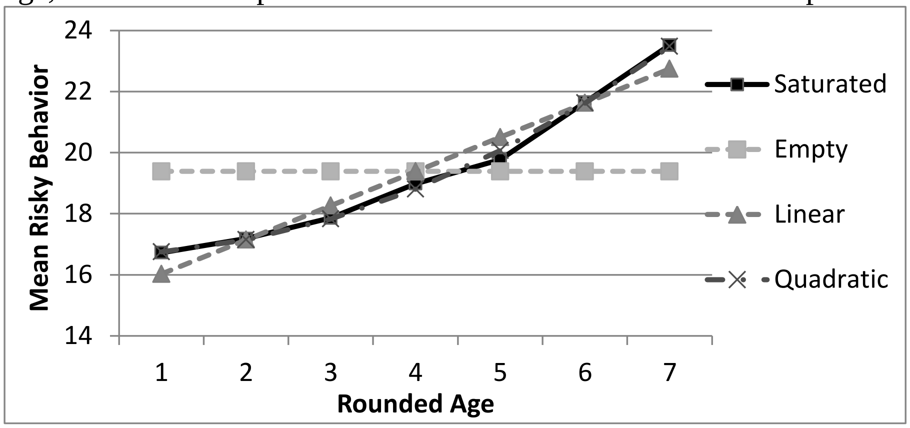
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	23.4688	0.3695	200	63.51	<.0001
agec18	1.9912	0.1577	193	12.62	<.0001
agec18*agec18	0.1451	0.02389	191	6.07	<.0001

```
* Call SAS macro to calculate LRT for nested models;
%FitTest(FitFewer=FitFixQuad, FitMore=FitRandQuad);


Likelihood Ratio Test for FitFixQuad vs. FitRandQuad
                    Neg2Log
   Name              Like        Parms         AIC         BIC      DevDiff      DFdiff       Pvalue
FitFixQuad          7634.8         7        7648.8      7671.9         .            .            .
FitRandQuad         7629.5        10        7649.5      7682.5      5.26439         3         0.15343
```

Is the random quadratic model (2d) better than the fixed quadratic, random linear model (2c)?
Nope, $-2\Delta LL = 5.26$, which is smaller than the critical value of 7.82ish on df= ~3ish

We will not calculate pseudo-R^2 for this random quadratic age model relative to the previous fixed linear age, random intercept model because random effects *do not* explain variable—they partition it instead.



Given how well the quadratic age model appears to fit the rounded means, we can move on by adding a time-invariant predictor of mothers' attitudes (as measured at age 12).

Model 3. Add Attitudes at Age 12 as Predictor of Intercept, Linear, and Quadratic Age Slopes

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}\left(Age_{ti} - 18\right) + \beta_{2i}\left(Age_{ti} - 18\right)^2 + e_{ti}$

Level 2:

Intercept: $\beta_{0i} = \gamma_{00} + \gamma_{01}\left(Attitudes_i - 4\right) + U_{0i}$

Linear: $\beta_{1i} = \gamma_{10} + \gamma_{11}\left(Attitudes_i - 4\right) + U_{1i}$

Quadratic: $\beta_{2i} = \gamma_{20} + \gamma_{21}\left(Attitudes_i - 4\right)$

Note that we have not included the random quadratic age slope variance given the results of model 2d. That means the quadratic effect of age is systematically varying in this model (instead of random, like the intercept and linear age effects).

```
TITLE1 'SAS Model 3: Fixed Quadratic, Random Linear Age Model';
TITLE2 'Attitudes Predicting Quadratic Age Slope';
PROC MIXED DATA=work.Example1 COVTEST NOCLPRINT NAMELEN=100 IC METHOD=ML;
     CLASS PersonID;
     MODEL risky = agec18 agec18*agec18
                   att4 agec18*att4 agec18*agec18*att4
                    / SOLUTION DDFM=Satterthwaite OUTPM=PredAttQuad;
     RANDOM INTERCEPT agec18 / TYPE=UN SUBJECT=PersonID;
     ODS OUTPUT CovParms=CovAttQuad; RUN; TITLE1; TITLE2;

ECHO 'SPSS Model 3: Fixed Quadratic, Random Linear Age Model;'.
ECHO 'Attitudes Predicting Quadratic Age Slope'.
MIXED risky BY PersonID WITH agec18 att4
     /METHOD    = ML
     /PRINT     = SOLUTION TESTCOV
     /FIXED     = agec18 agec18*agec18
                  att4 agec18*att4 agec18*agec18*att4
     /RANDOM    = INTERCEPT agec18 | COVTYPE(UN) SUBJECT(PersonID)
     /SAVE      = FIXPRED(PredAttQuad).
CORRELATIONS risky PredAttQuad.

display as result "STATA Model 3: Fixed Quadratic, Random Linear Age Model"
display as result "Attitudes Predicting Quadratic Age Slope"
mixed risky c.agec18 c.agec18#c.agec18 c.att4 c.agec18#c.att4 c.agec18#c.agec18#c.att4, ///
            || personid: agec18, variance mle covariance(unstructured),
      estat ic, n(200),
      predict PredAttQuad, xb,
corr risky PredAttQuad
```

SAS output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	PersonID	18.0843	2.2033	8.21	<.0001	L2 BP random intercept variance
UN(2,1)	PersonID	1.8873	0.3564	5.30	<.0001	L2 BP random int-linear covariance
UN(2,2)	PersonID	0.4895	0.07983	6.13	<.0001	L2 BP random linear age variance
Residual		8.3260	0.3724	22.36	<.0001	L1 WP residual variance

Information Criteria

Neg2LogLike	Parms	AIC	AICC	HQIC	BIC	CAIC
7599.4	10	7619.4	7619.6	7632.8	7652.4	7662.4

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	23.2993	0.3500	245	66.57	<.0001
agec18	1.9480	0.1461	1197	13.33	<.0001
agec18*agec18	0.1422	0.02199	1011	6.46	<.0001
att4*intercept	-3.4764	0.5805	245	-5.99	<.0001
agec18*att4	-0.9004	0.2421	1196	-3.72	0.0002
agec18*agec18*att4	-0.06409	0.03636	1011	-1.76	0.0783

```
* Call SAS macro to calculate pseudo R2;
  %PseudoR2(NCov=4, CovFewer=CovAttLin, CovMore=CovAttQuad);
```

PsuedoR2 (% Reduction) for CovFixQuad vs. CovAttQuad

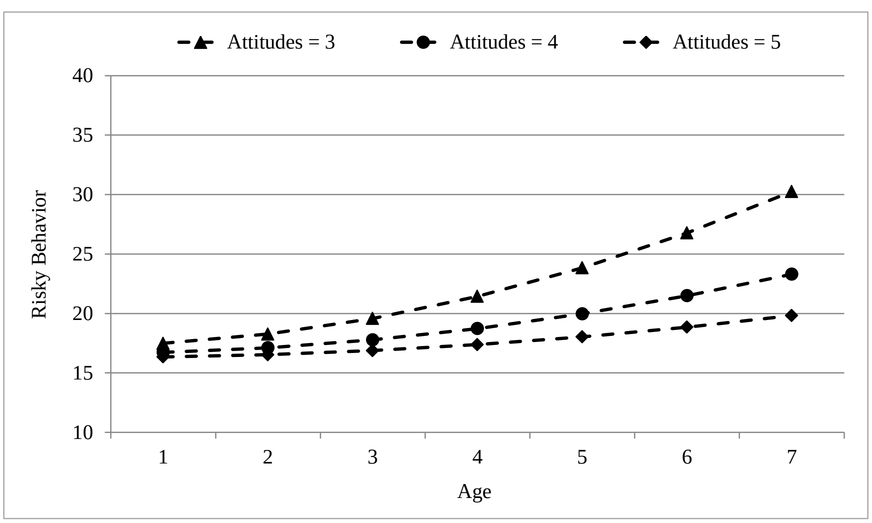
Name	CovParm	Subject	Estimate	StdErr	ZValue	ProbZ	PseudoR2
CovFixQuad	UN(1,1)	PersonID	21.6917	2.5636	8.46	<.0001	.
CovFixQuad	UN(2,2)	PersonID	0.5846	0.08927	6.55	<.0001	.
CovFixQuad	Residual		8.3520	0.3735	22.36	<.0001	.
CovAttQuad	UN(1,1)	PersonID	18.0843	2.2033	8.21	<.0001	0.16631 for L2 rand int
CovAttQuad	UN(2,2)	PersonID	0.4895	0.07983	6.13	<.0001	0.16261 for L2 rand lin
CovAttQuad	Residual		8.3260	0.3724	22.36	<.0001	0.00312 for L1 res

```
* Call SAS macro to calculate Total R2 for overall model;
  %TotalR2(DV=risky, PredFewer=PredAge, PredMore=PredAttQuad);
```

Total R2 (% Reduction) for PredAge vs. PredAttQuad

Name	Pred Corr	TotalR2	Total R2Diff
PredAge	0.43427	0.18859	.
PredAttQuad	0.48626	0.23645	0.047861

The fixed effects of age before accounted for 19% of the variance in risky behavior, so there is a net increase of ~5% due to attitudes.



The more conservative the mothers' attitude (higher values), the less positive (and almost less accelerating) the daughters' trajectory.

See chapter 7 for all the models and a results section.