Practice with Fixed and Random Effects of Time in Modeling Within-Person Change

The models for this example come from Hoffman (in preparation) chapter 5. We will be examining the extent to which change in a test score outcome across four annual occasions can be described with fixed and random linear effects of time (indexed by years in study, in which 0 is baseline) in a sample of 25 persons.

SAS Syntax and Output for Data Manipulation:

```
* Location for files to be saved - CHANGE THIS TO YOUR DIRECTORY;
%LET filesave=F:\Example Data\Chapter 5 Data;
LIBNAME filesave "&example.";

* Import data into work library, center time;
DATA example5; SET filesave.example5;
       time = wave - 1; LABEL time= "time: Time in Study (0=1)";
RUN;
```

The **ANSWER KEY** for both the model for the means (via saturated means) and the model for the variances (via unstructured **R** matrix of all possible variances and covariances) is possible to estimate in balanced data:

```
TITLE1 "Saturated Means, Unstructured Variance Model -- the ANSWER KEY";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT METHOD=REML;
       CLASS IDnum wave;
       MODEL outcome = wave / SOLUTION DDFM=Satterthwaite;
       REPEATED wave / R RCORR TYPE=UN SUBJECT=IDnum;
       LSMEANS wave;
RUN;
```

Dimensions

Covariance Parameters	10
Columns in X	5
Columns in Z	0
Subjects	25
Max Obs Per Subject	4

Estimated R Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	2.3618	2.7867	1.9566	2.4204
2	2.7867	4.8900	4.0440	5.5525
3	1.9566	4.0440	6.2172	7.7994
4	2.4204	5.5525	7.7994	11.7437

Estimated R Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.8200	0.5106	0.4596
2	0.8200	1.0000	0.7334	0.7327
3	0.5106	0.7334	1.0000	0.9128
4	0.4596	0.7327	0.9128	1.0000

Because this model uses REPEATED only (no RANDOM statement), the **R** matrix holds the total variances and covariances over waves directly. Likewise, **RCORR** holds the total correlations over waves directly.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	IDnum	2.3618	0.6818	3.46	0.0003
UN(2,1)	IDnum	2.7867	0.8971	3.11	0.0019
UN(2,2)	IDnum	4.8900	1.4116	3.46	0.0003
UN(3,1)	IDnum	1.9566	0.8783	2.23	0.0259
UN(3,2)	IDnum	4.0440	1.3958	2.90	0.0038
UN(3,3)	IDnum	6.2172	1.7947	3.46	0.0003
UN(4,1)	IDnum	2.4204	1.1831	2.05	0.0408
UN(4,2)	IDnum	5.5525	1.9176	2.90	0.0038
UN(4,3)	IDnum	7.7994	2.3615	3.30	0.0010
UN(4,4)	IDnum	11.7437	3.3901	3.46	0.0003

Fit Statistics	
-2 Res Log Likelihood	353.8
AIC (smaller is better)	373.8
AICC (smaller is better)	376.3
BIC (smaller is better)	385.9

Because we are in REML, only the variance model parameters count towards AIC and BIC. Thus, in REML we **cannot** use −2ΔLL to compare models that differ in their fixed effects (i.e., different models for the means).

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
9	108.30	<.0001

This is the test of whether we need *anything* beyond a constant residual variance σ_e^2 (df=9)… and we do.

Solution for Fixed Effects

Effect	wave: Occasion (1-4)	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		15.5516	0.6854	24	22.69	<.0001
wave	1	-5.1468	0.6088	24	-8.45	<.0001
wave	2	-3.6940	0.4703	24	-7.86	<.0001
wave	3	-1.9672	0.3074	24	-6.40	<.0001
wave	4	0	.	.	.	.

Type 3 Tests of Fixed Effects

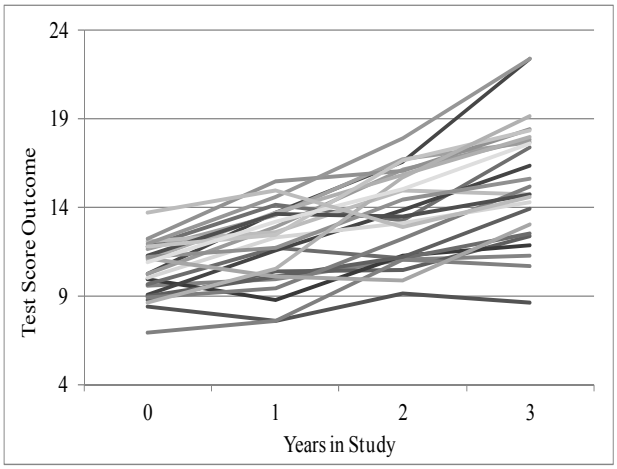
Effect	Num DF	Den DF	F Value	Pr > F
wave	3	24	23.86	<.0001

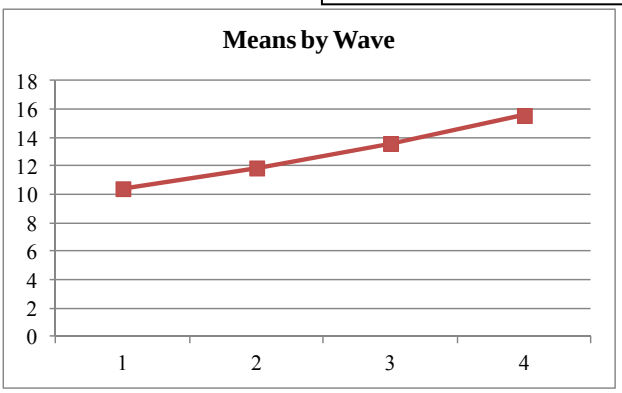
This is the ANOVA test of omnibus mean differences across wave (note df=3 for the 4 means across waves), assuming an unstructured **R** matrix (multivariate ANOVA).

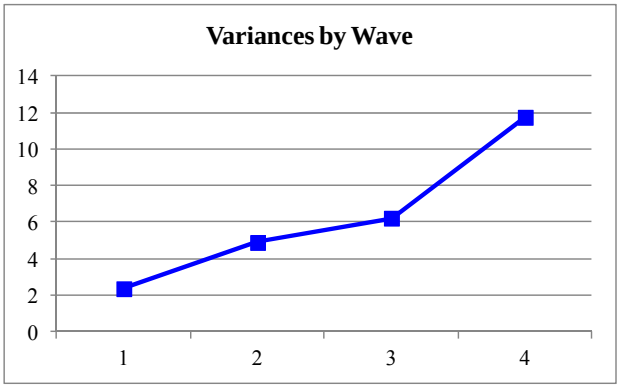
Least Squares Means

Effect	wave: Occasion (1-4)	Estimate	Standard Error	DF	t Value	Pr > |t|
wave	1	10.4048	0.3074	24	33.85	<.0001
wave	2	11.8576	0.4423	24	26.81	<.0001
wave	3	13.5844	0.4987	24	27.24	<.0001
wave	4	15.5516	0.6854	24	22.69	<.0001

Because *wave* is on the CLASS statement, the LSMEANS provides means per wave (as found from fixed intercept + difference for each wave in the solution for fixed effects).







These are the estimates from the Saturated Means, Unstructured Variance model, and here are the individual growth curves that these estimates summarize.

If an unstructured R matrix was not possible to estimate, I'd still examine the answer key for the model for the means (via a saturated means model), but estimate a random intercept only (which should always be possible):

```
TITLE1 "Saturated Means, Random Intercept Variance Model -- MEANS ANSWER KEY";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT METHOD=REML;
       CLASS IDnum wave;
       MODEL outcome = wave / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / G V VCORR TYPE=UN SUBJECT=IDnum;
       REPEATED wave / R TYPE=VC SUBJECT=IDnum;
       LSMEANS wave;
RUN;
```

Estimated V Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	6.3032	4.0933	4.0933	4.0933
2	4.0933	6.3032	4.0933	4.0933
3	4.0933	4.0933	6.3032	4.0933
4	4.0933	4.0933	4.0933	6.3032

Estimated V Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6494	0.6494	0.6494
2	0.6494	1.0000	0.6494	0.6494
3	0.6494	0.6494	1.0000	0.6494
4	0.6494	0.6494	0.6494	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
UN(1,1)	IDnum	4.0933	1.3443	3.04	0.0012
wave	IDnum	2.2099	0.3683	6.00	<.0001

Fit Statistics

-2 Res Log Likelihood	412.5
AIC (smaller is better)	416.5
AICC (smaller is better)	416.7
BIC (smaller is better)	419.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	49.51	<.0001

Solution for Fixed Effects

Effect	wave: Occasion (1-4)	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		15.5516	0.5021	42.4	30.97	<.0001
wave	1	-5.1468	0.4205	72	-12.24	<.0001
wave	2	-3.6940	0.4205	72	-8.79	<.0001
wave	3	-1.9672	0.4205	72	-4.68	<.0001
wave	4	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
wave	3	72	55.82	<.0001

This is the ANOVA test of omnibus mean differences across wave (note df=3 for the 4 means across waves), assuming a random intercept only (CS **V** matrix; univariate ANOVA).

Least Squares Means

Effect	wave: Occasion (1-4)	Estimate	Standard Error	DF	t Value	Pr > |t|
wave	1	10.4048	0.5021	42.4	20.72	<.0001
wave	2	11.8576	0.5021	42.4	23.62	<.0001
wave	3	13.5844	0.5021	42.4	27.05	<.0001
wave	4	15.5516	0.5021	42.4	30.97	<.0001

5.1: Empty Means, Random Intercept Model

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

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

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + e_{ti}$

```
TITLE1 "Eq 5.1: Empty Means, Random Intercept Model";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT
METHOD=REML;
       CLASS IDnum wave;
       MODEL outcome = / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT / G V VCORR TYPE=UN SUBJECT=IDnum;
       REPEATED wave / R TYPE=VC SUBJECT=IDnum;
RUN;
```

```
          Estimated R Matrix for IDnum 1
 Row        Col1        Col2        Col3        Col4
   1      7.0554
   2                  7.0554
   3                              7.0554
   4                                          7.0554
```

```
          Estimated G Matrix
                    IDnum:
                    Person ID
Row    Effect       number        Col1
   1    Intercept     1          2.8819
```

```
          Estimated V Matrix for IDnum 1
 Row        Col1        Col2        Col3        Col4
   1      9.9373      2.8819      2.8819      2.8819
   2      2.8819      9.9373      2.8819      2.8819
   3      2.8819      2.8819      9.9373      2.8819
   4      2.8819      2.8819      2.8819      9.9373
```

Because this model uses the REPEATED and RANDOM statements, the **V** matrix holds the total variances and covariances over waves (from putting **G** and **R** back together through the **Z** matrix). Likewise, **VCORR** holds the total correlations over waves.

```
    Estimated V Correlation Matrix for IDnum 1
 Row        Col1        Col2        Col3        Col4
   1      1.0000      0.2900      0.2900      0.2900
   2      0.2900      1.0000      0.2900      0.2900
   3      0.2900      0.2900      1.0000      0.2900
   4      0.2900      0.2900      0.2900      1.0000
```

VCORR provides the ICC as: IntVar/TotalVar

```
                Covariance Parameter Estimates
                                  Standard         Z
Cov Parm    Subject    Estimate       Error     Value      Pr > Z
UN(1,1)     IDnum        2.8819      1.3717      2.10      0.0178  Random intercept variance in G
wave        IDnum        7.0554      1.1521      6.12      <.0001  Residual variance in R
```

```
          Fit Statistics
-2 Res Log Likelihood           502.2
AIC (smaller is better)         506.2
AICC (smaller is better)        506.3
BIC (smaller is better)         508.7
```

```
  Null Model Likelihood Ratio Test
    DF    Chi-Square      Pr > ChiSq
     1          9.79          0.0018
```

This is the test of whether we need the random intercept variance (so df=1)… and we do.

```
                Solution for Fixed Effects
                      Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept     12.8496      0.4311      24      29.81      <.0001 This is gamma00
```

5.3: Fixed Linear Time, Random Intercept Model

```
TITLE1 "Eq 5.3: Fixed Linear Time, Random Intercept Model";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT
METHOD=REML;
  CLASS IDnum wave;
  MODEL outcome = time / SOLUTION DDFM=Satterthwaite;
  RANDOM INTERCEPT / G V VCORR TYPE=UN SUBJECT=IDnum;
  REPEATED wave / R TYPE=VC SUBJECT=IDnum;
  ESTIMATE "Intercept at Time 0" int 1 time 0;
  ESTIMATE "Intercept at Time 1" int 1 time 1;
  ESTIMATE "Intercept at Time 2" int 1 time 2;
  ESTIMATE "Intercept at Time 3" int 1 time 3;
RUN;
```

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(\text{Time}_{ti}) + e_{ti}$

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

$\beta_{1i} = \gamma_{10}$

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + \gamma_{10}(\text{Time}_{ti}) + e_{ti}$

Note the two different versions of the "time" variable in the syntax. Both are necessary here because they do different things. "**Wave**" is treated as a **categorical** predictor, and its role is to structure the **R** matrix in the event of missing data. Therefore, "wave" goes on the CLASS and REPEATED statements. In contrast, "**time**" is treated as a **continuous** predictor, and its role is to index linear effects of time (and it is centered such that wave 1 = time 0). Accordingly, in the ESTIMATE statements, only one value after "time" is needed.

Estimated R Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	2.1725			
2		2.1725		
3			2.1725	
4				2.1725

After controlling for the fixed linear effect of time, the residual variance was reduced from $\sigma_e^2 = 7.06$ in the empty means, random intercept model to $\sigma_e^2 = 2.17$ in this model. This is a reduction of (7.06 – 2.17) / 7.06 = .69 (or 69% of the residual variance is accounted for by a fixed linear time).

Estimated G Matrix

Row	Effect	IDnum: Person ID number	Col1
1	Intercept	1	4.1026

However, the random intercept variance actually increased from 2.88 to 4.10. This is because of how $\tau_{U_0}^2$ is found:

true $\tau_{U_0}^2$ = observed $\tau_{U_0}^2 - (\sigma_e^2 / n)$

So reducing σ_e^2 will make $\tau_{U_0}^2$ increase.

Estimated V Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	6.2751	4.1026	4.1026	4.1026
2	4.1026	6.2751	4.1026	4.1026
3	4.1026	4.1026	6.2751	4.1026
4	4.1026	4.1026	4.1026	6.2751

Estimated V Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6538	0.6538	0.6538
2	0.6538	1.0000	0.6538	0.6538
3	0.6538	0.6538	1.0000	0.6538
4	0.6538	0.6538	0.6538	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z	
UN(1,1)	IDnum	4.1026	1.3441	3.05	0.0011	Random intercept variance in G
wave	IDnum	2.1725	0.3572	6.08	<.0001	Residual variance in R

Fit Statistics

-2 Res Log Likelihood	415.1
AIC (smaller is better)	419.1
AICC (smaller is better)	419.2
BIC (smaller is better)	421.5

Are we allowed to examine the −2ΔLL to see if adding a fixed linear effect of time improved model fit in REML? If not, what do we do instead?

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	51.12	<.0001

This tests whether we need the random intercept variance (so df=1)… and we (still) do.

```
                  Solution for Fixed Effects
                        Standard
Effect       Estimate       Error      DF    t Value    Pr > |t|
Intercept     10.2745      0.4743    34.7      21.66      <.0001 this is gamma00
time           1.7167      0.1318      74      13.02      <.0001 this is gamma10

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
time            1      74     169.57    <.0001
```

Estimates → **These are the predicted outcome means from a fixed linear model**

```
                                Standard
Label                 Estimate       Error      DF    t Value    Pr > |t|
Intercept at Time 0    10.2745      0.4743    34.7      21.66      <.0001
Intercept at Time 1    11.9912      0.4361    25.1      27.50      <.0001
Intercept at Time 2    13.7080      0.4361    25.1      31.43      <.0001
Intercept at Time 3    15.4247      0.4743    34.7      32.52      <.0001
```

5.5: Random Linear Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(\text{Time}_{ti}) + e_{ti}$

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

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

Composite: $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(\text{Time}_{ti}) + e_{ti}$

```
TITLE1 "Eq 5.5: Random Linear Time Model";
PROC MIXED DATA=example5 COVTEST NOITPRINT
NOCLPRINT METHOD=REML;
  CLASS IDnum wave;
  MODEL outcome = time / SOLUTION DDFM=Satterthwaite;
  RANDOM INTERCEPT time / G V GCORR VCORR TYPE=UN SUBJECT=IDnum;
  REPEATED wave / R TYPE=VC SUBJECT=IDnum;
  ESTIMATE "Intercept at Time 0" int 1 time 0;
  ESTIMATE "Intercept at Time 1" int 1 time 1;
  ESTIMATE "Intercept at Time 2" int 1 time 2;
  ESTIMATE "Intercept at Time 3" int 1 time 3;
RUN;
```

Note that the "time" variable gets included in the RANDOM statement, not "wave"—including "wave" would result in model non-convergence, because it would try to estimate a random slope variance for each possible difference between waves (instead of a single variance for a continuous random slope through all the waves).

```
          Estimated R Matrix for IDnum 1
 Row        Col1        Col2        Col3        Col4
   1      0.6986
   2                  0.6986
   3                              0.6986
   4                                          0.6986
```

After adding a random linear effect of time, the residual variance is smaller, but it is not correct to say that it has been reduced. Random effects do not explain variance; they simply re-allocate it. Here, this means that part of what was residual is now individual differences in the linear effect of time as a new pile of variance in the **G** matrix below.

```
               Estimated G Matrix
                   IDnum:
                   Person ID
Row    Effect      number         Col1        Col2

   1   Intercept     1           2.2624     0.05454
   2   time          1          0.05454      0.9089
```

The **G** matrix provides the variances and covariances of the individual random effects. Now **G** is a 2x2 matrix because we have 2 random effects (intercept, linear slope).

The **GCORR** matrix provides the correlation(s) among the individual random effects. Here, the individual intercepts and slopes are correlated $r = .04$.

```
          Estimated G Correlation Matrix
                   IDnum:
                   Person ID
Row    Effect      number         Col1        Col2
   1   Intercept     1           1.0000     0.03803
   2   time          1          0.03803      1.0000
```

Estimated V Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	2.9611	2.3170	2.3715	2.4260
2	2.3170	3.9790	4.2438	5.2073
3	2.3715	4.2438	6.8148	7.9885
4	2.4260	5.2073	7.9885	11.4684

Estimated V Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6750	0.5279	0.4163
2	0.6750	1.0000	0.8150	0.7709
3	0.5279	0.8150	1.0000	0.9036
4	0.4163	0.7709	0.9036	1.0000

The **V** matrix holds the total variances and covariances over waves (from putting **G** and **R** back together through the **Z** matrix). Likewise, **VCORR** holds the total correlations over waves. Note that all of these are now predicted to differ as a function of which wave it is (see table 5.2 for a description of how this works).

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	
UN(1,1)	IDnum	2.2624	0.8003	2.83	0.0023	Random intercept variance in G
UN(2,1)	IDnum	0.05454	0.3507	0.16	0.8764	Random intercept-slope covariance in G
UN(2,2)	IDnum	0.9089	0.3040	2.99	0.0014	Random linear slope variance in G
wave	IDnum	0.6986	0.1397	5.00	<.0001	Residual variance in R

Fit Statistics

-2 Res Log Likelihood	366.7
AIC (smaller is better)	374.7
AICC (smaller is better)	375.2
BIC (smaller is better)	379.6

Are we allowed to examine the $-2\Delta LL$ to see if adding a random linear effect of time improved model fit in REML? If so, how many model parameters have we added?

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	99.47	<.0001

This tests whether we need *anything* in the **G** matrix (so df=3). Note this does NOT tell us if we need the random linear time slope specifically!

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	
Intercept	10.2745	0.3318	24	30.97	<.0001	this is gamma00
time	1.7167	0.2048	24	8.38	<.0001	this is gamma10

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	24	70.26	<.0001

Estimates → **These are the predicted outcome means from a random linear model**

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept at Time 0	10.2745	0.3318	24	30.97	<.0001
Intercept at Time 1	11.9912	0.3736	24	32.09	<.0001
Intercept at Time 2	13.7080	0.5030	24	27.25	<.0001
Intercept at Time 3	15.4247	0.6711	24	22.98	<.0001

We can use the model estimates to calculate 95% random effects confidence intervals that describe the *predicted* range of *individual* random effects:

Random Effect 95% CI = fixed effect $\pm \left(1.96*\sqrt{\text{Random Variance}}\right)$

Intercept 95% CI = $\gamma_{00} \pm \left(1.96*\sqrt{\tau^2_{U_0}}\right) = 10.27 \pm \left(1.96*\sqrt{2.26}\right) =$ 7.32 to 13.22

Linear Time Slope 95% CI = $\gamma_{10} \pm \left(1.96*\sqrt{\tau^2_{U_1}}\right) = 1.72 \pm \left(1.96*\sqrt{0.91}\right) = -0.15$ to 3.59

Last but not least: there may still be residual covariances after modeling individual differences in the linear effect of time (i.e., adding a random linear slope to the G matrix). We can test alternative **R** matrix assumptions besides VC (which assumes no residual covariance/correlation over time) to see if this is the case:

```
TITLE1 "Random Linear Time Model + AR1 R Matrix";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT METHOD=REML;
       CLASS IDnum wave;
       MODEL outcome = time / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT time / G GCORR V VCORR TYPE=UN SUBJECT=IDnum;
       REPEATED / R RCORR TYPE=AR(1) SUBJECT=IDnum;
RUN;
```

Estimated R Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	0.7193	0.01841	0.000471	0.000012
2	0.01841	0.7193	0.01841	0.000471
3	0.000471	0.01841	0.7193	0.01841
4	0.000012	0.000471	0.01841	0.7193

Estimated R Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.02560	0.000655	0.000017
2	0.02560	1.0000	0.02560	0.000655
3	0.000655	0.02560	1.0000	0.02560
4	0.000017	0.000655	0.02560	1.0000

Fit Statistics

-2 Res Log Likelihood	366.7
AIC (smaller is better)	376.7
AICC (smaller is better)	377.4
BIC (smaller is better)	382.8

The −2LL is not smaller than the random linear time model, so adding an AR1 correlation to the **R** matrix does not improve model fit.

```
TITLE1 "Random Linear Time Model + TOEP2 R Matrix";
PROC MIXED DATA=example5 COVTEST NOITPRINT NOCLPRINT METHOD=REML;
       CLASS IDnum wave;
       MODEL outcome = time / SOLUTION DDFM=Satterthwaite;
       RANDOM INTERCEPT time / G GCORR V VCORR TYPE=UN SUBJECT=IDnum;
       REPEATED wave / R RCORR TYPE=TOEP(2) SUBJECT=IDnum;
RUN;
```

Estimated R Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	0.7127	0.01259		
2	0.01259	0.7127	0.01259	
3		0.01259	0.7127	0.01259
4			0.01259	0.7127

Estimated R Correlation Matrix for IDnum 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.01766		
2	0.01766	1.0000	0.01766	
3		0.01766	1.0000	0.01766
4			0.01766	1.0000

Fit Statistics

-2 Res Log Likelihood	366.7
AIC (smaller is better)	376.7
AICC (smaller is better)	377.4
BIC (smaller is better)	382.8

The −2LL is not smaller than the random linear time model, so adding a lag-1 covariance to the **R** matrix does not improve model fit, either.