Tricks of the Trade: Getting the most out of `xtmixed`

Roberto G. Gutierrez

Director of Statistics
StataCorp LP

2008 UK Stata Users Group Meeting

Outline

- The Linear Mixed Model
- Example 1: Standard Random Coefficients
- Example 2: Grouped Covariance Structures
- Example 3: Heteroskedastic Residual Errors
- Example 4: Smoothing Via Penalized Splines
- Conclusions

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon$$

where

- $\mathbf{y}$ is the $n \times 1$ vector of responses
- $\mathbf{X}$ is the $n \times p$ fixed-effects design matrix
- β are the fixed effects
- $\mathbf{Z}$ is the $n \times q$ random-effects design matrix
- $\mathbf{u}$ are the random effects
- ϵ is the $n \times 1$ vector of errors such that

$$\begin{bmatrix} \mathbf{u} \\ \epsilon \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \mathbf{I}_n \end{bmatrix}\right)$$

- Random effects are not directly estimated, but instead characterized by the elements of **G**, known as *variance components*
- You can, however "predict" random effects. These are known as best linear unbiased predictions (BLUPs)
- As such, you fit a mixed model by estimating $\boldsymbol{\beta}$, σ_ϵ^2, and the variance components in **G**
- We can fit linear mixed models in Stata using `xtmixed` and `gllamm`. In the special case of a random-intercept model, we can also use `xtreg`

- Classical representation has roots in the design literature, but can make model specification difficult
- When the data can be thought of as M independent panels, it is more convenient to express the mixed model as (for $i = 1, \ldots, M$)

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{u}_i + \epsilon_i$$

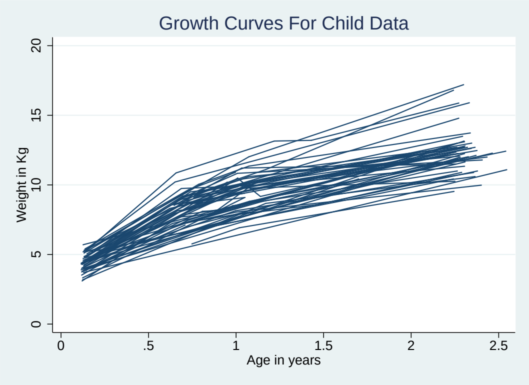
where $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{S})$, for $q \times q$ variance $\mathbf{S}$, and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Z}_M \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_M \end{bmatrix}; \quad \mathbf{G} = \mathbf{I}_M \otimes \mathbf{S}$$

Example

- Goldstein (1986) analyzed data on weight gain of Asian children in a British community (Rabe-Hesketh and Skrondal 2008, section 5.10)
- We analyze a subset of their data, namely 68 children weighed between one and five times inclusive
- The graph of growth curves will suggest the following model features:
  - overall quadratic growth
  - child-specific random intercepts
  - (perhaps) child-specific linear trends
  - child-specific quadratic components would perhaps be a bit much

```
. use http://www.stata.com/icpsr/mixed/child, clear
(Weight data on Asian children)
. sort id age
. graph twoway (line weight age, connect(ascending)), ///
>                xtitle(Age in years) ytitle(Weight in Kg) ///
>                title(Growth Curves For Child Data)
```



- Graphical features suggest the following model for the jth weighing of the ith child

$$\texttt{weight}_{ij} = \beta_0 + \beta_1 \texttt{age}_{ij} + \beta_2 \texttt{age}_{ij}^2 + u_{i0} + u_{i1} \texttt{age}_{ij} + \epsilon_{ij}$$

- This is a standard random-coefficients model, the bread and butter of `xtmixed`
- It is good practice to use `cov(unstructured)` and not assume the two random-effects terms are independent, the default
- You can always do an LR test to ensure that the added covariance term is significant

```
. gen age2 = age^2
. xtmixed weight age age2 || id: age, cov(unstructured) variance
Mixed-effects REML regression                  Number of obs      =       198
Group variable: id                             Number of groups   =        68
                                               Obs per group: min =         1
                                                              avg =       2.9
                                                              max =         5
                                               Wald chi2(2)       =   1940.65
Log restricted-likelihood =  -262.4327         Prob > chi2        =    0.0000
```

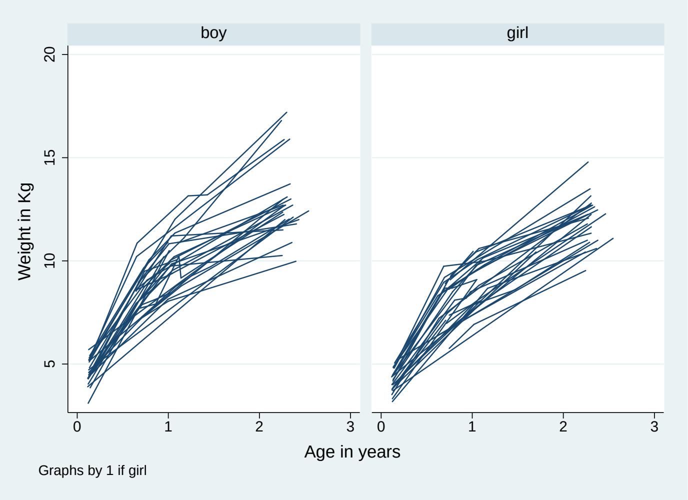
weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	7.703451	.2408987	31.98	0.000	7.231298	8.175604
age2	-1.66009	.0890272	-18.65	0.000	-1.834581	-1.4856
_cons	3.494664	.1384934	25.23	0.000	3.223222	3.766106

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
id: Unstructured				
var(age)	.2617525	.0912799	.1321462	.5184738
var(_cons)	.4172866	.1686882	.1889453	.9215797
cov(age,_cons)	.085354	.0904636	-.0919514	.2626593
var(Residual)	.3341601	.058922	.2365176	.4721128

```
LR test vs. linear regression:       chi2(3) =   114.39   Prob > chi2 = 0.0000
```

- The previous model grouped boys and girls together
- Is there a systematic difference (in the overall mean curve) between boys and girls?
- Do boys and girls demonstrate different variability about their respective average curves?
- We can certainly check graphically

```
. graph twoway (line weight age, connect(ascending)), by(girl) ///
>               xtitle(Age in years) ytitle(Weight in Kg)
```


boy
girl
Weight in Kg
20
15
10
5
0
1
2
3
0
1
2
3
Age in years
Graphs by 1 if girl

- The deficiency of our previous model is that it assumed the variance components were the same for both boys and girls

$$\texttt{weight}_{ij} = \beta_0 + \beta_1 \texttt{age}_{ij} + \beta_2 \texttt{age}_{ij}^2 + \beta_3 \texttt{girl}_{ij} + u_{i0} + u_{i1} \texttt{age}_{ij} + \epsilon_{ij}$$

- Our graph indicates that girls' curves are bunched closer together
- As such, a better model would be to have gender-specific random effects, i.e. distinct r.e. covariance matrices for boys and girls
- In other words we want the portion in red above replaced by

$$u_{i0}^{b} \texttt{boy}_{ij} + u_{i1}^{b} \left(\texttt{age}_{ij} \times \texttt{boy}_{ij} \right) + u_{i0}^{g} \texttt{girl}_{ij} + u_{i1}^{g} \left(\texttt{age}_{ij} \times \texttt{girl}_{ij} \right)$$

- In our new model, the covariance matrix of the random effects is block diagonal, i.e.

$$\mathrm{Var}\begin{bmatrix} u_{i0}^{b} \\ u_{i1}^{b} \\ u_{i0}^{g} \\ u_{i1}^{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_b & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_g \end{bmatrix}$$

  where both $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_g$ are 2×2 unstructured covariance matrices

- You can achieve this effect by "repeating level specifications"

- We will also add corresponding fixed-effects terms, boy/girl dummy variables and boy/girl interactions with age. Otherwise we would be imposing dubious constraints

- We wish to fit the following model

$$
\begin{aligned}
\texttt{weight}_{ij} \quad = \quad & \beta_2 \texttt{age}^2_{ij} + \\
& \beta_3 \texttt{boy}_{ij} + \beta_4 \left(\texttt{age}_{ij} \times \texttt{boy}_{ij}\right) + \\
& \beta_5 \texttt{girl}_{ij} + \beta_6 \left(\texttt{age}_{ij} \times \texttt{girl}_{ij}\right) + \\
& u^b_{i0} \texttt{boy}_{ij} + u^b_{i1} \left(\texttt{age}_{ij} \times \texttt{boy}_{ij}\right) + \\
& u^g_{i0} \texttt{girl}_{ij} + u^g_{i1} \left(\texttt{age}_{ij} \times \texttt{girl}_{ij}\right) + \epsilon_{ij}
\end{aligned}
$$

- At this point I recommend using ML instead of the default REML estimation. ML permits LR tests for models where the fixed-effects structures differ
- For example, say you wanted to test against a model with no interactions, fixed or random

```
. gen boy = !girl
. gen boyXage = boy*age
. gen girlXage = girl*age
. xtmixed weight age2 boy boyXage girl girlXage, nocons ///
>                || id: boy boyXage, nocons cov(un)     ///
>                || id: girl girlXage, nocons cov(un) mle var
Mixed-effects ML regression                     Number of obs      =       198
Group variable: id                              Number of groups   =        68
                                                Obs per group: min =         1
                                                               avg =       2.9
                                                               max =         5
                                                Wald chi2(5)       =   7095.79
Log likelihood = -248.70821                     Prob > chi2        =    0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age2	-1.641597	.0867182	-18.93	0.000	-1.811562	-1.471633
boy	3.766094	.1618969	23.26	0.000	3.448782	4.083406
boyXage	7.782752	.2609228	29.83	0.000	7.271353	8.294152
girl	3.257528	.178941	18.20	0.000	2.90681	3.608246
girlXage	7.538577	.2386229	31.59	0.000	7.070885	8.006269

--more--

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
id: Unstructured				
var(boy)	.2887796	.1915665	.078688	1.059801
var(boyXage)	.4557309	.1794435	.210644	.9859798
cov(boy,boyXage)	.0227221	.1373405	-.2464604	.2919046
id: Unstructured				
var(girl)	.4799603	.2223231	.1936061	1.189848
var(girlXage)	.0423413	.0608414	.0025331	.7077496
cov(girl,girlXage)	.0645366	.0869897	-.1059602	.2350333
var(Residual)	.3211566	.0555259	.2288493	.4506964

```
LR test vs. linear regression:      chi2(6) =   113.34   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

- It turns out the greater spread in the boys' curves is due to larger variability in the linear component, not the intercept
- Neither covariance appears to be significant. You can drop both by simply reverting to `xtmixed`'s default `independent` covariance structure
- The `identity` could be used to further restrict the model (equality constraints)
- Using repeated level specifications, each separated by `||`, for achieving gender-specific error structures is equivalent to using the GROUP option of some PROCedure for fitting MIXED models employed by Some Alternative Software

- What about heteroskedasticity in the residual errors?

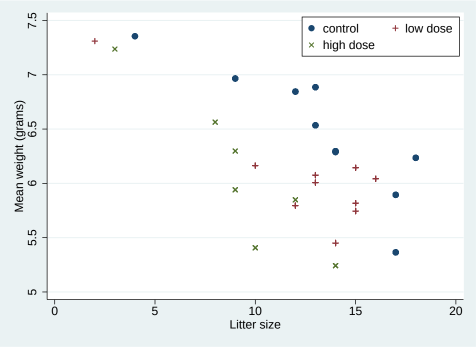
Example

- Dempster et al. (1984) analyze data from a reproductive study on rats to assess the effect of an experimental compound on pup weights (Rabe-Hesketh and Skrondal 2008, exercise 3.5)
- 27 litters were recorded over three treatment groups: control, low dose, and high dose
- Weight is related to dosage level and litter size, which are "litter-level" covariates
- Weight is also related to sex, a pup-level covariate

```
. use http://www.stata.com/icpsr/mixed/rats, clear
(Weights of rat pups)

. egen mnw = mean(weight), by(litter)

. twoway (scatter mnw size if dose==0)  ///
>        (scatter mnw size if dose==1, msymbol(plus)) ///
>        (scatter mnw size if dose==2, msymbol(x) msize(large)), ///
>                    ytitle(Mean weight (grams)) ///
>                    legend(order(1 "control" 2 "low dose" 3 "high dose")) ///
>                    legend(position(1) ring(0))
```



- Our initial model is

$$\texttt{weight}_{ij} = \beta_0 + \beta_1 \texttt{dose}_{1ij} + \beta_2 \texttt{dose}_{2ij} + \beta_3 \texttt{size}_{ij} + \beta_4 \texttt{female}_{ij} + u_i + \epsilon_{ij}$$

  for $i = 1, ..., 27$ litters and $j = 1, ..., n_i$ pups within litter

- This is a standard random-intercept model, fit by `xtmixed` or, even, `xtreg`

- Residual plots vs. the linear predictor are always a good idea. In our case, we produce these plots by variable `female` because we are curious about heteroskedasticity

```
. xi: xtmixed weight i.dose size female || litter:
i.dose            _Idose_0-2          (naturally coded; _Idose_0 omitted)
```

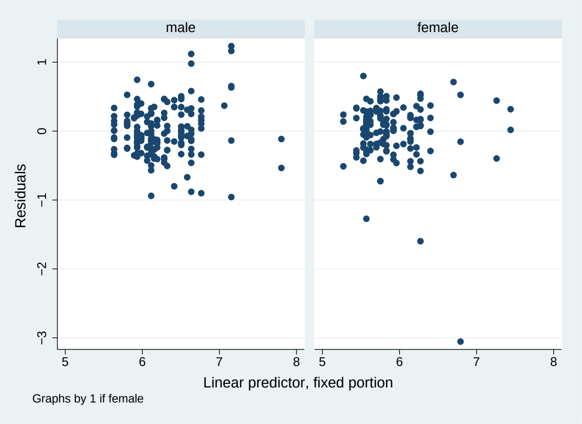
(output omitted)

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Idose_1	-.4416666	.1513553	-2.92	0.004	-.7383176	-.1450157
_Idose_2	-.8706054	.1830525	-4.76	0.000	-1.229382	-.511829
size	-.1299602	.0190485	-6.82	0.000	-.1672946	-.0926259
female	-.3626441	.0477374	-7.60	0.000	-.4562077	-.2690805
_cons	8.324096	.2770569	30.04	0.000	7.781074	8.867118

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
litter: Identity				
sd(_cons)	.3140074	.0532536	.2252069	.4378225
sd(Residual)	.4045051	.0166929	.3730758	.4385822

```
LR test vs. linear regression: chibar2(01) =    90.73 Prob >= chibar2 = 0.0000
```

```
. predict xbeta
(option xb assumed)
. predict r, residuals
. twoway (scatter r xbeta, by(female))
```



- In our previous model, we want ϵ_{ij} replaced by

$$\epsilon_{ij} = \epsilon_{ij}^{m}(1 - \texttt{female}_{ij}) + \epsilon_{ij}^{f}\texttt{female}_{ij}$$

- The bad news is that `xtmixed` will always produce a single, overall residual term. The good news is we can express the above instead as

$$\epsilon_{ij} = \epsilon_{ij}^{m} + (\epsilon_{ij}^{f} - \epsilon_{ij}^{m})\texttt{female}_{ij}$$

  and we can estimate the *additional* variability due to female

- This alternate form allows us to fit this model in `xtmixed`, provided we create a pseudo two-level model, with the lowest-level "groups" being the observations (pups) themselves, nested within litters

```
. gen pup = _n
. xi: xtmixed weight i.dose size female || litter: || pup: female, nocons var
Mixed-effects REML regression                   Number of obs      =       321
```

Group Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
litter	27	2	11.9	18
pup	321	1	1.0	1

```
                                                Wald chi2(4)       =    107.22
Log restricted-likelihood = -196.90368          Prob > chi2        =    0.0000
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Idose_1	-.4500473	.15523	-2.90	0.004	-.7542925	-.1458021
_Idose_2	-.8780883	.18757	-4.68	0.000	-1.245719	-.5104578
size	-.1307603	.0196311	-6.66	0.000	-.1692365	-.092284
female	-.3634425	.04821	-7.54	0.000	-.4579324	-.2689526
_cons	8.339868	.2845412	29.31	0.000	7.782177	8.897558

--more--

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
litter: Identity				
var(_cons)	.1046383	.035361	.053956	.2029279
pup: Identity				
var(female)	.0558646	.02933	.0199636	.1563272
var(Residual)	.1370851	.0161837	.108768	.1727743

```
LR test vs. linear regression:       chi2(2) =    94.55   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
. nlcom (  male: exp(2 * [lnsig_e]_cons)) ///
>       (female: exp(2 * [lnsig_e]_cons) + exp(2 * [lns2_1_1]_cons))
        male:  exp(2 * [lnsig_e]_cons)
      female:  exp(2 * [lnsig_e]_cons) + exp(2 * [lns2_1_1]_cons)
```

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
male	.1370851	.0161837	8.47	0.000	.1053657	.1688044
female	.1929497	.023584	8.18	0.000	.1467259	.2391734

- Fitting heteroskedastic-error models using this procedure will often result in non-convergent models

- The reason is that implicit in the above is the assumption that $\sigma^2_{f\epsilon} > \sigma^2_{m\epsilon}$

- If not true, the variance component representing added variability will tend towards zero and form a ridge in the likelihood surface

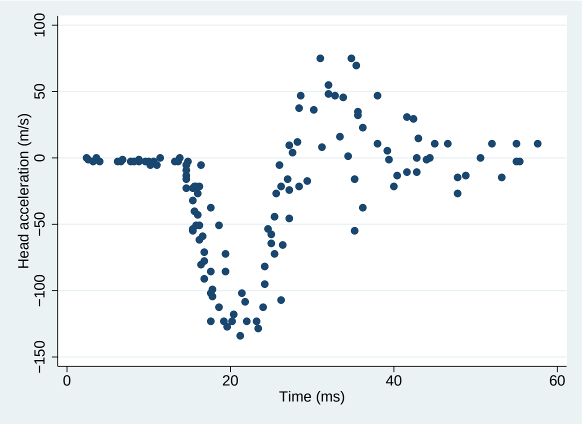
- The solution? Simply model the added variability as due to `male` rather than as due to `female`

- Finally, you can also use `xtmixed` for spline smoothing:

Example

- Silverman (1985) analyzed 133 measurements taken from a simulated motorcycle crash
- Head acceleration (y) was measured over time (x)
- Because of the changing nature of the curve over time and the heteroskedasticity of errors, these data are a staple of the smoothing literature

```
. use http://www.stata.com/icpsr/mixed/motor, clear
. graph twoway (scatter accel time)
```



- A linear-spline smoothing model has the form

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^{M} \gamma_j \left| x_i - \kappa_j \right|_+ + \epsilon_i$$

  for M knot points κ_j, usually chosen to form a grid

- Think of linear smoothing splines as just a series of interlocking line segments, the slopes of which need to be estimated

- The above suggests plain linear regression, with the appropriately-generated regressors, of course. Call this the "fixed-effects" approach

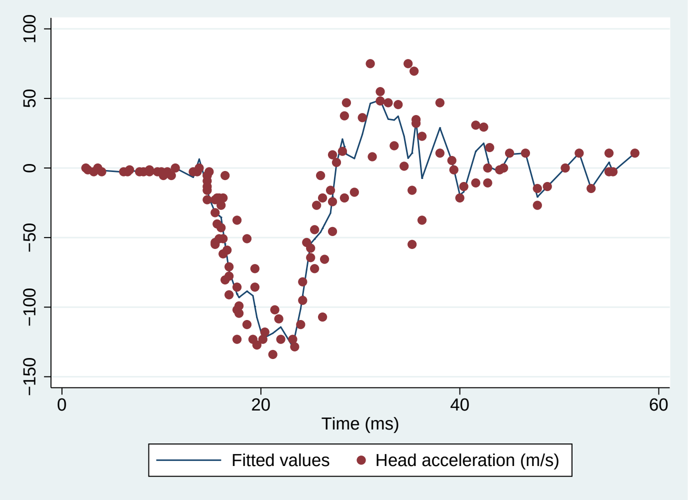
```
. local i 1

. forvalues k = 1(1)60 {
  2.          gen time_`i' = cond(time - `k' > 0, time - `k', 0)
  3.          local ++i
  4. }

. qui regress accel time time_*

. predict accel_fixed
(option xb assumed; fitted values)

. graph twoway (line accel_fixed time) (scatter accel time)
```


100
50
0
−50
−100
−150
0
20
40
60
Time (ms)
Fitted values
Head acceleration (m/s)

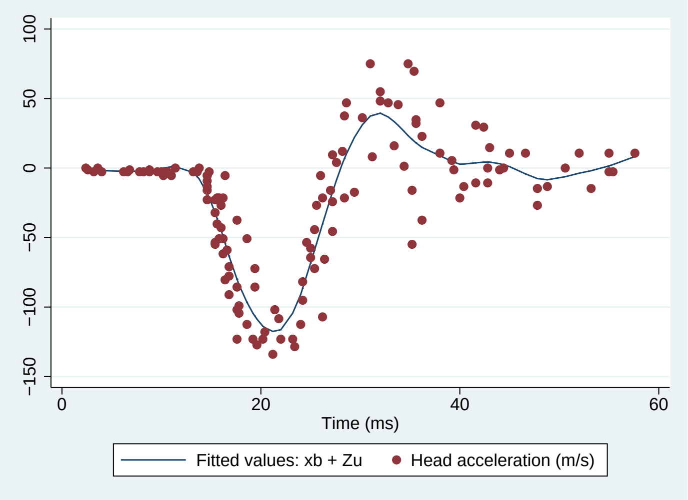
- As you may have noticed, the problem with the fixed-effects approach is that it tends to interpolate the data
- One solution is to use *penalized splines*, which adds a roughness penalty to the likelihood from the linear-regression approach
- Ruppert et al. (2003), among others, show that this is equivalent to treating the slopes as random rather than fixed, and estimating them as BLUPs of a mixed model
- As such, a "random-effects" approach yields a much nicer-looking smooth, and we can get `xtmixed` to do all the heavy lifting

```
. xtmixed accel time || _all: time_*, noconstant cov(identity)
  (output omitted)

------------------------------------------------------------------------------
       accel |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |  -.4672689   13.33173    -0.04   0.972    -26.59698    25.66244
       _cons |  -.0152613   34.32348    -0.00   1.000    -67.28805    67.25753
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
      sd(time_1..time_56)(1) |    7.01774   1.479116      4.642918    10.60727
-----------------------------+------------------------------------------------
               sd(Residual)  |   22.53256   1.462753      19.84051    25.58988
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   151.17 Prob >= chibar2 = 0.0000
(1) time_1 time_2 time_3 time_4 time_6 time_7 time_8 time_9 time_10 time_11
    time_12 time_13 time_14 time_15 time_16 time_17 time_18 time_19 time_20
    time_21 time_22 time_23 time_24 time_25 time_26 time_27 time_28 time_29
    time_30 time_31 time_32 time_33 time_34 time_35 time_36 time_37 time_38
    time_39 time_40 time_41 time_42 time_43 time_44 time_45 time_47 time_48
    time_49 time_50 time_52 time_53 time_55 time_56
```

```
. predict accel_random, fitted
. graph twoway (line accel_random time) (scatter accel time)
```



Conclusions

- `xtmixed` is versatile
- You can repeat level specifications to achieve structured covariance matrices
- When combined with `xtmixed` available structures, covariance matrices can be constrained even further
- BLUPs are a useful smoothing tool. Their shrinkage properties keep them from overfitting the data

Dempster, A. P., M. R. Selwyn, C. M. Patel and A. J. Roth. 1984. Statistical and computational aspects of mixed model analysis. *Journal of the Royal Statistical Society, Series C* **33**: 203–214.

Goldstein, H. 1986. Efficient statistical modelling of longitudinal data. *Human Biology* **13**: 129–142.

Laird, N. M. and J. H. Ware. 1982. Random-effects models for longitudinal data. *Biometrics* **38**: 963–974.

Rabe-Hesketh, S. and A. Skronal. 2008. *Multilevel and Longitudinal Modeling Using Stata, Second Edition*. College Station, TX: Stata Press.

Ruppert, D., M. P. Wand and R. J. Carroll. 2003. *Semiparametric Regression*. Cambridge: Cambridge University Press.

Silverman, B. W. 1985. Some aspects of the spline smoothing approach to nonparametric curve fitting. *Journal of the Royal Statistical Society, Series B* **47**: 1–52.