The Stata Journal (2011)
11, Number 1, pp. 145–148

Stata tip 95: Estimation of error covariances in a linear model

Nicholas J. Horton
Department of Mathematics and Statistics
Clark Science Center
Smith College
Northampton, MA
nhorton@smith.edu

1 Introduction

A recent review (Horton 2008) of the second edition of *Multilevel and Longitudinal Modeling Using Stata* (Rabe-Hesketh and Skrondal 2008) decried the lack of support in previous versions of Stata for models within the `xtmixed` command that directly estimate the variance–covariance matrix (akin to the REPEATED statement in SAS PROC MIXED). In this tip, I describe how support for these models is now available in Stata 11 (see also `help whatsnew10to11`) and demonstrate its use by replication of an analysis of a longitudinal dental study using an unstructured covariance matrix.

2 Model

I use the notation of Fitzmaurice, Laird, and Ware (2004, chap. 4 and 5) to specify linear models of the form $E(Y_i) = X_i\beta$, where Y_i and X_i denote the vector of responses and the matrix of covariates, respectively, for the ith subject, where $i = 1, \ldots, N$. Assume that each subject has up to n observations on a common set of times. The response vector Y_i is assumed to be multivariate normal with covariance given by $\Sigma_i(\theta)$, where θ is a vector of covariance parameters. If an unstructured covariance matrix is assumed, then there will be $n \times (n+1)/2$ covariance parameters. Restricted maximum-likelihood estimation is used.

3 Example

I consider data from an analysis of a study of dental growth, described on page 184 of Fitzmaurice, Laird, and Ware (2004). Measures of distances (in mm) were obtained on 27 subjects (11 girls and 16 boys) at ages 8, 10, 12, and 14 years.

3.1 Estimation in SAS

Below I give SAS code to fit a model with the mean response unconstrained over time (3 degrees of freedom) and main effect for gender as well as an unstructured working covariance matrix (10 parameters):

 st0222

```
proc mixed data=one;
        class id time;
        model y = time female / s;
        repeated time / type=un subject=id r;
run;
```

This code generates the following output:

```
The Mixed Procedure
                  Model Information
Data Set                     WORK.ONE
Dependent Variable           y
Covariance Structure         Unstructured
Subject Effect               id
Estimation Method            REML
Residual Variance Method     None
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Between-Within

           Dimensions
Covariance Parameters            10
Columns in X                      6
Columns in Z                      0
Subjects                         27
Max Obs Per Subject               4

            Estimated R Matrix for id 1
 Row        Col1        Col2        Col3        Col4
   1      5.3741      2.7887      3.8442      2.6242
   2      2.7887      4.2127      2.8832      3.1717
   3      3.8442      2.8832      6.4284      4.3024
   4      2.6242      3.1717      4.3024      5.3751

                       Solution for Fixed Effects
                                 Standard
Effect       time    Estimate       Error      DF    t Value    Pr > |t|
Intercept             26.9258      0.5376      25      50.08      <.0001
time          8       -3.9074      0.4514      25      -8.66      <.0001
time         10       -2.9259      0.3466      25      -8.44      <.0001
time         12       -1.4444      0.3442      25      -4.20      0.0003
time         14             0           .       .          .           .
female                -2.0452      0.7361      25      -2.78      0.0102
```

3.2 Estimation in Stata

The equivalent model can now be fit in Stata 11:

```
. use http://www.math.smith.edu/labs/denttall
. xtmixed y ib14.time female, || id:, nocons residuals(un, t(time)) var
```

The `xtmixed` command yields the equivalent output:

```
Mixed-effects REML regression                   Number of obs      =       108
Group variable: id                              Number of groups   =        27

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4

                                                Wald chi2(4)       =    101.50
Log restricted-likelihood =  -212.4093          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |
          8  |  -3.907407   .4513647    -8.66   0.000    -4.792066   -3.022749
         10  |  -2.925926   .3466401    -8.44   0.000    -3.605328   -2.246524
         12  |  -1.444444   .3441962    -4.20   0.000    -2.119057   -.7698322
             |
      female |  -2.045172    .736141    -2.78   0.005    -3.487982   -.6023627
       _cons |   26.92581   .5376092    50.08   0.000     25.87212    27.97951
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id:                  (empty) |
-----------------------------+------------------------------------------------
Residual: Unstructured       |
                     var(e8) |   5.374086   1.510892      3.097379    9.324271
                    var(e10) |    4.21272   1.201038      2.409277    7.366114
                    var(e12) |   6.428418   1.810989      3.700897    11.16609
                    var(e14) |   5.375108   1.608682      2.989761    9.663575
                 cov(e8,e10) |   2.788773   1.112924      .6074823    4.970064
                 cov(e8,e12) |   3.844272   1.392097      1.115811    6.572732
                 cov(e8,e14) |   2.624241   1.207689      .2572134    4.991268
                cov(e10,e12) |   2.883246   1.183372      .5638802    5.202612
                cov(e10,e14) |   3.171762   1.153809      .9103389    5.433186
                cov(e12,e14) |   4.302404   1.499388      1.363657     7.24115
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(9) =    54.59   Prob > chi2 = 0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

Several points are worth noting:

1. The default output from `xtmixed` provides estimates of variability as well as confidence intervals for the covariance parameter estimates.

2. Considerable flexibility regarding additional covariance structures is provided by the `residuals()` option (including exchangeable, autoregressive, and moving-average structures).

3. Specifying a `by()` variable within the `residuals()` option can allow separate estimation of error covariances by group (for example, in this setting, separate estimation of the structures for men and for women).

4. The `ib14` specification for the time factor variable facilitates changing the reference grouping to match the SAS defaults.

5. Dropping the `var` option will generate correlations (which may be more interpretable if the variances change over time).

For the dental example, we see that the estimated correlation is lowest between the observations that are farthest apart ($r = 0.49$) and generally higher for shorter intervals.

```
  corr(e8,e10) |   .5861106   .1306678      .2743855    .7863675
  corr(e8,e12) |   .6540481   .1129091      .3761828    .8239756
  corr(e8,e14) |   .4882675   .1518479      .1420355    .7280491
 corr(e10,e12) |   .5540493   .1370823      .2322075    .7665423
 corr(e10,e14) |   .6665393   .1115412      .3894063    .8330066
 corr(e12,e14) |   .7319232   .0930009      .4931844     .868134
```

4 Summary

Modeling the associations between observations on the same subject using mixed effects and an unstructured covariance matrix is a flexible and attractive alternative to a random-effects model with cluster–robust standard errors. This is particularly useful when the number of measurement occasions is relatively small, and measurements are taken at a common set of occasions for all subjects. The addition of support for this model within `xtmixed` in Stata 11 is a welcome development.

5 Acknowledgments

Thanks to Kristin MacDonald, Roberto Gutierrez, Garrett Fitzmaurice, and the anonymous reviewers for helpful comments on an earlier draft.

References

Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2004. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley.

Horton, N. J. 2008. Review of Multilevel and Longitudinal Modeling Using Stata, Second Edition, by Sophia Rabe-Hesketh and Anders Skrondal. *Stata Journal* 8: 579–582.

Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. College Station, TX: Stata Press.