xtgee postestimation — Postestimation tools for xtgee

Postestimation commands

The following postestimation command is of special interest after `xtgee`:

Command	Description
`estat wcorrelation`	estimated matrix of the within-group correlations

The following standard postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of parameters
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`etable`	table of estimation results
* `forecast`	dynamic forecasts and simulations
`hausman`	Hausman's specification test
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of parameters
`margins`	marginal means, predictive margins, marginal effects, and average marginal effects
`marginsplot`	graph the results from margins (profile plots, interaction plots, etc.)
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of parameters
`predict`	means, rates, probabilities, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of parameters
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

*`forecast` is not appropriate with `mi` estimation results.

predict

Description for predict

predict creates a new variable containing predictions such as predicted values, probabilities, linear predictions, standard errors, and the equation-level score.

Menu for predict

Statistics > Postestimation

Syntax for predict

predict [*type*] *newvar* [*if*] [*in*] [, *statistic* nooffset]

statistic	Description
Main	
mu	predicted value of *depvar*; considers the offset() or exposure(); the default
rate	predicted value of *depvar*
pr(*n*)	probability $\Pr(y_{it} = n)$ for family(poisson) link(log)
pr(*a*,*b*)	probability $\Pr(a \leq y_{it} \leq b)$ for family(poisson) link(log)
xb	linear prediction
stdp	standard error of the linear prediction
score	first derivative of the log likelihood with respect to $\mathbf{x}_{it}\boldsymbol{\beta}$

These statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample.

Options for predict

Main

mu, the default, and rate calculate the predicted value of *depvar*. mu takes into account the offset() or exposure() together with the denominator if the family is binomial; rate ignores those adjustments. mu and rate are equivalent if you did not specify offset() or exposure() when you fit the xtgee model and you did not specify family(binomial #) or family(binomial *varname*), meaning the binomial family and a denominator not equal to one.

Thus mu and rate are the same for family(gaussian) link(identity).

mu and rate are not equivalent for family(binomial pop) link(logit). Then mu would predict the number of positive outcomes and rate would predict the probability of a positive outcome.

mu and rate are not equivalent for family(poisson) link(log) exposure(time). Then mu would predict the number of events given exposure time and rate would calculate the incidence rate—the number of events given an exposure time of 1.

pr(*n*) calculates the probability $\Pr(y_{it} = n)$ for family(poisson) link(log), where *n* is a nonnegative integer that may be specified as a number or a variable.

pr(*a*,*b*) calculates the probability $\Pr(a \leq y_{it} \leq b)$ for family(poisson) link(log), where *a* and *b* are nonnegative integers that may be specified as numbers or variables;

b missing ($b \geq .$) means $+\infty$;
pr(20,.) calculates $\Pr(y_{it} \geq 20)$;
pr(20,*b*) calculates $\Pr(y_{it} \geq 20)$ in observations for which $b \geq .$ and calculates $\Pr(20 \leq y_{it} \leq b)$ elsewhere.

pr(.,*b*) produces a syntax error. A missing value in an observation of the variable *a* causes a missing value in that observation for pr(*a*,*b*).

xb calculates the linear prediction.

stdp calculates the standard error of the linear prediction.

score calculates the equation-level score, $u_{it} = \partial \ln L(\mathbf{x}_{it}\boldsymbol{\beta})/\partial(\mathbf{x}_{it}\boldsymbol{\beta})$.

nooffset is relevant only if you specified offset(*varname*), exposure(*varname*), family(binomial #), or family(binomial *varname*) when you fit the model. It modifies the calculations made by predict so that they ignore the offset or exposure variable and the binomial denominator. Thus predict ..., mu nooffset produces the same results as predict ..., rate.

margins

Description for margins

margins estimates margins of response for predicted values, probabilities, and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

margins [*marginlist*] [, *options*]

margins [*marginlist*] , predict(*statistic* ...) [predict(*statistic* ...) ...] [*options*]

statistic	Description
mu	predicted value of *depvar*; considers the offset() or exposure(); the default
rate	predicted value of *depvar*
pr(*n*)	probability $\Pr(y_{it} = n)$ for family(poisson) link(log)
pr(*a*,*b*)	probability $\Pr(a \leq y_{it} \leq b)$ for family(poisson) link(log)
xb	linear prediction
stdp	not allowed with margins
score	not allowed with margins

Statistics not allowed with margins are functions of stochastic quantities other than e(b).

For the full syntax, see [R] **margins**.

estat

Description for estat

`estat wcorrelation` displays the estimated matrix of the within-group correlations.

Menu for estat

Statistics > Postestimation

Syntax for estat

`estat wcorrelation` [, `compact` `format(`%*fmt*`)`]

`collect` is allowed with `estat wcorrelation`; see [U] **11.1.10 Prefix commands**.

Options for estat

`compact` specifies that only the parameters (alpha) of the estimated matrix of within-group correlations be displayed rather than the entire matrix.

`format(`%*fmt*`)` overrides the display format; see [D] **format**.

Remarks and examples

▷ Example 1

`xtgee` can estimate rich correlation structures. In example 2 of [XT] **xtgee**, we fit the model

```
. use https://www.stata-press.com/data/r19/nlswork2
(National Longitudinal Survey of Young Women, 14-24 years old in 1968)
. xtgee ln_w grade age c.age#c.age
  (output omitted)
```

After estimation, `estat wcorrelation` reports the working correlation matrix **R**:

```
. estat wcorrelation
Estimated within-idcode correlation matrix R:
```

	c1	c2	c3	c4	c5	c6
r1	1					
r2	.4851356	1				
r3	.4851356	.4851356	1			
r4	.4851356	.4851356	.4851356	1		
r5	.4851356	.4851356	.4851356	.4851356	1	
r6	.4851356	.4851356	.4851356	.4851356	.4851356	1
r7	.4851356	.4851356	.4851356	.4851356	.4851356	.4851356
r8	.4851356	.4851356	.4851356	.4851356	.4851356	.4851356
r9	.4851356	.4851356	.4851356	.4851356	.4851356	.4851356

	c7	c8	c9
r7	1		
r8	.4851356	1	
r9	.4851356	.4851356	1

The equal-correlation model corresponds to an exchangeable correlation structure, meaning that the correlation of observations within person is a constant. The working correlation estimated by `xtgee` is 0.4851. (`xtreg, re`, by comparison, reports 0.5141; see the `xtreg` command in example 2 of [XT] **xtgee**.) We constrained the model to have this simple correlation structure. What if we relaxed the constraint? To go to the other extreme, let's place no constraints on the matrix (other than its being symmetric). We do this by specifying `correlation(unstructured)`, although we can abbreviate the option.

```
. xtgee ln_w grade age c.age#c.age, corr(unstructured) nolog
GEE population-averaged model                    Number of obs     =   16,085
Group and time vars: idcode year                 Number of groups =    3,913
Family: Gaussian                                 Obs per group:
Link:   Identity                                              min =        1
Correlation: unstructured                                     avg =      4.1
                                                              max =        9
                                                 Wald chi2(3)      = 2405.20
Scale parameter = .1418513                       Prob > chi2       =  0.0000
```

ln_wage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
grade	.0720684	.002151	33.50	0.000	.0678525	.0762843
age	.1008095	.0081471	12.37	0.000	.0848416	.1167775
c.age#c.age	-.0015104	.0001617	-9.34	0.000	-.0018272	-.0011936
_cons	-.8645484	.1009488	-8.56	0.000	-1.062404	-.6666923

```
. estat wcorrelation
Estimated within-idcode correlation matrix R:
```

	c1	c2	c3	c4	c5	c6
r1	1					
r2	.4354838	1				
r3	.4280248	.5597329	1			
r4	.3772342	.5012129	.5475113	1		
r5	.4031433	.5301403	.502668	.6216227	1	
r6	.3663686	.4519138	.4783186	.5685009	.7306005	1
r7	.2819915	.3605743	.3918118	.4012104	.4642561	.50219
r8	.3162028	.3445668	.4285424	.4389241	.4696792	.5222537
r9	.2148737	.3078491	.3337292	.3584013	.4865802	.4613128

	c7	c8	c9
r7	1		
r8	.6475654	1	
r9	.5791417	.7386595	1

This correlation matrix looks different from the previously constrained one and shows, in particular, that the serial correlation of the residuals diminishes as the lag increases, although residuals separated by small lags are more correlated than, say, AR(1) would imply.

◁

▷ Example 2

In example 1 of [XT] **xtprobit**, we showed a random-effects model of unionization using the `union` data described in [XT] **xt**. We performed the estimation using `xtprobit` but said that we could have used `xtgee` as well. Here we fit a population-averaged (equal correlation) model for comparison:

```
. use https://www.stata-press.com/data/r19/union
(NLS Women 14-24 in 1968)

. xtgee union age grade i.not_smsa south##c.year, family(binomial) link(probit)

Iteration 1:  Tolerance = .12544249
Iteration 2:  Tolerance = .0034686
Iteration 3:  Tolerance = .00017448
Iteration 4:  Tolerance = 8.382e-06
Iteration 5:  Tolerance = 3.997e-07

GEE population-averaged model                     Number of obs    = 26,200
Group variable: idcode                            Number of groups =  4,434
Family: Binomial                                  Obs per group:
Link:   Probit                                                 min =      1
Correlation: exchangeable                                      avg =    5.9
                                                               max =     12
                                                  Wald chi2(6)     = 242.57
Scale parameter = 1                               Prob > chi2      = 0.0000

------------------------------------------------------------------------------
       union | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0089699   .0053208     1.69   0.092    -.0014586    .0193985
       grade |   .0333174   .0062352     5.34   0.000     .0210966    .0455382
  1.not_smsa |  -.0715717    .027543    -2.60   0.009    -.1255551   -.0175884
     1.south |  -1.017368    .207931    -4.89   0.000    -1.424905   -.6098308
        year |  -.0062708   .0055314    -1.13   0.257    -.0171122    .0045706
             |
south#c.year |
           1 |   .0086294     .00258     3.34   0.001     .0035727     .013686
             |
       _cons |  -.8670997    .294771    -2.94   0.003     -1.44484   -.2893592
------------------------------------------------------------------------------
```

Let's look at the correlation structure and then relax it:

```
. estat wcorrelation, format(%8.4f)

Estimated within-idcode correlation matrix R:

           |     c1       c2       c3       c4       c5       c6       c7
-----------+---------------------------------------------------------------
        r1 |  1.0000
        r2 |  0.4615   1.0000
        r3 |  0.4615   0.4615   1.0000
        r4 |  0.4615   0.4615   0.4615   1.0000
        r5 |  0.4615   0.4615   0.4615   0.4615   1.0000
        r6 |  0.4615   0.4615   0.4615   0.4615   0.4615   1.0000
        r7 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   1.0000
        r8 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   0.4615
        r9 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   0.4615
       r10 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   0.4615
       r11 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   0.4615
       r12 |  0.4615   0.4615   0.4615   0.4615   0.4615   0.4615   0.4615

           |     c8       c9      c10      c11      c12
-----------+---------------------------------------------
        r8 |  1.0000
        r9 |  0.4615   1.0000
       r10 |  0.4615   0.4615   1.0000
       r11 |  0.4615   0.4615   0.4615   1.0000
       r12 |  0.4615   0.4615   0.4615   0.4615   1.0000
```

We estimate the fixed correlation between observations within person to be 0.4615. We have many data (an average of 5.9 observations on 4,434 women), so estimating the full correlation matrix is feasible. Let's do that and then examine the results:

```
. xtgee union age grade i.not_smsa south##c.year, family(binomial) link(probit)
> corr(unstructured) nolog
GEE population-averaged model                        Number of obs    = 26,200
Group and time vars: idcode year                     Number of groups =  4,434
Family: Binomial                                     Obs per group:
Link:   Probit                                                    min =      1
Correlation: unstructured                                         avg =    5.9
                                                                  max =     12
                                                     Wald chi2(6)     = 198.45
Scale parameter = 1                                  Prob > chi2      = 0.0000

------------------------------------------------------------------------------
       union | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0096612   .0053366     1.81   0.070    -.0007984    .0201208
       grade |   .0352762   .0065621     5.38   0.000     .0224148    .0481377
  1.not_smsa |   -.093073   .0291971    -3.19   0.001    -.1502983   -.0358478
     1.south |  -1.028526    .278802    -3.69   0.000    -1.574968   -.4820839
        year |  -.0088187    .005719    -1.54   0.123    -.0200278    .0023904
             |
south#c.year |
          1  |   .0089824   .0034865     2.58   0.010      .002149    .0158158
             |
       _cons |  -.7306192    .316757    -2.31   0.021    -1.351451    -.109787
------------------------------------------------------------------------------

. estat wcorrelation, format(%8.4f)
Estimated within-idcode correlation matrix R:
          |     c1        c2        c3        c4        c5        c6        c7
----------+---------------------------------------------------------------------
       r1 | 1.0000
       r2 | 0.6667    1.0000
       r3 | 0.6151    0.6523    1.0000
       r4 | 0.5268    0.5717    0.6101    1.0000
       r5 | 0.3309    0.3669    0.4005    0.4783    1.0000
       r6 | 0.3000    0.3706    0.4237    0.4562    0.6426    1.0000
       r7 | 0.2995    0.3568    0.3851    0.4279    0.4931    0.6384    1.0000
       r8 | 0.2759    0.3021    0.3225    0.3751    0.4682    0.5597    0.7009
       r9 | 0.2989    0.2981    0.3021    0.3806    0.4605    0.5068    0.6090
      r10 | 0.2285    0.2597    0.2748    0.3637    0.3981    0.4909    0.5889
      r11 | 0.2325    0.2289    0.2696    0.3246    0.3551    0.4426    0.5103
      r12 | 0.2359    0.2351    0.2544    0.3134    0.3474    0.3822    0.4788
          |     c8        c9       c10       c11       c12
----------+-------------------------------------------------
       r8 | 1.0000
       r9 | 0.6714    1.0000
      r10 | 0.5973    0.6325    1.0000
      r11 | 0.5625    0.5756    0.5738    1.0000
      r12 | 0.4999    0.5412    0.5329    0.6428    1.0000
```

As before, we find that the correlation of residuals decreases as the lag increases, but more slowly than an AR(1) process.

◁

▷ Example 3

In this example, we examine injury incidents among 20 airlines in each of 4 years. The data are fictional, and, as a matter of fact, are really from a random-effects model.

```
. use https://www.stata-press.com/data/r19/airacc
. generate lnpm = ln(pmiles)
. xtgee i_cnt inprog, family(poisson) eform offset(lnpm) nolog
GEE population-averaged model                     Number of obs    =     80
Group variable: airline                           Number of groups =     20
Family: Poisson                                   Obs per group:
Link:   Log                                                    min =      4
Correlation: exchangeable                                      avg =    4.0
                                                               max =      4
                                                  Wald chi2(1)     =   5.27
Scale parameter = 1                               Prob > chi2      = 0.0217

------------------------------------------------------------------------------
       i_cnt |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      inprog |   .9059936   .0389528    -2.30   0.022     .8327758    .9856487
       _cons |   .0080065   .0002912  -132.71   0.000     .0074555    .0085981
        lnpm |          1  (offset)
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).
. estat wcorrelation
Estimated within-airline correlation matrix R:

           |        c1         c2         c3         c4
-----------+--------------------------------------------
        r1 |         1
        r2 |  .4606406          1
        r3 |  .4606406   .4606406          1
        r4 |  .4606406   .4606406   .4606406          1
```

Now there are not really enough data here to reliably estimate the correlation without any constraints of structure, but here is what happens if we try:

```
. xtgee i_cnt inprog, family(poisson) eform offset(lnpm) corr(unstructured) nolog
GEE population-averaged model                     Number of obs    =     80
Group and time vars: airline time                 Number of groups =     20
Family: Poisson                                   Obs per group:
Link:   Log                                                    min =      4
Correlation: unstructured                                      avg =    4.0
                                                               max =      4
                                                  Wald chi2(1)     =   0.36
Scale parameter = 1                               Prob > chi2      = 0.5496

------------------------------------------------------------------------------
       i_cnt |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      inprog |   .9791082   .0345486    -0.60   0.550     .9136826    1.049219
       _cons |   .0078716   .0002787  -136.82   0.000     .0073439    .0084373
        lnpm |          1  (offset)
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).
```

```
. estat wcorrelation

Estimated within-airline correlation matrix R:

             |        c1          c2          c3          c4
  -----------+------------------------------------------------
          r1 |         1
          r2 |  .5700298           1
          r3 |   .716356    .4192126           1
          r4 |  .2383264    .3839863    .3521287           1
```

There is no sensible pattern to the correlations.

We created this dataset from a random-effects Poisson model. We reran our data-creation program and this time had it create 400 airlines rather than 20, still with 4 years of data each. Here are the equal-correlation model and estimated correlation structure:

```
. use https://www.stata-press.com/data/r19/airacc2, clear

. xtgee i_cnt inprog, family(poisson) eform offset(lnpm) nolog

GEE population-averaged model                     Number of obs    =  1,600
Group variable: airline                           Number of groups =    400
Family: Poisson                                   Obs per group:
Link:   Log                                                    min =      4
Correlation: exchangeable                                      avg =    4.0
                                                               max =      4
                                                  Wald chi2(1)     = 111.80
Scale parameter = 1                               Prob > chi2      = 0.0000

------------------------------------------------------------------------------
       i_cnt |        IRR   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      inprog |   .8915304   .0096807   -10.57   0.000     .8727571    .9107076
       _cons |   .0071357   .0000629  -560.57   0.000     .0070134    .0072601
        lnpm |          1  (offset)
------------------------------------------------------------------------------

Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).

. estat wcorrelation

Estimated within-airline correlation matrix R:

             |        c1          c2          c3          c4
  -----------+------------------------------------------------
          r1 |         1
          r2 |  .5291707           1
          r3 |  .5291707    .5291707           1
          r4 |  .5291707    .5291707    .5291707           1
```

The following estimation results assume unstructured correlation:

```
. xtgee i_cnt inprog, family(poisson) corr(unstructured) eform offset(lnpm) nolog
GEE population-averaged model                        Number of obs    =  1,600
Group and time vars: airline time                    Number of groups =    400
Family: Poisson                                      Obs per group:
Link:   Log                                                       min =      4
Correlation: unstructured                                         avg =    4.0
                                                                  max =      4
                                                     Wald chi2(1)     = 113.43
Scale parameter = 1                                  Prob > chi2      = 0.0000

------------------------------------------------------------------------------
       i_cnt |        IRR   Std. err.       z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      inprog |   .8914155   .0096208   -10.65   0.000     .8727572    .9104728
       _cons |   .0071402   .0000628  -561.50   0.000     .0070181    .0072645
        lnpm |          1  (offset)
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate (conditional on zero random
      effects).

. estat wcorrelation
Estimated within-airline correlation matrix R:

      |        c1          c2          c3          c4
------+-----------------------------------------------
   r1 |         1
   r2 |  .4733189           1
   r3 |  .5240576    .5748868           1
   r4 |  .5139748    .5048895    .5840707           1
```

The equal-correlation model estimated a fixed correlation of 0.5292, and above we have correlations ranging between 0.4733 and 0.5841 with little pattern in their structure.

◁

Also see

[XT] **xtgee** — GEE population-averaged panel-data models

[U] 20 Estimation and postestimation commands

For suggested citations, see the FAQ on citing Stata documentation.