**Postestimation commands**

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

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat wcorrelation</code></td>
<td>estimated matrix of the within-group correlations</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

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

*`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]
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>predicted value of <em>depvar</em>; considers the offset() or exposure(); the default</td>
</tr>
<tr>
<td>rate</td>
<td>predicted value of <em>depvar</em></td>
</tr>
<tr>
<td>pr(<em>n</em>)</td>
<td>probability Pr((y_{it} = n)) for <em>family</em>(poisson) link(log)</td>
</tr>
<tr>
<td>pr(<em>a</em>,<em>b</em>)</td>
<td>probability Pr((a \leq y_{it} \leq b)) for <em>family</em>(poisson) link(log)</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
<tr>
<td>score</td>
<td>first derivative of the log likelihood with respect to (x_{it}\beta)</td>
</tr>
</tbody>
</table>

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

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

pr\( (\cdot, b) \) 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, \( u_{it} = \partial \ln L(x_{it}\beta)/\partial(x_{it}\beta) \).

nooffset is relevant only if you specified offset(\textit{varname}), exposure(\textit{varname}), family(binomial #), or family(binomial \textit{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**

\textit{margins} estimates margins of response for predicted values, probabilities, and linear predictions.

**Menu for margins**

Statistics > Postestimation

**Syntax for margins**

\begin{verbatim}
margins [marginlist] [ , options ]
margins [marginlist], predict(statistic ...) [predict(statistic ...) ...] [options]
\end{verbatim}

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

Statistics not allowed with \textit{margins} are functions of stochastic quantities other than \texttt{e(b)}.

For the full syntax, see [R] \textit{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

stata.com

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/r17/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   c7   c8   c9
----------|------|------|------|------|------|------|------|------|------|
r1        |  1.00 1.00 1.00 | 1.00 1.00 1.00 1.00 1.00 1.00
r2        | .4851356 1.00 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r3        | .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r4        | .4851356 .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r5        | .4851356 .4851356 .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r6        | .4851356 .4851356 .4851356 .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r7        | .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r8        | .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 1.00 | .4851356 1.00 1.00 1.00 1.00 1.00
r9        | .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 .4851356 1.00 | 1.00 1.00 1.00 1.00 1.00 1.00
```

```
```
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 \texttt{xtgee} is 0.4851. (\texttt{xtreg, re}, by comparison, reports 0.5141; see the \texttt{xtreg} command in example 2 of \texttt{[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 \texttt{correlation(unstructured)}, although we can abbreviate the option.

\begin{verbatim}
.xtgee ln_w grade age c.age#c.age, corr(unstr) nolog
\end{verbatim}

\begin{verbatim}
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: min =  1
Correlation: unstructured
avg =  4.1
max =  9
Wald chi2(3)   = 2405.20
Scale parameter = .1418513
Prob > chi2    =  0.0000
\end{verbatim}

\begin{verbatim}
| 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  |
\end{verbatim}

\begin{verbatim}
.estat wcorrelation
Estimated within-idcode correlation matrix R:
\end{verbatim}

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4354838</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.4280248</td>
<td>.5597329</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.3772342</td>
<td>.5012129</td>
<td>.5475113</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>.4031433</td>
<td>.5301403</td>
<td>.502668</td>
<td>.6216227</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r6</td>
<td>.3663686</td>
<td>.4519138</td>
<td>.4783186</td>
<td>.5685009</td>
<td>.7306005</td>
<td>1</td>
</tr>
<tr>
<td>r7</td>
<td>.2819915</td>
<td>.3605743</td>
<td>.3918118</td>
<td>.4012104</td>
<td>.4642601</td>
<td>.50219</td>
</tr>
<tr>
<td>r8</td>
<td>.3162028</td>
<td>.3445668</td>
<td>.4285424</td>
<td>.4389241</td>
<td>.4696792</td>
<td>.5222537</td>
</tr>
<tr>
<td>r9</td>
<td>.2148737</td>
<td>.3078491</td>
<td>.3337292</td>
<td>.3584013</td>
<td>.4865802</td>
<td>.4613128</td>
</tr>
<tr>
<td></td>
<td>c7</td>
<td>c8</td>
<td>c9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r8</td>
<td>.6475654</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r9</td>
<td>.5791417</td>
<td>.7386958</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

\section*{Example 2}

In example 1 of \texttt{[XT] xtprobit}, we showed a random-effects model of unionization using the \texttt{union} data described in \texttt{[XT] xt}. We performed the estimation using \texttt{xtprobit} but said that we could have used \texttt{xtgee} as well. Here we fit a population-averaged (equal correlation) model for comparison:
. use https://www.stata-press.com/data/r17/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 = .0034666
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
correlation matrix R:
max = 12
Wald chi2(6) = 242.57

Scale parameter = 1 Prob > chi2 = 0.0000

|        | Coefficient | Std. err. | z    | P>|z| | [95% conf. interval] |
|--------|-------------|-----------|------|------|---------------------|
| union  |             |           |      |      |                     |
| age    | .0089699    | .0053208  | 1.69 | 0.092| -.0014586           |
| grade  | .0333174    | .0062352  | 5.34 | 0.000| .0210966            |
| 1.not_smsa | -.0715717 | .027551   | -2.60 | 0.009| -.1255551           |
| 1.south | -1.017368   | .207931   | -4.89 | 0.000| -1.424905           |
| year   | -.0062708   | .0055314  | -1.13 | 0.257| -.0171122           |
| south#c.year | .0086294 | .00258    | 3.34 | 0.001| .0035727            |
| _cons  | -.8670997   | .294771   | -2.94 | 0.003| -1.44484            |

Let’s look at the correlation structure and then relax it:
. estat wcorrelation, format(%8.4f)
Estimated within-idcode correlation matrix R:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
<th>c7</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>0.4615</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>0.4615</td>
<td>0.4615</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r6</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>r7</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>1.0000</td>
</tr>
<tr>
<td>r8</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
</tr>
<tr>
<td>r9</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
</tr>
<tr>
<td>r10</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
</tr>
<tr>
<td>r11</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
</tr>
<tr>
<td>r12</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
<td>0.4615</td>
</tr>
<tr>
<td>c8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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(unstr) 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

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

. estat wcorrelation, format(%8.4f)
Estimated within-idcode correlation matrix R:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>c6</th>
<th>c7</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>0.6667</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>0.6151</td>
<td>0.6523</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>0.5268</td>
<td>0.5717</td>
<td>0.6101</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r5</td>
<td>0.3309</td>
<td>0.3669</td>
<td>0.4005</td>
<td>0.4783</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r6</td>
<td>0.3000</td>
<td>0.3706</td>
<td>0.4237</td>
<td>0.4562</td>
<td>0.6426</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>r7</td>
<td>0.2995</td>
<td>0.3568</td>
<td>0.3851</td>
<td>0.4279</td>
<td>0.4931</td>
<td>0.6384</td>
<td>1.0000</td>
</tr>
<tr>
<td>r8</td>
<td>0.2759</td>
<td>0.3021</td>
<td>0.3225</td>
<td>0.3751</td>
<td>0.4682</td>
<td>0.5597</td>
<td>0.7009</td>
</tr>
<tr>
<td>r9</td>
<td>0.2989</td>
<td>0.2981</td>
<td>0.3021</td>
<td>0.3806</td>
<td>0.4605</td>
<td>0.5068</td>
<td>0.6090</td>
</tr>
<tr>
<td>r10</td>
<td>0.2285</td>
<td>0.2597</td>
<td>0.2748</td>
<td>0.3637</td>
<td>0.3981</td>
<td>0.4909</td>
<td>0.5889</td>
</tr>
<tr>
<td>r11</td>
<td>0.2325</td>
<td>0.2289</td>
<td>0.2696</td>
<td>0.3246</td>
<td>0.3551</td>
<td>0.4426</td>
<td>0.5103</td>
</tr>
<tr>
<td>r12</td>
<td>0.2359</td>
<td>0.2351</td>
<td>0.2544</td>
<td>0.3134</td>
<td>0.3474</td>
<td>0.3822</td>
<td>0.4788</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c8</td>
<td>c9</td>
<td>c10</td>
<td>c11</td>
<td>c12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r8</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r9</td>
<td>0.6714</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r10</td>
<td>0.5973</td>
<td>0.6325</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r11</td>
<td>0.5625</td>
<td>0.5756</td>
<td>0.5738</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r12</td>
<td>0.4999</td>
<td>0.5412</td>
<td>0.5329</td>
<td>0.6428</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

```stata
use https://www.stata-press.com/data/r17/airacc
generate lnpm = ln(pmiles)
taxgee 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

|         | 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).

```stata
estat wcorrelation
```

Estimated within-airline correlation matrix R:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4606406</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.4606406</td>
<td>.4606406</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.4606406</td>
<td>.4606406</td>
<td>.4606406</td>
<td>1</td>
</tr>
</tbody>
</table>

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:

```stata
xtgee i_cnt inprog, family(poisson) eform offset(lnpm) corr(unstr) 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

|         | 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:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>0.5700298</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>0.716356</td>
<td>0.4192126</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>0.2383264</td>
<td>0.3839863</td>
<td>0.3521287</td>
<td>1</td>
</tr>
</tbody>
</table>

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/r17/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

|     | 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:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>0.5291707</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>0.5291707</td>
<td>0.5291707</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>0.5291707</td>
<td>0.5291707</td>
<td>0.5291707</td>
<td>1</td>
</tr>
</tbody>
</table>
The following estimation results assume unstructured correlation:

```
xtgee i_cnt inprog, family(poisson) corr(unstr) 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
Link: Log
Correlation: unstructured
Obs per group:
\[
\begin{align*}
\text{min} & = 4 \\
\text{avg} & = 4.0 \\
\text{max} & = 4 \\
\text{Wald chi2(1)} & = 113.43
\end{align*}
\]
Scale parameter = 1

|     | IRR      | Std. err. | z     | P>|z|    | [95% conf. interval] |
|-----|----------|-----------|-------|--------|----------------------|
| i_cnt |          |           |       |        |                      |
| 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:

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>.4733189</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3</td>
<td>.5240576</td>
<td>.5748868</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>r4</td>
<td>.5139748</td>
<td>.5048895</td>
<td>.5840707</td>
<td>1</td>
</tr>
</tbody>
</table>

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 — Fit population-averaged panel-data models by using GEE
[U] 20 Estimation and postestimation commands