Description

The following postestimation commands are available after `zip`:

<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 ic</code></td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</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>estat (svy)</code></td>
<td>postestimation statistics for survey data</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>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td><code>lrtest</code></td>
<td>likelihood-ratio test</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>predictions, residuals, influence statistics, and other diagnostic measures</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>suest</code></td>
<td>seemingly unrelated estimation</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>

1 `forecast` is not appropriate with `svy` estimation results.
2 `lrtest` is not appropriate with `svy` estimation results.
Syntax for predict

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

```
predict [type] { stub* | newvar_reg newvar_inflate } [if] [in] , scores
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>number of events; the default</td>
</tr>
<tr>
<td>ir</td>
<td>incidence rate</td>
</tr>
<tr>
<td>pr</td>
<td>probability of a degenerate zero</td>
</tr>
<tr>
<td>pr(n)</td>
<td>probability ( Pr(y_j = n) )</td>
</tr>
<tr>
<td>pr(a,b)</td>
<td>probability ( Pr(a \leq y_j \leq b) )</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</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.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

```
n, the default, calculates the predicted number of events, which is \( (1 - p_j) \exp(x_j \beta) \) if neither `offset()` nor `exposure()` was specified when the model was fit, where \( p_j \) is the predicted probability of a zero outcome; \( (1 - p_j) \exp \{ (x_j \beta) + \text{offset}_j \} \) if `offset()` was specified; or \( (1 - p_j) \{ \exp(x_j \beta) \times \text{exposure}_j \} \) if `exposure()` was specified.
```

```
ir calculates the incidence rate \( \exp(x_j \beta) \), which is the predicted number of events when exposure is 1. This is equivalent to specifying both the \( n \) and the `nooffset` options.
```

```
pr calculates the probability \( Pr(y_j = 0) \), where this zero was obtained from the degenerate distribution \( F(z_j \gamma) \). If `offset()` was specified within the `inflate()` option, then \( F(z_j \gamma + \text{offset}_j \gamma) \) is calculated.
```

```
pr(n) calculates the probability \( Pr(y_j = n) \), where \( n \) is a nonnegative integer that may be specified as a number or a variable. Note that `pr` is not equivalent to `pr(0)`.
```

```
pr(a,b) calculates the probability \( Pr(a \leq y_j \leq b) \), where \( a \) and \( b \) are nonnegative integers that may be specified as numbers or variables;
```

```
b missing \( (b \geq .) \) means \(+\infty\);
pr(20,.) calculates \( Pr(y_j \geq 20) \);
pr(20,b) calculates \( Pr(y_j \geq 20) \) in observations for which \( b \geq . \) and calculates \( Pr(20 \leq y_j \leq b) \) 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, which is $x_j \beta$ if neither `offset()` nor `exposure()` was specified; $x_j \beta + \text{offset}_j$ if `offset()` was specified; or $x_j \beta + \ln(\text{exposure}_j)$ if `exposure()` was specified; see `nooffset` below.

`stdp` calculates the standard error of the linear prediction.

`nooffset` is relevant only if you specified `offset()` or `exposure()` when you fit the model. It modifies the calculations made by `predict` so that they ignore the offset or exposure variable; the linear prediction is treated as $x_j \beta$ rather than as $x_j \beta + \text{offset}_j$ or $x_j \beta + \ln(\text{exposure}_j)$. Specifying `predict ...`, `nooffset` is equivalent to specifying `predict ...`, `ir`.

`scores` calculates equation-level score variables.

The first new variable will contain $\partial \ln L / \partial (x_j \beta)$.

The second new variable will contain $\partial \ln L / \partial (z_j \gamma)$.

## Remarks and examples

### Example 1

Continuing with example 1 from [R] `zip`, we will use `predict` to compute the predicted number of fish captured by each individual.

```
. use http://www.stata-press.com/data/r13/fish
. zip count persons livebait, inf(child camper) vuong
        (output omitted)
. predict numfished
        (option n assumed; predicted number of events)
```

`predict` with the `pr` option computes the probability that an individual does not fish.

```
. predict pr, pr
```

On the other hand, `predict` with the `pr(n)` option computes the probability of catching $n$ fish; particularly, the probability of catching zero fish will be

```
. predict pr0, pr(0)
. list pr pr0 in 1
    +--------+--------+
   | pr    | pr0    |
   +--------+--------+
   | .3793549| .8609267|
   +--------+--------+
```

Notice that `pr0` is always equal to or greater than `pr`. For example, for the first individual, the probability of not fishing is 0.38; on the other hand, the probability of catching zero fish (0.86) is equal to the sum of the probability of not fishing and the probability of fishing but not catching any fish. `pr0` can be also computed as one minus the probability of catching at least one fish, that is:

```
. predict pr_catch, pr(1,.)
. gen pr0b = 1-pr_catch
```
Methods and formulas

The probabilities calculated using the \texttt{pr} \( n \) option are the probability \( \Pr(y_i = n) \). These are calculated using

\[
\Pr(0|x_i) = \omega_i + (1 - \omega_i) \exp(-\lambda_i)
\]

\[
\Pr(n|x_i) = (1 - \omega_i) \frac{\lambda_i^n \exp(-\lambda_i)}{n!}
\]

for \( n = 1, 2, \ldots \)

where \( \omega_i \) is the probability of obtaining an observation from the degenerate distribution whose mass is concentrated at zero. \( \omega_i \) can be obtained from the \texttt{pr} option.

See Cameron and Trivedi (2013, sec. 4.6) for further details.

Reference


Also see

[R] \texttt{zip} — Zero-inflated Poisson regression

[U] 20 Estimation and postestimation commands