Title

stata.com

zinb postestimation — Postestimation tools for zinb

Postestimation commands

The following postestimation commands are available after `zinb`:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of estimates
`estat ic`	Akaike's and Schwarz's Bayesian information criteria (AIC and BIC)
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estat (svy)`	postestimation statistics for survey data
`estimates`	cataloging estimation results
* `forecast`	dynamic forecasts and simulations
* `hausman`	Hausman's specification test
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
* `lrtest`	likelihood-ratio test
`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 coefficients
`predict`	predictions, residuals, influence statistics, and other diagnostic measures
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`suest`	seemingly unrelated estimation
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

* `forecast`, `hausman`, and `lrtest` are not appropriate with `svy` estimation results.

predict

Description for predict

`predict` creates a new variable containing predictions such as numbers of events, incidence rates, probabilities, linear predictions, and standard errors.

Menu for predict

Statistics > Postestimation

Syntax for predict

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

`predict` [*type*] { *stub*`*` | *newvar*$_{\text{reg}}$ *newvar*$_{\text{inflate}}$ *newvar*$_{\text{lnalpha}}$ } [*if*] [*in*] , `scores`

statistic	Description
Main	
`n`	number of events; the default
`ir`	incidence rate
`pr`	probability of a degenerate zero
`pr(`*n*`)`	probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	probability $\Pr(a \leq y_j \leq b)$
`xb`	linear prediction
`stdp`	standard error of the linear prediction

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

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

`ir` calculates the incidence rate, 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 of a degenerate zero, predicted from the fitted degenerate distribution $F_j = F(\mathbf{z}_j\boldsymbol{\gamma})$. If `offset()` was specified within the `inflate()` option, then $F_j = F(\mathbf{z}_j\boldsymbol{\gamma} + \text{offset}_j^{\gamma})$ 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 $\mathbf{x}_j\boldsymbol{\beta}$ if neither `offset()` nor `exposure()` was specified; $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$ if `offset()` was specified; or $\mathbf{x}_j\boldsymbol{\beta} + \ln(\text{exposure}_j)$ 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 $\mathbf{x}_j\boldsymbol{\beta}$ rather than as $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$ or $\mathbf{x}_j\boldsymbol{\beta} + \ln(\text{exposure}_j)$. Specifying `predict ..., nooffset` is equivalent to specifying `predict ..., ir`.

`scores` calculates equation-level score variables.

The first new variable will contain $\partial \ln L/\partial(\mathbf{x}_j\boldsymbol{\beta})$.

The second new variable will contain $\partial \ln L/\partial(\mathbf{z}_j\boldsymbol{\gamma})$.

The third new variable will contain $\partial \ln L/\partial \ln\alpha$.

margins

Description for margins

`margins` estimates margins of response for number of events, incidence rates, probabilities, and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

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

statistic	Description
`n`	number of events; the default
`ir`	incidence rate
`pr`	probability of a degenerate zero
`pr(`*n*`)`	probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	probability $\Pr(a \leq y_j \leq b)$
`xb`	linear prediction
`stdp`	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**.

Methods and formulas

See *Methods and formulas* in [R] **zinb** for the model definition and notation.

The probabilities calculated using the `pr(`*n*`)` option are the probability $\Pr(y_j = n)$. These are calculated using

$$\Pr(y_j = 0|\mathbf{x}_j, \mathbf{z}_j) = F_j + (1 - F_j)\, p_2(0|\mathbf{x}_j)$$

$$\Pr(y_j = n|\mathbf{x}_j, \mathbf{z}_j) = (1 - F_j)\, p_2(n|\mathbf{x}_j) \qquad \text{for } n = 1, 2, \ldots$$

where F_j is the probability of obtaining an observation from the degenerate distribution whose mass is concentrated at zero, and $p_2(n|\mathbf{x}_j)$ is the probability of $y_j = n$ from the nondegenerate, negative binomial distribution. F_j can be obtained by using the `pr` option.

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

References

Cameron, A. C., and P. K. Trivedi. 2013. *Regression Analysis of Count Data*. 2nd ed. New York: Cambridge University Press.

Manjón, M., and O. Martínez. 2014. The chi-squared goodness-of-fit test for count-data models. *Stata Journal* 14: 798–816.

Also see

[R] **zinb** — Zero-inflated negative binomial regression

[U] 20 Estimation and postestimation commands