tnbreg postestimation — Postestimation tools for tnbreg

Postestimation commands

The following postestimation commands are available after `tnbreg`:

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, conditional means, probabilities, conditional probabilities, linear predictions, and standard errors.

Menu for predict

Statistics > Postestimation

Syntax for predict

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

predict [*type*] { *stub** | $newvar_{reg}$ $newvar_{disp}$ } [*if*] [*in*] , scores

statistic	Description
Main	
n	number of events; the default
ir	incidence rate
cm	conditional mean, $E(y_j \mid y_j > \tau_j)$
pr(*n*)	probability $\Pr(y_j = n)$
pr(*a*,*b*)	probability $\Pr(a \leq y_j \leq b)$
cpr(*n*)	conditional probability $\Pr(y_j = n \mid y_j > \tau_j)$
cpr(*a*,*b*)	conditional probability $\Pr(a \leq y_j \leq b \mid y_j > \tau_j)$
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 $\exp(\mathbf{x}_j\boldsymbol{\beta})$ if neither offset() nor exposure() was specified when the model was fit; $\exp(\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j)$ if offset() was specified; or $\exp(\mathbf{x}_j\boldsymbol{\beta}) \times \text{exposure}_j$ if exposure() was specified.

ir calculates the incidence rate $\exp(\mathbf{x}_j\boldsymbol{\beta})$, which is the predicted number of events when exposure is 1. This is equivalent to specifying both the n and the nooffset options.

cm calculates the conditional mean,

$$E(y_j \mid y_j > \tau_j) = \frac{E(y_j, y_j > \tau_j)}{\Pr(y_j > \tau_j)}$$

where τ_j is the truncation point found in e(llopt).

`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.

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

b missing ($b \ge .$) means $+\infty$;
`pr(20,.)` calculates $\Pr(y_j \ge 20)$;
`pr(20,`*b*`)` calculates $\Pr(y_j \ge 20)$ in observations for which $b \ge .$ and calculates $\Pr(20 \le y_j \le b)$ 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*`)`.

`cpr(`*n*`)` calculates the conditional probability $\Pr(y_j = n \,|\, y_j > \tau_j)$, where τ_j is the truncation point found in `e(llopt)`. n is an integer greater than the truncation point that may be specified as a number or a variable.

`cpr(`*a*`,`*b*`)` calculates the conditional probability $\Pr(a \le y_j \le b \,|\, y_j > \tau_j)$, where τ_j is the truncation point found in `e(llopt)`. The syntax for this option is analogous to that used for `pr(`*a*`,`*b*`)` except that a must be greater than the truncation point.

`xb` calculates the linear prediction, which is $\mathbf{x}_j\boldsymbol{\beta}$ if neither `offset()` nor `exposure()` was specified when the model was fit; $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j$ 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 $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j$ 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 $\partial \ln L / \partial(\ln\alpha)$ for `dispersion(mean)`.

The second new variable will contain $\partial \ln L / \partial(\ln\delta)$ for `dispersion(constant)`.

margins

Description for margins

`margins` estimates margins of response for numbers of events, incidence rates, conditional means, probabilities, conditional 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
`cm`	conditional mean, $E(y_j \mid y_j > \tau_j)$
`pr(`*n*`)`	probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	probability $\Pr(a \le y_j \le b)$
`cpr(`*n*`)`	conditional probability $\Pr(y_j = n \mid y_j > \tau_j)$
`cpr(`*a*`,`*b*`)`	conditional probability $\Pr(a \le y_j \le b \mid y_j > \tau_j)$
`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

In the following formulas, we use the same notation as in [R] **tnbreg**.

Methods and formulas are presented under the following headings:

Mean-dispersion model
Constant-dispersion model

Mean-dispersion model

The equation-level scores are given by

$$\text{score}(\mathbf{x}\boldsymbol{\beta})_j = p_j(y_j - \mu_j) - \frac{p_j^{(m+1)}\mu_j}{\Pr(Y > \tau_j \,|\, p_j, m)}$$

$$\text{score}(\omega)_j = -\,m\left\{\frac{\alpha(\mu_j - y_j)}{1+\alpha\mu_j} - \ln(1+\alpha\mu_j) + \psi(y_j+m) - \psi(m)\right\} - \frac{p_j^m}{\Pr(Y > \tau_j \,|\, p_j, m)}\{m\ln(p_j) + \mu_j p_j\}$$

where $\omega_j = \ln\alpha_j$, $\psi(z)$ is the digamma function, and τ_j is the truncation point found in `e(llopt)`.

Constant-dispersion model

The equation-level scores are given by

$$\text{score}(\mathbf{x}\boldsymbol{\beta})_j = m_j\left\{\psi(y_j+m_j) - \psi(m_j) + \ln(p) + \frac{p^{m_j}\ln(p)}{\Pr(Y > \tau_j \,|\, p, m_j)}\right\}$$

$$\text{score}(\omega)_j = y_j - (y_j+m_j)(1-p) - \text{score}(\mathbf{x}\boldsymbol{\beta})_j - \frac{\mu_j p}{\Pr(Y > \tau_j \,|\, p, m_j)}$$

where $\omega_j = \ln\delta_j$ and τ_j is the truncation point found in `e(llopt)`.

Also see

[R] **tnbreg** — Truncated negative binomial regression

[U] **20 Estimation and postestimation commands**