Title stata.com

zioprobit postestimation — Postestimation tools for zioprobit

Postestimation commands

The following postestimation commands are available after `zioprobit`:

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
`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 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`	probabilities, linear predictions and their SEs, etc.
`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 probabilities, linear predictions, and standard errors.

Menu for predict

Statistics > Postestimation

Syntax for predict

predict [*type*] { *stub** | *newvar* | *newvarlist* } [*if*] [*in*] [, *statistic* outcome(*outcome*) nooffset]

predict [*type*] *stub** [*if*] [*in*] , scores

statistic	Description
Main	
pmargin	marginal probabilities of levels, $\Pr(y_j = h)$; the default
pjoint1	joint probabilities of levels and participation, $\Pr(y_j = h, s_j = 1)$
pcond1	probabilities of levels conditional on participation, $\Pr(y_j = h \vert s_j = 1)$
ppar	probability of participation, $\Pr(s_j = 1)$
pnpar	probability of nonparticipation, $\Pr(s_j = 0)$
xb	linear prediction
xbinfl	linear prediction for inflation equation
stdp	standard error of the linear prediction
stdpinfl	standard error of the linear prediction for inflation equation

If you do not specify outcome(), pmargin, pjoint1, and pcond1 (with one new variable specified) assume outcome(#1).

You specify one or k new variables with pmargin, pjoint1, and pcond1, where k is the number of outcomes.

You specify one new variable with ppar, pnpar, xb, xbinfl, stdp, and stdpinfl.

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

pmargin, the default, calculates the predicted marginal probabilities of outcome levels, $\Pr(y_j = h)$.

pjoint1 calculates the predicted joint probabilities of outcome levels and participation, $\Pr(y_j = h, s_j = 1)$.

pcond1 calculates the predicted probabilities of outcome levels conditional on participation, $\Pr(y_j = h \vert s_j = 1)$.

With pmargin, pjoint1, and pcond1, you can compute predicted probabilities for one or for all outcome levels. When you specify one new variable, predict computes probabilities for the first outcome level. You can specify the outcome(#i) option to obtain probabilities for the ith level. When you specify multiple new variables or a stub, predict computes probabilities for all outcome levels. The behavior of predict with one new variable is equivalent to specifying outcome(#1).

ppar and pnpar calculate the predicted marginal probability of participation [$\Pr(s_j = 1)$] and of nonparticipation [$\Pr(s_j = 0)$], respectively.

In health-related fields, probabilities of participation and nonparticipation are known as probabilities of susceptibility and nonsusceptibility. Similarly to predict after ziologit, you can use options ps and pns to compute these probabilities. Options ps and pns produce identical results to the respective options ppar and pnpar but label new variables as Pr(susceptible) and Pr(nonsusceptible) instead of Pr(participation) and Pr(nonparticipation).

xb calculates the linear prediction for the ordered probit equation, which is $\mathbf{x}_j\boldsymbol{\beta}$ if offset() was not specified with zioprobit and is $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$ if offset() was specified.

xbinfl calculates the linear prediction for the inflation equation, which is $\mathbf{z}_j\boldsymbol{\gamma}$ if offset() was not specified in inflate() and is $\mathbf{z}_j\boldsymbol{\gamma} + \text{offset}_j^{\gamma}$ if offset() was specified in inflate().

stdp calculates the standard error of the linear prediction for the ordered probit equation.

stdpinfl calculates the standard error of the linear prediction for the inflation equation.

outcome(*outcome*) specifies the outcome for which predicted probabilities are to be calculated. outcome() should contain either one value of the dependent variable or one of #1, #2, ..., with #1 meaning the first category of the dependent variable, #2 meaning the second category, etc. outcome() is allowed only with pmargin, pjoint1, and pcond1.

nooffset is relevant only if you specified offset(*varname*) with zioprobit or within the inflate() option. It modifies the calculations made by predict so that they ignore the offset variable; that is, the linear prediction for the main regression equation is treated as $\mathbf{x}_j\boldsymbol{\beta}$ rather than as $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$ and the linear prediction for the inflation equation is treated as $\mathbf{z}_j\boldsymbol{\gamma}$ rather than as $\mathbf{z}_j\boldsymbol{\gamma} + \text{offset}_j^{\gamma}$.

scores calculates equation-level score variables.

The first new variable will contain $\partial \ln L/\partial(\mathbf{x}_j\boldsymbol{\beta})$. In the absence of independent variables in the main equation, this variable is not stored.

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

When the dependent variable takes k different values, the third new variable through new variable $k+1$ will contain $\partial \ln L/\partial(\kappa_h)$ for $h = 0, 1, \ldots, k-2$.

margins

Description for margins

`margins` estimates margins of response for probabilities and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

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

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

statistic	Description
default	marginal probabilities for each outcome
pmargin	marginal probabilities of levels, $\Pr(y_j = h)$; the default
pjoint1	joint probabilities of levels and participation, $\Pr(y_j = h, s_j = 1)$
pcond1	probabilities of levels conditional on participation, $\Pr(y_j = h \vert s_j = 1)$
ppar	probability of participation, $\Pr(s_j = 1)$
pnpar	probability of nonparticipation, $\Pr(s_j = 0)$
xb	linear prediction
xbinfl	linear prediction for inflation equation
stdp	not allowed with `margins`
stdpinfl	not allowed with `margins`

pmargin, pjoint1, and pcond1 default to the first outcome.

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

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

Remarks and examples

stata.com

Various sets of predictions and marginal effects may be of interest for the ZIOP model. For instance, we may want to investigate the marginal effects of a covariate on the probability of participation, or on the probabilities for levels of consumption conditional on participation, or on the overall probabilities for different consumption levels. We explore these options in greater detail in the following examples.

▷ Example 1: Average marginal effects on probability of nonparticipation

In example 1 of [R] **zioprobit**, we fit a model for level of cigarette consumption.

```
. use https://www.stata-press.com/data/r17/tobacco
(Fictional tobacco consumption data)
. zioprobit tobacco education income i.female age,
> inflate(education income i.parent age i.female i.religion)
  (output omitted )
```

We can use `margins` to estimate the expected marginal effect of gender for individuals with a college degree (17 years of education) and a smoking parent on the probability of nonparticipation (being a genuine nonsmoker). To do this, we specify `predict(pnpar)` with `margins` as follows:

```
. margins, predict(pnpar) dydx(female) at(education = 17 parent = 1)
Average marginal effects                          Number of obs = 15,000
Model VCE: OIM
Expression: Pr(nonparticipation), predict(pnpar)
dy/dx wrt:  1.female
At: education = 17
    parent    =  1
```

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
female						
Female	.0855995	.0100239	8.54	0.000	.0659531	.105246

```
Note: dy/dx for factor levels is the discrete change from the base level.
```

Women with a college degree and a smoking parent are expected to have about an 8.5% higher chance of being genuine nonsmokers than do men.

◁

▷ Example 2: Predicted probabilities of zero-valued outcomes

In example 1 of [R] **zioprobit**, we found that the coefficient on `income` was positive in the level equation but negative in the participation equation. In the case of our tobacco consumption example, economic theory offers a reasonable interpretation for this. Higher income may act as an indicator for health awareness, which accounts for its association with an increased probability of being a genuine nonsmoker. However, if cigarettes are a normal good—that is, something for which demand increases when income increases—then smokers with higher income should have a lower probability of having zero consumption at the time of the survey.

We first consider the effect of `income` at six prespecified values ranging from \$10,000 to \$60,000 on the probability of being a genuine nonsmoker (nonparticipation). Because we know the values 1 to 6 correspond to income in tens of thousands, we conserve space and suppress the default legend by using the `noatlegend` option.

```
. margins, predict(pnpar) at(income = (1/6)) noatlegend

Predictive margins                                      Number of obs = 15,000
Model VCE: OIM

Expression: Pr(nonparticipation), predict(pnpar)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1727928   .0056963    30.33   0.000     .1616283    .1839573
          2  |   .1944564   .0055236    35.20   0.000     .1836304    .2052824
          3  |   .2176787   .0052981    41.09   0.000     .2072945    .2280629
          4  |   .2424141   .0050413    48.09   0.000     .2325333    .2522948
          5  |   .2685948   .0047857    56.12   0.000      .259215    .2779745
          6  |   .2961309   .0045767    64.70   0.000     .2871607    .3051011
------------------------------------------------------------------------------
```

The probability of being a genuine nonsmoker increases with `income`. For instance, for individuals who earn \$10,000 a year, the expected increase in the probability of being a genuine nonsmoker is about $(0.1945 - 0.1728) \times 100\% = 2.17\%$ if they earn an additional \$10,000.

We next investigate the effect of `income` on the joint probability of being a smoker (participation equals 1) with zero consumption. We do this by specifying `predict(pjoint1 outcome(0))` with `margins`:

```
. margins, predict(pjoint1 outcome(0)) at(income = (1/6)) noatlegend

Predictive margins                                      Number of obs = 15,000
Model VCE: OIM

Expression: Pr(tobacco=0, participation=1), predict(pjoint1 outcome(0))

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .5595131   .0040167   139.30   0.000     .5516405    .5673856
          2  |   .5080066   .0037793   134.42   0.000     .5005993    .5154139
          3  |   .4558656   .0035553   128.22   0.000     .4488975    .4628338
          4  |   .4029089   .0033683   119.62   0.000     .3963072    .4095106
          5  |   .3485692   .0032481   107.32   0.000     .3422031    .3549352
          6  |   .2929155     .00317    92.40   0.000     .2867023    .2991287
------------------------------------------------------------------------------
```

For individuals who are smokers, the probability of zero consumption decreases as income increases, suggesting that tobacco is like a normal good for smokers. For example, for individuals who earn \$10,000 a year, earning an additional \$10,000 will decrease their probability of being a smoker with zero consumption by about $(0.5595 - 0.5080) \times 100\% = 5.15\%$.

If we wanted to compute the effect of income on the overall probability of zero consumption instead of the probability of zero consumption among smokers, we would omit `pjoint1` from within the `predict` option.

```
. margins, predict(outcome(0)) at(income = (1/6)) noatlegend
  (output omitted)
```

This version of the `margins` command gives the sum of the probability of nonparticipation and the joint probability of participation with zero consumption.

◁

Methods and formulas

The participation equation is

$$s_j = I(\mathbf{z}_j\boldsymbol{\gamma} + u_{1j} > 0) \qquad j = 1, 2, \ldots, n$$

where s_j is 1 if the jth subject belongs to the participation group (for example, smokers) and is 0 if the subject belongs to the nonparticipation group (for example, genuine nonsmokers), $\mathbf{z}_j$ are the covariates used to model the group membership, $\boldsymbol{\gamma}$ is a vector of coefficients, and u_{1j} is a random-error term following a standard normal distribution.

The ordinal outcome equation is

$$\widetilde{y}_j = \sum_{h=0}^{H} hI\left(\kappa_{h-1} < \mathbf{x}_j\boldsymbol{\beta} + u_{2j} \leq \kappa_h\right)$$

where $\widetilde{y}_j$ is the ordinal outcome conditional on participation, $\mathbf{x}_j$ are the outcome covariates, $\boldsymbol{\beta}$ are the coefficients, and u_{2j} is a random-error term following a standard normal distribution. The observed outcome values are 0, 1, ..., H. $\kappa_0, \kappa_1, \ldots, \kappa_{H-1}$ are real numbers such that $\kappa_i < \kappa_m$ for $i < m$. κ_{-1} is taken as $-\infty$ and κ_H is taken as $+\infty$. We assume that the error terms u_{1j} and u_{2j} are independent. We observe $y_j = s_j\widetilde{y}_j$.

The probability of participation is

$$\Pr\left(s_j = 1|\mathbf{z}_j\right) = \Phi\left(\mathbf{z}_j\boldsymbol{\gamma}\right) \tag{1}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

The probability of nonparticipation is

$$\Pr\left(s_j = 0|\mathbf{z}_j\right) = 1 - \Phi\left(\mathbf{z}_j\boldsymbol{\gamma}\right) \tag{2}$$

The probability of outcome $y_j = h$ given that the jth subject belongs to the participation group is

$$\begin{aligned}\Pr(y_j = h|s_j = 1, \mathbf{x}_j) &= \Pr(\widetilde{y}_j = h|\mathbf{x}_j) \\ &= \Phi\left(\kappa_h - \mathbf{x}_j\boldsymbol{\beta}\right) - \Phi\left(\kappa_{h-1} - \mathbf{x}_j\boldsymbol{\beta}\right)\end{aligned} \tag{3}$$

for $h = 0, 1, \ldots, H$, where $\Phi\left(\kappa_{-1} - \mathbf{x}_j\boldsymbol{\beta}\right) = 0$ and $\Phi\left(\kappa_H - \mathbf{x}_j\boldsymbol{\beta}\right) = 1$.

The joint probability of outcome $y_j = h$ and participation can be expressed as

$$\Pr(y_j = h,\ s_j = 1|\mathbf{z}_j, \mathbf{x}_j) = \Pr\left(s_j = 1|\mathbf{z}_j\right)\Pr(y_j = h|s_j = 1, \mathbf{x}_j)$$

for $h = 0, 1, \ldots, H$, and computed using (1) and (3).

The marginal probabilities of the outcome y_j are

$$\begin{aligned}\Pr(y_j = 0|\mathbf{z}_j, \mathbf{x}_j) &= \Pr\left(s_j = 0|\mathbf{z}_j\right) + \Pr\left(s_j = 1|\mathbf{z}_j\right)\Pr(\widetilde{y}_j = 0|\mathbf{x}_j) \\ \Pr(y_j = h|\mathbf{z}_j, \mathbf{x}_j) &= \Pr\left(s_j = 1|\mathbf{z}_j\right)\Pr(\widetilde{y}_j = h|\mathbf{x}_j) \qquad h = 1, 2, \ldots, H-1 \\ \Pr(y_j = H|\mathbf{z}_j, \mathbf{x}_j) &= \Pr\left(s_j = 1|\mathbf{z}_j\right)\Pr(\widetilde{y}_j = H|\mathbf{x}_j)\end{aligned}$$

and can be computed using (1), (2), and (3).

If the `offset()` option is specified with `zioprobit`, $\mathbf{x}_j\boldsymbol{\beta}$ is replaced with $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$. If the `offset()` option is specified within the `inflate()` option, $\mathbf{z}_j\boldsymbol{\gamma}$ is replaced with $\mathbf{z}_j\boldsymbol{\gamma} + \text{offset}_j^{\gamma}$.

Also see

[R] **zioprobit** — Zero-inflated ordered probit regression

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