Title

stata.com

ivtobit postestimation — Postestimation tools for ivtobit

Postestimation commands

The following postestimation commands are available after `ivtobit`:

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; not available with two-step estimator
`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

* `estat ic`, `forecast`, and `suest` are not appropriate after `ivtobit, twostep`.

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

predict

Description for predict

`predict` creates a new variable containing predictions such as linear predictions, standard errors, probabilities, and expected values.

Menu for predict

Statistics > Postestimation

Syntax for predict

After ML or twostep

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

After ML

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

statistic	Description
Main	
`xb`	linear prediction; the default
`stdp`	standard error of the linear prediction
`stdf`	standard error of the forecast; not available with two-step estimator
`pr(`*a*`,`*b*`)`	$\Pr(a < y_j < b)$ accounting for endogeneity; not available with two-step estimator
`e(`*a*`,`*b*`)`	$E(y_j \mid a < y_j < b)$ accounting for endogeneity; not available with two-step estimator
`ystar(`*a*`,`*b*`)`	$E(y_j^*)$, $y_j = \max\{a, \min(y_j, b)\}$ accounting for endogeneity; not available with two-step estimator

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

`stdf` is not allowed with `svy` estimation results.

where *a* and *b* may be numbers or variables; *a* missing ($a \geq .$) means $-\infty$, and *b* missing ($b \geq .$) means $+\infty$; see **[U] 12.2.1 Missing values**.

Options for predict

Main

`xb`, the default, calculates the linear prediction.

`stdp` calculates the standard error of the linear prediction. It can be thought of as the standard error of the predicted expected value or mean for the observation's covariate pattern. The standard error of the prediction is also referred to as the standard error of the fitted value.

`stdf` calculates the standard error of the forecast, which is the standard error of the point prediction for 1 observation. It is commonly referred to as the standard error of the future or forecast value. By construction, the standard errors produced by `stdf` are always larger than those produced by `stdp`; see *Methods and formulas* in [R] **regress postestimation**. `stdf` is not available with the two-step estimator.

`pr(`*a*`,`*b*`)` calculates $\Pr(a < \mathbf{x}_j\mathbf{b} + u_j < b)$, the probability that $y_j|\mathbf{x}_j$ would be observed in the interval (a, b) accounting for endogeneity.

a and *b* may be specified as numbers or variable names; *lb* and *ub* are variable names;
`pr(20,30)` calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < 30)$;
`pr(`*lb*`,`*ub*`)` calculates $\Pr(lb < \mathbf{x}_j\mathbf{b} + u_j < ub)$; and
`pr(20,`*ub*`)` calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < ub)$.

a missing ($a \geq .$) means $-\infty$; `pr(.,30)` calculates $\Pr(-\infty < \mathbf{x}_j\mathbf{b} + u_j < 30)$;
`pr(`*lb*`,30)` calculates $\Pr(-\infty < \mathbf{x}_j\mathbf{b} + u_j < 30)$ in observations for which $lb \geq .$
and calculates $\Pr(lb < \mathbf{x}_j\mathbf{b} + u_j < 30)$ elsewhere.

b missing ($b \geq .$) means $+\infty$; `pr(20,.)` calculates $\Pr(+\infty > \mathbf{x}_j\mathbf{b} + u_j > 20)$;
`pr(20,`*ub*`)` calculates $\Pr(+\infty > \mathbf{x}_j\mathbf{b} + u_j > 20)$ in observations for which $ub \geq .$
and calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < ub)$ elsewhere.

`pr(`*a*`,`*b*`)` is not available with the two-step estimator.

`e(`*a*`,`*b*`)` calculates $E(\boldsymbol{x}_j\boldsymbol{b} + u_j \mid a < \boldsymbol{x}_j\boldsymbol{b} + u_j < b)$, the expected value of $y_j|\boldsymbol{x}_j$ conditional on $y_j|\boldsymbol{x}_j$ being in the interval (a, b), meaning that $y_j|\boldsymbol{x}_j$ is truncated. a and b are specified as they are for `pr()`. Endogeneity is accounted for when calculating `e(`*a*`,`*b*`)`. `e(`*a*`,`*b*`)` is not available with the two-step estimator.

`ystar(`*a*`,`*b*`)` calculates $E(y_j^*)$, where $y_j^* = a$ if $\boldsymbol{x}_j\boldsymbol{b} + u_j \leq a$, $y_j^* = b$ if $\boldsymbol{x}_j\boldsymbol{b} + u_j \geq b$, and $y_j^* = \boldsymbol{x}_j\boldsymbol{b} + u_j$ otherwise, meaning that y_j^* is censored. a and b are specified as they are for `pr()`. Endogeneity is accounted for when calculating `ystar(`*a*`,`*b*`)`. `ystar(`*a*`,`*b*`)` is not available with the two-step estimator.

`scores`, not available with `twostep`, calculates equation-level score variables.

For models with one endogenous regressor, five new variables are created.

The first new variable will contain $\partial \ln L/\partial(\boldsymbol{z}_i\boldsymbol{\delta})$.

The second new variable will contain $\partial \ln L/\partial(\boldsymbol{x}_i\boldsymbol{\Pi})$.

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

The fourth new variable will contain $\partial \ln L/\partial \ln\sigma_{u|v}$.

The fifth new variable will contain $\partial \ln L/\partial \ln\sigma_v$.

For models with p endogenous regressors, $p + \{(p+1)(p+2)\}/2 + 1$ new variables are created.

The first new variable will contain $\partial \ln L/\partial(\boldsymbol{z}_i\boldsymbol{\delta})$.

The second through $(p+1)$th new score variables will contain $\partial \ln L/\partial(\boldsymbol{x}_i\boldsymbol{\Pi}_k)$, $k = 1, \ldots, p$, where $\boldsymbol{\Pi}_k$ is the kth column of $\boldsymbol{\Pi}$.

The remaining score variables will contain the partial derivatives of $\ln L$ with respect to s_{11}, s_{21}, …, $s_{p+1,1}$, s_{22}, …, $s_{p+1,2}$, …, $s_{p+1,p+1}$, where $s_{m,n}$ denotes the (m, n) element of the Cholesky decomposition of the error covariance matrix.

margins

Description for margins

`margins` estimates margins of response for linear predictions, probabilities, and expected values.

Menu for margins

Statistics > Postestimation

Syntax for margins

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

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

statistic	Description
`xb`	linear prediction; the default
`pr(`*a*`,`*b*`)`	$\Pr(a < y_j < b)$ accounting for endogeneity; not available with two-step estimator
`e(`*a*`,`*b*`)`	$E(y_j \mid a < y_j < b)$ accounting for endogeneity; not available with two-step estimator
`ystar(`*a*`,`*b*`)`	$E(y_j^*)$, $y_j = \max\{a, \min(y_j, b)\}$ accounting for endogeneity; not available with two-step estimator
`stdp`	not allowed with `margins`
`stdf`	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**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Marginal effects
Obtaining predicted values

Marginal effects

▷ Example 1

We can obtain average marginal effects by using the `margins` command after `ivtobit`. For the labor-supply model of example 1 in [R] **ivtobit**, suppose that we wanted to know the average marginal effects on the woman's expected income, conditional on her income being greater than $10,000.

```
. use http://www.stata-press.com/data/r14/laborsup
. ivtobit fem_inc fem_educ kids (other_inc = male_educ), ll
  (output omitted)
. margins, dydx(*) predict(e(10, .))
Average marginal effects                        Number of obs     =        500
Model VCE    : OIM
Expression   : E(fem_inc|fem_inc>10), predict(e(10, .))
dy/dx w.r.t. : other_inc fem_educ kids male_educ
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
other_inc	-.2251407	.0517915	-4.35	0.000	-.3266503	-.1236311
fem_educ	1.164607	.1460617	7.97	0.000	.8783312	1.450883
kids	-1.303852	.2588412	-5.04	0.000	-1.811171	-.7965323
male_educ	-.3034651	.0309838	-9.79	0.000	-.3641922	-.242738

In our sample, increasing the number of children in the family by one decreases the expected wage by \$1,304 on average (wages in our dataset are measured in thousands of dollars). `male_edu` has no effect because it appears only as an instrument.

◁

Obtaining predicted values

After fitting your model using `ivtobit`, you can obtain the linear prediction and its standard error for both the estimation sample and other samples using the `predict` command. If you used the maximum likelihood estimator, you can also obtain conditional expected values of the observed and latent dependent variables, the standard error of the forecast, and the probability of observing the dependent variable in a specified interval. See **[U] 20 Estimation and postestimation commands** and [R] **predict**.

Methods and formulas

The linear prediction is calculated as $\boldsymbol{z}_i\widehat{\boldsymbol{\delta}}$, where $\widehat{\boldsymbol{\delta}}$ is the estimated value of $\boldsymbol{\delta}$, and $\boldsymbol{z}_i$ and $\boldsymbol{\delta}$ are defined in (1a) of [R] **ivtobit**. Expected values and probabilities are calculated using the same formulas as those used by the standard tobit model. However, instead of evaluating the standard normal cumulative density and probability density at the linear prediction, expected values and probabilities are evaluated at m_i, where m_i is defined in *Methods and formulas* of [R] **ivtobit**. Using m_i instead of $\boldsymbol{z}_i\widehat{\boldsymbol{\delta}}$ in the formulas accounts for endogeneity.

Also see

[R] **ivtobit** — Tobit model with continuous endogenous covariates

[U] 20 Estimation and postestimation commands