Title stata.com

ivtobit postestimation — Postestimation tools for ivtobit

Postestimation commands

The following postestimation commands are of special interest after `ivtobit`:

Command	Description
`estat correlation`	report the correlation matrix of the errors of the dependent variable and the endogenous variables
`estat covariance`	report the covariance matrix of the errors of the dependent variable and the endogenous variables

These commands are not appropriate after the two-step estimator or with `svy` estimation results.

The following standard postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of estimates
* `estat ic`	Akaike's, consistent Akaike's, corrected Akaike's, and Schwarz's Bayesian information criteria (AIC, CAIC, AICc, 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; 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`	linear predictions and their SEs, probabilities, 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

*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 structural functions, linear predictions, standard errors, probabilities, and expected values.

Menu for predict

Statistics > Postestimation

Syntax for predict

After ML

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

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

After twostep

predict [*type*] *newvar* [*if*] [*in*] [, *twostep_statistic*]

statistic	Description
Main	
xb	linear prediction excluding endogeneity; the default
mean	linear prediction accounting for endogeneity
stdp	standard error of the linear prediction
stdf	standard error of the forecast
pr(*a*,*b*)	$\Pr(a < y_j < b)$ accounting for endogeneity
e(*a*,*b*)	$E(y_j \mid a < y_j < b)$ accounting for endogeneity
ystar(*a*,*b*)	$E(y_j^*)$, $y_j^* = \max\{a, \min(y_j, b)\}$ accounting for endogeneity

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.

asfmethod	Description
Main	
asf	average structural function; the default
fixedasf	fixed average structural function

twostep_statistic	Description
Main	
xb	linear prediction; the default
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

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

`mean` calculates the linear prediction. Results depend on how the endogeneity complication is handled, which is determined by the `asf` or `fixedasf` option. `mean` is not available with the two-step estimator.

`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**.

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

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

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

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

Results depend on how the endogeneity complication is handled, which is determined by the `asf` or `fixedasf` option.

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

`e(`*a*`,`*b*`)` calculates $E(y_j \mid a < y_j < b)$, the expected value of $y_j|\boldsymbol{z}_j$ conditional on $y_j|\boldsymbol{z}_j$ being in the interval (a, b), meaning that $y_j|\boldsymbol{z}_j$ is truncated. a and b are specified as they are for `pr()`. Results depend on how the endogeneity complication is handled, which is determined by the `asf` or `fixedasf` option. `e(`*a*`,`*b*`)` is not available with the two-step estimator.

`ystar(`*a*`,`*b*`)` calculates $E(y_j^*)$, where $y_j^* = a$ if $\boldsymbol{z}_i\boldsymbol{\delta} + u_j \leq a$, $y_j^* = b$ if $\boldsymbol{z}_i\boldsymbol{\delta} + u_j \geq b$, and $y_j^* = \boldsymbol{z}_i\boldsymbol{\delta} + u_j$ otherwise, meaning that y_j^* is censored. a and b are specified as they are for `pr()`. Results depend on how the endogeneity complication is handled, which is determined by the `asf` or `fixedasf` option. `ystar(`*a*`,`*b*`)` is not available with the two-step estimator.

`asf` and `fixedasf` determine how the average structural function (ASF) of the specified statistic is computed. These options are not allowed with `xb`, `stdp`, or `stdf`.

`asf` is the default for the ML estimator when the `mean`, `pr(`a`,`b`)`, `e(`a`,`b`)`, or `ystar(`a`,`b`)` statistic is specified. `asf` computes the ASF of the specified statistic. It is the statistic conditional on the errors of the endogenous variable equations. Put another way, it is the statistic accounting for the correlation of the endogenous covariates with the errors of the main equation. Derivatives and contrasts based on `asf` have a structural interpretation. See `margins` for computing derivatives and contrasts.

`fixedasf` calculates a fixed ASF. It is the specified statistic using only the coefficients and variables of the outcome equation. `fixedasf` does not condition on the errors of the endogenous variable equations. Contrasts between two fixed counterfactuals averaged over the whole sample have a potential-outcome interpretation. Intuitively, it is as if the values of the covariates were fixed at a value exogenously by fiat. See `margins` for computing derivatives and contrasts.

To be clear, derivatives and contrasts between two fixed counterfactuals using the default `asf` option also have a potential-outcome interpretation. And, unlike `fixedasf`, they retain that interpretation when computed over subpopulations for both linear and nonlinear models.

`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 $\partial \ln L/\partial(\boldsymbol{x}_i\boldsymbol{\Pi}_k)$, $k=1,\dots,p$, 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 the $(p+1)(p+2)/2$ ancillary parameters.

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*]

After ML

statistic	Description
Main	
`xb`	linear prediction excluding endogeneity; the default
`mean`	linear prediction accounting for endogeneity
`stdp`	not allowed with `margins`
`stdf`	not allowed with `margins`
`pr(`*a*`,`*b*`)`	$\Pr(a < y_j < b)$ accounting for endogeneity
`e(`*a*`,`*b*`)`	$E(y_j \vert a < y_j < b)$ accounting for endogeneity
`ystar(`*a*`,`*b*`)`	$E(y_j^*)$, $y_j^* = \max\{a, \min(y_j, b)\}$ accounting for endogeneity

After twostep

statistic	Description
Main	
`xb`	linear prediction; the default
`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**.

estat

Description for estat

`estat correlation` displays the correlation matrix of the errors of the dependent variable and the endogenous variables.

`estat covariance` displays the covariance matrix of the errors of the dependent variable and the endogenous variables.

`estat correlation` and `estat covariance` are not allowed after the `ivprobit` two-step estimator.

Menu for estat

Statistics > Postestimation

Syntax for estat

Correlation matrix

estat correlation [, border(*bspec*) left(#) format(%*fmt*)]

Covariance matrix

estat covariance [, border(*bspec*) left(#) format(%*fmt*)]

Options for estat

Main

border(*bspec*) sets border style of the matrix display. The default is border(all).

left(#) sets the left indent of the matrix display. The default is left(2).

format(%*fmt*) specifies the format for displaying the individual elements of the matrix. The default is format(%9.0g).

Remarks and examples

stata.com

Remarks are presented under the following headings:

Marginal effects

Below, we discuss the interpretation of predictions with the `asf` and `fixedasf` options for the ML estimator using `margins`.

The model is defined by two equations. The first is the equation for the outcome of interest, given by

$$y^*_{1i} = \boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i$$

where we do not observe y^*_{1i}; instead, we observe

$$y_{1i} = \begin{cases} a & y^*_{1i} < a \\ y^*_{1i} & a \leq y^*_{1i} \leq b \\ b & y^*_{1i} > b \end{cases}$$

The second is the equation for the endogenous covariates, $\boldsymbol{y}_{2i}$,

$$\boldsymbol{y}_{2i} = \boldsymbol{x}_{1i}\boldsymbol{\Pi}_1 + \boldsymbol{x}_{2i}\boldsymbol{\Pi}_2 + \boldsymbol{v}_i$$

This last equation is the difference between a standard tobit model and the model fit by `ivtobit`. $\boldsymbol{y}_{2i}$ is modeled by an exogenous component, $\boldsymbol{x}_{1i}\boldsymbol{\Pi}_1 + \boldsymbol{x}_{2i}\boldsymbol{\Pi}_2$, and a component that is correlated with u_i and causes the endogeneity problem, $\boldsymbol{v}_i$. The ASF linear prediction conditions on an estimate of $\widehat{\boldsymbol{v}}_i$. It is given by

$$\widehat{m}_i = \widehat{E}\left(y_{1i}|\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i}, y_{2i}, \widehat{\boldsymbol{v}}_i\right)$$
$$\widehat{m}_i = \boldsymbol{y}_{2i}\widehat{\boldsymbol{\theta}}_1 + \boldsymbol{x}_{1i}\widehat{\boldsymbol{\theta}}_2 + \widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{\theta}}_3$$

Because the correlation between $\boldsymbol{v}_i$ and u_i is the problem we intended to address, conditioning on $\boldsymbol{v}_i$ purges the model of endogeneity. Using the ASF, we can get derivatives and contrast. See Wooldridge (2010) and Blundell and Powell (2003) for an in-depth discussion of ASFs and their interpretation.

The fixed ASF, estimated when the `fixedasf` option is specified, has a different interpretation. Suppose we wanted to analyze $\boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i$ at two different values of $\mathbf{y_2}$, the original $\mathbf{y_2}$ and $\mathbf{y_2} + 1$. We want the average difference at these two points for the given values of the other covariates. The values of the covariates are not arrived at via the model; they are fixed by fiat. You can think of them as potential outcomes. The difference of the two values of $\mathbf{y_2}$ is given by

$$\left\{\left(\boldsymbol{y}_{2i} + 1\right)\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i\right\} - \left(\boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i\right)$$

If we average over the distribution of u, we obtain

$$E\left\{(\boldsymbol{y}_{2i}+1)\boldsymbol{\beta}+\boldsymbol{x}_{1i}\gamma\right\}-E\left(\boldsymbol{y}_{2i}\boldsymbol{\beta}+\boldsymbol{x}_{1i}\gamma\right)$$

We do not account for endogeneity because the values of the covariates are given and fixed. If the research question you are pursuing after fitting the model has this interpretation, `fixedasf` gives you an adequate prediction. If, however, you do not want to treat the covariates as fixed, you should account for endogeneity and use `asf` predictions.

▷ Example 1: Obtaining marginal effects

We can obtain average marginal effects by using the `margins` command after `ivtobit`. For the social-media model of example 1 in [R] **ivtobit**, suppose that we wanted to know the average marginal effects on the probability of spending more than 12 hours using social media. Average marginal effects are equivalent to estimating how a change in a covariate affects every individual in our sample and taking the average of these effects. The effect of each covariate is estimated with all other covariates kept at their observed values.

```
. use https://www.stata-press.com/data/r18/smedia
(Fictional data on hours spent on social media)
. ivtobit hsocial i.sphone i.computer i.year (hstudy = tvhours i.pedu), ul(12)
  (output omitted)
. margins, dydx(*) predict(p(12, .))
Average marginal effects                                 Number of obs = 1,324
Model VCE: OIM

Expression: Pr(hsocial>12), predict(p(12, .))
dy/dx wrt:  hstudy 1.sphone 1.computer 2.year 3.year 4.year
```

		Delta-method				
	dy/dx	std. err.	z	P>|z|	[95% conf.	interval]
hstudy	-.1508219	.0044544	-33.86	0.000	-.1595523	-.1420914
1.sphone	.3943071	.0085181	46.29	0.000	.3776119	.4110022
1.computer	.3993489	.0082484	48.42	0.000	.3831824	.4155154
year						
2	.0663908	.0120746	5.50	0.000	.042725	.0900567
3	.1334667	.012488	10.69	0.000	.1089908	.1579426
4	.2343393	.0123729	18.94	0.000	.2100888	.2585897

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

Having a smartphone increases the probability of spending more than 12 hours on social media by 0.39, on average. Any additional study time decreases the probability of spending more than 12 hours using social media by 0.15, on average. The other average marginal effects are interpreted similarly. All effects above have a structural interpretation because we are conditioning on the level of endogeneity. See Wooldridge (2010) and Blundell and Powell (2003) for an in-depth discussion of ASFs and their interpretation.

◁

Obtaining predicted values

After fitting your model with `ivtobit`, you can obtain the linear prediction and its standard error for both the estimation sample and other samples by using the `predict` command. If you used the ML estimator, you can also obtain the linear prediction, the conditional expected values of the observed and latent dependent variables, and the probability of observing the dependent variable in a specified interval—each of these can be computed with an ASF or a fixed ASF interpretation. In addition, with the ML estimator, you can obtain the standard error of the forecast. See **[U] 20 Estimation and postestimation commands** and **[R] predict**.

Stored results

`estat correlation` stores the following results in `r()`:

Matrices
`r(corr)` correlation matrix of the errors

`estat covariance` stores the following results in `r()`:

Matrices
`r(cov)` covariance matrix of the errors

Methods and formulas

The linear prediction is calculated as $\mathbf{z}_i\widehat{\boldsymbol{\delta}}$, where $\widehat{\boldsymbol{\delta}}$ is the estimated value of $\boldsymbol{\delta}$, and $\mathbf{z}_i$ 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, when we use the default `asf` option with `mean`, `pr(`a`,`b`)`, `e(`a`,`b`)`, or `ystar(`a`,`b`)`, instead of evaluating the standard normal cumulative density and probability density at the linear prediction, we evaluate expected values and probabilities at $\widehat{m}_i$, where $\widehat{m}_i$ is defined in *Methods and formulas* of [R] **ivtobit**. Using $\widehat{m}_i$ instead of $\mathbf{z}_i\widehat{\boldsymbol{\delta}}$ in the formulas produces the ASF, which accounts for endogeneity. The fixed ASF, obtained with the `fixedasf` option, evaluates the statistic at $\mathbf{z}_i\widehat{\boldsymbol{\delta}}$.

References

Blundell, R. W., and J. L. Powell. 2003. Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, ed. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 2, 312–357. Cambridge: Cambridge University Press.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

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

[U] 20 Estimation and postestimation commands