ivprobit postestimation — Postestimation tools for ivprobit

Postestimation commands predict margins estat
Remarks and examples Stored results Methods and formulas References
Also see

Postestimation commands

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

Command	Description
`estat classification`	report various summary statistics, including the classification table
`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
`lroc`	compute area under ROC curve and graph the curve
`lsens`	graph sensitivity and specificity versus probability cutoff

These commands are not appropriate after the two-step estimator or the `svy` prefix.

The following standard postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of parameters
* `estat ic`	Akaike's, consistent Akaike's, corrected Akaike's, and Schwarz's Bayesian information criteria (AIC, CAIC, AICc, and BIC, respectively)
`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 parameters
*† `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 parameters
`predict`	linear predictions and their SEs, probabilities, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of parameters
* `suest`	seemingly unrelated estimation

test	Wald tests of simple and composite linear hypotheses
testnl	Wald tests of nonlinear hypotheses

*estat ic, forecast, lrtest, and suest are not appropriate after ivprobit, 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, and probabilities.

Menu for predict

Statistics > Postestimation

Syntax for predict

After ML

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

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
pr	probability of a positive outcome
stdp	standard error of the linear prediction

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.

pr calculates the probability of a positive outcome. Results depend on how the endogeneity complication is handled, which is determined by the asf or fixedasf option. pr is not available with the two-step estimator.

stdp calculates the standard error of the linear prediction.

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

asf is the default for the ML estimator when the pr 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.

rules requests that Stata use any rules that were used to identify the model when making the prediction. By default, Stata calculates missing for excluded observations. rules is not available with the two-step estimator.

asif requests that Stata ignore the rules and the exclusion criteria and calculate predictions for all observations possible using the estimated parameters from the model. asif is not available with the two-step estimator.

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

For models with one endogenous regressor, four 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 $\partial \ln L / \partial \operatorname{atanh} \rho$.

The fourth new variable will contain $\partial \ln L / \partial \ln \sigma$.

For models with p endogenous regressors, $p + \{(p+1)(p+2)\}/2$ 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_{21}, s_{31}, $\ldots$, $s_{p+1,1}$, s_{22}, $\ldots$, $s_{p+1,2}$, $\ldots$, $s_{p+1,p+1}$, 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 and probabilities.

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
`pr`	probability of a positive outcome
`stdp`	not allowed with `margins`

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

Remarks are presented under the following headings:

Marginal effects
Obtaining predicted values

Marginal effects

Below, we discuss the interpretation of the predicted probability, `pr`, 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

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

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

$$y_{1i} = \begin{cases} 0 & y_{1i}^* < 0 \\ 1 & y_{1i}^* \geq 0 \end{cases}$$

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 probit model and the model fit by `ivprobit`. $\boldsymbol{y}_{2i}$ is modeled as 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 predicted probability conditions on an estimate of $\widehat{\boldsymbol{v}}_i$. It is given by

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

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 $1(\boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i > 0)$ at two different values of $\mathbf{y_2}$, the original $\mathbf{y_2}$ and $\mathbf{y_2} + 1$. $1(\cdot)$ is an indicator function that evaluates to 1 if the condition in parentheses is satisfied and 0 otherwise. 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

$$1\left\{\left(\boldsymbol{y}_{2i} + 1\right)\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i > 0\right\} - 1\left(\boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\gamma + u_i > 0\right)$$

If we average over the distribution of u, we obtain

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

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

We can obtain marginal effects by using the `margins` command after `ivprobit`. Continuing with example 1 in [R] **ivprobit**, we calculate the difference in the probability of a woman working, `fem_work`, if `other_inc` increases by 10% versus the probability when `other_inc` is unchanged. The effect we get has an ASF interpretation. The probabilities are estimated conditional on the residual from the endogenous variable. In other words, the computed effects condition on the level of endogeneity in the model. See Wooldridge (2010) for a discussion about the interpretation of the estimates and the computation of marginal effects of probit estimators under endogeneity.

```
. use https://www.stata-press.com/data/r19/laborsup
. ivprobit fem_work fem_educ kids (other_inc = male_educ)
  (output omitted)
. margins, at(other_inc = generate(other_inc))
>          at(other_inc = generate(other_inc*1.10))
>          contrast(at(r) nowald) predict(pr)

Contrasts of predictive margins                         Number of obs = 500
Model VCE: OIM

Expression: Average structural function probabilities, predict(pr)
1._at: other_inc =       other_inc
2._at: other_inc = other_inc*1.10
```

	Contrast	Delta-method std. err.	[95% conf.	interval]
_at (2 vs 1)	-.0762151	.0100472	-.0959073	-.0565229

Here we see that a 10% increase in `other_inc` leads to an average decrease of 0.076 in the probability of `fem_work`. The effect we get has an ASF interpretation. The probabilities are estimated conditional on the residual from the endogenous variable. In other words, the computed effects condition on the level of endogeneity in the model. See Wooldridge (2010) for a discussion about the interpretation of the estimates and the computation of marginal effects of probit estimators under endogeneity.

◁

Obtaining predicted values

After fitting your model with `ivprobit`, you can obtain the linear prediction and its standard error for both the estimation sample and other samples by using the `predict` command; see **[U] 20 Estimation and postestimation commands** and [R] **predict**. If you use the maximum likelihood estimator, you can also obtain the predicted probability or the linear prediction with an ASF or fixed ASF interpretation.

`predict`'s `pr` option calculates the probability of a positive outcome, remembering any rules used to identify the model, and calculates missing for excluded observations. `predict`'s `rules` option uses the rules in predicting probabilities, whereas `predict`'s `asif` option ignores both the rules and the exclusion criteria and calculates probabilities for all possible observations by using the estimated parameters from the model. See *Obtaining predicted values* in [R] **probit postestimation** for an example.

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 $\boldsymbol{z}_i\widehat{\boldsymbol{\delta}}$, where $\widehat{\boldsymbol{\delta}}$ is the estimated value of $\boldsymbol{\delta}$, and $\boldsymbol{z}_i$ and $\boldsymbol{\delta}$ are defined in (1*a*) of [R] **ivprobit**. The probability of a positive outcome is evaluated at m_i, $\Phi(m_i)$, where $\Phi(\cdot)$ is the standard normal distribution function and m_i is defined in *Methods and formulas* of [R] **ivprobit**. The ASF uses $\widehat{m}_i$ instead of $\mathbf{y}_{2i}\widehat{\boldsymbol{\beta}}+\mathbf{x}_{1i}\widehat{\gamma}$ to evaluate $\Phi(\cdot)$ and account for endogeneity in the model. The fixed ASF is evaluated at $\mathbf{y}_{2i}\widehat{\boldsymbol{\beta}}+\mathbf{x}_{1i}\widehat{\gamma}$.

References

Blundell, R. W., and J. L. Powell. 2003. "Endogeneity in nonparametric and semiparametric regression models". In *Theory and Applications: Eighth World Congress*. Advances in Economics and Econometrics, edited by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 2: 312–357. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511610257.011.

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

Also see

[R] **ivprobit** — Probit model with continuous endogenous covariates

[R] **estat classification** — Classification statistics and table

[R] **lroc** — Compute area under ROC curve and graph the curve

[R] **lsens** — Graph sensitivity and specificity versus probability cutoff

[U] 20 Estimation and postestimation commands

For suggested citations, see the FAQ on citing Stata documentation.