Title

stata.com

xtheckman postestimation — Postestimation tools for xtheckman

Postestimation commands

The following postestimation commands are available after `xtheckman`:

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)
`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`	linear predictions, probabilities, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

predict

Description for predict

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

Menu for predict

Statistics > Postestimation

Syntax for predict

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

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

statistic	Description
Main	
`xb`	linear prediction; the default
`xbsel`	linear prediction for selection equation
`pr(`*a*`,`*b*`)`	$\Pr(y_{it} \mid a < y_{it} < b)$
`e(`*a*`,`*b*`)`	$E(y_{it} \mid a < y_{it} < b)$
`ystar(`*a*`,`*b*`)`	$E(y^*_{it})$, $y^*_{it} = \max\{a, \min(y_{it}, b)\}$
`ycond`	$E(y_{it}\|y_{it}$ observed$)$
`psel`	$\Pr(y_{it}$ observed$)$

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

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 $\mathbf{x}_{it}\mathbf{b}$.

`xbsel` calculates the linear prediction for the selection equation.

`pr(`*a*`,`*b*`)` calculates $\Pr(a < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < b)$, the probability that $y_{it}|\mathbf{x}_{it}$ 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 $\Pr(20 < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < 30)$; `pr(`*lb*`,`*ub*`)` calculates $\Pr(lb < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < ub)$; and `pr(20,`*ub*`)` calculates $\Pr(20 < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < ub)$.

a missing ($a \geq .$) means $-\infty$; `pr(.,30)` calculates $\Pr(-\infty < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < 30)$; `pr(`*lb*`,30)` calculates $\Pr(-\infty < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < 30)$ in observations for which $lb \geq .$ and calculates $\Pr(lb < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < 30)$ elsewhere.

b missing ($b \geq .$) means $+\infty$; `pr(20,.)` calculates $\Pr(+\infty > \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} > 20)$; `pr(20,`*ub*`)` calculates $\Pr(+\infty > \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} > 20)$ in observations for which *ub* $\geq .$ and calculates $\Pr(20 < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < ub)$ elsewhere.

`e(`*a*`,`*b*`)` calculates $E(\mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} \mid a < \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it} < b)$, the expected value of $y_{it}|\mathbf{x}_{it}$ conditional on $y_{it}|\mathbf{x}_{it}$ being in the interval (a, b), meaning that $y_{it}|\mathbf{x}_{it}$ is truncated. *a* and *b* are specified as they are for `pr()`.

`ystar(`*a*`,`*b*`)` calculates $E(y^*_{it})$, where $y^*_{it} = a$ if $\mathbf{x}_{it}\mathbf{b}+\nu_{1i}+\epsilon_{1it} \leq a$, $y^*_{it} = b$ if $\mathbf{x}_{it}\mathbf{b}+\nu_{1i}+\epsilon_{1it} \geq b$, and $y^*_{it} = \mathbf{x}_{it}\mathbf{b} + \nu_{1i} + \epsilon_{1it}$ otherwise, meaning that y^*_{it} is not selected. *a* and *b* are specified as they are for `pr()`.

`ycond` calculates the expected value of the dependent variable conditional on the dependent variable being observed, that is, selected; $E(y_{it} \mid y_{it}$ observed$)$.

`psel` calculates the probability of selection (or being observed): $\Pr(y_{it}$ observed$) = \Pr(\mathbf{z}_{it}\boldsymbol{\alpha} + \nu_{2i} + \epsilon_{2it} > 0)$.

`nooffset` is relevant when you specify `offset(`*varname*`)` for `xtheckman`. It modifies the calculations made by `predict` so that they ignore the offset variable; the linear prediction is treated as $\mathbf{x}_{it}\mathbf{b}$ rather than as $\mathbf{x}_{it}\mathbf{b} + \text{offset}_{it}$.

`scores` calculates parameter-level score variables.

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
`xbsel`	linear prediction for selection equation
`pr(`*a*`,`*b*`)`	$\Pr(y_{it} \mid a < y_{it} < b)$
`e(`*a*`,`*b*`)`	$E(y_{it} \mid a < y_{it} < b)$
`ystar(`*a*`,`*b*`)`	$E(y^*_{it})$, $y^*_{it} = \max\{a, \min(y_{it}, b)\}$
`ycond`	$E(y_{it}|y_{it}$ observed$)$
`psel`	$\Pr(y_{it}$ observed$)$

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

The default statistic produced by `predict` after `xtheckman` is the expected value of the dependent variable from the underlying distribution of the regression model. See example 1 of [XT] **xtheckman** for an example where `margins` is used to predict the conditional mean.

Also see

[XT] **xtheckman** — Random-effects regression with sample selection

[U] 20 Estimation and postestimation commands