churdle postestimation — Postestimation tools for churdle

Postestimation commands

The following standard postestimation commands are available after `churdle`:

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
`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

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

predict

Description for predict

`predict` creates a new variable containing predictions such as conditional expectation of *depvar*, residuals, linear predictions, standard errors, and probabilities.

Menu for predict

Statistics > Postestimation

Syntax for predict

predict [*type*] *newvar* [*if*] [*in*] [, *statistic* equation(*eqno*)]

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

statistic	Description
Main	
`ystar`	conditional expectation of *depvar*; the default
`residuals`	residuals
`ystar(`*a*,*b*`)`	$E(y_j^*)$, $y_j^* = \max\{a, \min(y_j, b)\}$
`xb`	linear prediction
`stdp`	standard error of the linear prediction
`pr(`*a*,*b*`)`	$\Pr(a < y_j < b)$
`e(`*a*,*b*`)`	$E(y_j \mid a < y_j < b)$

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**. For `churdle exponential`, b is `.` (missing).

Options for predict

Main

`ystar`, the default, calculates the conditional expectation of the dependent variable.

`residuals` calculates the residuals.

`ystar(`*a*,*b*`)` calculates $E(y_j^*)$. a and b are specified as they are for `pr()`. If a and b are equal to the lower and upper bounds specified in `churdle`, then $E(y_j^*)$ is equivalent to the predicted value of the dependent variable `ystar`.

`xb` calculates the linear prediction.

`stdp` calculates the standard error of the linear prediction.

`pr(`*a*`,`*b*`)` calculates $\Pr(a < y_j < b)$, the probability that $y_j|\mathbf{x}_i$ 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 $\Pr(20 < y_j < 30)$; `pr(`*lb*`,` *ub*`)` calculates $\Pr(lb < y_j < ub)$; and `pr(20,` *ub*`)` calculates $\Pr(20 < y_j < ub)$.

a missing ($a \geq$ `.`) means `ll`; `pr(.,30)` calculates $\Pr(ll < y_j < 30)$; `pr(`*lb*`,30)` calculates $\Pr(ll < y_j < 30)$ in observations for which $lb \geq$ `.` and calculates $\Pr(lb < y_j < 30)$ elsewhere.

b missing ($b \geq$ `.`) means ∞; `pr(20,.)` calculates $\Pr(\infty > y_j > 20)$; `pr(20,`*ub*`)` calculates $\Pr(\infty > y_j > 20)$ in observations for which $ub \geq$ `.` and calculates $\Pr(ub > y_j > 20)$ elsewhere. For `churdle linear`, *ul* is ∞.

`e(`*a*`,`*b*`)` calculates $E(y_j \mid a < y_j < b)$, the expected value of $y_j|\mathbf{x}_j$ conditional on $y_j|\mathbf{x}_j$ being in the interval (a, b), meaning that $y_j|\mathbf{x}_j$ is bounded. *a* and *b* are specified as they are for `pr()`.

`equation(`*eqno*`)` specifies the equation for which predictions are to be made for the `xb` and `stdp` options. `equation()` should contain either one equation name or one of `#1`, `#2`, ... with `#1` meaning the first equation, `#2` meaning the second equation, etc.

If you do not specify `equation()`, the results are the same as if you specified `equation(# 1)`.

`scores` calculates the equation-level score variables. If you specify one new variable, the scores for the latent-variable equation are computed. If you specify a variable list, the scores for each equation are calculated. The following scores may be obtained:

the first new variable will contain $\partial \ln L/\partial(\mathbf{x}_j\boldsymbol{\beta})$,

the second new variable will contain $\partial \ln L/\partial(\mathbf{z}_j\gamma_{ll})$,

the third new variable will contain $\partial \ln L/\partial(\mathbf{z}_j\gamma_{ul})$,

the fourth new variable will contain $\partial \ln L/\partial(\sigma)$,

the fifth new variable will contain $\partial \ln L/\partial(\sigma_{ll})$, and

the sixth new variable will contain $\partial \ln L/\partial(\sigma_{ul})$.

margins

Description for margins

margins estimates margins of response for conditional expectations, linear predictions, and probabilities.

Menu for margins

Statistics > Postestimation

Syntax for margins

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

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

statistic	Description
Main	
ystar	conditional expectation of *depvar*; the default
ystar(*a*,*b*)	$E(y_j^*)$, $y_j^* = \max\{a, \min(y_j, b)\}$; for churdle exponential b is .
xb	linear prediction
pr(*a*,*b*)	$\Pr(a < y_j < b)$; for churdle exponential b is .
e(*a*,*b*)	$E(y_j \mid a < y_j < b)$; for churdle exponential b is .
residuals	not allowed with margins
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**.

Remarks and examples

▷ Example 1: Predictions for depvar

Below we use the parameters estimated in example 1 of [R] **churdle** to calculate the average hours exercised given the covariates.

```
. use http://www.stata-press.com/data/r15/fitness
. churdle linear hours age i.smoke distance i.single,
> select(commute whours age) ll(0)
  (output omitted )
. predict hourshat
(statistic ystar assumed)
```

We might also be interested in estimating the average number of hours exercised given that an individual exercises. Below we estimate this quantity and compare it to the predicted and true values of the dependent variable for all the observations.

```
. predict exercises, e(0,.)
. summarize hours hourshat exercises
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hours	19,831	.8800172	1.051221	0	5.308835
hourshat	19,831	.8786302	.4915214	.0754708	1.904694
exercises	19,831	1.580729	.3998335	.5630298	2.079012

As expected, we observe that the sample-average predictions are higher for those who exercise.

◁

▷ Example 2: Marginal effects

Suppose we want to study whether single individuals exercise more on average than married individuals. Below, we use `margins` to estimate the average effect of being single on hours spent exercising in the population.

```
. margins, dydx(1.single)
Average marginal effects                        Number of obs     =     19,831
Model VCE    : OIM

Expression   : Conditional mean estimates of dependent variable, predict()
dy/dx w.r.t. : 1.single
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
single						
single	.3858462	.0091398	42.22	0.000	.3679324	.4037599

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

The average effect of moving each individual from not single to single is an increase in exercise of 0.39 hours, or about 23 minutes.

◁

The statistics `pr(`a`,`b`)`, `e(`a`,`b`)`, and `ystar(`a`,`b`)` produce counterfactual predictions.

These statistics predict what would be observed if the limits $\ell\ell$ and $u\ell$ were the specified a and b and the estimated parameters did not change, even though the parameters of the model were estimated using the limits $\ell\ell$ and $u\ell$.

For example, suppose we model contributions to a retirement plan in a world where the government requires a minimum contribution of 2% so $\ell\ell = 2$. After estimating the model parameters, we could predict the average contribution, given the covariates, when the government raises the minimum contribution to 3% with the statistic `ystar(3,.)`.

Methods and formulas

Let $\ell\ell$ refer to the lower limit and $u\ell$ to the upper limit. Also let the probabilities of being at these limits be given by

$$\Pr\left(y_i = \ell\ell | \mathbf{z}_i\right) = \Phi\left(\ell\ell - \mathbf{z}_i'\gamma_{\ell\ell}\right)$$
$$\Pr\left(y_i = u\ell | \mathbf{z}_i\right) = \Phi\left(\mathbf{z}_i'\gamma_{u\ell} - u\ell\right)$$

where $\mathbf{z}_i$ are the covariates of the selection model for individual i, which may be distinct from the covariates $\mathbf{x}_i$ for the outcome model; $\mathbf{\Phi}$ is the standard normal cumulative distribution function; $\gamma_{\ell\ell}$ is the parameter vector for the lower-limit selection model; and $\gamma_{u\ell}$ is the parameter vector for the upper-limit selection model.

We will limit the exposition below to the case with a lower and upper limit.

In `churdle linear`, `ystar` is equivalent to $E(y_i|\mathbf{x}_i)$ and is given by

$$E(y_i|\mathbf{x}_i,\mathbf{z}_i) = \Phi(\mathbf{z}_i'\gamma_{u\ell} - u\ell)\, u\ell + \Phi(\ell\ell - \mathbf{z}_i'\gamma_{\ell\ell})\, \ell\ell$$
$$+ \{\Phi(u\ell - \mathbf{z}_i'\gamma) - \Phi(\ell\ell - \mathbf{z}_i'\gamma)\} \left\{ \mathbf{x}_i'\boldsymbol{\beta} + \sigma \frac{\phi\left(\frac{\ell\ell - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) - \phi\left(\frac{u\ell - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{u\ell - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{\ell\ell - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)} \right\}$$

`pr(`a`,`b`)` is given by

$$\Pr(a < y_i < b|\mathbf{z}_i) = \Phi(b - \mathbf{z}_i'\boldsymbol{\beta}) - \Phi(a - \mathbf{z}_i'\boldsymbol{\beta})$$

`e(`a`,`b`)` is given by

$$E(a < y_i < b|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma \frac{\phi\left(\frac{a - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) - \phi\left(\frac{b - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{b - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)}$$

and `ystar(`a`,`b`)` is given by

$$E(y^*) = \Phi(\mathbf{z}_i'\gamma_{u\ell} - b)\, b + \Phi(a - \mathbf{z}_i'\gamma_{\ell\ell})\, a$$
$$+ \Pr(a < y_i < b|\mathbf{x}_i)\, E(a < y_i < b|\mathbf{x}_i)$$

For `churdle exponential`, `ystar` is equivalent to

$$E(y_i|\mathbf{x}_i) = \Phi(\ell\ell - \mathbf{z}_i'\gamma_{\ell\ell})\, \ell\ell$$
$$+ \{1 - \Phi(\ell\ell - \mathbf{z}_i'\gamma)\} \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \sigma^2/2\right) \left[\frac{1 - \Phi\left\{\frac{\ln(\ell\ell) - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma} - \sigma\right\}}{1 - \Phi\left\{\frac{\ln(\ell\ell) - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right\}} \right]$$

`p(`a`,.)` is given by

$$\Pr(a < y_i|\mathbf{x}_i) = 1 - \Phi(a - \mathbf{z}_i'\gamma_{\ell\ell})$$

`e(`a`,.)` is given by

$$E(a < y_i|\mathbf{x}_i) = \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \sigma^2/2\right) \left[\frac{1 - \Phi\left\{\frac{\ln(a) - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma} - \sigma\right\}}{1 - \Phi\left\{\frac{\ln(a) - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right\}} \right]$$

and `ystar(`a`,.)` is given by

$$E(y^*) = a\Phi(a - \mathbf{z}_i'\gamma_{\ell\ell}) + \Pr(a < y_i|\mathbf{x}_i)\, E(a < y_i|\mathbf{x}_i)$$

Also see

[R] **churdle** — Cragg hurdle regression

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