cfregress postestimation — Postestimation tools for cfregress

Postestimation commands

The following postestimation command is of special interest after `cfregress`:

Command	Description
`estat endogenous`	perform tests of endogeneity

The following postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of parameters
`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 parameters
`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`	predictions and their SEs, residuals, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of parameters
`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, residuals, and standard errors.

Menu for predict

Statistics > Postestimation

Syntax for predict

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

statistic	Description
Main	
`xb`	linear prediction; the default
`xbv`	linear prediction that includes control functions
`e`	residuals
`ve`	residuals that includes control functions
`stdp`	standard error of the prediction
`stdf`	standard error of the forecast

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 for the main equation that does not include control-function terms, that is, $\mathbf{x}_i\mathbf{b}$.

`xbv` calculates the linear prediction for the main equation that includes the estimated control-function terms with their coefficients, that is, $\mathbf{x}_i\mathbf{b} + \widehat{\nu}_i\mathbf{p} + h(\widehat{\nu}_i, \mathbf{y}_i, \mathbf{x}_i)'\mathbf{p}_h$.

`e` calculates the residuals after removing the control functions, that is, $y_i - \mathbf{x}_i\mathbf{b} - \widehat{\nu}_i\mathbf{p} - h(\widehat{\nu}_i, \mathbf{y}_i, \mathbf{x}_i)'\mathbf{p}_h$.

`ve` calculates the residuals that include the control functions, that is, $y_i - \mathbf{x}_i\mathbf{b}$.

`stdp` calculates the standard error of the prediction, which can be thought of as the standard error of the predicted expected value or mean for the observation's covariate pattern. This 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 one 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**.

margins

Description for margins

`margins` estimates margins of response for linear predictions.

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
`xbv`	linear prediction that includes control functions
`e`	not allowed with `margins`
`ve`	not allowed with `margins`
`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**.

estat

Description for estat

`estat endogenous` performs tests to determine whether endogenous regressors in the model are in fact exogenous. Tests are performed as Wald tests on the coefficients of relevant control functions and their interactions in the model and account for the type of variance–covariance matrix used. This method of testing for endogeneity gives different results from that of `estat endogenous` after `ivregress` in finite samples, even when models are identical. See Hansen (2022) for a discussion.

Menu for estat

Statistics > Postestimation

Syntax for estat

`estat endogenous` [*varlist*]

`collect` is allowed with `estat endogenous`; see **[U] 11.1.10 Prefix commands**.

Remarks and examples

Control-function regression lends itself naturally to tests of endogeneity. Under the null hypothesis that an endogenous variable is in fact endogenous, the coefficient on its associated control function, as well as the coefficients on any interactions of the control function, will be zero. Accordingly, a test of these coefficients is a test of the endogeneity of the associated endogenous variable.

`estat endogenous` tests the endogeneity of all endogenous variables jointly if specified without a variable list. Otherwise, only the endogeneity of the listed variables is tested.

For an example of the use of `estat endogenous` after `cfregress`, see example 1 in [R] **cfregress**.

Stored results

`estat endogenous` stores the following in `r()`:

Scalars

`r(chi2)`	χ^2 statistic
`r(df)`	degrees of freedom
`r(p)`	p-value for χ^2 statistic

Methods and formulas

As discussed in [R] **cfregress**, the equation estimated by `cfregress` has the form

$$y_{i0} = \mathbf{y}_i\boldsymbol{\beta}_1 + \mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{w}_i\boldsymbol{\beta}_3 + \hat{\boldsymbol{\nu}}_i\boldsymbol{\rho} + h(\hat{\boldsymbol{\nu}}_i, \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{w}_i)'\boldsymbol{\rho}_h + \epsilon_i$$

where $\hat{\boldsymbol{\nu}}_i$ is a set of estimated control functions, one for each of the endogenous variables in $\mathbf{y}_i$, and $h(\cdot)$ is a known vector-valued function. $h(\cdot)$ can include, for our purposes, interactions between the control functions in $\hat{\boldsymbol{\nu}}_i$, as well as interactions between control functions and the exogenous and endogenous variables in the model.

`estat endogenous`, when specified without a variable list, conducts a joint Wald test of $\boldsymbol{\rho} = 0$ and $\boldsymbol{\rho}_h = 0$.

When a variable list is specified, `estat endogenous` conducts a Wald test for the null hypothesis that all the coefficients in $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_h$, which involve the control functions of the specified variables, are jointly equal to 0.

Reference

Hansen, B. E. 2022. *Econometrics*. Princeton, NJ: Princeton University Press.

Also see

[R] **cfregress** — Control-function linear regression

[U] 20 Estimation and postestimation commands