

ivqregress postestimation — Postestimation tools for ivqregress

- Postestimation commands

Remarks and examples

Also see
- predict

Stored results
- margins

Methods and formulas
- estat

References

## Postestimation commands

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

Command	Description
<code>estat coefplot</code>	plot coefficients and their confidence intervals at different quantiles
<code>estat endogeffects</code>	perform tests of endogeneity
* <code>estat dualci</code>	report dual confidence intervals for endogenous variable
* <code>estat waldplot</code>	plot Wald statistics corresponding to each grid point

\* `estat dualci` and `estat waldplot` work only after `ivqregress iqr`.

The following postestimation commands are also available:

Command	Description
<code>contrast</code>	contrasts and ANOVA-style joint tests of estimates
<code>estat summarize</code>	summary statistics for the estimation sample
<code>estat vce</code>	variance–covariance matrix of the estimators (VCE)
<code>estimates</code>	cataloging estimation results
<code>etable</code>	table of estimation results
<code>forecast</code>	dynamic forecasts and simulations
<code>lincom</code>	point estimates, standard errors, testing, and inference for linear combinations of coefficients
<code>margins</code>	marginal means, predictive margins, marginal effects, and average marginal effects
<code>marginsplot</code>	graph the results from margins (profile plots, interaction plots, etc.)
<code>nlcom</code>	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
<code>predict</code>	linear predictions and residuals
<code>predictnl</code>	point estimates, standard errors, testing, and inference for generalized predictions
<code>pwcompare</code>	pairwise comparisons of estimates
<code>test</code>	Wald tests of simple and composite linear hypotheses
<code>testnl</code>	Wald tests of nonlinear hypotheses

# predict

## Description for predict

`predict` creates a new variable containing predictions such as linear predictors and residuals.

## Menu for predict

Statistics > Postestimation

## Syntax for predict

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

<i>statistic</i>	Description
Main	
<code>xb</code>	linear predictor; the default
<code>residuals</code>	residuals

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 predictor.

`residuals` calculates the residuals, that is,  $y_j - \mathbf{x}_j\mathbf{b}$ .

`equation([eqno])` specifies the equation to which you are making the calculation.

`equation()` is filled in with one *eqno*. `equation(#1)` would mean that the calculation is to be made for the first equation, `equation(#2)` would mean the second, and so on. You could also refer to the equations by their names. `equation(p50)` would refer to the equation named p50 and `equation(p90)` to the equation named p90.

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

# margins

## Description for margins

margins estimates margins of response for linear predictors.

## Menu for margins

Statistics > Postestimation

## Syntax for margins

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

statistic	Description
xb	linear predictor; the default
<u>r</u> esiduals	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 coefplot` plots the estimated coefficients and their confidence intervals (CIs) after `ivqregress`.

`estat endogeneity` tests four hypotheses for the coefficients on the endogenous variables; see [Chernozhukov and Hansen \(2006\)](#). In particular, `estat endogeneity` provides tests for the following null hypotheses:

1. No effect: the endogenous variables do not affect the outcome variable.
2. Constant effect: the effects of the endogenous variables do not vary across estimated quantiles.
3. Dominance: the effects of the endogenous variables are greater than 0 across estimated quantiles.
4. Exogeneity: the variables are exogenous instead of endogenous.

`estat dualci` computes the dual CIs for the coefficients on the endogenous variable (one for each quantile) after `ivqregress iqr`; see [Chernozhukov and Hansen \(2008\)](#). The dual CI is robust to the weak instruments, and it is usually wider than the traditional CI.

`estat waldplot` plots the Wald statistic corresponding to each grid point after `ivqregress iqr`.

### Menu for estat

Statistics > Postestimation

### Syntax for estat

*Plot coefficients and their CIs at different quantiles*

```
estat coefplot [varname] [, coefplot_options]
```

*Perform tests of endogeneity*

```
estat endogeneity [varlist] [, endogeneity_options]
```

*Report dual CIs for endogenous variable*

```
estat dualci [, level(#) display_options]
```

*Plot Wald statistics corresponding to each grid point*

```
estat waldplot [, waldplot_options]
```

*varname* is one of the endogenous regressors or exogenous variables specified when fitting `ivqregress`. By default, *varname* is the endogenous variable specified with `ivqregress iqr` or the first endogenous variable specified with `ivqregress smooth`.

*varlist* contains one or more of the endogenous variables specified when fitting a model with `ivqregress`; the default is the first endogenous variable.

<i>coefplot_options</i>	Description
<code>nocl</code>	do not plot the CIs
<code>no2sls</code>	do not plot the 2SLS estimates
Plot	
<i><a href="#">connect_options</a></i>	change look of lines or connecting method
<i><a href="#">marker_options</a></i>	change look of markers (color, size, etc.)
CI plot	
<code>ciopts(<i>area_options</i>)</code>	affect rendition of the pointwise CIs
Line options	
<i><a href="#">lineopts(<i>cline_options</i>)</a></i>	affect rendition of reference line identifying the 2SLS estimates
Y axis, X axis, Titles, Legend, Overall	
<i><a href="#">twoway_options</a></i>	any options other than <code>by()</code> documented in [G-3] <i><a href="#">twoway_options</a></i>

---

<i>endogeffects_options</i>	Description
<code>all</code>	test four hypotheses; the default
<code>noeffect</code>	test of no effect
<code>constant</code>	test of constant effect
<code>dominance</code>	test of stochastic dominance
<code>exogeneity</code>	test of exogeneity
<code>level(#)</code>	confidence level of a test; default is <code>level(95)</code>
<code>rseed(#)</code>	set random-number seed to #
<code>reps(#)</code>	perform # bootstrap replications; default is <code>reps(100)</code>

---

<i>waldplot_options</i>	Description
<code>quantile(#)</code>	plot Wald statistics for the #th quantile estimation
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
Plot	
<i><a href="#">connect_options</a></i>	change look of lines or connecting method
<i><a href="#">marker_options</a></i>	change look of markers (color, size, etc.)
CI plot	
<code>ciopts(<i>area_options</i>)</code>	affect rendition of the dual CI plot
Y axis, X axis, Titles, Legend, Overall	
<i><a href="#">twoway_options</a></i>	any options other than <code>by()</code> documented in [G-3] <i><a href="#">twoway_options</a></i>

## Options for estat

Options for `estat` are presented under the following headings:

*Options for estat coefplot*  
*Options for estat endoeffects*  
*Options for estat dualci*  
*Options for estat waldplot*

### Options for estat coefplot

`nocl` removes plots of the pointwise CIs. The default is to plot the CIs.

`no2sls` removes the plot of the 2SLS estimates. The default is to plot the 2SLS reference line.

#### Plot

`connect_options` specify how points on a graph are to be connected; [G-3] [connect\\_options](#).

`marker_options` affect the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] [marker\\_options](#).

#### CI plot

`ciopts(area_options)` affects rendition of the pointwise CIs; see [G-3] [area\\_options](#).

#### Line options

`lineopts(cline_options)` affects rendition of reference line identifying the 2SLS estimates; see [G-3] [cline\\_options](#).

#### Y axis, X axis, Titles, Legend, Overall

`twoway_options` are any of the options documented in [G-3] [twoway\\_options](#), excluding `by()`. These include options for titling the graph (see [G-3] [title\\_options](#)) and for saving the graph to disk (see [G-3] [saving\\_option](#)).

### Options for estat endoeffects

`all` provides tests for all the four hypotheses, which is the default. If this option is specified with one of the options `noeffect`, `constant`, `dominance`, or `exogeneity`, then tests for all four hypotheses will be performed.

`noeffect` provides a test for the null hypothesis of no effect, that is, a test that the specified endogenous variables do not affect the outcome variable.

`constant` provides a test for the null hypothesis of constant effect, that is, a test that the effects of the specified endogenous variables do not vary across estimated quantiles.

`dominance` provides a test for the null hypothesis of dominance, that is, a test that the effects of the specified endogenous variables are greater than 0 across estimated quantiles.

`exogeneity` provides a test for the null hypothesis of exogeneity, that is, a test that the specified variables are exogenous instead of endogenous.

`level(#)` specifies the confidence level, as a percentage, for CIs. The default is `level(95)` or as set by `set level`; see [U] [20.8 Specifying the width of confidence intervals](#).

`rseed(#)` sets the random-number seed. Specifying this option makes the results reproducible because the critical values are drawn from a bootstrap sample.

`reps(#)` specifies the number of bootstrap replications to get the critical values of the test. The default is `reps(100)`.

### Options for estat dualci

`level(#)` specifies the confidence level, as a percentage, for the dual CIs. The default is `level(95)` or as set by `set level`; see [U] 20.8 Specifying the width of confidence intervals.

`display_options`: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] Estimation options.

### Options for estat waldplot

`quantile(#)` specifies to plot the Wald statistics for each grid points in the `#`th quantile estimation; the default is the first equation.

`level(#)` specifies the confidence level, as a percentage, for the dual CIs. The default is `level(95)` or as set by `set level`; see [U] 20.8 Specifying the width of confidence intervals.

#### Plot

`connect_options` specify how points on a graph are to be connected; [G-3] [connect\\_options](#).

`marker_options` affect the rendition of markers drawn at the plotted points, including their shape, size, color, and outline; see [G-3] [marker\\_options](#).

#### CI plot

`ciopts(area_options)` affects rendition of the pointwise dual CI plot; see [G-3] [area\\_options](#).

#### Y axis, X axis, Titles, Legend, Overall

`twoway_options` are any of the options documented in [G-3] [twoway\\_options](#), excluding `by()`. These include options for titling the graph (see [G-3] [title\\_options](#)) and for saving the graph to disk (see [G-3] [saving\\_option](#)).

## Remarks and examples

[stata.com](https://www.stata.com)

For examples of using `estat coefplot` and `estat endogeffects`, see [example 3](#) in [R] [ivqregress](#). For an example of using `estat dualci`, see [example 1](#) in [R] [ivqregress](#). For an example of using `estat waldplot`, see [example 4](#) in [R] [ivqregress](#).

### ► Example 1

In [example 1](#) in [R] [ivqregress](#), we fit an instrumental-variables quantile regression (IVQR) model to estimate the effects of 401(k) participation on the conditional median of the net financial assets. Suppose that now we want to know the median of the net financial assets when everyone does or does not participate in a 401(k) conditional on other covariates; we can use `margins` to find the answer.

We specify `i.p401k` immediately after `margins` to obtain the median of the assets under 401(k) participation and under no 401(k) participation. The `at()` option specifies the values of other covariates when computing the median. In particular, the continuous variables such as `income`, `age`, `familysize`, and `educ` are fixed at the sample mean, and people are assumed to be married, participate in an IRA, receive pension benefits, and own a home.

```
. use https://www.stata-press.com/data/r18/assets2
(Excerpt from Chernozhukov and Hansen (2004))

. ivqregress iqr assets (i.p401k = i.e401k) income age familysize
> i.married i.ira i.pension i.ownhome educ
(output omitted)

. margins i.p401k, at((mean) income age familysize educ
> married = 1 ira = 1 pension = 1 ownhome = 1)
Adjusted predictions                                Number of obs = 9,913
Model VCE: Robust

Expression: Linear predictor, predict()
At: income      = 37208.4 (mean)
    age         = 41.05891 (mean)
    familysize  = 2.865328 (mean)
    married     =      1
    ira         =      1
    pension     =      1
    ownhome     =      1
    educ        = 13.20629 (mean)
```

	Delta-method					
	Margin	std. err.	z	P> z	[95% conf. interval]	
p401k						
No	23681.37	1007.612	23.50	0.000	21706.49	25656.26
Yes	28994.77	1123.076	25.82	0.000	26793.58	31195.96

The results show that the conditional median of assets when everyone participates in a 401(k) is \$28,995. In contrast, the conditional median of assets when no one participates in a 401(k) is only \$23,681. The difference between these two medians is \$5,313, which is the quantile treatment effect of `p401k` and is the same as the coefficient’s value.



Stored results

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

- Scalars
  - `r(N_reps)` number of replications
  - `r(level)` confidence level
- Macros
  - `r(endog)` tested endogenous regressors
- Matrices
  - `r(table)` matrix containing test statistics and critical values

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

- Scalars
  - `r(N)` number of observations
  - `r(level)` confidence level for the dual CI
- Matrices
  - `r(table)` matrix containing test statistics, *p*-values, and confidence errors, test statistics, *p*-values, and confidence intervals



## Methods and formulas

Methods and formulas are presented under the following headings:

*Tests of effects of endogenous variables*  
*Dual CI*

## Tests of effects of endogenous variables

`estat endogeffects` implements the general inference procedure outlined in section 4 of [Chernozhukov and Hansen \(2006\)](#). It tests the following null hypotheses:

1. No effect: the endogenous variables do not affect the outcome variable.
2. Constant effect: the effects of the endogenous variables do not vary across estimated quantiles.
3. Dominance: the effects of the endogenous variables are greater than 0 across estimated quantiles.
4. Exogeneity: the variables are exogenous instead of endogenous.

It is convenient to write each null hypothesis in the following form:

$$\mathbf{R}(\tau)\{\boldsymbol{\theta}(\tau) - \mathbf{r}(\tau)\} = 0 \quad \text{for each } \tau \in T$$

where  $\mathbf{R}(\tau)$  is a  $q \times p$  matrix of rank  $q$  when  $q$  is smaller than the dimension of  $\boldsymbol{\theta}(\tau)$ .  $\boldsymbol{\theta}(\tau)$  is the  $p \times 1$  coefficient vector for the IVQR model in the  $\tau$ th quantile,  $\mathbf{r}(\tau) \in R^p$ , and  $T$  is a set of estimated quantile indexes. This form is different from the classical setting because  $\boldsymbol{\theta}(\cdot)$  and  $\mathbf{r}(\cdot)$  are functions, which need to be estimated in some cases.

Based on the IVQR model estimates  $\widehat{\boldsymbol{\theta}}(\cdot)$ , we focus on the basic inference process

$$\mathbf{v}_n(\cdot) = \mathbf{R}(\cdot)\{\widehat{\boldsymbol{\theta}}(\cdot) - \widehat{\mathbf{r}}(\cdot)\}$$

where  $\widehat{\mathbf{r}}(\cdot)$  is either a vector of constants or a vector of estimates from the classical quantile regression. We use the Kolmogorov–Smirnov statistic  $S_n = f\{\sqrt{n}\mathbf{v}_n(\cdot)\}$ , which is a function of  $\mathbf{v}_n(\cdot)$ .

$$S_n = \sqrt{n} \sup_{\tau \in T} \|\mathbf{v}_n(\tau)\|_{\widehat{\mathbf{A}}(\tau)}$$

where  $\|\mathbf{v}\|_{\mathbf{A}} = \sqrt{\mathbf{v}'\mathbf{A}\mathbf{v}}$ . For the choice of  $\widehat{\mathbf{A}}(\tau)$ , see section 4.4 in [Chernozhukov and Hansen \(2006\)](#). In essence,  $\widehat{\mathbf{A}}(\tau)$  is the empirical variance of the estimating scores implied by the IVQR model.

The null hypothesis is rejected if

$$S_n > c(1 - \alpha)$$

where the critical value  $c(1 - \alpha)$  with confidence level  $1 - \alpha$  can be obtained using the bootstrap resampling procedure described in section 4.3 of [Chernozhukov and Hansen \(2006\)](#).

Now we describe the formal definition of the four hypotheses. For notational simplicity, we assume that there is one endogenous variable but the case of multiple endogenous variables can be extended easily. Let  $\alpha(\tau)$  denote the endogenous coefficient for the  $\tau$ th quantile estimation. In this case,  $\mathbf{R} = (1, 0, \dots, 0)$  for the four hypotheses considered.

1. No effect: the null hypothesis is that the endogenous variable has no impact on the outcome:  $\alpha(\tau) = 0$ . In this case,

$$H_0 : \alpha(\tau) = 0 \quad \text{for all } \tau \in T$$

$$H_1 : \alpha(\tau) \neq 0 \quad \text{for some } \tau \in T$$

and  $\widehat{r}(\cdot) = 0$ .

2. Constant effect: the null hypothesis of a constant effect is that the endogenous variable only affects the location of the outcome but not other moments. That is,  $\alpha(\tau) = c$  for all  $\tau \in T$ , where  $c$  is a constant. In this case,

$$H_0 : \alpha(\tau) = c \quad \text{for all } \tau \in T$$

$$H_1 : \alpha(\tau) \neq c \quad \text{for some } \tau \in T$$

and  $\widehat{r}(\cdot)$  is the estimate of endogenous coefficient for one of the quantile indexes.

3. Dominance: the dominance test tests whether the effects of endogenous variable are unambiguously beneficial. That is,  $\alpha(\tau) > 0$  for all  $\tau \in T$ . For this hypothesis, we use the one-sided Kolmogorov–Smirnov statistic

$$S_n = \sqrt{n} \sup_{\tau \in T} \max(-\alpha(\tau), 0)$$

In this case,

$$H_0 : \alpha(\tau) > 0 \quad \text{for all } \tau \in T$$

$$H_1 : \alpha(\tau) \leq 0 \quad \text{for some } \tau \in T$$

and  $\widehat{r}(\cdot) = 0$ .

4. Exogeneity: if all the covariates are exogenous, we can fit the model by the regular quantile regression and denote  $\eta(\tau)$  as the quantile regression estimator. The difference between  $\theta(\tau)$  and  $\eta(\tau)$  can be used to formulate a Hausman test of exogeneity. In this case, the null and alternative are defined as

$$H_0 : \alpha(\tau) = \mathbf{R}\eta(\tau) \quad \text{for all } \tau \in T$$

$$H_1 : \alpha(\tau) \neq \mathbf{R}\eta(\tau) \quad \text{for some } \tau \in T$$

and  $\widehat{\mathbf{r}}(\tau) = \widehat{\eta}(\tau)$ , where  $\widehat{\eta}(\tau)$  is the regular quantile regression estimate.

## Dual CI

`estat dualci` computes the dual CI proposed in [Chernozhukov and Hansen \(2008\)](#) for the coefficient on the endogenous variable in the IVQR model. The dual CI is robust to the weak instrument, and it is usually wider than the classical CI. `estat dualci` is allowed only after `ivqregress iqr`. If you have not read about the methods for the inverse quantile regression (IQR) estimator, see [The IQR estimator](#) in *Methods and formulas* of [\[R\] ivqregress](#).

Suppose we know the true value of the coefficient on the endogenous covariates, which we denote as  $\alpha(\tau)$ , and let  $W_n\{\alpha(\tau)\}$  be the Wald statistic for the coefficient on the instruments in the auxiliary quantile regression. Then by proposition 1 in [Chernozhukov and Hansen \(2008\)](#),

$$W_n\{\alpha(\tau)\} \rightarrow_d \chi^2(1)$$

and for the confidence region  $\text{CR}_p\{\alpha(\tau)\} = \{\alpha \in A : W_n(\alpha) < c_p\}$ , where  $P\{\chi^2(1) < c_p\} = p$ , and

$$P[\alpha(\tau) \in \text{CR}_p\{\alpha(\tau)\}] = P[W_n\{\alpha(\tau)\} < c_p] = p$$

Intuitively,  $W_n\{\alpha(\tau)\}$  is the Wald statistic for testing whether the coefficients for the instruments are 0. When  $\alpha$  equals the true value  $\alpha(\tau)$ ,  $W(\cdot)$  is  $\chi^2$  distributed with the degree of freedom of 1. Thus, a valid CI for  $\alpha$  can be constructed by the inversion of the Wald statistic. That is,

$$\text{CR}_p\{\alpha(\tau)\} = \{\alpha \in A : W_n(\alpha) < c_p\}$$

covers the true value of  $\alpha(\tau)$  with probability approaching  $p$ .

In practice, the dual CI is constructed by the lower and upper limits of the grid points such that the corresponding Wald statistic is smaller than the critical value  $c_p$ .

## References

- Chernozhukov, V., and C. B. Hansen. 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132: 491–525. <https://doi.org/10.1016/j.jeconom.2005.02.009>.
- . 2008. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics* 142: 379–398. <https://doi.org/10.1016/j.jeconom.2007.06.005>.

## Also see

- [R] **ivqregress** — Instrumental-variables quantile regression
- [U] **20 Estimation and postestimation commands**

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

