

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

Postestimation commands

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

Command	Description
<code>estat endogenous</code>	perform tests of endogeneity
<code>estat firststage</code>	report “first-stage” regression statistics
<code>estat overid</code>	perform tests of overidentifying restrictions
<code>estat weakrobust</code>	perform inference robust to weak instruments
* <code>estat sbknown</code>	perform tests for a structural break with a known break date
* <code>estat single</code>	perform tests for a structural break with an unknown break date

These commands are not appropriate with `svy` estimation results.

*`estat sbknown` and `estat sbsingle` work only after `ivregress 2s1s`.

The following postestimation commands are also available:

Command	Description
<code>contrast</code>	contrasts and ANOVA-style joint tests of parameters
<code>estat summarize</code>	summary statistics for the estimation sample
<code>estat vce</code>	variance–covariance matrix of the estimators (VCE)
<code>estat (svy)</code>	postestimation statistics for survey data
<code>estimates</code>	cataloging estimation results
<code>etable</code>	table of estimation results
† <code>forecast</code>	dynamic forecasts and simulations
† <code>hausman</code>	Hausman’s specification test
<code>lincom</code>	point estimates, standard errors, testing, and inference for linear combinations of parameters
<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 parameters
<code>predict</code>	linear predictions and their SEs, probabilities, etc.
<code>predictnl</code>	point estimates, standard errors, testing, and inference for generalized predictions
<code>pwcompare</code>	pairwise comparisons of parameters
<code>test</code>	Wald tests of simple and composite linear hypotheses
<code>testnl</code>	Wald tests of nonlinear hypotheses

†`forecast` and `hausman` are not appropriate with `svy` estimation results.

predict

Description for predict

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

Menu for predict

Statistics > Postestimation

Syntax for predict

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

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

<i>statistic</i>	Description
------------------	-------------

Main

`xb` $\mathbf{x}_j\mathbf{b}$, fitted values; the default

`stdp` standard error of the prediction

`stdf` standard error of the forecast

`dresiduals` $d_{\text{absorbvars}} + e_j = y_j - \mathbf{x}_j\mathbf{b}$

* `xbd` $\mathbf{x}_j\mathbf{b} + d_{\text{absorbvars}}$

* `d` $d_{\text{absorbvars}}$

`residuals` residuals

† `pr(a,b)` $\Pr(a < y_j < b)$ under exogeneity and normal errors

† `e(a,b)` $E(y_j | a < y_j < b)$ under exogeneity and normal errors

† `ystar(a,b)` $E(y_j^*), y_j^* = \max\{a, \min(y_j, b)\}$ under exogeneity and normal errors

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

†Not allowed if `absorb()` is used at estimation.

`xb`, `stdp`, and `residuals` are the only options allowed with `svy` estimation results.

where *a* and *b* may be numbers or variables; *a* missing (*a* ≥ .) means $-\infty$, and *b* missing (*b* ≥ .) means $+\infty$; see [U] 12.2.1 Missing values.

Options for predict

Main

`xb`, the default, calculates the linear prediction, $\mathbf{x}_j\mathbf{b}$.

`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 1 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**.

`dresiduals` calculates $y_j - \mathbf{x}_j\mathbf{b}$, which are the residuals plus the effects of the absorbed variables. These are based on the estimated equation when the observed values of the endogenous variables are used—not the projections of the instruments onto the endogenous variables.

`xbd` calculates $\mathbf{x}_j\mathbf{b} + d_{\text{absorbvar}}$, which are the fitted values including the individual effects of the absorbed variables.

`d` calculates $d_{\text{absorbvar}}$, the sum of individual coefficients for the absorbed variables.

`residuals` calculates the residuals, that is, $y_j - (\mathbf{x}_j\mathbf{b} + d_{\text{absorbvar}})$.

`pr(a,b)` calculates $\Pr(a < \mathbf{x}_j\mathbf{b} + u_j < b)$, the probability that $y_j|\mathbf{x}_j$ would be observed in the interval (a,b) under exogeneity and assuming errors are normally distributed.

a and b may be specified as numbers or variable names; lb and ub are variable names;

`pr(20,30)` calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < 30)$;

`pr(lb,ub)` calculates $\Pr(lb < \mathbf{x}_j\mathbf{b} + u_j < ub)$; and

`pr(20,ub)` calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < ub)$.

a missing ($a \geq .$) means $-\infty$; `pr(. ,30)` calculates $\Pr(-\infty < \mathbf{x}_j\mathbf{b} + u_j < 30)$;

`pr(lb,30)` calculates $\Pr(-\infty < \mathbf{x}_j\mathbf{b} + u_j < 30)$ in observations for which $lb \geq .$ and calculates $\Pr(lb < \mathbf{x}_j\mathbf{b} + u_j < 30)$ elsewhere.

b missing ($b \geq .$) means $+\infty$; `pr(20, .)` calculates $\Pr(+\infty > \mathbf{x}_j\mathbf{b} + u_j > 20)$;

`pr(20,ub)` calculates $\Pr(+\infty > \mathbf{x}_j\mathbf{b} + u_j > 20)$ in observations for which $ub \geq .$ and calculates $\Pr(20 < \mathbf{x}_j\mathbf{b} + u_j < ub)$ elsewhere.

`e(a,b)` calculates $E(\mathbf{x}_j\mathbf{b} + u_j | a < \mathbf{x}_j\mathbf{b} + u_j < b)$, 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 truncated. a and b are specified as they are for `pr()`. Exogeneity and normally distributed errors are assumed.

`ystar(a,b)` calculates $E(y_j^*)$, where $y_j^* = a$ if $\mathbf{x}_j\mathbf{b} + u_j \leq a$, $y_j^* = b$ if $\mathbf{x}_j\mathbf{b} + u_j \geq b$, and $y_j^* = \mathbf{x}_j\mathbf{b} + u_j$ otherwise, meaning that y_j^* is censored. a and b are specified as they are for `pr()`. Exogeneity and normally distributed errors are assumed.

`scores` calculates the scores for the model. A new score variable is created for each endogenous regressor, as well as an equation-level score that applies to all exogenous variables and constant term (if present). `scores` is not allowed if the `absorb()` option is used at estimation.

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]
```

<i>statistic</i>	Description
<code>xb</code>	linear prediction; the default
<code>stdp</code>	not allowed with <code>margins</code>
<code>stdf</code>	not allowed with <code>margins</code>
<code>dresiduals</code>	not allowed with <code>margins</code>
<code>xbd</code>	not allowed with <code>margins</code>
<code>d</code>	not allowed with <code>margins</code>
<code>residuals</code>	not allowed with <code>margins</code>
<code>pr(<i>a</i>,<i>b</i>)</code>	$\Pr(a < y_j < b)$ under exogeneity and normal errors
<code>e(<i>a</i>,<i>b</i>)</code>	$E(y_j a < y_j < b)$ under exogeneity and normal errors
<code>ystar(<i>a</i>,<i>b</i>)</code>	$E(y_j^*), y_j^* = \max\{a, \min(y_j, b)\}$ under exogeneity and normal errors

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. After GMM estimation, the C (difference-in-Sargan) statistic is reported. After 2SLS estimation with an unadjusted VCE, the Durbin (1954) and Wu–Hausman (Wu 1974; Hausman 1978) statistics are reported. After 2SLS estimation with a robust VCE, Wooldridge’s (1995) robust score test and a robust regression-based test are reported. In all cases, if the test statistic is significant, then the variables being tested must be treated as endogenous. `estat endogenous` is not available after LIML estimation.

`estat firststage` reports various statistics that measure the relevance of the excluded exogenous variables. By default, which statistics are reported depends on whether the equation has one or more than one endogenous regressor.

`estat overid` performs tests of overidentifying restrictions. If the 2SLS estimator was used, Sargan’s (1958) and Basman’s (1960) χ^2 tests are reported, as is Wooldridge’s (1995) robust score test; if the LIML estimator was used, Anderson and Rubin’s (1950) χ^2 test and Basman’s F test are reported; and if the GMM estimator was used, Hansen’s (1982) J statistic χ^2 test is reported. In all of these cases, a rejection of the test indicates that the instruments may not be valid.

`estat weakrobust` performs hypothesis tests on the coefficients of endogenous regressors that are robust to weak instruments. For just-identified models, the Anderson–Rubin (1949) test is reported. For overidentified models, if a homoskedastic VCE was used in estimation, the conditional likelihood-ratio (CLR) test of Moreira (2003) is reported. If a heteroskedastic or cluster–robust VCE was used instead, an appropriate generalization of the CLR test is used (see Finlay and Magnusson [2009]). When there is only one endogenous regressor, the confidence intervals associated with any of these tests can be requested.

Menu for estat

Statistics > Postestimation

Syntax for estat

Perform tests of endogeneity

```
estat endogenous [ varlist ] [ , lags(#) forceweights forcenonrobust ]
```

Report “first-stage” regression statistics

```
estat firststage [ , all forcenonrobust ]
```

Perform tests of overidentifying restrictions

```
estat overid [ , lags(#) forceweights forcenonrobust ]
```

Perform inference robust to weak instruments

```
estat weakrobust [ , weak_options ]
```

<i>weak_options</i>	Description
<code>betanull(<i>null_spec</i>)</code>	specify the null hypothesis on the endogenous regressors
<code>ar</code>	perform Anderson–Rubin test
<code>clr</code>	perform conditional likelihood-ratio test
<code>ci</code>	report confidence interval instead of test
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
<code>reps(#)</code>	perform # replications for simulations; default is <code>reps(25000)</code>
<code>rseed(#)</code>	set random-number seed for simulations to #
<code>bound(#_{min} #_{max})</code>	specify bounds of confidence interval grid
<code>ngrid(#)</code>	use # grid points for the confidence interval; default is <code>ngrid(500)</code>
<code>noadaptgrid</code>	suppress adaptive grid expansion
<code>[no]log</code>	suppress or display the iteration log
<code>iterate(#)</code>	perform maximum of # adaptive iterations; default is <code>iterate(20)</code>

`collect` is allowed with `estat endogenous`, `estat firststage`, `estat overid`, and `estat weakrobust`; see [U] 11.1.10 Prefix commands.

Options for estat

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

[Options for estat endogenous](#)

[Options for estat firststage](#)

[Options for estat overid](#)

[Options for estat weakrobust](#)

Options for estat endogenous

`lags(#)` specifies the number of lags to use for prewhitening when computing the heteroskedasticity- and autocorrelation-consistent (HAC) version of the score test of endogeneity. Specifying `lags(0)` requests no prewhitening. This option is valid only when the model was fit via 2SLS and a HAC covariance matrix was requested when the model was fit. The default is `lags(1)`.

`forceweights` requests that the tests of endogeneity be computed even though `aweight`s, `pweight`s, or `iweight`s were used in the previous estimation. By default, these tests are conducted only after unweighted or frequency-weighted estimation. The reported critical values may be inappropriate for weighted data, so the user must determine whether the critical values are appropriate for a given application.

`forcenonrobust` requests that the Durbin and Wu–Hausman tests be performed after 2SLS estimation even though a robust VCE was used at estimation time. This option is available only if the model was fit by 2SLS.

Options for estat firststage

`all` requests that all first-stage goodness-of-fit statistics be reported regardless of whether the model contains one or more endogenous regressors. By default, if the model contains one endogenous regressor, then the first-stage R^2 , adjusted R^2 , partial R^2 , and F statistics are reported, whereas if the model contains multiple endogenous regressors, then Shea's partial R^2 and adjusted partial R^2 are reported instead.

`forcenonrobust` requests that the minimum eigenvalue statistic and its critical values be reported even though a robust VCE was used at estimation time. The reported critical values assume that the errors are independent and identically distributed (i.i.d.) normal, so the user must determine whether the critical values are appropriate for a given application.

Options for `estat overid`

`lags(#)` specifies the number of lags to use for prewhitening when computing the heteroskedasticity- and autocorrelation-consistent (HAC) version of the score test of overidentifying restrictions. Specifying `lags(0)` requests no prewhitening. This option is valid only when the model was fit via 2SLS and a HAC covariance matrix was requested when the model was fit. The default is `lags(1)`.

`forceweights` requests that the tests of overidentifying restrictions be computed even though `aweights`, `pweights`, or `iweights` were used in the previous estimation. By default, these tests are conducted only after unweighted or frequency-weighted estimation. The reported critical values may be inappropriate for weighted data, so the user must determine whether the critical values are appropriate for a given application.

`forcenonrobust` requests that the Sargan and Basman tests of overidentifying restrictions be performed after 2SLS or LIML estimation even though a robust VCE was used at estimation time. These tests assume that the errors are i.i.d. normal, so the user must determine whether the critical values are appropriate for a given application.

Options for `estat weakrobust`

`betanull(null_spec)` specifies the null hypothesis of the test on the coefficients of the endogenous regressors. By default, the null hypothesis is that the coefficients of all endogenous regressors are zero. The null hypothesis can be specified using `numlist`, a list of equalities, or a matrix. When using `numlist`, you must specify the null values for all endogenous regressors in the order given in the estimation command and specify the `copy` suboption if there is more than one endogenous regressor. When you use a list of equalities or a matrix, null values are set to zero for endogenous regressors you did not specify. When you use a matrix, null values are assigned to coefficients by matrix column names.

`null_spec` is one of

`# [#... , copy]`

`varname = # [varname = # [...]] [, skip]`

`matname [, skip copy]`

`skip` specifies that any variables that are not endogenous regressors in the model be ignored. The default action is to issue an error message.

`copy` specifies that the list of null values be assigned into the null hypothesis by position rather than by name.

`ar` requests that an Anderson–Rubin test be performed. The associated confidence interval of the test can be requested with the `ci` option. By default, the Anderson–Rubin test is performed when the model is just identified. If both `ar` and `clr` are specified, both tests are performed.

`clr` requests that a CLR test be performed. The associated confidence interval of the test can be requested with the `ci` option. By default, a CLR test is performed when the model is overidentified. If both `ar` and `clr` are specified, both tests are performed.

- `ci` requests that a confidence interval be returned instead of a test statistic and a p -value. This applies to all tests specified by the user. Option `ci` is available only when there is a single endogenous regressor.
- `level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] 20.8 Specifying the width of confidence intervals.
- `reps(#)` specifies the number of replications to use in simulating critical values when p -values cannot be computed analytically. The default is `reps(25000)`.
- `rseed(#)` sets the random-number seed. This option can be used to reproduce results for simulating critical values when p -values cannot be computed analytically. `rseed(#)` is equivalent to typing `set seed #` prior to calling `estat weakrobust`; see [R] set seed.
- `bound(#min #max)` specifies the lower (`#min`) and the upper bound (`#max`) for the grid used to approximate confidence intervals when they cannot be computed analytically. The specified bounds will be the starting bounds for the adaptive grid expansion procedure or, if `noadaptgrid` is specified, the fixed bounds of the grid. By default, the lower bound is set to five standard errors below the coefficient estimated by `ivregress`, and the upper bound is five standard errors above.
- `ngrid(#)` specifies the starting number of gridpoints for approximating confidence intervals. The default is `ngrid(500)`. Higher values will increase computation time but improve precision, which may be helpful when the confidence interval is wide. If `noadaptgrid` is specified, `ngrid()` specifies the fixed number of gridpoints.
- `noadaptgrid` requests that the default method for computing confidence intervals, where the grid is adaptively expanded, be turned off and the interval be estimated with a fixed grid. `noadaptgrid` is equivalent to specifying `iterate(0)`.
- `log` and `nolog` specify whether to display the iteration log showing the progress of the adaptive gridding procedure. The iteration log is displayed by default unless you used `set iterlog off` to suppress it; see `set iterlog` in [R] set iter.
- `iterate(#)` specifies the maximum number of iterations for the adaptive grid expansion procedure. Each iteration expands the grid by half the width of the starting bounds in each direction, and the procedure stops when the maximum number of iterations is reached or when the confidence interval has finite endpoints and was unchanged by the most recent expansion. The default is `iterate(20)`.

Remarks and examples

Remarks are presented under the following headings:

estat endogenous
estat firststage
estat overid
estat weakrobust

estat endogenous

A natural question to ask is whether a variable presumed to be endogenous in the previously fit model could instead be treated as exogenous. If the endogenous regressors are in fact exogenous, then the OLS estimator is more efficient; and depending on the strength of the instruments and other factors, the sacrifice in efficiency by using an instrumental-variables estimator can be significant. Thus, unless an instrumental-variables estimator is really needed, OLS should be used instead. `estat endogenous` provides several tests of endogeneity after 2SLS and GMM estimation.

▷ Example 1

In [example 1](#) of [R] `ivregress`, we fit a model of the average rental rate for housing in a state as a function of the percentage of the population living in urban areas and the average value of houses. We treated `hsngval` as endogenous because unanticipated shocks that affect rental rates probably affect house prices as well. We used family income and region dummies as additional instruments for `hsngval`. Here we test whether we could treat `hsngval` as exogenous.

```
. use https://www.stata-press.com/data/r19/hsng
(1980 Census housing data)
. ivregress 2sls rent pcturban (hsngval = faminc i.region)
(output omitted)
. estat endogenous
Tests of endogeneity
H0: Variables are exogenous
Durbin (score) chi2(1)          = 12.8473 (p = 0.0003)
Wu-Hausman F(1,46)             = 15.9067 (p = 0.0002)
```

Because we did not specify any variable names after the `estat endogenous` command, Stata by default tested all the endogenous regressors (namely, `hsngval`) in our model. The null hypothesis of the Durbin and Wu–Hausman tests is that the variable under consideration can be treated as exogenous. Here both test statistics are highly significant, so we reject the null of exogeneity; we must continue to treat `hsngval` as endogenous.

◀

The difference between the Durbin and Wu–Hausman tests of endogeneity is that the former uses an estimate of the error term’s variance based on the model assuming the variables being tested are exogenous, while the latter uses an estimate of the error variance based on the model assuming the variables being tested are endogenous. Under the null hypothesis that the variables being tested are exogenous, both estimates of the error variance are consistent. What we label the Wu–Hausman statistic is Wu’s (1974) “ T_2 ” statistic, which Hausman (1978) showed can be calculated very easily via linear regression. Baum, Schaffer, and Stillman (2003, 2007) provide a lucid discussion of these tests.

When you fit a model with multiple endogenous regressors, you can test the exogeneity of a subset of the regressors while continuing to treat the others as endogenous. For example, say you have three endogenous regressors, y_1 , y_2 , and y_3 , and you fit your model by typing

```
. ivregress depvar ... (y1 y2 y3 = ...)
```

Suppose you are confident that y_1 must be treated as endogenous, but you are undecided about y_2 and y_3 . To test whether y_2 and y_3 can be treated as exogenous, you would type

```
. estat endogenous y2 y3
```

The Durbin and Wu–Hausman tests assume that the error term is i.i.d. Therefore, if you requested a robust VCE at estimation time, `estat endogenous` will instead report Wooldridge’s (1995) score test and a regression-based test of exogeneity. Both these tests can tolerate heteroskedastic and autocorrelated errors, while only the regression-based test is amenable to clustering.

▷ Example 2

We refit our housing model, requesting robust standard errors, and then test the exogeneity of `hsngval`:

```
. use https://www.stata-press.com/data/r19/hsng
(1980 Census housing data)
. ivregress 2sls rent pcturban (hsngval = faminc i.region), vce(robust)
(output omitted)
. estat endogenous
Tests of endogeneity
H0: Variables are exogenous
Robust score chi2(1)      =  2.10428   (p = 0.1469)
Robust regression F(1,46) =  4.31101   (p = 0.0435)
```

Wooldridge’s score test does not reject the null hypothesis that `hsngval` is exogenous at conventional significance levels ($p = 0.1469$). However, the regression-based test does reject the null hypothesis at the 5% significance level ($p = 0.0435$). Typically, these two tests yield the same conclusion; the fact that our dataset has only 50 observations could be contributing to the discrepancy. Here we would be inclined to continue to treat `hsngval` as endogenous. Even if `hsngval` is exogenous, the 2SLS estimates are still consistent. On the other hand, if `hsngval` is in fact endogenous, the OLS estimates would not be consistent. Moreover, as we will see in our discussion of the `estat overid` command, our additional instruments may be invalid. To test whether an endogenous variable can be treated as exogenous, we must have a valid set of instruments to use to fit the model in the first place!

◀

Unlike the Durbin and Wu–Hausman tests, Wooldridge’s score and the regression-based tests do not allow you to test a subset of the endogenous regressors in the model; you can test only whether all the endogenous regressors are in fact exogenous.

After GMM estimation, `estat endogenous` calculates what Hayashi (2000, 220) calls the C statistic, also known as the difference-in-Sargan statistic. The C statistic can be made robust to heteroskedasticity, autocorrelation, and clustering; and the version reported by `estat endogenous` is determined by the weight matrix requested via the `wmatrix()` option used when fitting the model with `ivregress`. Additionally, the test can be used to determine the exogeneity of a subset of the endogenous regressors, regardless of the type of weight matrix used.

If you fit your model using the LIML estimator, you can use the `hausman` command to carry out a traditional Hausman (1978) test between the OLS and LIML estimates.

estat firststage

For an excluded exogenous variable to be a valid instrument, it must be sufficiently correlated with the included endogenous regressors but uncorrelated with the error term. In recent decades, researchers have paid considerable attention to the issue of instruments that are only weakly correlated with the endogenous regressors. In such cases, the usual 2SLS, GMM, and LIML estimators are biased toward the OLS estimator, and inference based on the standard errors reported by, for example, `ivregress` can be severely misleading. For more information on the theory behind instrumental-variables estimation with weak instruments, see Nelson and Startz (1990); Staiger and Stock (1997); Hahn and Hausman (2003); the survey article by Stock, Wright, and Yogo (2002); and Angrist and Pischke (2009, chap. 4).

When the instruments are only weakly correlated with the endogenous regressors, some Monte Carlo evidence suggests that the LIML estimator performs better than the 2SLS and GMM estimators; see, for example, [Poi \(2006\)](#) and [Stock, Wright, and Yogo \(2002\)](#) (and the papers cited therein). On the other hand, the LIML estimator often results in confidence intervals that are somewhat larger than those from the 2SLS estimator.

Moreover, using more instruments is not a solution, because the biases of instrumental-variables estimators increase with the number of instruments. See [Hahn and Hausman \(2003\)](#).

`estat firststage` produces several statistics for judging the explanatory power of the instruments and is most easily explained with examples.

▷ Example 3

Again building on the model fit in [example 1](#) of [\[R\] ivregress](#), we now explore the degree of correlation between the additional instruments `faminc`, `2.region`, `3.region`, and `4.region` and the endogenous regressor `hsngval`:

```
. use https://www.stata-press.com/data/r19/hsng
(1980 Census housing data)
. ivregress 2sls rent pcturban (hsngval = faminc i.region)
(output omitted)
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,44)	Prob > F
hsngval	0.6908	0.6557	0.5473	13.2978	0.0000

Minimum eigenvalue statistic = 13.2978

Critical Values	# of endogenous regressors:	1
H0: Instruments are weak	# of excluded instruments:	4

	5%	10%	20%	30%
2SLS relative bias	16.85	10.27	6.71	5.34
2SLS size of nominal 5% Wald test	24.58	13.96	10.26	8.31
LIML size of nominal 5% Wald test	5.44	3.87	3.30	2.98

To understand these results, recall that the first-stage regression is

$$\text{hsngval}_i = \pi_0 + \pi_1 \text{pcturban}_i + \pi_2 \text{faminc} + \pi_3 \text{2.region} + \pi_4 \text{3.region} + \pi_5 \text{4.region} + v_i$$

where v_i is an error term. The column marked “R-sq.” is the simple R^2 from fitting the first-stage regression by OLS, and the column marked “Adjusted R-sq.” is the adjusted R^2 from that regression. Higher values purportedly indicate stronger instruments, and instrumental-variables estimators exhibit less bias when the instruments are strongly correlated with the endogenous variable.

Looking at just the R^2 and adjusted R^2 can be misleading, however. If `hsngval` were strongly correlated with the included exogenous variable `pcturban` but only weakly correlated with the additional instruments, then these statistics could be large even though a weak-instrument problem is present.

The partial R^2 statistic measures the correlation between `hsgnval` and the additional instruments after *partialing out* the effect of `pcturban`. Unlike the R^2 and adjusted R^2 statistics, the partial R^2 statistic will not be inflated because of strong correlation between `hsgnval` and `pcturban`. Bound, Jaeger, and Baker (1995) and others have promoted using this statistic.

The column marked “F(4, 44)” is an F statistic for the joint significance of π_2 , π_3 , π_4 , and π_5 , the coefficients on the additional instruments. Its p -value is listed in the column marked “Prob > F”. If the F statistic is not significant, then the additional instruments have no significant explanatory power for `hsgnval` after controlling for the effect of `pcturban`. However, Hall, Rudebusch, and Wilcox (1996) used Monte Carlo simulation to show that simply having an F statistic that is significant at the typical 5% or 10% level is not sufficient. Stock, Wright, and Yogo (2002) suggest that the F statistic should exceed 10 for inference based on the 2SLS estimator to be reliable when there is one endogenous regressor.

`estat firststage` also presents the Cragg and Donald (1993) minimum eigenvalue statistic as a further test of weak instruments. Stock and Yogo (2005) discuss two characterizations of weak instruments: first, weak instruments cause instrumental-variables estimators to be biased; second, hypothesis tests of parameters estimated by instrumental-variables estimators may suffer from severe size distortions. The test statistic in our example is 13.30, which is identical to the F statistic just discussed because our model contains one endogenous regressor.

The null hypothesis of each of Stock and Yogo’s tests is that the set of instruments is weak. To perform these tests, we must first choose either the largest relative bias of the 2SLS estimator we are willing to tolerate or the largest rejection rate of a nominal 5% Wald test we are willing to tolerate. If the test statistic exceeds the critical value, we can conclude that our instruments are not weak.

The row marked “2SLS relative bias” contains critical values for the test that the instruments are weak based on the bias of the 2SLS estimator *relative to* the bias of the OLS estimator. For example, from past experience we might know that the OLS estimate of a parameter β may be 50% too high. Saying that we are willing to tolerate a 10% relative bias means that we are willing to tolerate a bias of the 2SLS estimator no greater than 5% (that is, 10% of 50%). In our rental rate model, if we are willing to tolerate a 10% relative bias, then we can conclude that our instruments are not weak because the test statistic of 13.30 exceeds the critical value of 10.27. However, if we were willing to tolerate only a relative bias of 5%, we would conclude that our instruments are weak because $13.30 < 16.85$.

The rows marked “2SLS Size of nominal 5% Wald test” and “LIML Size of nominal 5% Wald test” contain critical values pertaining to Stock and Yogo’s (2005) second characterization of weak instruments. This characterization defines a set of instruments to be weak if a Wald test at the 5% level can have an actual rejection rate of no more than 10%, 15%, 20%, or 25%. Using the current example, suppose that we are willing to accept a rejection rate of at most 10%. Because $13.30 < 24.58$, we cannot reject the null hypothesis of weak instruments. On the other hand, if we use the LIML estimator instead, then we can reject the null hypothesis because $13.30 > 5.44$.

◀

□ Technical note

Stock and Yogo (2005) tabulated critical values for 2SLS relative biases of 5%, 10%, 20%, and 30% for models with 1, 2, or 3 endogenous regressors and between 3 and 30 excluded exogenous variables (instruments). They also provide critical values for worst-case rejection rates of 5%, 10%, 20%, and 25% for nominal 5% Wald tests of the endogenous regressors with 1 or 2 endogenous regressors and between 1 and 30 instruments. If the model previously fit by `ivregress` has more instruments or endogenous regressors than these limits, the critical values are not shown. Stock and Yogo did not consider GMM estimators.

□

When the model being fit contains more than one endogenous regressor, the R^2 and F statistics described above can overstate the relevance of the excluded instruments. Suppose that there are two endogenous regressors, Y_1 and Y_2 , and that there are two additional instruments, z_1 and z_2 . Say that z_1 is highly correlated with both Y_1 and Y_2 but z_2 is not correlated with either Y_1 or Y_2 . Then, the first-stage regression of Y_1 on z_1 and z_2 (along with the included exogenous variables) will produce large R^2 and F statistics, as will the regression of Y_2 on z_1 , z_2 , and the included exogenous variables. Nevertheless, the lack of correlation between z_2 and Y_1 and Y_2 is problematic. Here, although the order condition indicates that the model is just identified (the number of excluded instruments equals the number of endogenous regressors), the irrelevance of z_2 implies that the model is in fact not identified. Even if the model is overidentified, including irrelevant instruments can adversely affect the properties of instrumental-variables estimators, because their biases increase as the number of instruments increases.

▷ Example 4

`estat firststage` presents different statistics when the model contains multiple endogenous regressors. For illustration, we refit our model of rental rates, assuming that both `hsngval` and `faminc` are endogenously determined. We use `i.region` along with `popden`, a measure of population density, as additional instruments.

```
. ivregress 2sls rent pcturban (hsngval faminc = i.region popden)
(output omitted)
```

```
. estat firststage
```

Shea's partial R-squared

Variable	Shea's partial R-sq.	Shea's adj. partial R-sq.
hsngval	0.3477	0.2897
faminc	0.1893	0.1173

Minimum eigenvalue statistic = 2.51666

Critical Values	# of endogenous regressors:	2
H0: Instruments are weak	# of excluded instruments:	4

	5%	10%	20%	30%
2SLS relative bias	11.04	7.56	5.57	4.73
2SLS size of nominal 5% Wald test	16.87	9.93	7.54	6.28
LIML size of nominal 5% Wald test	4.72	3.39	2.99	2.79

Consider the endogenous regressor `hsngval`. Part of its variation is attributable to its correlation with the other regressors `pcturban` and `faminc`. The other component of `hsngval`'s variation is peculiar to it and orthogonal to the variation in the other regressors. Similarly, we can think of the instruments as predicting the variation in `hsngval` in two ways, one stemming from the fact that the predicted values of `hsngval` are correlated with the predicted values of the other regressors and one from the variation in the predicted values of `hsngval` that is orthogonal to the variation in the predicted values of the other regressors.

What really matters for instrumental-variables estimation is whether the component of `hsngval` that is orthogonal to the other regressors can be explained by the component of the predicted value of `hsngval` that is orthogonal to the predicted values of the other regressors in the model. Shea's (1997) partial

R^2 statistic measures this correlation. Because the bias of instrumental-variables estimators increases as more instruments are used, Shea's adjusted partial R^2 statistic is often used instead, as it makes a degrees-of-freedom adjustment for the number of instruments, analogous to the adjusted R^2 measure used in OLS regression. Although what constitutes a "low" value for Shea's partial R^2 depends on the specifics of the model being fit and the data used, these results, taken in isolation, do not strike us as being a particular cause for concern.

However, with this specification the minimum eigenvalue statistic is low. We cannot reject the null hypothesis of weak instruments for either of the characterizations we have discussed.

◀

By default, `estat firststage` determines which statistics to present based on the number of endogenous regressors in the model previously fit. However, you can specify the `all` option to obtain all the statistics.

□ Technical note

If the previous estimation was conducted using `aweight`s, `pweight`s, or `iweight`s, then the first-stage regression summary statistics are computed using those weights. However, in these cases the minimum eigenvalue statistic and its critical values are not available.

If the previous estimation included a robust VCE, then the first-stage F statistic is based on a robust VCE as well; for example, if you fit your model with a HAC VCE using the Bartlett kernel and four lags, then the F statistic reported is based on regression results using a HAC VCE using the Bartlett kernel and four lags. By default, the minimum eigenvalue statistic and its critical values are not displayed. You can use the `forcenonrobust` option to obtain them in these cases; the minimum eigenvalue statistic is computed using the weights, though the critical values reported may not be appropriate.

□

estat overid

In addition to the requirement that instrumental variables be correlated with the endogenous regressors, the instruments must also be uncorrelated with the structural error term. If the model is overidentified, meaning that the number of additional instruments exceeds the number of endogenous regressors, then we can test whether the instruments are uncorrelated with the error term. If the model is just identified, then we cannot perform a test of overidentifying restrictions.

The estimator you used to fit the model determines which tests of overidentifying restrictions `estat overid` reports. If you used the 2SLS estimator without a robust VCE, `estat overid` reports Sargan's (1958) and Basman's (1960) χ^2 tests. If you used the 2SLS estimator and requested a robust VCE, Wooldridge's robust score test of overidentifying restrictions is performed instead; without a robust VCE, Wooldridge's test statistic is identical to Sargan's test statistic. If you used the LIML estimator, `estat overid` reports the Anderson–Rubin (1950) likelihood-ratio test and Basman's (1960) F test. `estat overid` reports Hansen's (1982) J statistic if you used the GMM estimator. Davidson and MacKinnon (1993, 235–236) give a particularly clear explanation of the intuition behind tests of overidentifying restrictions. Also see Judge et al. (1985, 614–616) for a summary of tests of overidentifying restrictions for the 2SLS and LIML estimators.

Tests of overidentifying restrictions actually test two different things simultaneously. One, as we have discussed, is whether the instruments are uncorrelated with the error term. The other is that the equation is misspecified and that one or more of the excluded exogenous variables should in fact be included in the structural equation. Thus, a significant test statistic could represent either an invalid instrument or an incorrectly specified structural equation.

▶ Example 5

Here we refit the model that treated just `hsgnval` as endogenous using 2SLS, and then we perform tests of overidentifying restrictions:

```
. ivregress 2sls rent pcturban (hsgnval = faminc i.region)
(output omitted)
. estat overid
Tests of overidentifying restrictions:
Sargan (score) chi2(3) = 11.2877 (p = 0.0103)
Basman chi2(3) = 12.8294 (p = 0.0050)
```

Both test statistics are significant at the 5% test level, which means that either one or more of our instruments are invalid or that our structural model is specified incorrectly.

One possibility is that the error term in our structural model is heteroskedastic. Both Sargan's and Basman's tests assume that the errors are i.i.d.; if the errors are not i.i.d., then these tests are not valid. Here we refit the model by requesting heteroskedasticity-robust standard errors, and then we use `estat overid` to obtain Wooldridge's score test of overidentifying restrictions, which is robust to heteroskedasticity.

```
. ivregress 2sls rent pcturban (hsgnval = faminc i.region), vce(robust)
(output omitted)
. estat overid
Test of overidentifying restrictions:
Score chi2(3) = 6.8364 (p = 0.0773)
```

Here we no longer reject the null hypothesis that our instruments are valid at the 5% significance level, though we do reject the null at the 10% level. You can verify that the robust standard error on the coefficient for `hsgnval` is more than twice as large as its nonrobust counterpart and that the robust standard error for `pcturban` is nearly 50% larger.

◀

□ Technical note

The test statistic for the test of overidentifying restrictions performed after GMM estimation is simply the sample size times the value of the objective function $Q(\beta_1, \beta_2)$ defined in (5) of [R] `ivregress`, evaluated at the GMM parameter estimates. If the weight matrix \mathbf{W} is optimal, meaning that $\mathbf{W} = \text{Var}(\mathbf{z}_i u_i)$, then $Q(\beta_1, \beta_2) \overset{A}{\sim} \chi^2(q)$, where q is the number of overidentifying restrictions. However, if the estimated \mathbf{W} is not optimal, then the test statistic will not have an asymptotic χ^2 distribution.

Like the Sargan and Basman tests of overidentifying restrictions for the 2SLS estimator, the Anderson–Rubin and Basman tests after LIML estimation are predicated on the errors' being i.i.d. If the previous LIML results were reported with robust standard errors, then `estat overid` by default issues an error message and refuses to report the Anderson–Rubin and Basman test statistics. You can use the `forcenonrobust` option to override this behavior. You can also use `forcenonrobust` to obtain the Sargan and Basman test statistics after 2SLS estimation with robust standard errors.

□

By default, `estat overid` issues an error message if the previous estimation was conducted using `aweight`s, `pweight`s, or `iweight`s. You can use the `forceweights` option to override this behavior, though the test statistics may no longer have the expected χ^2 distributions.

estat weakrobust

Instrumental-variables methods require that the instruments be correlated with the endogenous regressors. In principle, this is a low bar: even instruments only weakly correlated with the endogenous regressors allow for valid asymptotic inference using standard instrumental-variables methods. But weak instruments can lead to misleading inference in practice, even in relatively large samples. This is especially the case when models are overidentified or there is a high degree of endogeneity.

When instruments are weak, first-stage coefficients are small relative to the variance of their estimates. This causes the distribution of instrumental-variables estimators to be highly nonnormal and thus poorly approximated by standard inference methods. (See [Andrews, Stock, and Sun \[2019\]](#).)

The linear instrumental-variables model fit by `ivregress` is written in matrix form as

$$\begin{aligned} \mathbf{y} &= \mathbf{Y}\beta_1 + \mathbf{X}_1\beta_2 + \mathbf{u} \\ \mathbf{Y} &= \mathbf{X}_1\Pi_1 + \mathbf{X}_2\Pi_2 + \mathbf{V} \end{aligned}$$

where \mathbf{y} is a vector with the values of the dependent variable in the sample, \mathbf{Y} is a matrix of endogenous regressors, \mathbf{X}_1 is a matrix of included exogenous regressors, and \mathbf{X}_2 is a matrix of excluded exogenous regressors. The first equation is referred to as the structural equation, and the second equation is referred to as the first-stage equation. By substituting the endogenous regressors \mathbf{Y} into the structural equation and reparameterizing, we derive the reduced-form equation

$$\mathbf{y} = \mathbf{X}_1\delta_1 + \mathbf{X}_2\delta_2 + \epsilon$$

where $\delta_1 \equiv \Pi_1\beta_1 + \beta_2$ and $\delta_2 \equiv \Pi_2\beta_1$.

The principle of the tests performed by `estat weakrobust` is that the reduced-form and first-stage equations are free of endogeneity. The estimates of δ_1 , δ_2 , Π_1 , and Π_2 can be tested for consistency with a null hypothesis about β_1 , our parameter of interest. (In practice, the exogenous covariates \mathbf{X}_2 are partialled out for simplicity.) `estat weakrobust` performs tests of the form $\beta_1 = \mathbf{r}_1$, where \mathbf{r}_1 is a constant vector of the same dimension as β_1 .

When the model is just identified, meaning there are no more instruments than there are endogenous regressors, the test of [Anderson and Rubin \(1949\)](#), or the associated confidence interval, is reported by default. When the model has been fit with a heteroskedastic, cluster-robust, or HAC VCE, an appropriate version of the Anderson–Rubin test or confidence interval is reported. In all cases, the Anderson–Rubin test statistic takes on a χ^2 distribution, and critical values and p -values can be exactly computed.

When the model is overidentified and has been fit with a homoskedastic VCE, the CLR test of [Moreira \(2003\)](#), or its associated confidence interval, is reported by default. The CLR test has better power properties than the Anderson–Rubin test in the overidentified case. When the model is fit with a nonhomoskedastic VCE, a generalization of the CLR test proposed by [Finlay and Magnusson \(2009\)](#) is used. This test is equivalent to the GMM quasi-LR statistic of [Kleibergen \(2007\)](#) in the linear instrumental-variables case considered here.

Regardless of the VCE specified, the CLR test statistic has a nonstandard distribution. However, when the model has been fit with a homoskedastic VCE and there is only a single endogenous regressor, the p -value for the CLR statistic can be computed exactly using the method of [Andrews, Moreira, and Stock \(2007\)](#). Otherwise, the p -value for the CLR test statistic is obtained by simulation.

Both the Anderson–Rubin and the CLR tests (and their CIs) can be requested for any model, even though only one test (or CI) is computed and reported by default. In the just-identified case, the CLR test is equivalent to the Anderson–Rubin test.

▷ Example 6

We revisit the model of average rental rate as a function of average housing values and the proportion of the population living in urban areas, this time supposing we do not have access to `faminc`, the variable for median family income. We may suspect that our region indicator variables are weak instruments for `hsngval` and therefore perform inference robust to weak instruments using `estat weakrobust`.

```
. ivregress 2sls rent pcturban (hsngval = i.region), vce(robust)
(output omitted)
. estat weakrobust, rseed(12345)
Test robust to weak instruments
Model VCE: Robust
( 1) hsngval = 0
Cond. likelihood-ratio (CLR) test = 5.48
      Prob > CLR = 0.0255
Notes: CLR test reported by default because
       model is overidentified.
       p-value computed by simulation
       (25,000 replications).
```

Here `estat weakrobust` reports a CLR test because there are more instruments than there are endogenous regressors. We find evidence to reject the null hypothesis that the coefficient on `hsngval` is zero, but the p -value is substantially larger than the p -value of 0.003 that is reported by `ivregress`. Note that because the model is fit with a nonhomoskedastic VCE and a CLR test is being reported, the p -value is computed by simulation. We have included a random-number seed using the `rseed()` option for reproducibility, but this is optional.

We can request a confidence interval that is robust to weak instruments for `hsngval` with the `ci` option:

```
. estat weakrobust, ci rseed(12345)
Searching for CI bounds:
Iteration 0: Grid points = 500
Iteration 1: Grid points = 1,000
(CI computed using 1,000 grid points on [-.003591, .006664])
Confidence interval robust to weak instruments
Model VCE: Robust
```

	Coefficient	CLR	
		[95% conf. interval]	
hsngval	.0015365	.0002263	.002857

```
Notes: CLR CI reported by default because model is
       overidentified.
       Computed using simulation (25,000
       replications).
```

The confidence interval comes close to containing zero. Indeed, if we specified a confidence level of 99% using `level(99)`, zero would be included. Accounting for the possibility of weak instruments, we may conclude that we cannot rule out a zero coefficient on `hsngval`.

◀

The `ci` option, which computes Anderson–Rubin and CLR confidence intervals, is available only in the single-endogenous regressor case. When the model has been fit with a homoskedastic VCE or is just identified, confidence intervals can be computed directly by inverting the relevant test. When the

model has been fit with a nonhomoskedastic VCE and is overidentified, there is no closed form for the endpoints of the conditional likelihood confidence interval, so a gridding procedure is used to estimate the confidence intervals.

The coefficient on the endogenous regressor, as fit by `ivregress`, is reported next to the confidence interval in the output of `estat weakrobust`, `ci` for reference. However, there is generally no guarantee that the estimated coefficient will be contained in the reported confidence interval. An exception is that when the model has been fit with homoskedastic errors, the CLR confidence set will contain the LIML estimator. In the just-identified case, the Anderson–Rubin confidence set will also contain the LIML estimator. See [Moreira \(2003\)](#) for a discussion.

□ Technical note

Unlike conventional confidence intervals, the confidence “intervals” produced by inverting Anderson–Rubin and CLR tests are not always finite intervals and may not be intervals at all. Confidence intervals can take one of five forms:

1. Finite interval $[a, b]$. When instruments are strong and the model is well identified, both Anderson–Rubin and CLR confidence intervals typically take this form.
2. Union of finite intervals $[a_1, b_1] \cup [a_2, b_2] \cup \dots \cup [a_m, b_m]$. Confidence intervals can take this form when the model VCE is nonhomoskedastic.
3. Union of (possibly infinite) intervals $(-\infty, b_1] \cup \dots \cup [a_m, +\infty)$. Confidence sets may take this form when instruments are weak. When the model VCE is homoskedastic, this form may only be the union of two infinite intervals.
4. Real line $(-\infty, +\infty)$. Confidence sets may take this form when instruments are weak.
5. Empty set \emptyset . An empty confidence interval means the test rejects every possible null value for the endogenous regressor, thus rejecting the model. The only case in which the model can be rejected in this way is when the model is overidentified and an Anderson–Rubin confidence interval is used. In other cases, an empty confidence interval means the grid used to estimate the confidence interval was too narrow, and the bounds should be widened using the `bound()` option.

For detailed discussion of unconventional confidence intervals in inference that is robust to weak instruments, see [Mikusheva and Poi \(2006\)](#), [Mikusheva \(2010\)](#), and [Kleibergen \(2007\)](#).

□

Stored results

After 2SLS estimation, `estat endogenous` stores the following in `r()`:

Scalars

<code>r(durbin)</code>	Durbin χ^2 statistic
<code>r(p_durbin)</code>	p -value for Durbin χ^2 statistic
<code>r(wu)</code>	Wu–Hausman F statistic
<code>r(p_wu)</code>	p -value for Wu–Hausman F statistic
<code>r(df)</code>	degrees of freedom
<code>r(wudf_r)</code>	denominator degrees of freedom for Wu–Hausman F
<code>r(r_score)</code>	robust score statistic
<code>r(p_r_score)</code>	p -value for robust score statistic
<code>r(hac_score)</code>	HAC score statistic
<code>r(p_hac_score)</code>	p -value for HAC score statistic
<code>r(lags)</code>	lags used in prewhitening
<code>r(regF)</code>	regression-based F statistic

r(p_regF)	<i>p</i> -value for regression-based <i>F</i> statistic
r(regFdf_n)	regression-based <i>F</i> numerator degrees of freedom
r(regFdf_r)	regression-based <i>F</i> denominator degrees of freedom

After GMM estimation, `estat endogenous` stores the following in `r()`:

Scalars

r(C)	<i>C</i> χ^2 statistic
r(p_C)	<i>p</i> -value for <i>C</i> χ^2 statistic
r(df)	degrees of freedom

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

Scalars

r(mineig)	minimum eigenvalue statistic
-----------	------------------------------

Matrices

r(mineigcv)	critical values for minimum eigenvalue statistic
r(multiresults)	Shea's partial R^2 statistics
r(singleresults)	first-stage R^2 and <i>F</i> statistics

After 2SLS estimation, `estat overid` stores the following in `r()`:

Scalars

r(lags)	lags used in prewhitening
r(df)	χ^2 degrees of freedom
r(score)	score χ^2 statistic
r(p_score)	<i>p</i> -value for score χ^2 statistic
r(basmann)	Basmann χ^2 statistic
r(p_basmann)	<i>p</i> -value for Basmann χ^2 statistic
r(sargan)	Sargan χ^2 statistic
r(p_sargan)	<i>p</i> -value for Sargan χ^2 statistic

After LIML estimation, `estat overid` stores the following in `r()`:

Scalars

r(ar)	Anderson–Rubin (1950) χ^2 statistic
r(p_ar)	<i>p</i> -value for Anderson–Rubin (1950) χ^2 statistic
r(ar_df)	χ^2 degrees of freedom
r(basmann)	Basmann <i>F</i> statistic
r(p_basmann)	<i>p</i> -value for Basmann <i>F</i> statistic
r(basmann_df_n)	<i>F</i> numerator degrees of freedom
r(basmann_df_d)	<i>F</i> denominator degrees of freedom

After GMM estimation, `estat overid` stores the following in `r()`:

Scalars

r(HansenJ)	Hansen's <i>J</i> χ^2 statistic
r(p_HansenJ)	<i>p</i> -value for Hansen's <i>J</i> χ^2 statistic
r(J_df)	χ^2 degrees of freedom

After estimation of a just-identified model, or when `ar` is specified, `estat weakrobust` stores the following in `r()`:

Scalars

r(archi2)	Anderson–Rubin (1949) χ^2 statistic
r(p_archi2)	<i>p</i> -value for Anderson–Rubin (1949) χ^2 statistic
r(archi2_df)	χ^2 degrees of freedom
r(arF)	Anderson–Rubin (1949) <i>F</i> statistic
r(p_arF)	<i>p</i> -value for Anderson–Rubin (1949) <i>F</i> statistic
r(arF_df_n)	<i>F</i> numerator degrees of freedom

<code>r(arF_df_d)</code>	F denominator degrees of freedom
<code>r(ar_ngrid)</code>	number of gridpoints used for Anderson–Rubin (1949) confidence interval
Macros	
<code>r(bound)</code>	user-specified grid bounds
Matrices	
<code>r(betanull)</code>	values of the null hypothesis for endogenous regressors
<code>r(table)</code>	matrix containing coefficient with reported confidence interval (omitted if CI is empty)
<code>r(ar_ci)</code>	matrix containing the bounds of the Anderson–Rubin (1949) confidence interval
<code>r(ar_finalbound)</code>	matrix containing the final upper and lower bounds used by the gridding procedure

After estimation of an overidentified model, or when `clr` is specified, `estat weakrobust` stores the following in `r()`:

Scalars	
<code>r(clr)</code>	CLR statistic of Moreira (2003)
<code>r(p_clr)</code>	p -value for CLR statistic
<code>r(J)</code>	J statistic of Finlay and Magnusson (2009) (when model VCE is nonhomoskedastic)
<code>r(rk)</code>	rank statistic of Kleibergen and Paap (2006) (when model VCE is nonhomoskedastic)
<code>r(reps)</code>	number of simulation replications (when results are simulated)
<code>r(rseed)</code>	random seed used (when results are simulated)
<code>r(clr_ngrid)</code>	number of gridpoints used for CLR confidence interval
Macros	
<code>r(bound)</code>	user-specified grid bounds
<code>r(rngstate)</code>	random-number state used (when results are simulated)
Matrices	
<code>r(betanull)</code>	values of the null hypothesis for endogenous regressors
<code>r(Sbar)</code>	\bar{S} statistic of Moreira (2003) (when model VCE is homoskedastic)
<code>r(Tbar)</code>	\bar{T} statistic of Moreira (2003) (when model VCE is homoskedastic)
<code>r(table)</code>	matrix containing coefficient with reported confidence interval (omitted if CI is empty)
<code>r(clr_ci)</code>	matrix containing the bounds of the CLR confidence interval
<code>r(clr_finalbound)</code>	matrix containing the final upper and lower bounds used by the gridding procedure

Methods and formulas

Methods and formulas are presented under the following headings:

Notation
estat endogenous
estat firststage
estat overid
estat weakrobust
Homoskedastic errors
Nonhomoskedastic errors
Confidence intervals

Notation

Recall from [\[R\] ivregress](#) that the model is

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta}_1 + \mathbf{X}_1\boldsymbol{\beta}_2 + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$\mathbf{Y} = \mathbf{X}_1\Pi_1 + \mathbf{X}_2\Pi_2 + \mathbf{V} = \mathbf{Z}\Pi + \mathbf{V}$$

where \mathbf{y} is an $N \times 1$ vector of the left-hand-side variable, N is the sample size, \mathbf{Y} is an $N \times p$ matrix of p endogenous regressors, \mathbf{X}_1 is an $N \times k_1$ matrix of k_1 included exogenous regressors, \mathbf{X}_2 is an $N \times k_2$ matrix of k_2 excluded exogenous variables, $\mathbf{X} = [\mathbf{Y} \ \mathbf{X}_1]$, $\mathbf{Z} = [\mathbf{X}_1 \ \mathbf{X}_2]$, \mathbf{u} is an $N \times 1$ vector of errors, \mathbf{V} is an $N \times p$ matrix of errors, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2]$ is a $k = (p + k_1) \times 1$ vector of parameters, and $\boldsymbol{\Pi}$ is a $(k_1 + k_2) \times p$ vector of parameters. If a constant term is included in the model, then one column of \mathbf{X}_1 contains all ones.

estat endogenous

Partition \mathbf{Y} as $\mathbf{Y} = [\mathbf{Y}_1 \ \mathbf{Y}_2]$, where \mathbf{Y}_1 represents the p_1 endogenous regressors whose endogeneity is being tested and \mathbf{Y}_2 represents the p_2 endogenous regressors whose endogeneity is not being tested. If the endogeneity of all endogenous regressors is being tested, $\mathbf{Y} = \mathbf{Y}_1$ and $p_2 = 0$. After GMM estimation, `estat endogenous` refits the model treating \mathbf{Y}_1 as exogenous using the same type of weight matrix as requested at estimation time with the `wmatrix()` option; denote the Sargan statistic from this model by J_e and the estimated weight matrix by \mathbf{W}_e . Let $\mathbf{S}_e = \mathbf{W}_e^{-1}$. `estat endogenous` removes from \mathbf{S}_e the rows and columns corresponding to the variables represented by \mathbf{Y}_1 ; denote the inverse of the resulting matrix by \mathbf{W}'_e . Next, `estat endogenous` fits the model treating both \mathbf{Y}_1 and \mathbf{Y}_2 as endogenous, using the weight matrix \mathbf{W}'_e ; denote the Sargan statistic from this model by J_c . Then, $C = (J_e - J_c) \sim \chi^2(p_1)$. If one simply used the J statistic from the original model fit by `ivregress` in place of J_c , then in finite samples $J_e - J$ might be negative. The procedure used by `estat endogenous` is guaranteed to yield $C \geq 0$; see [Hayashi \(2000, 220\)](#).

Let $\hat{\mathbf{u}}_c$ denote the residuals from the model treating both \mathbf{Y}_1 and \mathbf{Y}_2 as endogenous, and let $\hat{\mathbf{u}}_e$ denote the residuals from the model treating only \mathbf{Y}_2 as endogenous. Then, Durbin's (1954) statistic is

$$D = \frac{\hat{\mathbf{u}}'_e \mathbf{P}_{Z\mathbf{Y}_1} \hat{\mathbf{u}}_e - \hat{\mathbf{u}}'_c \mathbf{P}_Z \hat{\mathbf{u}}_c}{\hat{\mathbf{u}}'_e \hat{\mathbf{u}}_e / N}$$

where $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{P}_{Z\mathbf{Y}_1} = [\mathbf{Z} \ \mathbf{Y}_1][(\mathbf{Z} \ \mathbf{Y}_1)'(\mathbf{Z} \ \mathbf{Y}_1)]^{-1}[\mathbf{Z} \ \mathbf{Y}_1]'$. $D \sim \chi^2(p_1)$. The Wu–Hausman ([Wu 1974](#); [Hausman 1978](#)) statistic is

$$WH = \frac{(\hat{\mathbf{u}}'_e \mathbf{P}_{Z\mathbf{Y}_1} \hat{\mathbf{u}}_e - \hat{\mathbf{u}}'_c \mathbf{P}_Z \hat{\mathbf{u}}_c) / p_1}{\{\hat{\mathbf{u}}'_e \hat{\mathbf{u}}_e - (\hat{\mathbf{u}}'_c \mathbf{P}_{Z\mathbf{Y}_1} \hat{\mathbf{u}}_e - \hat{\mathbf{u}}'_c \mathbf{P}_Z \hat{\mathbf{u}}_c)\} / (N - k_1 - p - p_1)}$$

$WH \sim F(p_1, N - k_1 - p - p_1)$. [Baum, Schaffer, and Stillman \(2003, 2007\)](#) discuss these tests in more detail.

Next, we describe Wooldridge's (1995) score test. The nonrobust version of Wooldridge's test is identical to Durbin's test. Suppose a robust covariance matrix was used at estimation time. Let $\hat{\mathbf{e}}$ denote the sample residuals obtained by fitting the model via OLS, treating \mathbf{Y} as exogenous. We then regress each variable represented in \mathbf{Y} on \mathbf{Z} ; call the residuals for the j th regression $\hat{\mathbf{r}}_j$, $j = 1, \dots, p$. Define $\hat{k}_{ij} = \hat{e}_i \hat{r}_{ij}$, $i = 1, \dots, N$. We then run the regression

$$\mathbf{1} = \theta_1 \hat{\mathbf{k}}_1 + \dots + \theta_p \hat{\mathbf{k}}_p + \boldsymbol{\epsilon}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones and $\boldsymbol{\epsilon}$ is a regression error term. $N - \text{RSS} \sim \chi^2(p)$, where RSS is the residual sum of squares from the regression just described. If instead a HAC VCE was used at estimation time, then before running the final regression we prewhiten the $\hat{\mathbf{k}}_j$ series by using a `VAR(q)` model, where q is the number of lags specified with the `lags()` option.

The regression-based test proceeds as follows. Following Hausman (1978, 1259), we regress \mathbf{Y} on \mathbf{Z} and obtain the residuals $\widehat{\mathbf{V}}$. Next, we fit the augmented regression

$$\mathbf{y} = \mathbf{Y}\beta_1 + \mathbf{X}_1\beta_2 + \widehat{\mathbf{V}}\gamma + \epsilon$$

by OLS regression, where ϵ is a regression error term. A test of the exogeneity of \mathbf{Y} is equivalent to a test of $\gamma = \mathbf{0}$. As Cameron and Trivedi (2005, 276) suggest, this test can be made robust to heteroskedasticity, autocorrelation, or clustering by using the appropriate robust VCE when testing $\gamma = \mathbf{0}$. When a nonrobust VCE is used, this test is equivalent to the Wu–Hausman test described earlier. One cannot simply fit this augmented regression via 2SLS to test the endogeneity of a subset of the endogenous regressors; Davidson and MacKinnon (1993, 229–231) discuss a test of $\gamma = \mathbf{0}$ for the homoskedastic version of the augmented regression fit by 2SLS, but an appropriate robust test is not apparent.

estat firststage

When the structural equation includes one endogenous regressor, `estat firststage` fits the regression

$$\mathbf{Y} = \mathbf{X}_1\pi_1 + \mathbf{X}_2\pi_2 + \mathbf{v}$$

via OLS. The R^2 and adjusted R^2 from that regression are reported in the output, as well as the F statistic from the Wald test of $H_0: \pi_2 = \mathbf{0}$. To obtain the partial R^2 statistic, `estat firststage` fits the regression

$$\mathbf{M}_{\mathbf{X}_1}\mathbf{y} = \mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2\xi + \epsilon$$

by OLS, where ϵ is a regression error term, ξ is a $k_2 \times 1$ parameter vector, and $\mathbf{M}_{\mathbf{X}_1} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$; that is, the partial R^2 is the R^2 between \mathbf{y} and \mathbf{X}_2 after eliminating the effects of \mathbf{X}_1 . If the model contains multiple endogenous regressors and the `all` option is specified, these statistics are calculated for each endogenous regressor in turn.

To calculate Shea’s partial R^2 , let \mathbf{y}_1 denote the endogenous regressor whose statistic is being calculated and \mathbf{Y}_0 denote the other endogenous regressors. Define $\tilde{\mathbf{y}}_1$ as the residuals obtained from regressing \mathbf{y}_1 on \mathbf{Y}_0 and \mathbf{X}_1 . Let $\hat{\mathbf{y}}_1$ denote the fitted values obtained from regressing \mathbf{y}_1 on \mathbf{X}_1 and \mathbf{X}_2 ; that is, $\hat{\mathbf{y}}_1$ are the fitted values from the first-stage regression for \mathbf{y}_1 , and define the columns of $\widehat{\mathbf{Y}}_0$ analogously. Finally, let $\tilde{\tilde{\mathbf{y}}}_1$ denote the residuals from regressing $\hat{\mathbf{y}}_1$ on $\widehat{\mathbf{Y}}_0$ and \mathbf{X}_1 . Shea’s partial R^2 is the simple R^2 from the regression of $\tilde{\tilde{\mathbf{y}}}_1$ on $\tilde{\tilde{\mathbf{y}}}_1$; denote this as R_S^2 . Shea’s adjusted partial R^2 is equal to $1 - (1 - R_S^2)(N - 1)/(N - k_Z + 1)$ if a constant term is included and $1 - (1 - R_S^2)(N - 1)/(N - k_Z)$ if there is no constant term included in the model, where $k_Z = k_1 + k_2$. For one endogenous regressor, one instrument, no exogenous regressors, and a constant term, R_S^2 equals the adjusted R_S^2 .

The Stock and Yogo minimum eigenvalue statistic, first proposed by Cragg and Donald (1993) as a test for underidentification, is the minimum eigenvalue of the matrix

$$\mathbf{G} = \frac{1}{k_Z} \widehat{\Sigma}_{\mathbf{V}\mathbf{V}}^{-1/2} \mathbf{Y}'\mathbf{M}'_{\mathbf{X}_1} \mathbf{X}_2(\mathbf{X}'_2\mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{M}_{\mathbf{X}_1}\mathbf{Y}\widehat{\Sigma}_{\mathbf{V}\mathbf{V}}^{-1/2}$$

where

$$\widehat{\Sigma}_{\mathbf{V}\mathbf{V}} = \frac{1}{N - k_Z} \mathbf{Y}'\mathbf{M}_Z\mathbf{Y}$$

$\mathbf{M}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, and $\mathbf{Z} = [\mathbf{X}_1 \ \mathbf{X}_2]$. Critical values are obtained from the tables in Stock and Yogo (2005).

estat overid

The Sargan (1958) and Basman (1960) χ^2 statistics are calculated by running the auxiliary regression

$$\hat{\mathbf{u}} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$$

where $\hat{\mathbf{u}}$ are the sample residuals from the model and \mathbf{e} is an error term. Then, Sargan's statistic is

$$S = N \left(1 - \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \right)$$

where $\hat{\mathbf{e}}$ are the residuals from that auxiliary regression. Basman's statistic is calculated as

$$B = S \frac{N - k_Z}{N - S}$$

Both S and B are distributed $\chi^2(m)$, where m , the number of overidentifying restrictions, is equal to $k_Z - k$, where k is the number of endogenous regressors.

Wooldridge's (1995) score test of overidentifying restrictions is identical to Sargan's (1958) statistic under the assumption of i.i.d. and therefore is not recomputed unless a robust VCE was used at estimation time. If a heteroskedasticity-robust VCE was used, Wooldridge's test proceeds as follows. Let $\hat{\mathbf{Y}}$ denote the $N \times k$ matrix of fitted values obtained by regressing the endogenous regressors on \mathbf{X}_1 and \mathbf{X}_2 . Let \mathbf{Q} denote an $N \times m$ matrix of excluded exogenous variables; the test statistic to be calculated is invariant to whichever m of the k_2 excluded exogenous variables is chosen. Define the i th element of $\hat{\mathbf{k}}_j$, $i = 1, \dots, N$, $j = 1, \dots, m$, as

$$k_{ij} = \hat{q}_{ij} \hat{u}_i$$

where \hat{q}_{ij} is the i th element of $\hat{\mathbf{q}}_j$, the residuals from regressing the j th column of \mathbf{Q} on $\hat{\mathbf{Y}}$ and \mathbf{X}_1 . Finally, fit the regression

$$\mathbf{1} = \theta_1 \hat{\mathbf{k}}_1 + \dots + \theta_m \hat{\mathbf{k}}_m + \boldsymbol{\epsilon}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones and $\boldsymbol{\epsilon}$ is a regression error term, and calculate the residual sum of squares, RSS. Then, the test statistic is $W = N - \text{RSS}$. $W \sim \chi^2(m)$. If a HAC VCE was used at estimation, then the $\hat{\mathbf{k}}_j$ are prewhitened using a VAR(p) model, where p is specified using the `lags()` option.

The Anderson–Rubin (AR; 1950) test of overidentifying restrictions for use after the LIML estimator is calculated as $\text{AR} = N(\kappa - 1)$, where κ is the minimal eigenvalue of a certain matrix defined in *Methods and formulas* of [R] **ivregress**. $\text{AR} \sim \chi^2(m)$. (Some texts define this statistic as $N \ln(\kappa)$ because $\ln(x) \approx (x - 1)$ for x near 1.) Basman's F statistic for use after the LIML estimator is calculated as $B_F = (\kappa - 1)(N - k_Z)/m$. $B_F \sim F(m, N - k_Z)$.

Hansen's J statistic is simply the sample size times the value of the GMM objective function defined in (5) of [R] **ivregress**, evaluated at the estimated parameter values. Under the null hypothesis that the overidentifying restrictions are valid, $J \sim \chi^2(m)$.

John Denis Sargan (1924–1996) was born in Yorkshire, UK. He pioneered the theory of instrumental-variables (IV) estimation in an article published in 1958. In the article, he also developed overidentification tests, developed significance tests, and discussed possible instruments for applied work. A year later, he wrote an article extending the theory to models containing autoregressive errors. This extension was one of his many contributions to time-series econometric analysis. For example, in 1964 he published a paper in which he developed misspecification tests for dynamic equations, along with an IV estimator for models with nonlinear parameters, and a model with a long-run equilibrium. His paper laid the foundation for other econometric methods, such as cointegration analysis, and established what would be known as the London School of Economics (LSE) approach to econometric modeling. He spent twenty years at this institution, supervising the doctoral work of many econometricians who themselves made important contributions to econometrics. In addition to Sargan’s many lasting contributions to econometrics, he also left a lasting impression on his students and colleagues through his generosity.

estat weakrobust

In *estat weakrobust* of *Remarks and examples*, we reintroduced the model fit by `ivregress`, written in matrix form as

$$\begin{aligned} \mathbf{y} &= \mathbf{Y}\boldsymbol{\beta}_1 + \mathbf{X}_1\boldsymbol{\beta}_2 + \mathbf{u} \\ \mathbf{Y} &= \mathbf{X}_1\boldsymbol{\Pi}_1 + \mathbf{X}_2\boldsymbol{\Pi}_2 + \mathbf{V} \end{aligned}$$

The first equation is referred to as the structural equation, and the second equation is referred to as the first-stage equation.

We then derived the reduced-form equation. Below, we write the model in the form of the reduced-form equation and the first-stage equations

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_1\boldsymbol{\delta}_1 + \mathbf{X}_2\boldsymbol{\delta}_2 + \boldsymbol{\epsilon} \\ \mathbf{Y} &= \mathbf{X}_1\boldsymbol{\Pi}_1 + \mathbf{X}_2\boldsymbol{\Pi}_2 + \mathbf{V} \end{aligned}$$

where $\boldsymbol{\delta}_1 \equiv \boldsymbol{\Pi}_1\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$ and $\boldsymbol{\delta}_2 \equiv \boldsymbol{\Pi}_2\boldsymbol{\beta}_1$. The reduced-form coefficients $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are $k_1 \times 1$ and $k_2 \times 1$, respectively. Without loss of generality, we can rewrite the model to omit included exogenous regressors \mathbf{X}_1 by considering \mathbf{y} , \mathbf{Y} , \mathbf{X}_2 , $\boldsymbol{\epsilon}$, and \mathbf{V} to have been replaced by their partialled-out equivalents:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_2\boldsymbol{\delta}_2 + \boldsymbol{\epsilon} \\ \mathbf{Y} &= \mathbf{X}_2\boldsymbol{\Pi}_2 + \mathbf{V} \end{aligned}$$

Our null hypothesis $\boldsymbol{\beta}_1 = \mathbf{r}_1$ implies that $\boldsymbol{\delta}_2 - \boldsymbol{\Pi}_2\mathbf{r}_1 = 0$.

Homoskedastic errors

Let $\widehat{\boldsymbol{\Omega}} = 1/N \begin{bmatrix} \tilde{\boldsymbol{\epsilon}}'\tilde{\boldsymbol{\epsilon}} & \tilde{\boldsymbol{\epsilon}}'\widehat{\mathbf{V}} \\ \widehat{\mathbf{V}}'\tilde{\boldsymbol{\epsilon}} & \widehat{\mathbf{V}}'\widehat{\mathbf{V}} \end{bmatrix}$ be the estimated $(p+1) \times (p+1)$ covariance matrix of $(\boldsymbol{\epsilon}', \mathbf{V}')'$. When the model is assumed to have homoskedastic errors, we can proceed by computing the statistics

$$\bar{S} = (\mathbf{X}_2'\mathbf{X}_2)^{-1/2}\mathbf{X}_2'(\mathbf{y} - \mathbf{Y}\mathbf{r}_1)(\mathbf{b}_0'\widehat{\boldsymbol{\Omega}}\mathbf{b}_0)^{-1/2}$$

and

$$\bar{T} = (\mathbf{X}'_2 \mathbf{X}_2)^{-1/2} \mathbf{X}'_2 [\mathbf{y}, \mathbf{Y}']' \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{A}_0 (\mathbf{A}'_0 \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{A}_0)^{-1/2}$$

where $\mathbf{b}_0 = [1, -\mathbf{r}_1]$ and $\mathbf{A}_0 = [\mathbf{r}_1, \mathbf{I}_p]'$.

The Anderson–Rubin (1949) statistic is computed as $AR = \bar{S}' \bar{S}$ and follows a $\chi^2(k_2)$ distribution. When the model has been fit with the `small` option in `ivregress`, an F version of the statistic is returned, which is distributed $F(k_2, N - k_1 - k_2)$. This F statistic is computed using $N/(N - k_1 - k_2) \widehat{\boldsymbol{\Omega}}$ in place of $\widehat{\boldsymbol{\Omega}}$.

The CLR statistic is computed as $CLR = \bar{S}' \bar{S} - \bar{\lambda}^{\min}$, where $\bar{\lambda}^{\min}$ is the minimum eigenvalue of the matrix $(\bar{S}, \bar{T})' (\bar{S}, \bar{T})$. When the model is fit with the `small` option in `ivregress`, the statistic is computed using $N/(N - k_1 - k_2) \widehat{\boldsymbol{\Omega}}$. When the model is just identified, $CLR \sim \chi^2(k_2)$. When the model is overidentified, CLR has a nonstandard distribution. The numerical method of Andrews, Moreira, and Stock (2007) is used to obtain p -values in the single-endogenous regressor case. In the multiple-endogenous regressor case, p -values are obtained by simulation conditional on \bar{T} , as outlined in Moreira (2003).

Nonhomoskedastic errors

When the model is fit with a nonhomoskedastic VCE, let

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{\delta} & \widehat{\boldsymbol{\Sigma}}_{\delta \boldsymbol{\Pi}} \\ \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Pi} \delta} & \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Pi}} \end{bmatrix}$$

be the $k_2(p + 1) \times k_2(p + 1)$ estimated covariance matrix of $\{\delta'_2, \text{Vec}(\boldsymbol{\Pi}_2)'\}'$, where $\text{Vec}(\cdot)$ is the column-major vectorization operator.

We compute the Anderson–Rubin statistic as

$$AR(\mathbf{r}_1) = (\mathbf{y} - \mathbf{Y}\mathbf{r}_1)' \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \widehat{\boldsymbol{\Psi}}(\mathbf{r}_1)^{-1} (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 (\mathbf{y} - \mathbf{Y}\mathbf{r}_1)$$

where $\widehat{\boldsymbol{\Psi}}(\mathbf{r}_1) = (\mathbf{b}_0 \otimes \mathbf{I}_k)' \widehat{\boldsymbol{\Sigma}} (\mathbf{b}_0 \otimes \mathbf{I}_k)$.

As in the homoskedastic case, AR takes on a $\chi^2(k_2)$ distribution under the null. When the model is fit with the `small` option, the statistic is computed using a small-sample adjustment, and an F version of the statistic is returned.

To compute the CLR test statistic, let $J(\mathbf{r}_1)$ be defined as in Finlay and Magnusson (2009) (or, equivalently, as in Kleibergen [2007]), and let $\text{rk}(\mathbf{r}_1)$ be the rank statistic of Kleibergen and Paap (2006) (the forms of these statistics are omitted here for brevity). The test statistic is given by

$$CLR(\mathbf{r}_1) = \frac{1}{2} \left[AR(\mathbf{r}_1) - \text{rk}(\mathbf{r}_1) + \sqrt{\{AR(\mathbf{r}_1) + \text{rk}(\mathbf{r}_1)\}^2 - 4J(\mathbf{r}_1)\text{rk}(\mathbf{r}_1)} \right]$$

In the just-identified case, CLR is equal to AR and has a $\chi^2(k_2)$ distribution. Otherwise, CLR has a nonstandard distribution, and p -values are obtained by simulation, as outlined in Kleibergen (2007). When the model is fit with the `small` option, the statistic is computed using a small-sample adjustment.

Confidence intervals

Confidence intervals are obtained by inverting the relevant tests. When the model is fit with a homoskedastic VCE, or when it is just identified, the Anderson–Rubin test can be inverted analytically, so the bounds of the resulting interval (or union of intervals) are computed as a closed form (see Mikushcheva [2010] for details). When the model VCE is nonhomoskedastic and the model is overidentified, the Anderson–Rubin confidence interval is obtained by a gridding procedure.

The CLR test is equivalent to the Anderson–Rubin test in the just-identified case, so its confidence intervals can be computed in the same way. In the overidentified case when errors are homoskedastic, the CLR test is inverted numerically using the method of Mikusheva (2010). When errors are nonhomoskedastic, a gridding procedure is used to obtain a confidence interval.

Acknowledgments

We thank Keith Finlay of the US Census Bureau, Leandro Magnusson of the University of Western Australia, and Mark Schaffer of Heriot-Watt University for their Stata commands `rivtest`, discussed in Finlay and Magnusson (2009), and `weakiv`.

We thank Anna Mikusheva of MIT and Brian Poi of Poi Consulting LLC for their Stata command `condivreg`, discussed in Mikusheva and Poi (2006).

References

- Anderson, T. W., and H. Rubin. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20: 46–63. <https://doi.org/10.1214/aoms/1177730090>.
- . 1950. The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 21: 570–582. <https://doi.org/10.1214/aoms/1177729752>.
- Andrews, D. W. K., M. J. Moreira, and J. H. Stock. 2007. Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics* 139: 116–132. <https://doi.org/10.1016/j.jeconom.2006.06.007>.
- Andrews, I., J. H. Stock, and L. Sun. 2019. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11: 727–753. <https://doi.org/10.1146/annurev-economics-080218-025643>.
- Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.
- Basmann, R. L. 1960. On finite sample distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association* 55: 650–659. <https://doi.org/10.2307/2281588>.
- Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1–31.
- . 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7: 465–506.
- Bound, J., D. A. Jaeger, and R. M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443–450. <https://doi.org/10.2307/2291055>.
- Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Chernozhukov, V., and C. B. Hansen. 2008. The reduced form: A simple approach to inference with weak instruments. *Economics Letters* 100: 68–71. <https://doi.org/10.1016/j.econlet.2007.11.012>.
- Cragg, J. G., and S. G. Donald. 1993. Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9: 222–240. <https://doi.org/10.1017/S0266466600007519>.
- Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Durbin, J. 1954. Errors in variables. *Review of the International Statistical Institute* 22: 23–32. <https://doi.org/10.2307/1401917>.
- Finlay, K., and L. M. Magnusson. 2009. Implementing weak-instrument robust tests for a general class of instrumental-variables models. *Stata Journal* 9: 398–421.
- Hahn, J., and J. A. Hausman. 2003. Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review Papers and Proceedings* 93: 118–125.

- Hall, A. R., G. D. Rudebusch, and D. W. Wilcox. 1996. Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37: 283–298. <https://doi.org/10.2307/2527324>.
- Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054. <https://doi.org/10.2307/1912775>.
- Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271. <https://doi.org/10.2307/1913827>.
- Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.
- Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: Wiley.
- Kleibergen, F. 2007. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* 139: 181–216. <https://doi.org/10.1016/j.jeconom.2006.06.010>.
- Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133: 97–126. <https://doi.org/10.1016/j.jeconom.2005.02.011>.
- Mikusheva, A. 2010. Robust confidence sets in the presence of weak instruments. *Journal of Econometrics* 157: 236–247. <https://doi.org/10.1016/j.jeconom.2009.12.003>.
- Mikusheva, A., and B. P. Poi. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 6: 335–347.
- Moreira, M. J. 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71: 1027–1048. <https://doi.org/10.1111/1468-0262.00438>.
- Nelson, C. R., and R. Startz. 1990. The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one. *Journal of Business* 63: S125–S140.
- Pflueger, C. E., and S. Wang. 2015. A robust test for weak instruments in Stata. *Stata Journal* 15: 216–225.
- Poi, B. P. 2006. Jackknife instrumental variables estimation in Stata. *Stata Journal* 6: 364–376.
- Sargan, J. D. 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26: 393–415. <https://doi.org/10.2307/1907619>.
- Shea, J. S. 1997. Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics* 79: 348–352. <https://doi.org/10.1162/rest.1997.79.2.348>.
- Staiger, D. O., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65: 557–586. <https://doi.org/10.2307/2171753>.
- Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518–529. <https://doi.org/10.1198/073500102288618658>.
- Stock, J. H., and M. Yogo. 2005. “Testing for weak instruments in linear IV regression”. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. W. K. Andrews and J. H. Stock, 80–108. New York: Cambridge University Press.
- Sun, L. 2018. Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models. *Stata Journal* 18: 803–825.
- Wooldridge, J. M. 1995. “Score diagnostics for linear models estimated by two stage least squares”. In *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*, edited by G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66–87. Oxford: Blackwell.
- Wu, D.-M. 1974. Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. *Econometrica* 42: 529–546. <https://doi.org/10.2307/1911789>.

Also see

[R] **ivregress** — Single-equation instrumental-variables regression

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

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).