## Postestimation commands

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

| Command | Description |
| --- | --- |
| estat overid | perform test of overidentifying restrictions |

The following standard postestimation commands are also available:

| Command | Description |
| --- | --- |
| contrast | contrasts and ANOVA-style joint tests of parameters |
| estat vce | variance–covariance matrix of the estimators (VCE) |
| estimates | cataloging estimation results |
| etable | table of estimation results |
| hausman | Hausman's specification test |
| lincom | point estimates, standard errors, testing, and inference for linear combinations of parameters |
| margins | marginal means, predictive margins, marginal effects, and average marginal effects |
| marginsplot | graph the results from margins (profile plots, interaction plots, etc.) |
| nlcom | point estimates, standard errors, testing, and inference for nonlinear combinations of parameters |
| predict | residuals |
| predictnl | point estimates, standard errors, testing, and inference for generalized predictions |
| pwcompare | pairwise comparisons of parameters |
| test | Wald tests of simple and composite linear hypotheses |
| testnl | Wald tests of nonlinear hypotheses |

# predict

## Description for predict

predict creates a new variable containing predictions such as linear predictions, residuals, and scores.

## Menu for predict

Statistics > Postestimation

## Syntax for predict

> predict [ *type* ] *newvar* [ *if* ] [ *in* ] [ , xb <u>eq</u>uation(*eqno* | *eqname*) ]

> predict [ *type* ] *newvar* [ *if* ] [ *in* ], <u>r</u>esiduals [ <u>eq</u>uation(*eqno* | *eqname*) ]

> predict [ *type* ] { *stub*\* | *newvar*₁ ... *newvar*$_q$ } [ *if* ] [ *in* ] [ , <u>r</u>esiduals ]

> predict [ *type* ] *stub*\* [ *if* ] [ *in* ], <u>sc</u>ores

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

Scores are available only for observations within the estimation sample.

You specify one new variable and (optionally) equation(), or you specify *stub*\* or $q$ or $p$ new variables, where $q$ is the number of moment equations and $p$ is the number of parameters in the model.

## Option for predict

    Main

xb, the default, calculates the linear prediction.

residuals calculates the residuals, the predicted values of the moment equations. This option requires that the length of the new variable list be equal to the number of moment equations, $q$. Otherwise, use *stub*\* to have predict generate enumerated variables with prefix *stub*. If equation() is not specified, the $j$th new variable will contain the residuals for the $j$th moment equation.

equation(*eqno* | *eqname*) with xb refers to a linear prediction in the model, whereas equation(*eqno* | *eqname*) with residuals refers to a moment equation.

For xb, specifying equation(#1) indicates that the calculation is to be made for the first linear prediction within curly braces. Specifying equation(xb1) would indicate that the calculation is to be made for the linear prediction xb1, assuming there is a linear prediction named xb1 in the model, for instance, {xb1: *varlist*}.

For `residuals`, specifying `equation(#1)` indicates that the calculation is to be made for the first moment equation. Specifying `equation(demand)` would indicate that the calculation is to be made for the moment equation named `demand`, assuming there is a moment equation named `demand` in the model.

If you specify one new variable name and omit `equation()`, results are the same as if you had specified `equation(#1)`.

For more information on using `predict` after multiple-equation estimation commands, see [R] **predict**.

`scores` calculates the parameter-level score equations, the first derivatives of the GMM criterion function with respect to the parameters scaled by $-0.5$. This option requires that the length of the new variable list be equal to the number of parameters, $p$. Otherwise, use *stub\** to have `predict` generate enumerated variables with prefix *stub*. The $j$th new variable will contain the $j$th score of the model.

# margins

## Description for margins

`margins` estimates margins of response for linear predictions.

## Menu for margins

Statistics > Postestimation

## Syntax for margins

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

margins [ *marginlist* ] , <u>pre</u>dict(*statistic* ...) [ <u>pre</u>dict(*statistic* ...) ... ] [ *options* ]

| *statistic* | Description |
|---|---|
| xb | linear prediction; the default |
| <u>residuals</u> | not allowed with `margins` |

xb defaults to the first equation.

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 overid reports Hansen's $J$ statistic, which is used to determine the validity of the overidentifying restrictions in a GMM model. If the model is correctly specified in the sense that $E\{\mathbf{z}_i u_i(\boldsymbol{\beta})\} = \mathbf{0}$, then the sample analog to that condition should hold at the estimated value of $\boldsymbol{\beta}$. Hansen's $J$ statistic is valid only if the weight matrix is optimal, meaning that it equals the inverse of the covariance matrix of the moment conditions. Therefore, estat overid only reports Hansen's $J$ statistic after two-step or iterated estimation, or if you specified winitial(*matname*) when calling gmm. In the latter case, it is your responsibility to determine the validity of the $J$ statistic.

## Menu for estat

Statistics > Postestimation

## Syntax for estat

```
estat overid
```

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

# Remarks and examples

As we noted in *Introduction* of [R] **gmm,** underlying generalized method of moments (GMM) estimators is a set of $l$ moment conditions, $E\{\mathbf{z}_i u_i(\boldsymbol{\beta})\} = \mathbf{0}$. When $l$ is greater than the number of parameters, $k$, any size-$k$ subset of the moment conditions would yield a consistent parameter estimate. We remarked that the parameter estimates we would obtain would in general depend on which $k$ moment conditions we used. However, if all our moment conditions are indeed valid, then the parameter estimates should not differ too much regardless of which $k$ moment conditions we used to estimate the parameters. The test of overidentifying restrictions is a model specification test based on this observation. The test of overidentifying restrictions requires that the number of moment conditions be greater than the number of parameters in the model.

Recall that the GMM criterion function is

$$Q = \left\{ \frac{1}{N} \sum_i \mathbf{z}_i u_i(\boldsymbol{\beta}) \right\}' \mathbf{W} \left\{ \frac{1}{N} \sum_i \mathbf{z}_i u_i(\boldsymbol{\beta}) \right\}$$

The test of overidentifying restrictions is remarkably simple. If $\mathbf{W}$ is an optimal weight matrix, under the null hypothesis $H_0 \colon E\{\mathbf{z}_i u_i(\boldsymbol{\beta})\} = \mathbf{0}$, the test statistic $J = N \times Q \sim \chi^2(l-k)$. A large test statistic casts doubt on the null hypothesis.

For the test to be valid, $\mathbf{W}$ must be optimal, meaning that $\mathbf{W}$ must be the inverse of the covariance matrix of the moment conditions:

$$\mathbf{W}^{-1} = E\{\mathbf{z}_i u_i(\boldsymbol{\beta}) u_i'(\boldsymbol{\beta}) \mathbf{z}_i'\}$$

Therefore, `estat overid` works only after the two-step and iterated estimators, or if you supplied your own initial weight matrix by using the `winitial(matname)` option to `gmm` and used the one-step estimator.

Often the overidentifying restrictions test is interpreted as a test of the validity of the instruments $\mathbf{z}$. However, other forms of model misspecification can sometimes lead to a significant test statistic. See Hall (2005, sec. 5.1) for a discussion of the overidentifying restrictions test and its behavior in correctly and misspecified models.

## ▷ Example 1

In example 6 of [R] **gmm**, we fit an exponential regression model of the number of doctor visits based on the person's gender, income, possession of private health insurance, and presence of a chronic disease. We argued that the variable `income` may be endogenous; we used the person's age and race as additional instrumental variables. Here we refit the model and test the specification of the model. We type

```
. use https://www.stata-press.com/data/r19/docvisits
. gmm (docvis - exp({xb:private chronic female income} + {b0})),
> instruments(private chronic female age black hispanic)
  (output omitted)
. estat overid
  Test of overidentifying restriction:
  Hansen's J chi2(2) = 9.52598 (p = 0.0085)
```

The $J$ statistic is significant even at the 1% significance level, so we conclude that our model is misspecified. One possibility is that age and race directly affect the number of doctor visits, so we are not justified in excluding them from the model.

A simple technique to explore whether any of the instruments is invalid is to examine the statistics

$$r_j = \mathbf{W}_{jj}^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^{N} z_{ij} u_i(\widehat{\boldsymbol{\beta}}) \right\}$$

for $j = 1, \ldots, k$, where $\mathbf{W}_{jj}$ denotes the $j$th diagonal element of $\mathbf{W}$, $u_i(\widehat{\boldsymbol{\beta}})$ denotes the sample residuals, and $k$ is the number of instruments. If all the instruments are valid, then the scaled sample moments should at least be on the same order of magnitude. If one (or more) instrument's $r_j$ is large in absolute value relative to the others, then that could be an indication that instrument is not valid.

In Stata, we type

```
. predict double r if e(sample), residuals // obtain residual from the model
. matrix W = e(W)                           // retrieve weight matrix
. local i 1
. // loop over each instrument and compute r_j
. foreach var of varlist private chronic female age black hispanic {
  2.          generate double r'var' = r*'var'*sqrt(W['i', 'i'])
  3.          local '++i'
  4. }
. summarize r*
```

| Variable | Obs | Mean | Std. dev. | Min | Max |
|---|---|---|---|---|---|
| r | 4,412 | .0344373 | 8.26176 | -151.1847 | 113.059 |
| rprivate | 4,412 | .007988 | 3.824118 | -72.66254 | 54.33852 |
| rchronic | 4,412 | .0026947 | 2.0707 | -43.7311 | 32.703 |
| rfemale | 4,412 | .0028168 | 1.566397 | -12.7388 | 24.43621 |
| rage | 4,412 | .0360978 | 4.752986 | -89.74112 | 55.58143 |
| rblack | 4,412 | -.0379317 | 1.062027 | -24.39747 | 27.34512 |
| rhispanic | 4,412 | -.017435 | 1.08567 | -5.509386 | 31.53512 |

We notice that the $r_j$ statistics for age, black, and hispanic are larger than those for the other instruments in our model, supporting our suspicion that age and race may have a direct impact on the number of doctor visits.

◁

# Stored results

estat overid stores the following in r():

Scalars
| | |
|---|---|
| r(J) | Hansen's $J$ statistic |
| r(J_df) | $J$ statistic degrees of freedom |
| r(J_p) | $J$ statistic $p$-value |

# References

Hall, A. R. 2005. *Generalized Method of Moments.* Oxford: Oxford University Press.

Lindsey, C. 2016. Testing model specification and using the program version of gmm. *The Stata Blog: Not Elsewhere Classified.* https://blog.stata.com/2016/02/11/testing-model-specification-and-using-the-program-version-of-gmm/.

# Also see

[R] **gmm** — Generalized method of moments estimation

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