Title

gmm postestimation — Postestimation tools for gmm

Description

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
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	residuals
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

Special-interest postestimation command

`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.

Syntax for predict

`predict` [*type*] *newvar* [*if*] [*in*] [, `equation(`#*eqno* | *eqname*`)`]

`predict` [*type*] { *stub** | *newvar*$_1$ ... *newvar*$_q$ } [*if*] [*in*]

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

You specify one new variable and (optionally) `equation()`, or you specify *stub*`*` or q new variables, where q is the number of moment equations.

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Option for predict

Main

`equation(#`*eqno* | *eqname*`)` specifies the equation for which residuals are desired. 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 an 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**.

Syntax for estat overid

```
estat overid
```

Menu for estat

Statistics > Postestimation > Reports and statistics

Remarks and examples

stata.com

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 $H_0: E\{\mathbf{z}_i u_i(\boldsymbol{\beta})\} = \mathbf{0}$, 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 http://www.stata-press.com/data/r13/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 jth 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)    // 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	4412	.0344373	8.26176	-151.1847	113.059
rprivate	4412	.007988	3.824118	-72.66254	54.33852
rchronic	4412	.0026947	2.0707	-43.7311	32.703
rfemale	4412	.0028168	1.566397	-12.7388	24.43621
rage	4412	.0360978	4.752986	-89.74112	55.58143
rblack	4412	-.0379317	1.062027	-24.39747	27.34512
rhispanic	4412	-.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

Reference

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

Also see

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

[U] 20 Estimation and postestimation commands