Title

ivpoisson postestimation — Postestimation tools for ivpoisson

Description

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

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 estimates
`estat summarize`	summary statistics for the estimation sample
`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
`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 coefficients
`predict`	predictions and probabilities
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`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. `ivpoisson gmm` uses GMM estimation to obtain parameter estimates. Under additive and multiplicative errors, Hansen's J statistic can be accurately reported when more instruments than endogenous regressors are specified. It is not appropriate to report the J statistic after `ivpoisson cfunction`, because a just-identified model is fit.

If the model is correctly specified in the sense that $E\{\widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta})\} = 0$, then the sample analog to that condition should hold at the estimated value of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The $\widetilde{\mathbf{z}}_i$ variables are the exogenous regressors $\mathbf{x}_i$ and instrumental variables $\mathbf{z}_i$ used in `ivpoisson gmm`. The $\mathbf{y}_{2,i}$ are the endogenous regressors. The u function is the error function, which will have a different form for multiplicative and additive errors in the regression.

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 `ivpoisson 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*] [, *statistic* `nooffset`]

statistic	Description
Main	
`n`	number of events; the default
`xbtotal`	linear prediction, using residual estimates for `ivpoisson cfunction`
`xb`	linear prediction
`residuals`	residuals

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

Menu for predict

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main

`n`, the default, calculates the predicted number of events via the exponential-form estimate. This is $\exp(\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2)$ if neither `offset()` nor `exposure()` was specified, $\exp(\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2 + \text{offset}_j)$ if `offset()` was specified, or $\exp(\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2) \times \text{exposure}_j$ if `exposure()` was specified.

After generalized method of moments estimation, the exponential-form estimate is not a consistent estimate of the conditional mean of y_j, because it is not corrected for $E(\epsilon_j|\mathbf{y}_{2,j})$. More details are found in *Methods and formulas*.

After control-function estimation, we correct the exponential-form estimate for $E(\epsilon_j|\mathbf{y}_{2,j})$ by using the estimated residuals of $\mathbf{y}_{2,j}$ and the `c_*` auxiliary parameters. This supplements the direct effect of $\mathbf{y}_{2,j}$ and $\mathbf{x}_j$ through $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with the indirect effects of $\mathbf{y}_{2,j}$, $\mathbf{x}_j$, and the instruments $\mathbf{z}_j$ through the endogenous error ϵ_j. Thus the exponential-form estimate consistently estimates the conditional mean of y_j.

`xbtotal` calculates the linear prediction, which is $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2$ if neither `offset()` nor `exposure()` was specified, $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2 + \text{offset}_j$ if `offset()` was specified, or $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2 + \ln(\text{exposure}_j)$ if `exposure()` was specified.

After control-function estimation, the estimate of the linear form $\mathbf{x}_j'\boldsymbol{\beta}_1$ includes the estimated residuals of the endogenous regressors with coefficients from the `c_*` auxiliary parameters.

`xb` calculates the linear prediction, which is $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2$ if neither `offset()` nor `exposure()` was specified, $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2 + \text{offset}_j$ if `offset()` was specified, or $\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2 + \ln(\text{exposure}_j)$ if `exposure()` was specified.

`residuals` calculates the residuals. Under additive errors, these are calculated as $y_j - \exp(\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2)$. Under multiplicative errors, they are calculated as $y_j / \exp(\mathbf{x}_j'\boldsymbol{\beta}_1 + \mathbf{y}_{2,j}'\boldsymbol{\beta}_2) - 1$.

When `offset()` or `exposure()` is specified, $\mathbf{x}_j'\boldsymbol{\beta}_1$ is not used directly in the residuals. $\mathbf{x}_j'\boldsymbol{\beta}_1 + \text{offset}_j$ is used if `offset()` was specified. $\mathbf{x}_j'\boldsymbol{\beta}_1 + \ln(\text{exposure}_j)$ is used if `exposure()` was specified. See `nooffset` below.

After control-function estimation, the estimate of the linear form $\mathbf{x}_j'\boldsymbol{\beta}_1$ includes the estimated residuals of the endogenous regressors with coefficients from the `c_*` auxiliary parameters.

`nooffset` is relevant only if you specified `offset()` or `exposure()` when you fit the model. It modifies the calculations made by `predict` so that they ignore the offset or exposure variable. `nooffset` removes the offset from calculations involving both the `treat()` equation and the dependent count variable.

Syntax for estat overid

```
estat overid
```

Menu for estat

Statistics > Postestimation > Reports and statistics

Remarks and examples

stata.com

`estat overid` reports Hansen's J statistic, which is used to determine the validity of the overidentifying restrictions in a GMM model. It is not appropriate to use it after `ivpoisson cfunction`, because a just-identified model is fit.

Recall that the GMM criterion function is

$$Q(\boldsymbol{\beta}) = \left\{\frac{1}{N}\sum_i \widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\right\}' \mathbf{W}_N \left\{\frac{1}{N}\sum_i \widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\right\} \qquad \text{(A1)}$$

Our u function within this formula will change depending on whether we use additive or multiplicative errors. The $\widetilde{\mathbf{z}}$ vector contains the exogenous regressors and instrumental variables used. `ivpoisson gmm` estimates regression coefficients to minimize Q.

Let l be the dimension of $\widetilde{\mathbf{z}}$ and k the number of regressors. If $\mathbf{W}_N$ is an optimal weight matrix, under the null hypothesis $H_0\colon E\left\{\widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\right\} = \mathbf{0}$, the test statistic $J = N \times Q \sim \chi^2(l-k)$. A large test statistic casts doubt on the null hypothesis.

Because the weight matrix $\mathbf{W}_N$ must be optimal, `estat overid` works only after the two-step and iterated estimation or if you supplied your own initial weight matrix by using the `winitial(`*matname*`)` option of `ivpoisson gmm` and used the one-step estimator.

Often the overidentifying restrictions test is interpreted as a test of the validity of the instruments **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 specified and misspecified models.

Note that `ivpoisson gmm` defaults to the two-step estimator when other options are not specified to override the default. Thus it is appropriate to perform the J test after the regression of example 1 in [R] **ivpoisson**.

▷ Example 1: Specification test

Recall example 1 of [R] **ivpoisson**. We estimated the parameters of an exponential conditional mean model for the number of visits to a website. Additive errors were used. Exogenous regressors included the gender of an individual and the number of ads received from the website.

An endogenous regressor, time spent on the Internet, was also included in the model. Two instruments were used. One of the instruments measured the time spent interacting with friends and out-of-town family. The other measured the time spent on the phone.

We will reestimate the parameters of the regression here and then test the specification.

```
. use http://www.stata-press.com/data/r13/website
(Visits to website)
. ivpoisson gmm visits ad female (time = phone frfam)
  (output omitted )
. estat overid
  Test of overidentifying restriction:
  Hansen's J chi2(1) = .129055 (p = 0.7194)
```

We have two instruments for one endogenous variable, so the J statistic has one degree of freedom. The J statistic is not significant. We fail to reject the null hypothesis that the model is correctly specified.

◁

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

Methods and formulas

The vector $\mathbf{x}_i$ contains the exogenous regressors, and $\mathbf{z}_i$ the instruments. The vector $\widetilde{\mathbf{z}}_i$ is partitioned as $(\mathbf{x}_i, \mathbf{z}_i)$. The vector $\mathbf{y}_{2,i}$ contains the endogenous regressors.

Under multiplicative errors, the conditional mean of y_i is

$$\begin{aligned} E(y_i|\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i) &= E\{E(y_i|\mathbf{x}_i, \mathbf{y}_{2,i}, \epsilon_i)|\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i\} \\ &= E\left\{\exp\left(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2\right)\epsilon_i|\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i\right\} \\ &= \exp\left(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2\right) E(\epsilon_i|\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i) \end{aligned}$$

Under the CF estimator,

$$\begin{aligned} E(\epsilon_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i) &= E\left\{E(\epsilon_i|\mathbf{v}_i,c_i)|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i\right\} \\ &= E\left\{\exp(\mathbf{v}_i'\boldsymbol{\rho}+c_i)|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i\right\} \\ &= \exp\left\{(\mathbf{y}_{2,i}-\mathbf{B}\widetilde{\mathbf{z}}_i')'\boldsymbol{\rho}\right\} E(c_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i) \\ &= \exp\left\{(\mathbf{y}_{2,i}-\mathbf{B}\widetilde{\mathbf{z}}_i')'\boldsymbol{\rho}\right\} \end{aligned}$$

Thus under the CF estimator, we estimate the conditional mean of y_i as

$$E(y_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i) = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}_1+\mathbf{y}_{2,i}'\boldsymbol{\beta}_2+(\mathbf{y}_{2,i}-\mathbf{B}\widetilde{\mathbf{z}}_i')'\boldsymbol{\rho}\right)$$

The CF estimator explicitly models the functional form of the endogeneity of $\mathbf{y}_{2,i}$ and ϵ_i with the instruments and exogenous regressors $\widetilde{\mathbf{z}}_i$. This allows it to correct the exponential-form estimator for the $E(\epsilon_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i)$ term.

In contrast, the GMM estimator does not model the functional form of the endogeneity of $\mathbf{y}_{2,i}$ and ϵ_i. Therefore, $E(\epsilon_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i)$ is not estimated, and the exponential-form estimator under GMM estimation simply ignores this term. Noting that because $\widetilde{\mathbf{z}}_i$ and ϵ_i are independent, $E(\epsilon_i|\mathbf{y}_{2,i},\widetilde{\mathbf{z}}_i) = E(\epsilon_i|\mathbf{y}_{2,i})$, we can obviously see that ignoring the term will lead to inconsistent estimation of the conditional mean of y_i. $y_{2,i}$ and ϵ_i are not independent, so $E(\epsilon_i|y_{2,i})$ may vary based on $y_{2,i}$.

In the additive errors setting, a similar derivation will show that the exponential-form estimator obtained from GMM estimation is inconsistent for the conditional mean of y_i.

Reference

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

Also see

[R] **ivpoisson** — Poisson regression with endogenous regressors

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