Title

> **ivpoisson** — Poisson regression with endogenous regressors

Syntax

Generalized method of moments estimator

> ivpoisson gmm *depvar* $[\textit{varlist}_1]$ $[(\textit{varlist}_2 = \textit{varlist}_{iv})]$ $[\textit{if}]$ $[\textit{in}]$ $[\textit{weight}]$
>
> $[$, *reg_err_opt options* $]$

Control-function estimator

> ivpoisson <u>cfunc</u>tion *depvar* $[\textit{varlist}_1]$ $(\textit{varlist}_2 = \textit{varlist}_{iv})$ $[\textit{if}]$ $[\textit{in}]$ $[\textit{weight}]$
>
> $[$, *options* $]$

reg_err_opt	Description
Model	
<u>add</u>itive	add regression errors to the conditional mean term; the default
<u>mult</u>iplicative	multiply regression errors by the conditional mean term

options	Description
[Model]	
noconstant	suppress constant term
exposure(*varname_e*)	include ln(*varname_e*) in model with coefficient constrained to 1
offset(*varname_o*)	include *varname_o* in model with coefficient constrained to 1
* twostep	use two-step GMM estimator; the default for ivpoisson gmm
* onestep	use one-step GMM estimator; the default for ivpoisson cfunction
* igmm	use iterative GMM estimator
[Weight matrix]	
wmatrix(*wmtype*)	specify weight matrix; *wmtype* may be robust, cluster *clustvar*, or unadjusted
center	center moments in weight-matrix computation
winitial(*iwtype*[, independent])	
	specify initial weight matrix; *iwtype* may be unadjusted, identity, or the name of a Stata matrix (independent may not be specified with ivpoisson gmm)
[SE/Robust]	
vce(*vcetype*)	*vcetype* may be robust, cluster *clustvar*, bootstrap, jackknife, or unadjusted
[Reporting]	
level(*#*)	set confidence level; default is level(95)
irr	report incidence-rate ratios
display_options	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
[Optimization]	
from(*initial_values*)	specify initial values for parameters
‡ igmmiterate(*#*)	specify maximum number of iterations for iterated GMM estimator
‡ igmmeps(*#*)	specify *#* for iterated GMM parameter convergence criterion; default is igmmeps(1e-6)
‡ igmmweps(*#*)	specify *#* for iterated GMM weight-matrix convergence criterion; default is igmmweps(1e-6)
optimization_options	control the optimization process; seldom used

* You can specify at most one of these options.

‡ These options may be specified only when igmm is specified.

varlist₁ and *varlist_{iv}* may contain factor variables; see [U] 11.4.3 Factor variables.

depvar, *varlist₁*, *varlist₂*, and *varlist_{iv}* may contain time-series operators; see [U] 11.4.4 Time-series varlists.

bootstrap, by, jackknife, rolling, and statsby are allowed; see [U] 11.1.10 Prefix commands.

Weights are not allowed with the bootstrap prefix; see [R] bootstrap.

aweights are not allowed with the jackknife prefix; see [R] jackknife.

aweights, fweights, iweights, and pweights are allowed; see [U] 11.1.6 weight.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

Menu

Statistics > Endogenous covariates > Poisson regression with endogenous regressors

Description

ivpoisson estimates the parameters of a Poisson regression model in which some of the regressors are endogenous. The model is also known as an exponential conditional mean model in which some of the regressors are endogenous. The model may be specified using either additive or multiplicative error terms. The model is frequently used to model count outcomes and is also used to model nonnegative outcome variables.

Options

> [Model]

noconstant, exposure(*varname_e*), offset(*varname_o*); see [R] **estimation options**.

additive, the default, specifies that the regression errors be added to the conditional mean term and have mean 0.

multiplicative specifies that the regression errors be multiplied by the conditional mean term and have mean 1.

twostep, onestep, and igmm specify which estimator is to be used.

> twostep requests the two-step GMM estimator. gmm obtains parameter estimates based on the initial weight matrix, computes a new weight matrix based on those estimates, and then reestimates the parameters based on that weight matrix. twostep is the default for ivpoisson gmm.

> onestep requests the one-step GMM estimator. The parameters are estimated based on an initial weight matrix, and no updating of the weight matrix is performed except when calculating the appropriate variance–covariance (VCE) matrix. onestep is the default for ivpoisson cfunction.

> igmm requests the iterative GMM estimator. gmm obtains parameter estimates based on the initial weight matrix, computes a new weight matrix based on those estimates, reestimates the parameters based on that weight matrix, computes a new weight matrix, and so on, to convergence. Convergence is declared when the relative change in the parameter vector is less than igmmeps(), the relative change in the weight matrix is less than igmmweps(), or igmmiterate() iterations have been completed. Hall (2005, sec. 2.4 and 3.6) mentions that there may be gains to finite-sample efficiency from using the iterative estimator.

> [Weight matrix]

wmatrix(*wmtype*) specifies the type of weight matrix to be used in conjunction with the two-step and iterated GMM estimators.

> Specifying wmatrix(robust) requests a weight matrix that is appropriate when the errors are independent but not necessarily identically distributed. wmatrix(robust) is the default.

> Specifying wmatrix(cluster *clustvar*) requests a weight matrix that accounts for arbitrary correlation among observations within clusters identified by *clustvar*.

> Specifying wmatrix(unadjusted) requests a weight matrix that is suitable when the errors are homoskedastic.

> wmatrix() cannot be specified if onestep is also specified.

center requests that the sample moments be centered (demeaned) when computing GMM weight matrices. By default, centering is not done.

winitial(*wmtype*[, independent]) specifies the weight matrix to use to obtain the first-step parameter estimates.

Specifying winitial(unadjusted) requests a weighting matrix that assumes the error functions are independent and identically distributed. This matrix is of the form $(\mathbf{Z}'\mathbf{Z})^{-1}$, where $\mathbf{Z}$ represents all the exogenous and instrumental variables.

winitial(identity) requests that the identity matrix be used.

winitial(*matname*) requests that Stata matrix *matname* be used.

Including the independent suboption creates a weight matrix that assumes error functions are independent. Elements of the weight matrix corresponding to covariances between any two error functions are set equal to zero. This suboption only applies to ivpoisson cfunction.

winitial(unadjusted) is the default for ivpoisson gmm.

winitial(unadjusted, independent) is the default for ivpoisson cfunction.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (robust), that allow for intragroup correlation (cluster *clustvar*), and that use bootstrap or jackknife methods (bootstrap, jackknife); see [R] *vce_option*.

vce(unadjusted) specifies that an unadjusted (nonrobust) VCE matrix be used; this, along with the twostep option, results in the "optimal two-step GMM" estimates often discussed in textbooks. vce(unadjusted) may not be set in ivpoisson cfunction.

The default *vcetype* is based on the *wmtype* specified in the wmatrix() option. If wmatrix() is specified but vce() is not, then *vcetype* is set equal to *wmtype*. To override this behavior in ivpoisson gmm and obtain an unadjusted (nonrobust) VCE matrix, specify vce(unadjusted). The default *vcetype* for ivpoisson cfunction is robust.

Specifying vce(bootstrap) or vce(jackknife) results in standard errors based on the bootstrap or jackknife, respectively. See [R] *vce_option*, [R] **bootstrap**, and [R] **jackknife** for more information on these VCEs.

The syntax for *vcetype*s is identical to those for wmatrix().

Reporting

level(*#*); see [R] **estimation options**.

irr reports estimated coefficients transformed to incidence-rate ratios, that is, e^{β_i} rather than β_i. Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated or stored. irr may be specified at estimation or when replaying previously estimated results. irr is not allowed with additive.

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

Optimization

from(*initial_values*) specifies the initial values to begin the estimation. You can specify a $1 \times k$ matrix, where k is the number of parameters in the model, or you can specify a parameter name,

its initial value, another parameter name, its initial value, and so on. For example, to initialize the coefficient for male to 1.23 and the constant _cons to 4.57, you would type

```
ivpoisson ..., from(male 1.23 _cons 4.57) ...
```

Initial values declared using this option override any that are declared within substitutable expressions. If you specify a parameter that does not appear in your model, ivpoisson exits with error code 480. If you specify a matrix, the values must be in the same order in which the parameters are declared in your model. ivpoisson ignores the row and column names of the matrix.

igmmiterate(#), igmmeps(#), and igmmweps(#) control the iterative process for the iterative GMM estimator for ivpoisson. These options can be specified only if you also specify igmm.

  igmmiterate(#) specifies the maximum number of iterations to perform with the iterative GMM estimator. The default is the number set using set maxiter (set [R] **maximize**), which is 16,000 by default.

  igmmeps(#) specifies the convergence criterion used for successive parameter estimates when the iterative GMM estimator is used. The default is igmmeps(1e-6). Convergence is declared when the relative difference between successive parameter estimates is less than igmmeps() and the relative difference between successive estimates of the weight matrix is less than igmmweps().

  igmmweps(#) specifies the convergence criterion used for successive estimates of the weight matrix when the iterative GMM estimator is used. The default is igmmweps(1e-6). Convergence is declared when the relative difference between successive parameter estimates is less than igmmeps() and the relative difference between successive estimates of the weight matrix is less than igmmweps().

optimization_options: <u>tech</u>nique(), conv_maxiter(), conv_ptol(), conv_vtol(), conv_nrtol(), and tracelevel(). technique() specifies the optimization technique to use; gn (the default), nr, dfp, and bfgs are allowed. conv_maxiter() specifies the maximum number of iterations; conv_ptol(), conv_vtol(), and conv_nrtol() specify the convergence criteria for the parameters, gradient, and scaled Hessian, respectively. tracelevel() allows you to obtain additional details during the iterative process. See [M-5] **optimize( )**.

Remarks and examples

ivpoisson estimates the parameters of a Poisson regression model in which some of the regressors are endogenous. A regressor is endogenous if it is related to the unobserved error term. The model is also known as an exponential conditional mean model in which some of the regressors are endogenous. The model may be specified using either additive or multiplicative error terms.

The model is frequently used to model count outcomes and is also used to model nonnegative outcome variables. Poisson regression is a special exponential conditional mean model. See [R] **poisson** for more information on Poisson regression.

The exponential conditional mean model has an error form representation in which the dependent variable y is a function of the exogenous regressors $\mathbf{x}$, endogenous regressors $\mathbf{y}_2$, and an error ϵ. The regressors $\mathbf{x}$ are independent of ϵ, while $\mathbf{y}_2$ are not.

ivpoisson allows ϵ to enter either additively,

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2) + \epsilon_i$$

or multiplicatively,

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2)\epsilon_i$$

Mullahy (1997), Cameron and Trivedi (2013), Windmeijer and Santos Silva (1997), and Wooldridge (2010) discuss the generalized method of moments (GMM) estimators implemented in `ivpoisson`. GMM is frequently used in modern econometrics. Many econometric and statistical models can be expressed as conditions on the population moments. The parameter estimates produced by GMM estimators make the sample-moment conditions as true as possible given the data. See [R] **gmm** for further information on GMM estimation and how Stata performs it.

The rest of the discussion is presented under the following headings:

> GMM estimator for additive model
> GMM estimator for multiplicative model
> CF estimator for multiplicative model

GMM estimator for additive model

The GMM estimator uses additional variables, known as instruments and denoted by $\mathbf{z}_i$, to specify moment conditions that hold in the population. The GMM parameter estimates make the sample versions of these population-moment conditions as close to true as possible. The instrumental variables are assumed to be correlated with the endogenous regressors $\mathbf{y}_{2,i}$ but independent of the errors ϵ_i.

Under additive errors, the dependent variable y_i is determined by exogenous regressors $\mathbf{x}_i$, endogenous regressors $\mathbf{y}_{2,i}$, and zero-mean error ϵ_i as

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2) + \epsilon_i$$

This leads to the following error function

$$u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2)$$

The population-moment conditions for GMM estimation are $E\left\{\widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\right\} = \mathbf{0}$, where the vector $\widetilde{\mathbf{z}}_i$ is partitioned as $(\mathbf{x}_i', \mathbf{z}_i')$. The sample-moment conditions are formed by replacing the expectation with the corresponding sample mean. The GMM estimator solves a minimization problem to make the sample-moment conditions as close to zero as possible. Details on how estimation is performed are given in *Methods and formulas*.

Now we will demonstrate how `ivpoisson gmm` works in the additive error setting with an example.

▷ Example 1: ivpoisson gmm with additive errors

This example uses simulated data based on the following story. A news website randomly samples 500 young adults in a major city. The website wants to model the number of times the sampled individuals visit its website (`visits`) based on their overall time spent on the Internet (`time`) and the number of times they receive an ad for the website through email or viewing another website (`ad`). The website also suspects the gender of the individual may matter, so an exogenous dummy variable, `female`, is included in the model.

We suspect time spent on the Internet is correlated with unobserved factors that additively affect the number of times an individual visits the website. So we treat `time` as an endogenous regressor. Two instruments are used for this variable. The time spent on the phone (`phone`) is one instrument. The other instrument is the time spent interacting with friends and family that live out of town (`frfam`).

We model the number of visits the website receives using an exponential conditional mean model with additive errors and use `ivpoisson gmm` to estimate the parameters of the regression in the output below. To allow for heteroskedasticity of the errors, we use robust standard errors, which is the default; see *Obtaining standard errors* in [R] **gmm** for a discussion of why robust standard errors is the default.

```
.use http://www.stata-press.com/data/r13/website
(Visits to website)

. ivpoisson gmm visits ad female (time = phone frfam)

Step 1
Iteration 0:    GMM criterion Q(b) =  .33829416
Iteration 1:    GMM criterion Q(b) =  .00362656
Iteration 2:    GMM criterion Q(b) =  .00131886
Iteration 3:    GMM criterion Q(b) =  .00131876

Step 2
Iteration 0:    GMM criterion Q(b) =  .00027102
Iteration 1:    GMM criterion Q(b) =  .00025811
Iteration 2:    GMM criterion Q(b) =  .00025811

Exponential mean model with endogenous regressors

Number of parameters =    4                          Number of obs  =      500
Number of moments    =    5
Initial weight matrix: Unadjusted
GMM weight matrix:     Robust
```

visits	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
time	.0589294	.0107942	5.46	0.000	.0377732	.0800857
ad	.137344	.010157	13.52	0.000	.1174366	.1572515
female	-.0247707	.0376218	-0.66	0.510	-.098508	.0489666
_cons	1.041505	.0385848	26.99	0.000	.9658807	1.11713

```
Instrumented:  time
Instruments:   ad female phone frfam
```

We find significant coefficients for all regressors but `female`. At fixed values of the other regressors, increased time spent on the Internet will raise the expected number of website visits. Receiving additional advertisements will also cause an increase in the expected number of website visits.

◁

GMM estimator for multiplicative model

Under multiplicative errors, the dependent variable y_i is determined by exogenous regressors $\mathbf{x}_i$, endogenous regressors $\mathbf{y}_{2,i}$, and unit-mean errors ϵ_i as

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2)\epsilon_i$$

This setting yields a different error function than the additive error case. This ratio formulation is

$$u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = y_i/\exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2) - 1$$

Given the instrumental variables $\mathbf{z}$, the population-moment conditions for GMM estimation are $E\{\widetilde{\mathbf{z}}_i u(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\} = \mathbf{0}$. The vector $\widetilde{\mathbf{z}}_i$ is partitioned as $(\mathbf{x}_i', \mathbf{z}_i')$. As above, the sample-moment conditions are the sample analogs of the population-moment conditions, and the GMM estimator solves a minimization problem to make the sample-moment conditions as close to zero as possible. Details on how estimation is performed are given in *Methods and formulas*.

▷ Example 2: ivpoisson gmm with multiplicative errors

In this example, we observe a simulated random sample of 5,000 households. We model the number of trips taken by members of the household in the 24-hour period immediately prior to the interview time by using an exponential conditional mean model with multiplicative errors. Exogenous regressors include the distance to the central business district from the household (cbd), the distance from the household to a public transit node (ptn), whether there is a full-time worker in the household (worker), and whether the examined period is on a weekend (weekend). We suspect that the endogenous regressor, the transportation cost of the household in the prior week (tcost), is correlated with unobserved factors that affect the number of trips taken. This transportation cost includes gasoline and bus, train tickets, etc.

The ratio of the cost of a public transit day pass in the sampled area to the national average cost of such a pass (pt) is also observed. This is used as an instrument for transportation cost.

In the output below, we estimate the parameters of the regression with ivpoisson gmm. To allow for heteroskedasticity of the errors, we use robust standard errors, which is the default.

```
. use http://www.stata-press.com/data/r13/trip
(Household trips)

. ivpoisson gmm trips cbd ptn worker weekend (tcost = pt), multiplicative

Step 1
Iteration 0:   GMM criterion Q(b) =  .04949852
Iteration 1:   GMM criterion Q(b) =  .00011194
Iteration 2:   GMM criterion Q(b) =  1.563e-08
Iteration 3:   GMM criterion Q(b) =  3.685e-16

Step 2
Iteration 0:   GMM criterion Q(b) =  2.287e-16
Iteration 1:   GMM criterion Q(b) =  1.413e-31

Exponential mean model with endogenous regressors

Number of parameters =   6                          Number of obs  =    5000
Number of moments    =   6
Initial weight matrix: Unadjusted
GMM weight matrix:     Robust
```

trips	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
tcost	.0352185	.0098182	3.59	0.000	.0159752	.0544617
cbd	-.008398	.0020172	-4.16	0.000	-.0123517	-.0044444
ptn	-.0113146	.0021819	-5.19	0.000	-.015591	-.0070383
worker	.6623018	.0519909	12.74	0.000	.5604015	.764202
weekend	.3009323	.0362682	8.30	0.000	.2298479	.3720167
_cons	.2654423	.1550127	1.71	0.087	-.0383769	.5692616

```
Instrumented:  tcost
Instruments:   cbd ptn worker weekend pt
```

We find that all coefficients are significant. At fixed values of the other regressors, we see that additional mileage from the central business district and public transit nodes reduces the expected number of trips taken. Individuals who live farther away from the central business district may still be out of the house the same amount of time, but they will take fewer trips because the transit time has increased. The situation is similar for those who live farther from public transit.

To interpret the other parameters, we will look at the partial effects of their respective independent variables. The partial effects of a change in an independent variable on the modeled conditional expectation function vary over the data because the model is nonlinear. However, under the multiplicative

error model, the ratio of the new value to the old value after a discrete change in an independent variable is constant over the data.

Let $\mathbf{w} = (\mathbf{x}', \mathbf{y}_2')'$. If we add 1 to the jth independent variable in $\mathbf{w}$, the functional form of the model implies that

$$
\frac{E(y|\mathbf{w}, w_j + 1, \epsilon)}{E(y|\mathbf{w}, w_j, \epsilon)} = \frac{E(y|w_1, \ldots, w_j + 1, \ldots, w_k, \epsilon)}{E(y|w_1, \ldots, w_j, \ldots, w_k, \epsilon)} = e^{\beta_j}
$$

When y is a count variable, this normalized effect is called the incidence-rate ratio (IRR) for a one-unit change in w_j.

More generally, the IRR for a Δw_j change in w_j is $e^{\beta_j \Delta w_j}$ under a multiplicative-error exponential conditional mean model. We can calculate incidence-rate ratios for different changes in the regressors by using lincom; see [R] **lincom**.

Here we replay the ivpoisson results by typing the command name and we specify the irr option to get the incidence-rate ratios. Each significance test for a coefficient equaling zero becomes a test for the incidence-rate ratio equaling one.

```
. ivpoisson, irr
Exponential mean model with endogenous regressors
Number of parameters =    6                         Number of obs  =    5000
Number of moments    =    6
Initial weight matrix: Unadjusted
GMM weight matrix:     Robust
```

trips	IRR	Robust Std. Err.	z	P>|z|	[95% Conf. Interval]	
tcost	1.035846	.0101701	3.59	0.000	1.016103	1.055972
cbd	.9916371	.0020003	-4.16	0.000	.9877243	.9955655
ptn	.9887491	.0021573	-5.19	0.000	.9845299	.9929864
worker	1.939251	.1008234	12.74	0.000	1.751376	2.14728
weekend	1.351118	.0490026	8.30	0.000	1.258409	1.450657
_cons	1.304008	.2021377	1.71	0.087	.9623501	1.766962

```
Instrumented:  tcost
Instruments:   cbd ptn worker weekend pt
```

Holding other regressors and the error constant, the expected number of trips made from houses with a full-time worker is nearly twice that of those houses without a full-time worker. Similarly, the expected number of trips made during a weekend day is close to 35% higher than the expected number of trips made on other days. For each additional dollar of weekly transportation cost, the expected number of household trips is increased by approximately 3.6%.

◁

CF estimator for multiplicative model

Control-function (CF) estimators can be used to account for endogenous regressors. As Wooldridge (2010, sec. 18.5) describes, CF estimators assume a certain structural relationship between the endogenous regressors and the exogenous regressors and use functions of first-stage parameter estimates to control for the endogeneity in the second stage.

Wooldridge (2010, sec. 18.5) notes that the VCE of the second-stage estimator must be adjusted to account for estimates from the first stage. `ivpoisson cfunction` solves this problem by stacking the moment conditions that define each stage and applying a single GMM estimator. See Newey (1984) and Wooldridge (2010, sec. 14.2) for a description of this technique. No adjustment to the VCE is necessary because there is only one stage.

The CF estimator augments the original multiplicative model with an estimated term that controls for the endogeneity of $y_{2,i}$. When $y_{2,i}$ is exogenous, the coefficient on this control term is zero. Let $\mathbf{z}$ be instrumental variables, and the vector $\widetilde{\mathbf{z}}_i$ be $(\mathbf{x}_i', \mathbf{z}_i')$.

The augmented model is

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2 + \mathbf{v}_i'\boldsymbol{\rho} + c_i)$$

where

$$\mathbf{y}_{2,i} = \mathbf{B}\widetilde{\mathbf{z}}_i' + \mathbf{v}_i$$

The term $\mathbf{v}_i'\boldsymbol{\rho}$ controls for the endogeneity of $\mathbf{y}_{2,i}$, and we normalize $E\{\exp(c_i)\} = 1$. The coefficient vector $\boldsymbol{\rho}$ measures the strength of the endogeneity of $\mathbf{y}_{2,i}$; $\mathbf{y}_{2,i}$ is exogenous when $\boldsymbol{\rho} = \mathbf{0}$.

`ivpoisson cfunction` estimates $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and the auxiliary parameters $\boldsymbol{\rho}$ and $\mathbf{B}$ by GMM; see *Methods and formulas* for details.

▷ Example 3: Control-function estimator

We return to the previous example, where we estimated the parameters of an exponential conditional mean model for the number of trips taken by a household in a 24-hour period. We will estimate the parameters of the regression with the CF estimator method and compare our results with those obtained with the GMM estimator in example 2.

In the output below, we estimate the parameters of the regression with the `ivpoisson cfunction` command.

```
. ivpoisson cfunction trips cbd ptn worker weekend (tcost = pt)
Step 1
Iteration 0:   GMM criterion Q(b) =   .00056156
Iteration 1:   GMM criterion Q(b) =   2.366e-07
Iteration 2:   GMM criterion Q(b) =   5.552e-14
Iteration 3:   GMM criterion Q(b) =   9.772e-27

Exponential mean model with endogenous regressors

Number of parameters =   13                         Number of obs   =     5000
Number of moments     =   13
Initial weight matrix: Unadjusted
GMM weight matrix:     Robust
```

		Robust				
trips	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
trips						
cbd	-.0082567	.0020005	-4.13	0.000	-.0121777	-.0043357
ptn	-.0113719	.0021625	-5.26	0.000	-.0156102	-.0071335
worker	.6903044	.0521642	13.23	0.000	.5880645	.7925444
weekend	.2978149	.0356474	8.35	0.000	.2279472	.3676825
tcost	.0320718	.0092738	3.46	0.001	.0138955	.0502481
_cons	.2145986	.1359327	1.58	0.114	-.0518246	.4810218
tcost						
cbd	.0165466	.0043693	3.79	0.000	.0079829	.0251102
ptn	-.040652	.0045946	-8.85	0.000	-.0496573	-.0316467
worker	1.550985	.0996496	15.56	0.000	1.355675	1.746294
weekend	.0423009	.0779101	0.54	0.587	-.1104002	.1950019
pt	.7739176	.0150072	51.57	0.000	.7445041	.8033312
_cons	12.13934	.1123471	108.05	0.000	11.91915	12.35954
/c_tcost	.1599984	.0111752	14.32	0.000	.1380954	.1819014

```
Instrumented:  tcost
Instruments:   cbd ptn worker weekend pt
```

The output table presents results for the estimated coefficients in each of three equations. First, in the `trips` equation, we see the results for the estimated coefficients in the equation for the dependent variable `trips`. Second, in the `tcost` equation, we see the estimated coefficients in the regression of `tcost` on the instrumental and exogenous variables. Third, the `/c_tcost` ancillary parameter corresponds to the estimate of ρ, the coefficient on the residual variable included to control for the endogeneity of `tcost`.

We find that all coefficients are significant in the exponential conditional mean equation, `trips`. The coefficient estimates in the `trips` equation are similar to the estimates obtained by the GMM estimator in example 2. That the estimated coefficient on the `tcost` control variable is significantly different from zero suggests that `tcost` is endogenous.

◁

Stored results

ivpoisson stores the following in e():

Scalars
e(N)	number of observations
e(k)	number of parameters
e(k_eq)	number of equations
e(k_aux)	number of auxiliary parameters
e(k_dv)	number of dependent variables
e(Q)	criterion function
e(J)	Hansen J χ^2 statistic
e(J_df)	J statistic degrees of freedom
e(N_clust)	number of clusters
e(rank)	rank of e(V)
e(ic)	number of iterations used by iterative GMM estimator
e(converged)	1 if converged, 0 otherwise

Macros
e(cmd)	ivpoisson
e(cmdline)	command as typed
e(depvar)	dependent variable of regression
e(instd)	instrumented variable
e(insts)	instruments
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(clustvar)	name of cluster variable
e(offset1)	offset variable for first equation
e(winit)	initial weight matrix used
e(winitname)	name of user-supplied initial weight matrix
e(estimator)	gmm or cfunction
e(additive)	additive if additive errors specified
e(multiplicative)	multiplicative if multiplicative errors specified
e(gmmestimator)	onestep, twostep, or igmm
e(wmatrix)	*wmtype* specified in wmatrix()
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. Err.
e(technique)	optimization technique
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
e(footnote)	program used to implement footnote display
e(marginsok)	predictions allowed by margins
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved

Matrices
e(b)	coefficient vector
e(V)	variance–covariance matrix
e(init)	initial values of the estimators
e(Wuser)	user-supplied initial weight matrix
e(W)	weight matrix used for final round of estimation
e(S)	moment covariance matrix used in robust VCE computations
e(V_modelbased)	model-based variance

Functions
e(sample)	marks estimation sample

Methods and formulas

The estimators in `ivpoisson` are GMM estimators that can be expressed in terms of error functions and the instruments that are used to form the moment conditions. When offsets o_j^β are used in the outcome variable equation, the following formulas apply with $\mathbf{x}_j'\boldsymbol{\beta}_1$ changed to $\mathbf{x}_j'\boldsymbol{\beta}_1 + o_j^\beta$.

The error functions for the GMM estimators are given in the text.

Here we provide some details about the form of the error function used by the CF estimator.

Recall that the multiplicative model is

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2)\epsilon_i$$

We parameterize the endogenous variables in the form

$$\mathbf{y}_{2,i} = \mathbf{B}\widetilde{\mathbf{z}}_i' + \mathbf{v}_i$$

This allows us to decompose ϵ_i as

$$\epsilon_i = \exp(\mathbf{v}_i'\boldsymbol{\rho} + c_i)$$

Given this setup, we obtain the following conditional mean:

$$E(y|\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_{2,i}, \mathbf{v}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2 + \mathbf{v}_i'\boldsymbol{\rho})$$

We estimate $\mathbf{v}_i$ as the residuals of the linear regression of $\mathbf{y}_{2,i}$ on $\widetilde{\mathbf{z}}_i$. The estimates of $\mathbf{v}_i$ are used as additional regressors in the exponential conditional mean model for y to estimate $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\rho}$. In essence, the estimates of $\mathbf{v}_i$ control for the endogeneity.

The error functions for the endogenous regressors are defined as

$$\mathbf{u}_{en,i}(\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i, \mathbf{B}) = \mathbf{y}_{2,i} - \mathbf{B}\widetilde{\mathbf{z}}_i'$$

Now we define the error function for the dependent variable as

$$u_y(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \mathbf{u}_{en,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) = y_i / \exp(\mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{y}_{2,i}'\boldsymbol{\beta}_2 + \mathbf{u}_{en,i}'\boldsymbol{\rho}) - 1$$

$\mathbf{u}_{en,i}$ will be vector valued if we have multiple endogenous regressors $\mathbf{y}_{2,i}$. Call the dimension of $\mathbf{y}_{2,i}$ g. $\mathbf{u}_{en,i}$ and $u_{y,i}$ define $g+1$ separate error functions. We will use variables $\widetilde{\mathbf{z}}_i$ to instrument each error function in $\mathbf{u}_{en,i}$. So for error function $j = 1, \ldots, g$, we have the error function $u_{en,i,j}$ and the population-moment conditions $E(\widetilde{\mathbf{z}}_i u_{en,i,j}) = \mathbf{0}$.

We calculate $\widehat{\mathbf{v}}_{oi}$ previous to estimation as the residuals of the linear regression of $\mathbf{y}_{2,i}$ on $\widetilde{\mathbf{z}}_i$. We use variables $\mathbf{x}_i$, $\mathbf{y}_{2,i}$, and $\widehat{\mathbf{v}}_{oi}$ to instrument the error function u_y. This leads to the population-moment conditions $E\left\{(\mathbf{x}_i', \mathbf{y}_{2,i}', \widehat{\mathbf{v}}_{oi}')u_{y,i}\right\} = \mathbf{0}$

Details of GMM estimation can be found in *Methods and formulas* of [R] **gmm**. Determination of the weight matrix $\mathbf{W}_N$ is discussed there.

Under the GMM estimation, the GMM estimators $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are the values of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ that minimize

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

for $q \times q$ weight matrix $\mathbf{W}_N$, where q is the dimension of $\widetilde{\mathbf{z}}_i$. The error functions u_i were defined in the text.

In the CF method, we have multiple error functions as defined above. We can stack the moment conditions and write them more compactly as $\mathbf{Z}_i' \mathbf{u}_i(\mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho})$, where

$$
\mathbf{Z}_i = \begin{bmatrix}
\mathbf{x}_i' & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{y}_{2,i}' & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \widetilde{\mathbf{z}}_i & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \widetilde{\mathbf{z}}_i
\end{bmatrix}
$$

and

$$
\mathbf{u}_i(\mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) = \begin{bmatrix}
u_y(y_i, \mathbf{x}_i, \mathbf{y}_{2,i}, \mathbf{u}_{en,i}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) \\
\mathbf{u}_{en}(\mathbf{y}_{2,i}, \widetilde{\mathbf{z}}_i, \mathbf{B})
\end{bmatrix}
$$

The matrix $\mathbf{Z}_i$ has $g+1$ rows and $k+gz$ columns, where k is the number of regressors for y_i and z is the number of exogenous regressors in $\widetilde{\mathbf{z}}_i$.

The GMM estimators $\widehat{\mathbf{B}}$, $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$, and $\widehat{\boldsymbol{\rho}}$ are the values of $\mathbf{B}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\rho}$ that minimize

$$
Q(\mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) =
\left\{ N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i(\mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) \right\}' \mathbf{W}_N \left\{ N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i(\mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\rho}) \right\} \tag{2}
$$

for $(k+gz) \times (k+gz)$ weight matrix $\mathbf{W}_N$.

By default, ivpoisson minimizes (1) and (2) using the Gauss–Newton method. See Hayashi (2000, 498) for a derivation. This technique is typically faster than quasi-Newton methods and does not require second-order derivatives.

References

Cameron, A. C., and P. K. Trivedi. 2013. *Regression Analysis of Count Data*. 2nd ed. New York: Cambridge University Press.

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

Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.

Mullahy, J. 1997. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics* 79: 586–593.

Newey, W. K. 1984. A method of moments interpretation of sequential estimators. *Economics Letters* 14: 201–206.

Windmeijer, F., and J. M. C. Santos Silva. 1997. Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics* 12: 281–294.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

[R] **ivpoisson postestimation** — Postestimation tools for ivpoisson

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

[R] **ivprobit** — Probit model with continuous endogenous regressors

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

[R] **ivtobit** — Tobit model with continuous endogenous regressors

[R] **nl** — Nonlinear least-squares estimation

[R] **nlsur** — Estimation of nonlinear systems of equations

[R] **poisson** — Poisson regression

[R] **regress** — Linear regression

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