Title

stata.com

ivprobit — Probit model with continuous endogenous regressors

Syntax

Maximum likelihood estimator

`ivprobit` *depvar* [*varlist*$_1$] (*varlist*$_2$ = *varlist*$_{\text{iv}}$) [*if*] [*in*] [*weight*] [, *mle_options*]

Two-step estimator

`ivprobit` *depvar* [*varlist*$_1$] (*varlist*$_2$ = *varlist*$_{\text{iv}}$) [*if*] [*in*] [*weight*], `twostep`
[*tse_options*]

mle_options	Description
Model	
`mle`	use conditional maximum-likelihood estimator; the default
`asis`	retain perfect predictor variables
`constraints(`*constraints*`)`	apply specified linear constraints
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `oim`, `robust`, `cluster` *clustvar*, `opg`, `bootstrap`, or `jackknife`
Reporting	
`level(#)`	set confidence level; default is `level(95)`
`first`	report first-stage regression
`nocnsreport`	do not display constraints
display_options	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process
`coeflegend`	display legend instead of statistics

tse_options	Description
Model	
* `twostep`	use Newey's two-step estimator; the default is `mle`
`asis`	retain perfect predictor variables
SE	
`vce(`*vcetype*`)`	*vcetype* may be `twostep`, `bootstrap`, or `jackknife`
Reporting	
`level(`*#*`)`	set confidence level; default is `level(95)`
`first`	report first-stage regression
display_options	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
`coeflegend`	display legend instead of statistics

* `twostep` is required.

varlist$_1$ and *varlist*$_{\mathrm{iv}}$ may contain factor variables; see **[U] 11.4.3 Factor variables**.

depvar, *varlist*$_1$, *varlist*$_2$, and *varlist*$_{\mathrm{iv}}$ may contain time-series operators; see **[U] 11.4.4 Time-series varlists**.

`bootstrap`, `by`, `jackknife`, `rolling`, `statsby`, and `svy` are allowed; see **[U] 11.1.10 Prefix commands**. `fp` is allowed with the maximum likelihood estimator.

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

`vce()`, `first`, `twostep`, and weights are not allowed with the `svy` prefix; see **[SVY] svy**.

`fweight`s, `iweight`s, and `pweight`s are allowed with the maximum likelihood estimator. `fweight`s are allowed with Newey's two-step estimator. See **[U] 11.1.6 weight**.

`coeflegend` does not appear in the dialog box.

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

Menu

Statistics > Endogenous covariates > Probit model with endogenous covariates

Description

`ivprobit` fits probit models where one or more of the regressors are endogenously determined. By default, `ivprobit` uses maximum likelihood estimation. Alternatively, Newey's (1987) minimum chi-squared estimator can be invoked with the `twostep` option. Both estimators assume that the endogenous regressors are continuous and are not appropriate for use with discrete endogenous regressors. See **[R] ivtobit** for tobit estimation with endogenous regressors and **[R] probit** for probit estimation when the model contains no endogenous regressors.

Options for ML estimator

Model

`mle` requests that the conditional maximum-likelihood estimator be used. This is the default.

`asis` requests that all specified variables and observations be retained in the maximization process. This option is typically not used and may introduce numerical instability. Normally, `ivprobit` drops any endogenous or exogenous variables that perfectly predict success or failure in the dependent variable. The associated observations are also dropped. For more information, see *Model identification* in [R] **probit**.

`constraints(`*constraints*`)`; see [R] **estimation options**.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`, `opg`), 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***.

Reporting

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

`first` requests that the parameters for the reduced-form equations showing the relationships between the endogenous variables and instruments be displayed. For the two-step estimator, `first` shows the first-stage regressions. For the maximum likelihood estimator, these parameters are estimated jointly with the parameters of the probit equation. The default is not to show these parameter estimates.

`nocnsreport`; see [R] **estimation options**.

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

Maximization

maximize_options: `difficult`, `technique(`*algorithm_spec*`)`, `iterate(`*#*`)`, [`no`]`log`, `trace`, `gradient`, `showstep`, `hessian`, `showtolerance`, `tolerance(`*#*`)`, `ltolerance(`*#*`)`, `nrtolerance(`*#*`)`, `nonrtolerance`, and `from(`*init_specs*`)`; see [R] **maximize**. This model's likelihood function can be difficult to maximize, especially with multiple endogenous variables. The `difficult` and `technique(bfgs)` options may be helpful in achieving convergence.

Setting the optimization type to `technique(bhhh)` resets the default *vcetype* to `vce(opg)`.

The following option is available with `ivprobit` but is not shown in the dialog box:

`coeflegend`; see [R] **estimation options**.

Options for two-step estimator

Model

`twostep` is required and requests that Newey's (1987) efficient two-step estimator be used to obtain the coefficient estimates.

`asis` requests that all specified variables and observations be retained in the maximization process. This option is typically not used and may introduce numerical instability. Normally, `ivprobit` drops any endogenous or exogenous variables that perfectly predict success or failure in the dependent variable. The associated observations are also dropped. For more information, see *Model identification* in [R] **probit**.

SE

`vce(vcetype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`twostep`) and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] ***vce_option***.

Reporting

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

`first` requests that the parameters for the reduced-form equations showing the relationships between the endogenous variables and instruments be displayed. For the two-step estimator, `first` shows the first-stage regressions. For the maximum likelihood estimator, these parameters are estimated jointly with the parameters of the probit equation. The default is not to show these parameter estimates.

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

The following option is available with `ivprobit` but is not shown in the dialog box:

`coeflegend`; see [R] **estimation options**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Model setup

`ivprobit` fits models with dichotomous dependent variables and endogenous regressors. You can use it to fit a probit model when you suspect that one or more of the regressors are correlated with the error term. `ivprobit` is to probit modeling what `ivregress` is to linear regression analysis; see [R] **ivregress** for more information.

Formally, the model is

$$\begin{aligned} y_{1i}^* &= \boldsymbol{y}_{2i}\boldsymbol{\beta} + \boldsymbol{x}_{1i}\boldsymbol{\gamma} + u_i \\ \boldsymbol{y}_{2i} &= \boldsymbol{x}_{1i}\boldsymbol{\Pi}_1 + \boldsymbol{x}_{2i}\boldsymbol{\Pi}_2 + \boldsymbol{v}_i \end{aligned}$$

where $i = 1, \ldots, N$, $\boldsymbol{y}_{2i}$ is a $1 \times p$ vector of endogenous variables, $\boldsymbol{x}_{1i}$ is a $1 \times k_1$ vector of exogenous variables, $\boldsymbol{x}_{2i}$ is a $1 \times k_2$ vector of additional instruments, and the equation for $\boldsymbol{y}_{2i}$ is written in reduced form. By assumption, $(u_i, \boldsymbol{v}_i) \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where σ_{11} is normalized to one to identify the model. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of structural parameters, and $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ are matrices of reduced-form parameters. This is a recursive model: $\mathbf{y}_{2i}$ appears in the equation for y_{1i}^*, but y_{1i}^* does not appear in the equation for $\mathbf{y}_{2i}$. We do not observe y_{1i}^*; instead, we observe

$$y_{1i} = \begin{cases} 0 & y_{1i}^* < 0 \\ 1 & y_{1i}^* \geq 0 \end{cases}$$

The order condition for identification of the structural parameters requires that $k_2 \geq p$. Presumably, $\boldsymbol{\Sigma}$ is not block diagonal between u_i and $\boldsymbol{v}_i$; otherwise, $\boldsymbol{y}_{2i}$ would not be endogenous.

❑ Technical note

This model is derived under the assumption that $(u_i, \boldsymbol{v}_i)$ is independent and identically distributed multivariate normal for all i. The `vce(cluster` *clustvar*`)` option can be used to control for a lack of independence. As with most probit models, if u_i is heteroskedastic, point estimates will be inconsistent. ❑

▷ Example 1

We have hypothetical data on 500 two-parent households, and we wish to model whether the woman is employed. We have a variable, `fem_work`, that is equal to one if she has a job and zero otherwise. Her decision to work is a function of the number of children at home (`kids`), number of years of schooling completed (`fem_educ`), and other household income measured in thousands of dollars (`other_inc`). We suspect that unobservable shocks affecting the woman's decision to hold a job also affect the household's other income. Therefore, we treat `other_inc` as endogenous. As an instrument, we use the number of years of schooling completed by the man (`male_educ`).

The syntax for specifying the exogenous, endogenous, and instrumental variables is identical to that used in `ivregress`; see [R] **ivregress** for details.

```
. use http://www.stata-press.com/data/r13/laborsup
. ivprobit fem_work fem_educ kids (other_inc = male_educ)
Fitting exogenous probit model
Iteration 0:   log likelihood = -344.63508
Iteration 1:   log likelihood = -255.36855
Iteration 2:   log likelihood = -255.31444
Iteration 3:   log likelihood = -255.31444
Fitting full model
Iteration 0:   log likelihood = -2371.4753
Iteration 1:   log likelihood = -2369.3178
Iteration 2:   log likelihood = -2368.2198
Iteration 3:   log likelihood = -2368.2062
Iteration 4:   log likelihood = -2368.2062
Probit model with endogenous regressors           Number of obs   =        500
                                                  Wald chi2(3)    =     163.88
Log likelihood = -2368.2062                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   other_inc |  -.0542756   .0060854    -8.92   0.000    -.0662027   -.0423485
    fem_educ |    .211111   .0268648     7.86   0.000     .1584569    .2637651
        kids |  -.1820929   .0478267    -3.81   0.000    -.2758316   -.0883543
       _cons |   .3672083   .4480724     0.82   0.412    -.5109975    1.245414
-------------+----------------------------------------------------------------
     /athrho |   .3907858   .1509443     2.59   0.010     .0949403    .6866313
    /lnsigma |   2.813383   .0316228    88.97   0.000     2.751404    2.875363
-------------+----------------------------------------------------------------
         rho |   .3720374   .1300519                      .0946561    .5958135
       sigma |   16.66621   .5270318                      15.66461    17.73186
------------------------------------------------------------------------------
Instrumented:  other_inc
Instruments:   fem_educ kids male_educ
------------------------------------------------------------------------------
Wald test of exogeneity (/athrho = 0): chi2(1) =     6.70 Prob > chi2 = 0.0096
```

Because we did not specify `mle` or `twostep`, `ivprobit` used the maximum likelihood estimator by default. At the top of the output, we see the iteration log. `ivprobit` fits a probit model ignoring endogeneity to obtain starting values for the endogenous model. The header of the output contains the sample size as well as a Wald statistic and p-value for the test of the hypothesis that all the slope coefficients are jointly zero. Below the table of coefficients, Stata reminds us that the endogenous variable is `other_inc` and that `fem_educ`, `kids`, and `male_educ` were used as instruments.

At the bottom of the output is a Wald test of the exogeneity of the instrumented variables. We reject the null hypothesis of no endogeneity. However, if the test statistic is not significant, there is not sufficient information in the sample to reject the null, so a regular probit regression may be appropriate. The point estimates from `ivprobit` are still consistent, though those from `probit` (see [R] **probit**) are likely to have smaller standard errors.

◁

Various two-step estimators have also been proposed for the endogenous probit model, and Newey's (1987) minimum chi-squared estimator is available with the `twostep` option.

▷ Example 2

Refitting our labor-supply model with the two-step estimator yields

```
. ivprobit fem_work fem_educ kids (other_inc = male_educ), twostep
Checking reduced-form model...
Two-step probit with endogenous regressors        Number of obs   =        500
                                                  Wald chi2(3)    =      93.97
                                                  Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   other_inc |   -.058473   .0093364    -6.26   0.000    -.0767719   -.040174
    fem_educ |    .227437   .0281628     8.08   0.000     .1722389    .282635
        kids |  -.1961748   .0496323    -3.95   0.000    -.2934522   -.0988973
       _cons |   .3956061   .4982649     0.79   0.427    -.5809752    1.372187
------------------------------------------------------------------------------
Instrumented:  other_inc
Instruments:   fem_educ kids male_educ
------------------------------------------------------------------------------
Wald test of exogeneity:    chi2(1) =     6.50          Prob > chi2 = 0.0108
```

All the coefficients have the same signs as their counterparts in the maximum likelihood model. The Wald test at the bottom of the output confirms our earlier finding of endogeneity.

◁

❑ Technical note

In a standard probit model, the error term is assumed to have a variance of one. In the probit model with endogenous regressors, we assume that $(u_i, \boldsymbol{v}_i)$ is multivariate normal with covariance matrix

$$\mathrm{Var}(u_i, \boldsymbol{v}_i) = \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \boldsymbol{\Sigma}'_{21} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

With the properties of the multivariate normal distribution, $\mathrm{Var}(u_i|\boldsymbol{v}_i) = 1 - \boldsymbol{\Sigma}'_{21}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. As a result, Newey's estimator and other two-step probit estimators do not yield estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ but rather $\boldsymbol{\beta}/\sigma$ and $\boldsymbol{\gamma}/\sigma$, where σ is the square root of $\mathrm{Var}(u_i|\boldsymbol{v}_i)$. Hence, we cannot directly compare the estimates obtained from Newey's estimator with those obtained via maximum likelihood or with those obtained from `probit`. See Wooldridge (2010, 585–594) for a discussion of Rivers and Vuong's (1988) two-step estimator. The issues raised pertaining to the interpretation of the coefficients of that estimator are identical to those that arise with Newey's estimator. Wooldridge also discusses ways to obtain marginal effects from two-step estimators.

❑

Despite the coefficients not being directly comparable to their maximum likelihood counterparts, the two-step estimator is nevertheless useful. The maximum likelihood estimator may have difficulty converging, especially with multiple endogenous variables. The two-step estimator, consisting of nothing more complicated than a probit regression, will almost certainly converge. Moreover, although the coefficients from the two models are not directly comparable, the two-step estimates can still be used to test for statistically significant relationships.

Model identification

As in the linear simultaneous-equation model, the order condition for identification requires that the number of excluded exogenous variables (that is, the additional instruments) be at least as great as the number of included endogenous variables. `ivprobit` checks this for you and issues an error message if the order condition is not met.

Like `probit`, `logit`, and `logistic`, `ivprobit` checks the exogenous and endogenous variables to see if any of them predict the outcome variable perfectly. It will then drop offending variables and observations and fit the model on the remaining data. Instruments that are perfect predictors do not affect estimation, so they are not checked. See *Model identification* in [R] **probit** for more information.

`ivprobit` will also occasionally display messages such as

```
Note: 4 failures and 0 successes completely determined.
```

For an explanation of this message, see [R] **logit**.

Stored results

`ivprobit, mle` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(N_cds)`	number of completely determined successes
`e(N_cdf)`	number of completely determined failures
`e(k)`	number of parameters
`e(k_eq)`	number of equations in `e(b)`
`e(k_eq_model)`	number of equations in overall model test
`e(k_aux)`	number of auxiliary parameters
`e(k_dv)`	number of dependent variables
`e(df_m)`	model degrees of freedom
`e(ll)`	log likelihood
`e(N_clust)`	number of clusters
`e(endog_ct)`	number of endogenous regressors
`e(p)`	model Wald p-value
`e(p_exog)`	exogeneity test Wald p-value
`e(chi2)`	model Wald χ^2
`e(chi2_exog)`	Wald χ^2 test of exogeneity
`e(rank)`	rank of `e(V)`
`e(ic)`	number of iterations
`e(rc)`	return code
`e(converged)`	1 if converged, 0 otherwise

Macros

e(cmd)	ivprobit
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(instd)	instrumented variables
e(insts)	instruments
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(clustvar)	name of cluster variable
e(chi2type)	Wald; type of model χ^2 test
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. Err.
e(asis)	asis, if specified
e(method)	ml
e(opt)	type of optimization
e(which)	max or min; whether optimizer is to perform maximization or minimization
e(ml_method)	type of ml method
e(user)	name of likelihood-evaluator program
e(technique)	maximization 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 the 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(Cns)	constraints matrix
e(rules)	information about perfect predictors
e(ilog)	iteration log (up to 20 iterations)
e(gradient)	gradient vector
e(Sigma)	$\widehat{\boldsymbol{\Sigma}}$
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample

`ivprobit, twostep` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(N_cds)`	number of completely determined successes
`e(N_cdf)`	number of completely determined failures
`e(df_m)`	model degrees of freedom
`e(df_exog)`	degrees of freedom for χ^2 test of exogeneity
`e(p)`	model Wald p-value
`e(p_exog)`	exogeneity test Wald p-value
`e(chi2)`	model Wald χ^2
`e(chi2_exog)`	Wald χ^2 test of exogeneity
`e(rank)`	rank of `e(V)`
Macros	
`e(cmd)`	`ivprobit`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(instd)`	instrumented variables
`e(insts)`	instruments
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(chi2type)`	`Wald`; type of model χ^2 test
`e(vce)`	*vcetype* specified in `vce()`
`e(vcetype)`	title used to label Std. Err.
`e(asis)`	`asis`, if specified
`e(method)`	`twostep`
`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 the 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(Cns)`	constraints matrix
`e(rules)`	information about perfect predictors
`e(V)`	variance–covariance matrix of the estimators
Functions	
`e(sample)`	marks estimation sample

Methods and formulas

Fitting limited-dependent variable models with endogenous regressors has received considerable attention in the econometrics literature. Building on the results of Amemiya (1978, 1979), Newey (1987) developed an efficient method of estimation that encompasses both Rivers and Vuong's (1988) simultaneous-equations probit model and Smith and Blundell's (1986) simultaneous-equations tobit model. With modern computers, maximum likelihood estimation is feasible as well. For compactness, we write the model as

$$y_{1i}^* = \boldsymbol{z}_i\boldsymbol{\delta} + u_i \tag{1a}$$

$$\boldsymbol{y}_{2i} = \boldsymbol{x}_i\boldsymbol{\Pi} + \boldsymbol{v}_i \tag{1b}$$

where $\boldsymbol{z}_i = (\boldsymbol{y}_{2i}, \boldsymbol{x}_{1i})$, $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$, $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, and $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_1', \boldsymbol{\Pi}_2')'$.

Deriving the likelihood function is straightforward because we can write the joint density $f\left(y_{1i}, \boldsymbol{y}_{2i}|\boldsymbol{x}_i\right)$ as $f\left(y_{1i}|\boldsymbol{y}_{2i}, \boldsymbol{x}_i\right) f\left(\boldsymbol{y}_{2i}|\boldsymbol{x}_i\right)$. When there is an endogenous regressor, the log likelihood for observation i is

$$\ln L_i = w_i \left[y_{1i} \ln\Phi\left(m_i\right) + \left(1 - y_{1i}\right) \ln\left\{1 - \Phi\left(m_i\right)\right\} + \ln\phi\left(\frac{y_{2i} - \boldsymbol{x}_i\boldsymbol{\Pi}}{\sigma}\right) - \ln\sigma \right]$$

where

$$m_i = \frac{\boldsymbol{z}_i\boldsymbol{\delta} + \rho\left(y_{2i} - \boldsymbol{x}_i\boldsymbol{\Pi}\right)/\sigma}{\left(1 - \rho^2\right)^{\frac{1}{2}}}$$

$\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions, respectively; σ is the standard deviation of v_i; ρ is the correlation coefficient between u_i and v_i; and w_i is the weight for observation i or one if no weights were specified. Instead of estimating σ and ρ, we estimate $\ln\sigma$ and $\operatorname{atanh}\rho$, where

$$\operatorname{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

For multiple endogenous regressors, let

$$\operatorname{Var}(u_i, \boldsymbol{v}_i) = \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \boldsymbol{\Sigma}_{21}' \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

As in any probit model, we have imposed the normalization $\operatorname{Var}(u_i) = 1$ to identify the model. The log likelihood for observation i is

$$\ln L_i = w_i \left[y_{1i} \ln\Phi\left(m_i\right) + \left(1 - y_{1i}\right) \ln\left\{1 - \Phi\left(m_i\right)\right\} + \ln f(\boldsymbol{y}_{2i}|\boldsymbol{x}_i) \right]$$

where

$$\ln f(\boldsymbol{y}_{2i}|\boldsymbol{x}_i) = -\frac{p}{2}\ln 2\pi - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_{22}\right| - \frac{1}{2}\left(\boldsymbol{y}_{2i} - \boldsymbol{x}_i\boldsymbol{\Pi}\right)\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{y}_{2i} - \boldsymbol{x}_i\boldsymbol{\Pi}\right)'$$

and

$$m_i = \left(1 - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-\frac{1}{2}}\left\{\boldsymbol{z}_i\boldsymbol{\delta} + \left(\boldsymbol{y}_{2i} - \boldsymbol{x}_i\boldsymbol{\Pi}\right)\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right\}$$

Instead of maximizing the log-likelihood function with respect to $\boldsymbol{\Sigma}$, we maximize with respect to the Cholesky decomposition $\boldsymbol{S}$ of $\boldsymbol{\Sigma}$; that is, there exists a lower triangular matrix, $\boldsymbol{S}$, such that $\boldsymbol{S}\boldsymbol{S}' = \boldsymbol{\Sigma}$. This maximization ensures that $\boldsymbol{\Sigma}$ is positive definite, as a covariance matrix must be. Let

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ s_{21} & s_{22} & 0 & \ldots & 0 \\ s_{31} & s_{32} & s_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{p+1,1} & s_{p+1,2} & s_{p+1,3} & \ldots & s_{p+1,p+1} \end{bmatrix}$$

With maximum likelihood estimation, this command supports the Huber/White/sandwich estimator of the variance and its clustered version using `vce(robust)` and `vce(cluster` *clustvar*`)`, respectively. See [P] **_robust**, particularly *Maximum likelihood estimators* and *Methods and formulas*.

The maximum likelihood version of `ivprobit` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] **variance estimation**.

The two-step estimates are obtained using Newey's (1987) minimum chi-squared estimator. The reduced-form equation for y_{1i}^* is

$$\begin{aligned} y_{1i}^* &= (\boldsymbol{x}_i\boldsymbol{\Pi} + \boldsymbol{v}_i)\boldsymbol{\beta} + \boldsymbol{x}_{1i}\boldsymbol{\gamma} + u_i \\ &= \boldsymbol{x}_i\boldsymbol{\alpha} + \boldsymbol{v}_i\boldsymbol{\beta} + u_i \\ &= \boldsymbol{x}_i\boldsymbol{\alpha} + \nu_i \end{aligned}$$

where $\nu_i = \boldsymbol{v}_i\boldsymbol{\beta} + u_i$. Because u_i and $\boldsymbol{v}_i$ are jointly normal, ν_i is also normal. Note that

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{\gamma} = D(\boldsymbol{\Pi})\boldsymbol{\delta}$$

where $D(\boldsymbol{\Pi}) = (\boldsymbol{\Pi}, \boldsymbol{I}_1)$ and $\boldsymbol{I}_1$ is defined such that $\boldsymbol{x}_i\boldsymbol{I}_1 = \boldsymbol{x}_{1i}$. Letting $\widehat{\boldsymbol{z}}_i = (\boldsymbol{x}_i\widehat{\boldsymbol{\Pi}}, \boldsymbol{x}_{1i})$, $\widehat{\boldsymbol{z}}_i\boldsymbol{\delta} = \boldsymbol{x}_i D(\widehat{\boldsymbol{\Pi}})\boldsymbol{\delta}$, where $D(\widehat{\boldsymbol{\Pi}}) = (\widehat{\boldsymbol{\Pi}}, \boldsymbol{I}_1)$. Thus one estimator of $\boldsymbol{\alpha}$ is $D(\widehat{\boldsymbol{\Pi}})\boldsymbol{\delta}$; denote this estimator by $\widehat{\boldsymbol{D}}\boldsymbol{\delta}$.

$\boldsymbol{\alpha}$ could also be estimated directly as the solution to

$$\max_{\boldsymbol{\alpha},\boldsymbol{\lambda}} \sum_{i=1}^{N} l(y_{1i}, \boldsymbol{x}_i\boldsymbol{\alpha} + \widehat{\boldsymbol{v}}_i\boldsymbol{\lambda}) \tag{2}$$

where $l(\cdot)$ is the log likelihood for probit. Denote this estimator by $\widetilde{\boldsymbol{\alpha}}$. The inclusion of the $\widehat{\boldsymbol{v}}_i\boldsymbol{\lambda}$ term follows because the multivariate normality of $(u_i, \boldsymbol{v}_i)$ implies that, conditional on $\boldsymbol{y}_{2i}$, the expected value of u_i is nonzero. Because $\boldsymbol{v}_i$ is unobservable, the least-squares residuals from fitting (1*b*) are used.

Amemiya (1978) shows that the estimator of $\boldsymbol{\delta}$ defined by

$$\max_{\boldsymbol{\delta}} \; (\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{D}}\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}^{-1}(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{D}}\boldsymbol{\delta})$$

where $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of the covariance of $\sqrt{N}(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{D}}\boldsymbol{\delta})$, is asymptotically efficient relative to all other estimators that minimize the distance between $\widetilde{\boldsymbol{\alpha}}$ and $D(\widehat{\boldsymbol{\Pi}})\boldsymbol{\delta}$. Thus an efficient estimator of $\boldsymbol{\delta}$ is

$$\widehat{\boldsymbol{\delta}} = (\widehat{\boldsymbol{D}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{D}}'\widehat{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\alpha}} \tag{3}$$

and

$$\mathrm{Var}(\widehat{\boldsymbol{\delta}}) = (\widehat{\boldsymbol{D}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{D}})^{-1} \tag{4}$$

To implement this estimator, we need $\widehat{\boldsymbol{\Omega}}^{-1}$.

Consider the two-step maximum likelihood estimator that results from first fitting (1*b*) by OLS and computing the residuals $\widehat{\boldsymbol{v}}_i = \boldsymbol{y}_{2i} - \boldsymbol{x}_i\widehat{\boldsymbol{\Pi}}$. The estimator is then obtained by solving

$$\max_{\boldsymbol{\delta},\boldsymbol{\lambda}} \sum_{i=1}^{N} l(y_{1i}, \boldsymbol{z}_i\boldsymbol{\delta} + \widehat{\boldsymbol{v}}_i\boldsymbol{\lambda})$$

This is the two-step instrumental variables (2SIV) estimator proposed by Rivers and Vuong (1988), and its role will become apparent shortly.

From Proposition 5 of Newey (1987), $\sqrt{N}(\widetilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{D}}\boldsymbol{\delta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$, where

$$\boldsymbol{\Omega} = \boldsymbol{J}_{\alpha\alpha}^{-1} + (\boldsymbol{\lambda} - \boldsymbol{\beta})'\boldsymbol{\Sigma}_{22}(\boldsymbol{\lambda} - \boldsymbol{\beta})\boldsymbol{Q}^{-1}$$

and $\boldsymbol{\Sigma}_{22} = E\{\boldsymbol{v}_i'\boldsymbol{v}_i\}$. $\boldsymbol{J}_{\alpha\alpha}^{-1}$ is simply the covariance matrix of $\widetilde{\boldsymbol{\alpha}}$, ignoring that $\widehat{\boldsymbol{\Pi}}$ is an estimated parameter matrix. Moreover, Newey shows that the covariance matrix from an OLS regression of $\boldsymbol{y}_{2i}(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\beta}})$ on $\boldsymbol{x}_i$ is a consistent estimator of the second term. $\widehat{\boldsymbol{\lambda}}$ can be obtained from solving (2), and the 2SIV estimator yields a consistent estimate, $\widehat{\boldsymbol{\beta}}$.

Mechanically, estimation proceeds in several steps.

1. Each of the endogenous right-hand-side variables is regressed on all the exogenous variables, and the fitted values and residuals are calculated. The matrix $\widehat{D} = D(\widehat{\boldsymbol{\Pi}})$ is assembled from the estimated coefficients.
2. `probit` is used to solve (2) and obtain $\widetilde{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\lambda}}$. The portion of the covariance matrix corresponding to $\boldsymbol{\alpha}$, $\boldsymbol{J}_{\alpha\alpha}^{-1}$, is also saved.
3. The 2SIV estimator is evaluated, and the parameters $\widehat{\boldsymbol{\beta}}$ corresponding to $\boldsymbol{y}_{2i}$ are collected.
4. $\boldsymbol{y}_{2i}(\widehat{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\beta}})$ is regressed on $\boldsymbol{x}_i$. The covariance matrix of the parameters from this regression is added to $\boldsymbol{J}_{\alpha\alpha}^{-1}$, yielding $\widehat{\boldsymbol{\Omega}}$.
5. Evaluating (3) and (4) yields the estimates $\widehat{\boldsymbol{\delta}}$ and $\mathrm{Var}(\widehat{\boldsymbol{\delta}})$.
6. A Wald test of the null hypothesis $H_0 : \boldsymbol{\lambda} = \mathbf{0}$, using the 2SIV estimates, serves as our test of exogeneity.

The two-step estimates are not directly comparable to those obtained from the maximum likelihood estimator or from `probit`. The argument is the same for Newey's efficient estimator as for Rivers and Vuong's (1988) 2SIV estimator, so we consider the simpler 2SIV estimator. From the properties of the normal distribution,

$$E(u_i|\boldsymbol{v}_i) = \boldsymbol{v}_i\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \qquad \text{and} \qquad \mathrm{Var}(u_i|\boldsymbol{v}_i) = 1 - \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

We write u_i as $u_i = \boldsymbol{v}_i\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} + e_i = \boldsymbol{v}_i\boldsymbol{\lambda} + e_i$, where $e_i \sim \mathrm{N}(0, 1-\rho^2)$, $\rho^2 = \boldsymbol{\Sigma}_{21}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, and e_i is independent of $\boldsymbol{v}_i$. In the second stage of 2SIV, we use a probit regression to estimate the parameters of

$$y_{1i} = \boldsymbol{z}_i\boldsymbol{\delta} + \boldsymbol{v}_i\boldsymbol{\lambda} + e_i$$

Because $\boldsymbol{v}_i$ is unobservable, we use the sample residuals from the first-stage regressions.

$$\Pr(y_{1i} = 1|\boldsymbol{z}_i, \boldsymbol{v}_i) = \Pr(\boldsymbol{z}_i\boldsymbol{\delta} + \boldsymbol{v}_i\boldsymbol{\lambda} + e_i > 0|\boldsymbol{z}_i, \boldsymbol{v}_i) = \Phi\left\{(1-\rho^2)^{-\frac{1}{2}}(\boldsymbol{z}_i\boldsymbol{\delta} + \boldsymbol{v}_i\boldsymbol{\lambda})\right\}$$

Hence, as mentioned previously, 2SIV and Newey's estimator do not estimate $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ but rather

$$\boldsymbol{\delta}_\rho = \frac{1}{(1-\rho^2)^{\frac{1}{2}}}\boldsymbol{\delta} \qquad \text{and} \qquad \boldsymbol{\lambda}_\rho = \frac{1}{(1-\rho^2)^{\frac{1}{2}}}\boldsymbol{\lambda}$$

Acknowledgments

The two-step estimator is based on the `probitiv` command written by Jonah Gelbach of the Department of Economics at Yale University and the `ivprob` command written by Joe Harkness of the Institute of Policy Studies at Johns Hopkins University.

References

Amemiya, T. 1978. The estimation of a simultaneous equation generalized probit model. *Econometrica* 46: 1193–1205.

——. 1979. The estimation of a simultaneous-equation tobit model. *International Economic Review* 20: 169–181.

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.

Miranda, A., and S. Rabe-Hesketh. 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *Stata Journal* 6: 285–308.

Newey, W. K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36: 231–250.

Rivers, D., and Q. H. Vuong. 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39: 347–366.

Smith, R. J., and R. Blundell. 1986. An exogeneity test for the simultaneous equation tobit model with an application to labor supply. *Econometrica* 54: 679–685.

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

Also see

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

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

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

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

[R] **probit** — Probit regression

[SVY] **svy estimation** — Estimation commands for survey data

[XT] **xtprobit** — Random-effects and population-averaged probit models

[U] 20 Estimation and postestimation commands