ivpanel: IV estimation of panel-data models

Allin Cottrell

February 15, 2021

1 Introduction

This package estimates three variants of panel data models with instrumental variables, namely fixed-effects, the "between" model, and random effects (Generalized 2-Stage Least Squares or G2SLS). An extended discussion of these models can be found in chapter 7 of Baltagi (2005). The sample script provided with this package replicates results obtained by Baltagi using Stata for each of the three supported models.

The required arguments to the main public function `ivpanel()` are as follows:

1. y (series), the dependent variable
2. X (list), the regressors, both exogenous and endogenous
3. Z (list), the instruments, including any exogenous regressors

Note that exogenous variables in list X should be repeated in Z, as in gretl's `tsls` command. The constant is automatically added to X and Z if it's not already present.

An optional fourth argument, *case*, can be used to specify the type of model. The argument is an integer switch that takes values 1, 2 or 3 (this may be extended in future). A value of 1 means fixed effects, and is implicit if no fourth argument is given; a value of 2 means to use the between estimator; and a value of 3 means G2SLS. Leaving aside the constant, all the members of X and Z must be time-varying when the fixed effects case is selected.

An optional final boolean argument, *quiet*, controls the printing of output: if *quiet* is set to a non-zero value, printing is suppressed.

2 The ivpanel bundle

By way of return value, `ivpanel()` offers a gretl bundle containing the items shown in Table 1.

Some comments on the bundle members follow.

- The R^2 value, `rsq`, is calculated as the square of the correlation between the dependent variable and the fitted values.
- The `dims` vector, if present, contains three elements holding, respectively, the number of groups used and the minimum and maximum time-series spans.
- The `Fpool` statistic for fixed effects is calculated as per Wooldridge (1990), adjusted for the panel case. The formula is

$$\frac{\text{SSRr} - \text{SSRu}}{\text{SSRf}} \times \frac{\text{dfd}}{\text{dfn}}$$

  where SSRf is the sum of squared residuals from the fixed-effects IV model and the other two SSR values are calculated thus:

name	*type*	*description*
`case`	scalar	the value of the *case* argument on input
`nobs`	scalar	the total number of observations used
`coeff`	matrix	column vector of coefficients
`stderr`	matrix	column vector of standard errors
`vcv`	matrix	the coefficient covariance matrix
`uhat`	matrix	the vector of residuals
`yhat`	matrix	the vector of fitted values
`SSR`	scalar	sum of squared residuals
`sigma`	scalar	standard error of the regression
`df`	scalar	degrees of the freedom for the regression
`rsq`	scalar	correlation-based R^2
`wald`	scalar	Wald joint χ^2 test for all regressors
`modstr`	string	short description of the model
`ystr`	string	the name of the dependent variable
`Estr`	string	the names of the endogenous regressor(s)
`Istr`	string	the names of the (additional) instruments
`dims`	matrix	data dimensions (cases 1 and 3 only)
`Fpool`	scalar	F-test for joint significance of fixed effects (case 1 only)

Table 1: Items in ivpanel bundle

1. We run the fixed-effects first-stage regressions and save the fitted values; we then use these to replace the endogenous regressors.
2. SSRr is then obtained via OLS, and SSRu from a fixed-effects regression. SSRu differs from SSRf in that it uses the "raw" residuals, without correction as one usually applies with two stage least squares (i.e. replacing the first-stage fitted values with the actual data for the endogenous regressors when computing the fitted values).

3 Sample script

The sample script estimates three models of the (log) crime rate across the counties of North Carolina over the years 1981 to 1987, using data from Cornwell and Trumbull (1994). The endogenous regressors are `lpolpc` (log of police per capita) and `lprbarr` (log of the estimated probability of arrest). The instruments (besides the exogenous regressors) are `ltaxpc` (log of tax revenue per capita), `lmix` (log of the offense "mix", face-to-face versus other) and a set of year dummies.

It's perhaps noteworthy that despite the IV approach, the coefficient on `lpolpc` is positive in all the models, the opposite of what would be expected if the estimator had succeeded in picking up a causal relationship. At least in the fixed effects specification the positive `lpolpc` coefficient is not statistically significant (P-value > 0.4).

4 Graphical interface

An entry-point for ivpanel can be found under the Panel sub-menu of gretl's Model menu: the label is "Panel IV model." See Figure 1. The specification of the endogenous regressors is implicit: any variables that appear in the "regressors" list and not in the "instruments" list are taken to be endogenous.

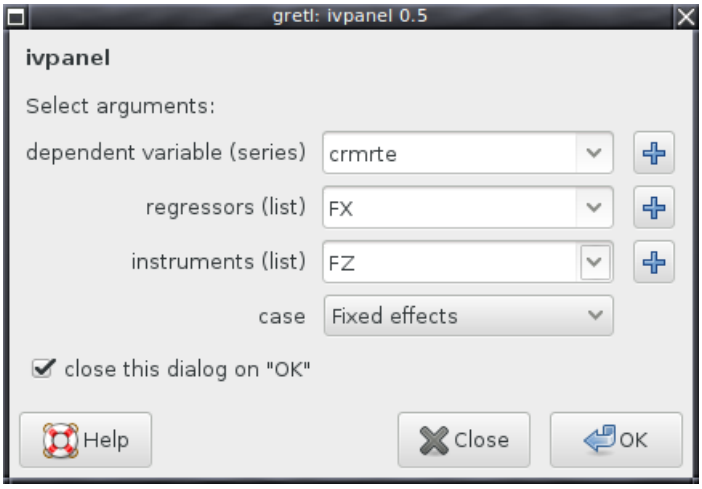

Figure 1: Specify arguments for ivpanel

5 Auxiliary printing function

The auxiliary public function `ivp_print()` is provided to "pretty-print" the results contained in the bundle provided by `ivpanel()`; `ivp_print()` takes a pointer to the bundle as its sole argument.

6 Change log

2021-02-15: add guard against excessive collinearity.
2020-09-22: avoid use of deprecated `isnull` function.
2015-10-17: first release as addon.

References

Baltagi, B. H. (2005) *Econometric Analysis of Panel Data, 3e*, Chichester: Wiley.

Cornwell, C. and W. Trumbull (1994) 'Estimating the economic model of crime with panel data', *Review of Economics and Statistics* 76: 360–366.

Wooldridge, J. M. (1990) 'A note on the Lagrange multiplier and F-statistics for two stage least squares regressions', *Economics Letters* 34: 151–155.