Title

xtivreg — Instrumental variables and two-stage least squares for panel-data models

Description

`xtivreg` offers five different estimators for fitting panel-data models in which some of the right-hand-side covariates are endogenous. These estimators are two-stage least-squares generalizations of simple panel-data estimators for exogenous variables. `xtivreg` with the `be` option uses the two-stage least-squares between estimator. `xtivreg` with the `fe` option uses the two-stage least-squares within estimator. `xtivreg` with the `re` option uses a two-stage least-squares random-effects estimator. There are two implementations: G2SLS from Balestra and Varadharajan-Krishnakumar (1987) and EC2SLS from Baltagi. The Balestra and Varadharajan-Krishnakumar G2SLS is the default because it is computationally less expensive. Baltagi's EC2SLS can be obtained by specifying the `ec2sls` option. `xtivreg` with the `fd` option requests the two-stage least-squares first-differenced estimator.

See Baltagi (2013) for an introduction to panel-data models with endogenous covariates. For the derivation and application of the first-differenced estimator, see Anderson and Hsiao (1981).

Quick start

Random-effects linear panel-data model with outcome `y`, exogenous `x1`, and `x2` instrumented by `x3` using `xtset` data

```
xtivreg y x1 (x2 = x3)
```

Use fixed-effects estimator and include indicators for each level of categorical variable `a`

```
xtivreg y x1 i.a (x2 = x3), fe
```

Use between-effects estimator and include indicators for levels of `b` as instruments

```
xtivreg y x1 i.a (x2 = x3 i.b), be
```

First-differenced model of `y` as a function of `x1` and `x2` and the lag of `y` instrumented by its third lag

```
xtivreg y x1 x2 (L.y = L3.y), fd
```

Menu

Statistics > Longitudinal/panel data > Endogenous covariates > Instrumental-variables regression (FE, RE, BE, FD)

Syntax

GLS random-effects (RE) model

`xtivreg` *depvar* [*varlist*$_1$] `(`*varlist*$_2$ `=` *varlist*$_{iv}$`)` [*if*] [*in*] [`, re` *RE_options*]

Between-effects (BE) model

`xtivreg` *depvar* [*varlist*$_1$] `(`*varlist*$_2$ `=` *varlist*$_{iv}$`)` [*if*] [*in*] `, be` [*BE_options*]

Fixed-effects (FE) model

`xtivreg` *depvar* [*varlist*$_1$] `(`*varlist*$_2$ `=` *varlist*$_{iv}$`)` [*if*] [*in*] `, fe` [*FE_options*]

First-differenced (FD) estimator

`xtivreg` *depvar* [*varlist*$_1$] `(`*varlist*$_2$ `=` *varlist*$_{iv}$`)` [*if*] [*in*] `, fd` [*FD_options*]

RE_options	Description
Model	
`re`	use random-effects estimator; the default
`ec2sls`	use Baltagi's EC2SLS random-effects estimator
`nosa`	use the Baltagi–Chang estimators of the variance components
`regress`	treat covariates as exogenous and ignore instrumental variables
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `conventional`, `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
Reporting	
`level(`*#*`)`	set confidence level; default is `level(95)`
`first`	report first-stage estimates
`small`	report t and F statistics instead of Z and χ^2 statistics
`theta`	report θ
display_options	control columns and 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

BE_options	Description
Model	
be	use between-effects estimator
regress	treat covariates as exogenous and ignore instrumental variables
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, robust, cluster *clustvar*, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
first	report first-stage estimates
small	report t and F statistics instead of Z and χ^2 statistics
display_options	control columns and 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

FE_options	Description
Model	
fe	use fixed-effects estimator
regress	treat covariates as exogenous and ignore instrumental variables
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, robust, cluster *clustvar*, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
first	report first-stage estimates
small	report t and F statistics instead of Z and χ^2 statistics
display_options	control columns and 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

FD_options	Description
Model	
noconstant	suppress constant term
fd	use first-differenced estimator
regress	treat covariates as exogenous and ignore instrumental variables
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, robust, cluster *clustvar*, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
first	report first-stage estimates
small	report t and F statistics instead of Z and χ^2 statistics
display_options	control columns and column formats, row spacing, line width, and display of omitted variables
coeflegend	display legend instead of statistics

A panel variable must be specified. For xtivreg, fd, a time variable must also be specified. Use xtset; see [XT] **xtset**.

varlist$_1$ and *varlist*$_{\text{iv}}$ may contain factor variables, except for the fd estimator; see **[U] 11.4.3 Factor variables**.

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

by, collect, and statsby are allowed; see **[U] 11.1.10 Prefix commands**.

coeflegend does not appear in the dialog box.

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

Options for RE model

Model

re requests the G2SLS random-effects estimator. re is the default.

ec2sls requests Baltagi's EC2SLS random-effects estimator instead of the default Balestra and Varadharajan-Krishnakumar estimator.

nosa specifies that the Baltagi–Chang estimators of the variance components be used instead of the default adapted Swamy–Arora estimators.

regress specifies that all the covariates be treated as exogenous and that the instrument list be ignored. Specifying regress causes xtivreg to fit the requested panel-data regression model of *depvar* on *varlist*$_1$ and *varlist*$_2$, ignoring *varlist*$_{\text{iv}}$.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), 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 [XT] ***vce_options***.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Specifying vce(robust) is equivalent to specifying vce(cluster *panelvar*); see *xtreg, re* in *Methods and formulas* of [XT] **xtreg**.

Reporting

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

first specifies that the first-stage regressions be displayed.

small specifies that t statistics be reported instead of Z statistics and that F statistics be reported instead of χ^2 statistics.

theta specifies that the output include the estimated value of θ used in combining the between and fixed estimators. For balanced data, this is a constant, and for unbalanced data, a summary of the values is presented in the header of the output.

display_options: noci, nopvalues, 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 xtivreg but is not shown in the dialog box:

coeflegend; see [R] **Estimation options**.

Options for BE model

Model

be requests the between regression estimator.

regress specifies that all the covariates be treated as exogenous and that the instrument list be ignored. Specifying regress causes xtivreg to fit the requested panel-data regression model of *depvar* on $varlist_1$ and $varlist_2$, ignoring $varlist_{\text{iv}}$.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (conventional), 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 [XT] ***vce_options***.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Specifying vce(robust) is equivalent to specifying vce(cluster *panelvar*); see *xtreg, fe* in *Methods and formulas* of [XT] **xtreg**.

Reporting

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

first specifies that the first-stage regressions be displayed.

small specifies that t statistics be reported instead of Z statistics and that F statistics be reported instead of χ^2 statistics.

display_options: noci, nopvalues, 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 `xtivreg` but is not shown in the dialog box:

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

Options for FE model

Model

`fe` requests the fixed-effects (within) regression estimator.

`regress` specifies that all the covariates be treated as exogenous and that the instrument list be ignored. Specifying `regress` causes `xtivreg` to fit the requested panel-data regression model of *depvar* on $varlist_1$ and $varlist_2$, ignoring $varlist_{iv}$.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`), 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 [XT] ***vce_options***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Specifying `vce(robust)` is equivalent to specifying `vce(cluster` *panelvar*`)`; see *xtreg, fe* in *Methods and formulas* of [XT] **xtreg**.

Reporting

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

`first` specifies that the first-stage regressions be displayed.

`small` specifies that t statistics be reported instead of Z statistics and that F statistics be reported instead of χ^2 statistics.

display_options: `noci`, `nopvalues`, `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 `xtivreg` but is not shown in the dialog box:

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

Options for FD model

Model

`noconstant`; see [R] **Estimation options**.

`fd` requests the first-differenced regression estimator.

`regress` specifies that all the covariates be treated as exogenous and that the instrument list be ignored. Specifying `regress` causes `xtivreg` to fit the requested panel-data regression model of *depvar* on $varlist_1$ and $varlist_2$, ignoring $varlist_{iv}$.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`conventional`), 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 [XT] ***vce_options***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Specifying `vce(robust)` is equivalent to specifying `vce(cluster` *panelvar*`)`; see *xtreg, fe* in *Methods and formulas* of [XT] **xtreg**.

Reporting

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

`first` specifies that the first-stage regressions be displayed.

`small` specifies that t statistics be reported instead of Z statistics and that F statistics be reported instead of χ^2 statistics.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `cformat(`*%fmt*`)`, `pformat(`*%fmt*`)`, `sformat(`*%fmt*`)`, and `nolstretch`; see [R] **Estimation options**.

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

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

Remarks and examples

stata.com

If you have not read [XT] **xt**, please do so.

Consider an equation of the form

$$y_{it} = \mathbf{Y}_{it}\gamma + \mathbf{X}_{1it}\boldsymbol{\beta} + \mu_i + \nu_{it} = \mathbf{Z}_{it}\boldsymbol{\delta} + \mu_i + \nu_{it} \tag{1}$$

where

y_{it} is the dependent variable;

$\mathbf{Y}_{it}$ is an $1 \times g_2$ vector of observations on g_2 endogenous variables included as covariates, and these variables are allowed to be correlated with the ν_{it};

$\mathbf{X}_{1it}$ is an $1 \times k_1$ vector of observations on the exogenous variables included as covariates;

$\mathbf{Z}_{it} = [\mathbf{Y}_{it}\ \mathbf{X}_{it}]$;

γ is a $g_2 \times 1$ vector of coefficients;

$\boldsymbol{\beta}$ is a $k_1 \times 1$ vector of coefficients; and

$\boldsymbol{\delta}$ is a $K \times 1$ vector of coefficients, where $K = g_2 + k_1$.

Assume that there is a $1 \times k_2$ vector of observations on the k_2 instruments in $\mathbf{X}_{2it}$. The order condition is satisfied if $k_2 \geq g_2$. Let $\mathbf{X}_{it} = [\mathbf{X}_{1it}\ \mathbf{X}_{2it}]$. `xtivreg` handles exogenously unbalanced panel data. Thus define T_i to be the number of observations on panel i, n to be the number of panels and N to be the total number of observations; that is, $N = \sum_{i=1}^{n} T_i$.

xtivreg offers five different estimators, which may be applied to models having the form of (1). The first-differenced estimator (FD2SLS) removes the μ_i by fitting the model in first differences. The within estimator (FE2SLS) fits the model after sweeping out the μ_i by removing the panel-level means from each variable. The between estimator (BE2SLS) models the panel averages. The two random-effects estimators, G2SLS and EC2SLS, treat the μ_i as random variables that are independent and identically distributed (i.i.d.) over the panels. Except for (FD2SLS), all of these estimators are generalizations of estimators in xtreg. See [XT] **xtreg** for a discussion of these estimators for exogenous covariates.

Although the estimators allow for different assumptions about the μ_i, all the estimators assume that the idiosyncratic error term ν_{it} has zero mean and is uncorrelated with the variables in $\mathbf{X}_{it}$. Just as when there are no endogenous covariates, as discussed in [XT] **xtreg**, there are various perspectives on what assumptions should be placed on the μ_i. If they are assumed to be fixed, the μ_i may be correlated with the variables in $\mathbf{X}_{it}$, and the within estimator is efficient within a class of limited information estimators. Alternatively, if the μ_i are assumed to be random, they are also assumed to be i.i.d. over the panels. If the μ_i are assumed to be uncorrelated with the variables in $\mathbf{X}_{it}$, the GLS random-effects estimators are more efficient than the within estimator. However, if the μ_i are correlated with the variables in $\mathbf{X}_{it}$, the random-effects estimators are inconsistent but the within estimator is consistent. The price of using the within estimator is that it is not possible to estimate coefficients on time-invariant variables, and all inference is conditional on the μ_i in the sample. See Mundlak (1978) and Hsiao (2014) for discussions of this interpretation of the within estimator.

▷ Example 1: Fixed-effects model

For the within estimator, consider another version of the wage equation discussed in [XT] **xtreg**. The data for this example come from an extract of women from the National Longitudinal Survey of Youth that was described in detail in [XT] **xt**. Restricting ourselves to only time-varying covariates, we might suppose that the log of the real wage was a function of the individual's age, age^2, her tenure in the observed place of employment, whether she belonged to union, whether she lives in metropolitan area, and whether she lives in the south. The variables for these are, respectively, age, c.age#c.age, tenure, union, not_smsa, and south. If we treat all the variables as exogenous, we can use the one-stage within estimator from xtreg, yielding

```
. use https://www.stata-press.com/data/r18/nlswork
(National Longitudinal Survey of Young Women, 14-24 years old in 1968)

. xtreg ln_w age c.age#c.age tenure not_smsa union south, fe

Fixed-effects (within) regression               Number of obs     =     19,007
Group variable: idcode                          Number of groups  =      4,134

R-squared:                                      Obs per group:
     Within  = 0.1333                                         min =          1
     Between = 0.2375                                         avg =        4.6
     Overall = 0.2031                                         max =         12

                                                F(6, 14867)       =     381.19
corr(u_i, Xb) = 0.2074                          Prob > F          =     0.0000

------------------------------------------------------------------------------
     ln_wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0311984   .0033902     9.20   0.000     .0245533    .0378436
             |
 c.age#c.age |  -.0003457   .0000543    -6.37   0.000    -.0004522   -.0002393
             |
      tenure |   .0176205   .0008099    21.76   0.000     .0160331    .0192079
    not_smsa |  -.0972535   .0125377    -7.76   0.000    -.1218289    -.072678
       union |   .0975672   .0069844    13.97   0.000     .0838769    .1112576
       south |  -.0620932    .013327    -4.66   0.000    -.0882158   -.0359706
       _cons |   1.091612   .0523126    20.87   0.000     .9890729    1.194151
-------------+----------------------------------------------------------------
     sigma_u |   .3910683
     sigma_e |  .25545969
         rho |  .70091004   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(4133, 14867) = 8.31                 Prob > F = 0.0000
```

All the coefficients are statistically significant and have the expected signs.

Now suppose that we wish to model tenure as a function of union and south and that we believe that the errors in the two equations are correlated. Because we are still interested in the within estimates, we now need a two-stage least-squares estimator. The following output shows the command and the results from fitting this model:

```
. xtivreg ln_w age c.age#c.age not_smsa (tenure =  union south), fe
Fixed-effects (within) IV regression            Number of obs     =     19,007
Group variable: idcode                          Number of groups  =      4,134

R-squared:                                      Obs per group:
     Within  =      .                                        min =          1
     Between = 0.1304                                        avg =        4.6
     Overall = 0.0897                                        max =         12

                                                Wald chi2(4)      =  147926.58
corr(u_i, Xb) = -0.6843                         Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     ln_wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      tenure |   .2403531   .0373419     6.44   0.000     .1671643    .3135419
         age |   .0118437   .0090032     1.32   0.188    -.0058023    .0294897
             |
 c.age#c.age |  -.0012145   .0001968    -6.17   0.000    -.0016003   -.0008286
             |
    not_smsa |  -.0167178   .0339236    -0.49   0.622    -.0832069    .0497713
       _cons |   1.678287   .1626657    10.32   0.000     1.359468    1.997106
-------------+----------------------------------------------------------------
     sigma_u |  .70661941
     sigma_e |  .63029359
         rho |  .55690561   (fraction of variance due to u_i)
------------------------------------------------------------------------------
 F test that all u_i=0: F(4133,14869) =     1.44          Prob > F    = 0.0000
------------------------------------------------------------------------------
Endogenous: tenure
Exogenous:  age c.age#c.age not_smsa union south
```

Although all the coefficients still have the expected signs, the coefficients on age and not_smsa are no longer statistically significant. Given that these variables have been found to be important in many other studies, we might want to rethink our specification.

◁

If we are willing to assume that the μ_i are uncorrelated with the other covariates, we can fit a random-effects model. The model is frequently known as the variance-components or error-components model. xtivreg has estimators for two-stage least-squares one-way error-components models. In the one-way framework, there are two variance components to estimate, the variance of the μ_i and the variance of the ν_{it}. Because the variance components are unknown, consistent estimates are required to implement feasible GLS. xtivreg offers two choices: a Swamy–Arora method and simple consistent estimators from Baltagi and Chang (2000).

Baltagi and Chang (1994) derived the Swamy–Arora estimators of the variance components for unbalanced panels. By default, xtivreg uses estimators that extend these unbalanced Swamy–Arora estimators to the case with instrumental variables. The default Swamy–Arora method contains a degree-of-freedom correction to improve its performance in small samples. Baltagi and Chang (2000) use variance-components estimators, which are based on the ideas of Amemiya (1971) and Swamy and Arora (1972), but they do not attempt to make small-sample adjustments. These consistent estimators of the variance components will be used if the nosa option is specified.

Using either estimator of the variance components, xtivreg offers two GLS estimators of the random-effects model. These two estimators differ only in how they construct the GLS instruments from the exogenous and instrumental variables contained in $\mathbf{X}_{it} = [\mathbf{X}_{1it}\ \mathbf{X}_{2it}]$. The default method, G2SLS, which is from Balestra and Varadharajan-Krishnakumar, uses the exogenous variables after they have been passed through the feasible GLS transform. In math, G2SLS uses $\mathbf{X}^*_{it}$ for the GLS instruments, where $\mathbf{X}^*_{it}$ is constructed by passing each variable in $\mathbf{X}_{it}$ through the GLS transform in (3) given in *Methods and formulas*. If the ec2sls option is specified, xtivreg performs Baltagi's

EC2SLS. In EC2SLS, the instruments are $\widetilde{\mathbf{X}}_{it}$ and $\overline{\mathbf{X}}_{it}$, where $\widetilde{\mathbf{X}}_{it}$ is constructed by passing each of the variables in $\mathbf{X}_{it}$ through the within transform, and $\overline{\mathbf{X}}_{it}$ is constructed by passing each variable through the between transform. The within and between transforms are given in the *Methods and formulas* section. Baltagi and Li (1992) show that, although the G2SLS instruments are a subset of those contained in EC2SLS, the extra instruments in EC2SLS are redundant in the sense of White (2001). Given the extra computational cost, G2SLS is the default.

▷ Example 2: GLS random-effects model

Here is the output from applying the G2SLS estimator to this model:

```
. xtivreg ln_w age c.age#c.age not_smsa 2.race (tenure = union birth south), re

G2SLS random-effects IV regression              Number of obs     =     19,007
Group variable: idcode                          Number of groups  =      4,134

R-squared:                                      Obs per group:
     Within  = 0.0664                                         min =          1
     Between = 0.2098                                         avg =        4.6
     Overall = 0.1463                                         max =         12

                                                Wald chi2(5)      =    1446.37
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     ln_wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      tenure |   .1391798   .0078756    17.67   0.000      .123744    .1546157
         age |   .0279649   .0054182     5.16   0.000     .0173454    .0385843
             |
 c.age#c.age |  -.0008357   .0000871    -9.60   0.000    -.0010063    -.000665
             |
    not_smsa |  -.2235103   .0111371   -20.07   0.000    -.2453386   -.2016821
             |
        race |
       Black |  -.2078613   .0125803   -16.52   0.000    -.2325183   -.1832044
       _cons |   1.337684   .0844988    15.83   0.000     1.172069    1.503299
-------------+----------------------------------------------------------------
     sigma_u |  .36582493
     sigma_e |  .63031479
         rho |  .25197078   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Endogenous: tenure
Exogenous:  age c.age#c.age not_smsa 2.race union birth_yr south
```

We have included two time-invariant covariates, `birth_yr` and `2.race`. All the coefficients are statistically significant and are of the expected sign.

Applying the EC2SLS estimator yields similar results:

```
. xtivreg ln_w age c.age#c.age not_smsa 2.race (tenure = union birth south), re
> ec2sls

EC2SLS random-effects IV regression             Number of obs     =     19,007
Group variable: idcode                          Number of groups  =      4,134

R-squared:                                      Obs per group:
     Within  = 0.0898                                         min =          1
     Between = 0.2608                                         avg =        4.6
     Overall = 0.1926                                         max =         12

                                                Wald chi2(5)      =    2721.92
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     ln_wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      tenure |    .064822   .0025647    25.27   0.000     .0597953    .0698486
         age |   .0380048   .0039549     9.61   0.000     .0302534    .0457562
             |
 c.age#c.age |  -.0006676   .0000632   -10.56   0.000    -.0007915   -.0005438
             |
    not_smsa |  -.2298961   .0082993   -27.70   0.000    -.2461625   -.2136297
             |
        race |
      Black  |  -.1823627   .0092005   -19.82   0.000    -.2003954     -.16433
       _cons |   1.110564   .0606538    18.31   0.000     .9916849    1.229443
-------------+----------------------------------------------------------------
     sigma_u |  .36582493
     sigma_e |  .63031479
         rho |  .25197078   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Endogenous: tenure
Exogenous:  age c.age#c.age not_smsa 2.race union birth_yr south
```

Fitting the same model as above with the G2SLS estimator and the consistent variance components estimators yields

```
. xtivreg ln_w age c.age#c.age not_smsa 2.race (tenure = union birth south),
> re nosa

G2SLS random-effects IV regression              Number of obs     =     19,007
Group variable: idcode                          Number of groups  =      4,134

R-squared:                                      Obs per group:
     Within  = 0.0664                                         min =          1
     Between = 0.2098                                         avg =        4.6
     Overall = 0.1463                                         max =         12

                                                Wald chi2(5)      =    1446.93
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     ln_wage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      tenure |   .1391859    .007873    17.68   0.000     .1237552    .1546166
         age |   .0279697    .005419     5.16   0.000     .0173486    .0385909
             |
 c.age#c.age |  -.0008357   .0000871    -9.60   0.000    -.0010064    -.000665
             |
    not_smsa |  -.2235738   .0111344   -20.08   0.000    -.2453967   -.2017508
             |
        race |
      Black  |  -.2078733   .0125751   -16.53   0.000    -.2325201   -.1832265
       _cons |   1.337522   .0845083    15.83   0.000     1.171889    1.503155
-------------+----------------------------------------------------------------
     sigma_u |  .36535633
     sigma_e |  .63020883
         rho |   .2515512   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Endogenous: tenure
Exogenous:  age c.age#c.age not_smsa 2.race union birth_yr south
```

◁

▷ Example 3: First-differenced estimator

The two-stage least-squares first-differenced estimator (FD2SLS) has been used to fit both fixed-effect and random-effect models. If the μ_i are truly fixed-effects, the FD2SLS estimator is not as efficient as the two-stage least-squares within estimator for finite T_i. Similarly, if none of the endogenous variables are lagged dependent variables, the exogenous variables are all strictly exogenous, and the random effects are i.i.d. and independent of the $\mathbf{X}_{it}$, the two-stage GLS estimators are more efficient than the FD2SLS estimator. However, the FD2SLS estimator has been used to obtain consistent estimates when one of these conditions fails. Anderson and Hsiao (1981) used a version of the FD2SLS estimator to fit a panel-data model with a lagged dependent variable.

Arellano and Bond (1991) develop new one-step and two-step GMM estimators for dynamic panel data. See [XT] **xtabond** for a discussion of these estimators and Stata's implementation of them. In their article, Arellano and Bond (1991) apply their new estimators to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986). They also compare the results of their estimators with those from the Anderson–Hsiao estimator using data from an unbalanced panel of firms from the United Kingdom. As is conventional, all variables are indexed over the firm i and time t. In this dataset, $\texttt{n}_{it}$ is the log of employment in firm i inside the United Kingdom at time t, $\texttt{w}_{it}$ is the natural log of the real product wage, $\texttt{k}_{it}$ is the natural log of the gross capital stock, and $\texttt{ys}_{it}$ is the natural log of industry output. The model also includes time dummies `yr1980`, `yr1981`, `yr1982`, `yr1983`, and `yr1984`. In Arellano and Bond (1991, table 5, column e), the authors present the results from applying one version of the Anderson–Hsiao estimator to these data. This example reproduces their results for the coefficients, though standard errors are slightly different because Arellano and Bond are using robust standard errors from GMM while we obtain our robust standard errors from 2SLS.

```
. use https://www.stata-press.com/data/r18/abdata
. xtivreg n l2.n l(0/1).w l(0/2).(k ys) yr1981-yr1984 (l.n = l3.n), fd vce(robust)
First-differenced IV regression
Group variable: id                              Number of obs     =        471
Time variable: year                             Number of groups  =        140
R-squared:                                      Obs per group:
     Within  = 0.0141                                         min =          3
     Between = 0.9165                                         avg =        3.4
     Overall = 0.9892                                         max =          5
                                                Wald chi2(14)     =     259.49
corr(u_i, Xb) = 0.9239                          Prob > chi2       =     0.0000
                                     (Std. err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         D.n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         LD. |   1.422765   1.019992     1.39   0.163    -.5763824    3.421913
        L2D. |  -.1645517   .1300598    -1.27   0.206    -.4194643    .0903609
             |
           w |
         D1. |  -.7524675   .2341305    -3.21   0.001    -1.211355     -.29358
         LD. |   .9627611   .7828358     1.23   0.219    -.5715688    2.497091
             |
           k |
         D1. |   .3221686   .1066645     3.02   0.003     .1131099    .5312273
         LD. |  -.3248778   .3933448    -0.83   0.409    -1.095819    .4460637
        L2D. |  -.0953947   .1257672    -0.76   0.448    -.3418938    .1511045
             |
          ys |
         D1. |   .7660906   .3172664     2.41   0.016       .14426    1.387921
         LD. |  -1.361881   .8980497    -1.52   0.129    -3.122026    .3982639
        L2D. |   .3212993   .4234835     0.76   0.448     -.508713    1.151312
             |
      yr1981 |
         D1. |  -.0574197   .0323419    -1.78   0.076    -.1208088    .0059693
             |
      yr1982 |
         D1. |  -.0882952   .0580339    -1.52   0.128    -.2020395    .0254491
             |
      yr1983 |
         D1. |  -.1063153   .0934136    -1.14   0.255    -.2894026    .0767719
             |
      yr1984 |
         D1. |  -.1172108   .1150944    -1.02   0.308    -.3427917    .1083701
             |
       _cons |   .0161204    .025376     0.64   0.525    -.0336155    .0658564
-------------+----------------------------------------------------------------
     sigma_u |  .29069213
     sigma_e |  .34152632
         rho |  .42011045   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Endogenous: L.n
Exogenous:  L2.n w L.w k L.k L2.k ys L.ys L2.ys yr1981 yr1982 yr1983 yr1984
            L3.n
```

◁

Stored results

xtivreg, re stores the following in e():

Scalars

e(N)	number of observations
e(N_g)	number of groups
e(df_m)	model degrees of freedom
e(df_rz)	residual degrees of freedom
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(Tcon)	1 if panels balanced, 0 otherwise
e(N_clust)	number of clusters
e(sigma)	ancillary parameter (gamma, lnormal)
e(sigma_u)	panel-level standard deviation
e(sigma_e)	standard deviation of ϵ_{it}
e(r2_w)	R^2 for within model
e(r2_o)	R^2 for overall model
e(r2_b)	R^2 for between model
e(chi2)	χ^2
e(rho)	ρ
e(F)	model F (small only)
e(m_p)	p-value from model test
e(thta_min)	minimum θ
e(thta_5)	θ, 5th percentile
e(thta_50)	θ, 50th percentile
e(thta_95)	θ, 95th percentile
e(thta_max)	maximum θ
e(rank)	rank of e(V)

Macros

e(cmd)	xtivreg
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(endog)	names of endogenous variables
e(exog)	names of exogenous variables
e(model)	g2sls or ec2sls
e(small)	small, if specified
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(properties)	b V
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed 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 of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

xtivreg, be stores the following in e():

Scalars

e(N)	number of observations
e(N_g)	number of groups
e(mss)	model sum of squares
e(df_m)	model degrees of freedom
e(rss)	residual sum of squares
e(df_r)	residual degrees of freedom
e(df_rz)	residual degrees of freedom for the between-transformed regression
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(rs_a)	adjusted R^2
e(r2_w)	R^2 for within model
e(r2_o)	R^2 for overall model
e(r2_b)	R^2 for between model
e(N_clust)	number of clusters
e(chi2)	model Wald
e(chi2_p)	p-value for model χ^2 test
e(F)	F statistic (small only)
e(rmse)	root mean squared error
e(rank)	rank of e(V)

Macros

e(cmd)	xtivreg
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(endog)	names of endogenous variables
e(exog)	names of exogenous variables
e(model)	be
e(small)	small, if specified
e(clustvar)	name of cluster variable
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed 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 of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

xtivreg, fe stores the following in e():

Scalars

e(N)	number of observations
e(N_g)	number of groups
e(df_m)	model degrees of freedom
e(rss)	residual sum of squares
e(df_r)	residual degrees of freedom (small only)
e(df_rz)	residual degrees of freedom for the within-transformed regression
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(N_clust)	number of clusters
e(sigma)	ancillary parameter (gamma, lnormal)
e(corr)	corr(u_i, Xb)
e(sigma_u)	panel-level standard deviation
e(sigma_e)	standard deviation of ϵ_{it}
e(r2_w)	R^2 for within model
e(r2_o)	R^2 for overall model
e(r2_b)	R^2 for between model
e(chi2)	model Wald (not small)
e(chi2_p)	p-value for model χ^2 test
e(rho)	ρ
e(F)	F statistic (small only)
e(F_f)	F for H_0: u_i=0
e(F_fp)	p-value for F for H_0:u_i=0
e(df_a)	degrees of freedom for absorbed effect
e(rank)	rank of e(V)

Macros

e(cmd)	xtivreg
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(endog)	names of endogenous variables
e(exog)	names of exogenous variables
e(model)	fe
e(small)	small, if specified
e(clustvar)	name of cluster variable
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed 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 of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices

r(table)	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

xtivreg, fd stores the following in e():

Scalars

e(N)	number of observations
e(N_g)	number of groups
e(df_m)	model degrees of freedom
e(rss)	residual sum of squares
e(df_r)	residual degrees of freedom (small only)
e(df_rz)	residual degrees of freedom for first-differenced regression
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(N_clust)	number of clusters
e(sigma)	ancillary parameter (gamma, lnormal)
e(corr)	corr(u_i, Xb)
e(sigma_u)	panel-level standard deviation
e(sigma_e)	standard deviation of ϵ_{it}
e(r2_w)	R^2 for within model
e(r2_o)	R^2 for overall model
e(r2_b)	R^2 for between model
e(chi2)	model Wald (not small)
e(chi2_p)	p-value for model χ^2 test
e(rho)	ρ
e(F)	F statistic (small only)
e(rank)	rank of e(V)

Macros

e(cmd)	xtivreg
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(endog)	names of endogenous variables
e(exog)	names of exogenous variables
e(model)	fd
e(small)	small, if specified
e(clustvar)	name of cluster variable
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample

In addition to the above, the following is stored in `r()`:

Matrices

`r(table)`	matrix containing the coefficients with their standard errors, test statistics, p-values, and confidence intervals

Note that results stored in `r()` are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

Consider an equation of the form

$$y_{it} = \mathbf{Y}_{it}\gamma + \mathbf{X}_{1it}\boldsymbol{\beta} + \mu_i + \nu_{it} = \mathbf{Z}_{it}\boldsymbol{\delta} + \mu_i + \nu_{it} \qquad (2)$$

where

y_{it} is the dependent variable;
$\mathbf{Y}_{it}$ is an $1 \times g_2$ vector of observations on g_2 endogenous variables included as covariates, and these variables are allowed to be correlated with the ν_{it};
$\mathbf{X}_{1it}$ is an $1 \times k_1$ vector of observations on the exogenous variables included as covariates;
$\mathbf{Z}_{it} = [\mathbf{Y}_{it}\ \mathbf{X}_{it}]$;
γ is a $g_2 \times 1$ vector of coefficients;
$\boldsymbol{\beta}$ is a $k_1 \times 1$ vector of coefficients; and
$\boldsymbol{\delta}$ is a $K \times 1$ vector of coefficients, where $K = g_2 + k_1$.

Assume that there is a $1 \times k_2$ vector of observations on the k_2 instruments in $\mathbf{X}_{2it}$. The order condition is satisfied if $k_2 \geq g_2$. Let $\mathbf{X}_{it} = [\mathbf{X}_{1it}\ \mathbf{X}_{2it}]$. `xtivreg` handles exogenously unbalanced panel data. Thus define T_i to be the number of observations on panel i, n to be the number of panels, and N to be the total number of observations; that is, $N = \sum_{i=1}^{n} T_i$.

Methods and formulas are presented under the following headings:

xtivreg, fd
xtivreg, fe
xtivreg, be
xtivreg, re

xtivreg, fd

As the name implies, this estimator obtains its estimates and conventional VCE from an instrumental-variables regression on the first-differenced data. Specifically, first differencing the data yields

$$y_{it} - y_{it-1} = (\mathbf{Z}_{it} - \mathbf{Z}_{i,t-1})\,\delta + \nu_{it} - \nu_{i,t-1}$$

With the μ_i removed by differencing, we can obtain the estimated coefficients and their estimated variance–covariance matrix from a two-stage least-squares regression of Δy_{it} on $\Delta\mathbf{Z}_{it}$ with instruments $\Delta\mathbf{X}_{it}$.

R^2 within is reported as $\left[\mathrm{corr}\left\{(\mathbf{Z}_{it} - \overline{\mathbf{Z}}_i)\widehat{\boldsymbol{\delta}}, y_{it} - \overline{y}_i\right\}\right]^2$.

R^2 between is reported as $\left\{\mathrm{corr}(\overline{\mathbf{Z}}_i\widehat{\boldsymbol{\delta}}, \overline{y}_i)\right\}^2$.

R^2 overall is reported as $\left\{\mathrm{corr}(\mathbf{Z}_{it}\widehat{\boldsymbol{\delta}}, y_{it})\right\}^2$.

xtivreg, fe

At the heart of this model is the within transformation. The within transform of a variable w is

$$\widetilde{w}_{it} = w_{it} - \overline{w}_{i.} + \overline{w}$$

where

$$\overline{w}_{i.} = \frac{1}{T_i}\sum_{t=1}^{T_i} w_{it}$$

$$\overline{w} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T_i} w_{it}$$

and n is the number of groups and N is the total number of observations on the variable.

The within transform of (2) is

$$\widetilde{y}_{it} = \widetilde{\mathbf{Z}}_{it} + \widetilde{\nu}_{it}$$

The within transform has removed the μ_i. With the μ_i gone, the within 2SLS estimator can be obtained from a two-stage least-squares regression of $\widetilde{y}_{it}$ on $\widetilde{\mathbf{Z}}_{it}$ with instruments $\widetilde{\mathbf{X}}_{it}$.

Suppose that there are K variables in $\mathbf{Z}_{it}$, including the mandatory constant. There are $K+n-1$ parameters estimated in the model, and the conventional VCE for the within estimator is

$$\frac{N-K}{N-n-K+1}V_{IV}$$

where V_{IV} is the VCE from the above two-stage least-squares regression. The robust and cluster–robust variance–covariance matrices are the robust and cluster–robust variance–covariance matrices from a two-stage least-squares regression of $\widetilde{y}_{it}$ on $\widetilde{\mathbf{Z}}_{it}$ with instruments $\widetilde{\mathbf{X}}_{it}$.

From the estimate of $\widehat{\boldsymbol{\delta}}$, estimates $\widehat{\mu}_i$ of μ_i are obtained as $\widehat{\mu}_i = \overline{y}_i - \overline{\mathbf{Z}}_i\widehat{\boldsymbol{\delta}}$. Reported from the calculated $\widehat{\mu}_i$ is its standard deviation and its correlation with $\overline{\mathbf{Z}}_i\widehat{\boldsymbol{\delta}}$. Reported as the standard deviation of ν_{it} is the regression's estimated root mean squared error, s^2, which is adjusted (as previously stated) for the $n-1$ estimated means.

R^2 within is reported as the R^2 from the mean-deviated regression.

R^2 between is reported as $\left\{\mathrm{corr}(\overline{\mathbf{Z}}_i\widehat{\boldsymbol{\delta}}, \overline{y}_i)\right\}^2$.

R^2 overall is reported as $\left\{\mathrm{corr}(\mathbf{Z}_{it}\widehat{\boldsymbol{\delta}}, y_{it})\right\}^2$.

At the bottom of the output, an F statistics against the null hypothesis that all the μ_i are zero is reported. This F statistic is an application of the results in Wooldridge (1990).

xtivreg, be

After passing (2) through the between transform, we are left with

$$\overline{y}_i = \alpha + \overline{\mathbf{Z}}_i\boldsymbol{\delta} + \mu_i + \overline{\nu}_i \quad (3)$$

where

$$\overline{w}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} w_{it} \qquad \text{for } w \in \{y, \mathbf{Z}, \nu\}$$

Similarly, define $\overline{\mathbf{X}}_i$ as the matrix of instruments $\mathbf{X}_{it}$ after they have been passed through the between transform.

The BE2SLS estimator of (3) obtains its coefficient estimates and its VCE, a two-stage least-squares regression of $\overline{y}_i$ on $\overline{Z}_i$ with instruments $\overline{\mathbf{X}}_i$ in which each average appears T_i times.

R^2 between is reported as the R^2 from the fitted regression.

R^2 within is reported as $\left[\text{corr}\left\{(\mathbf{Z}_{it} - \overline{\mathbf{Z}}_i)\widehat{\boldsymbol{\delta}}, y_{it} - \overline{y}_i\right\}\right]^2$.

R^2 overall is reported as $\left\{\text{corr}(\mathbf{Z}_{it}\widehat{\boldsymbol{\delta}}, y_{it})\right\}^2$.

xtivreg, re

Per Baltagi and Chang (2000), let

$$u = \mu_i + \nu_{it}$$

be the $N \times 1$ vector of combined errors. Then under the assumptions of the random-effects model,

$$E(uu') = \sigma_\nu^2 \text{diag}\left[I_{T_i} - \frac{1}{T_i}\boldsymbol{\iota}_{T_i}\boldsymbol{\iota}_{T_i}'\right] + \text{diag}\left[w_i \frac{1}{T_i}\boldsymbol{\iota}_{T_i}\boldsymbol{\iota}_{T_i}'\right]$$

where

$$\omega_i = T_i\sigma_\mu^2 + \sigma_\nu^2$$

and $\boldsymbol{\iota}_{T_i}$ is a vector of ones of dimension T_i.

Because the variance components are unknown, consistent estimates are required to implement feasible GLS. `xtivreg` offers two choices. The default is a simple extension of the Swamy–Arora method for unbalanced panels.

Let

$$u_{it}^w = \widetilde{y}_{it} - \widetilde{\mathbf{Z}}_{it}\widehat{\boldsymbol{\delta}}_w$$

be the combined residuals from the within estimator. Let $\widetilde{u}_{it}$ be the within-transformed u_{it}. Then

$$\widehat{\sigma}_\nu = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \widetilde{u}_{it}^2}{N - n - K + 1}$$

Let

$$u_{it}^b = y_{it} - \mathbf{Z}_{it}\boldsymbol{\delta}_b$$

be the combined residual from the between estimator. Let $\overline{u}_{i.}^b$ be the between residuals after they have been passed through the between transform. Then

$$\widehat{\sigma}_\mu^2 = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \overline{u}_{it}^2 - (n-K)\widehat{\sigma}_\nu^2}{N - r}$$

where

$$r = \text{trace}\left\{ \left(\overline{\mathbf{Z}}_i' \overline{\mathbf{Z}}_i \right)^{-1} \overline{\mathbf{Z}}_i' \mathbf{Z}_\mu \mathbf{Z}_\mu' \overline{\mathbf{Z}}_i \right\}$$

where

$$\mathbf{Z}_\mu = \text{diag}\left(\boldsymbol{\iota}_{T_i} \boldsymbol{\iota}_{T_i}' \right)$$

If the `nosa` option is specified, the consistent estimators described in Baltagi and Chang (2000) are used. These are given by

$$\widehat{\sigma}_\nu = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \widetilde{u}_{it}^2}{N-n}$$

and

$$\widehat{\sigma}_\mu^2 = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \overline{u}_{it}^2 - n\widehat{\sigma}_\nu^2}{N}$$

The default Swamy–Arora method contains a degree-of-freedom correction to improve its performance in small samples.

Given estimates of the variance components, $\widehat{\sigma}_\nu^2$ and $\widehat{\sigma}_\mu^2$, the feasible GLS transform of a variable w is

$$w^* = w_{it} - \widehat{\theta}_{it} \overline{w}_{i.} \tag{4}$$

where

$$\overline{w}_{i.} = \frac{1}{T_i} \sum_{t=1}^{T_i} w_{it}$$

$$\widehat{\theta}_{it} = 1 - \left(\frac{\widehat{\sigma}_\nu^2}{\widehat{\omega}_i} \right)^{-\frac{1}{2}}$$

and

$$\widehat{\omega}_i = T_i \widehat{\sigma}_\mu^2 + \widehat{\sigma}_\nu^2$$

Using either estimator of the variance components, `xtivreg` contains two GLS estimators of the random-effects model. These two estimators differ only in how they construct the GLS instruments from the exogenous covariates and instrumental variables contained in $\mathbf{X}_{it} = [\mathbf{X}_{1it}\mathbf{X}_{2it}]$. The default method, G2SLS, which is from Balestra and Varadharajan-Krishnakumar, uses the exogenous variables after they have been passed through the feasible GLS transform. Mathematically, G2SLS uses $\mathbf{X}^*$ for the GLS instruments, where $\mathbf{X}^*$ is constructed by passing each variable in $\mathbf{X}$ though the GLS transform in (4). The G2SLS estimator obtains its coefficient estimates and VCE from an instrumental variable regression of y_{it}^* on $\mathbf{Z}_{it}^*$ with instruments $\mathbf{X}_{it}^*$.

If the `ec2sls` option is specified, `xtivreg` performs Baltagi's EC2SLS. In EC2SLS, the instruments are $\widetilde{\mathbf{X}}_i t$ and $\overline{\mathbf{X}}_{it}$, where $\widetilde{X}_{it}$ is constructed by each of the variables in $\mathbf{X}_{it}$ throughout the GLS transform in (4), and $\overline{\mathbf{X}}_{it}$ is made of the group means of each variable in $\mathbf{X}_{it}$. The EC2SLS estimator obtains its coefficient estimates and its VCE from an instrumental variables regression of y_{it}^* on $\mathbf{Z}_{it}^*$ with instruments $\widetilde{\mathbf{X}}_{it}$ and $\overline{\mathbf{X}}_{it}$.

Baltagi and Li (1992) show that although the G2SLS instruments are a subset of those in EC2SLS, the extra instruments in EC2SLS are redundant in the sense of White (2001). Given the extra computational cost, G2SLS is the default.

The standard deviation of $\mu_i + \nu_{it}$ is calculated as $\sqrt{\widehat{\sigma}^2_\mu + \widehat{\sigma}^2_\nu}$.

R^2 between is reported as $\left\{\text{corr}(\overline{\mathbf{Z}}_i\widehat{\boldsymbol{\delta}}, \overline{y}_i)\right\}^2$.

R^2 within is reported as $\left[\text{corr}\left\{(\mathbf{Z}_{it} - \overline{\mathbf{Z}}_i)\widehat{\boldsymbol{\delta}}, y_{it} - \overline{y}_i\right\}\right]^2$.

R^2 overall is reported as $\left\{\text{corr}(\mathbf{Z}_{it}\widehat{\boldsymbol{\delta}}, y_{it})\right\}^2$.

Acknowledgment

We thank Mead Over of the Center for Global Development, who wrote an early implementation of `xtivreg`.

References

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12: 1–13. https://doi.org/10.2307/2525492.

Anderson, T. W., and C. Hsiao. 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76: 598–606. https://doi.org/10.2307/2287517.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297. https://doi.org/10.2307/2297968.

Balestra, P., and J. Varadharajan-Krishnakumar. 1987. Full information estimations of a system of simultaneous equations with error component structure. *Econometric Theory* 3: 223–246. https://doi.org/10.1017/S0266466600010318.

Baltagi, B. H. 2009. *A Companion to Econometric Analysis of Panel Data*. Chichester, UK: Wiley.

——. 2013. *Econometric Analysis of Panel Data*. 5th ed. Chichester, UK: Wiley.

Baltagi, B. H., and Y.-J. Chang. 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics* 62: 67–89. https://doi.org/10.1016/0304-4076(94)90017-5.

——. 2000. Simultaneous equations with incomplete panels. *Econometric Theory* 16: 269–279. https://doi.org/10.1017/S0266466600162073.

Baltagi, B. H., and Q. Li. 1992. A note on the estimation of simultaneous equations with error components. *Econometric Theory* 8: 113–119. https://doi.org/10.1017/S0266466600010768.

Du, K., Y. Zhang, and Q. Zhou. 2020. Fitting partially linear functional-coefficient panel-data models with Stata. *Stata Journal* 20: 976–998.

Hsiao, C. 2014. *Analysis of Panel Data*. 3rd ed. New York: Cambridge University Press.

Layard, R., and S. J. Nickell. 1986. Unemployment in Britain. *Economica* 53: S121–S169. https://doi.org/10.2307/2554377.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46: 69–85. https://doi.org/10.2307/1913646.

Swamy, P. A. V. B., and S. S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40: 261–275. https://doi.org/10.2307/1909405.

White, H. L., Jr. 2001. *Asymptotic Theory for Econometricians*. Rev. ed. New York: Academic Press.

Wooldridge, J. M. 1990. A note on the Lagrange multiplier and F-statistics for two stage least squares regressions. *Economics Letters* 34: 151–155. https://doi.org/10.1016/0165-1765(90)90236-T.

Also see

[XT] **xtivreg postestimation** — Postestimation tools for xtivreg

[XT] **xtabond** — Arellano–Bond linear dynamic panel-data estimation

[XT] **xteregress** — Extended random-effects linear regression

[XT] **xthtaylor** — Hausman–Taylor estimator for error-components models

[XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models+

[XT] **xtset** — Declare data to be panel data

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

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

For suggested citations, see the FAQ on citing Stata documentation.