xtabond — Arellano–Bond linear dynamic panel-data estimation

Description

`xtabond` fits a linear dynamic panel-data model where the unobserved panel-level effects are correlated with the lags of the dependent variable, known as the Arellano–Bond estimator. This estimator is designed for datasets with many panels and few periods, and it requires that there be no autocorrelation in the idiosyncratic errors.

Quick start

Arellano–Bond estimation of `y` on `x1` and `x2` using `xtset` data

```
xtabond y x1 x2
```

One-step estimator with robust standard errors

```
xtabond y x1 x2, vce(robust)
```

Two-step estimator with bias-corrected robust standard errors

```
xtabond y x1 x2, vce(robust) twostep
```

Arellano–Bond estimation also including 2 lagged values of `y`

```
xtabond y x1 x2, lags(2)
```

Menu

Statistics > Longitudinal/panel data > Dynamic panel data (DPD) > Arellano–Bond estimation

Syntax

xtabond *depvar* [*indepvars*] [*if*] [*in*] [, *options*]

options	Description
Model	
noconstant	suppress constant term
inst(*varlist*)	additional instrument variables
lags(#)	use # lags of dependent variable as covariates; default is lags(1)
maxldep(#)	maximum lags of dependent variable for use as instruments
maxlags(#)	maximum lags of predetermined and endogenous variables for use as instruments
twostep	compute the two-step estimator instead of the one-step estimator
Predetermined	
pre(*varlist*[...])	predetermined variables; can be specified more than once
Endogenous	
endogenous(*varlist*[...])	endogenous variables; can be specified more than once
SE/Robust	
vce(*vcetype*)	*vcetype* may be gmm or robust
Reporting	
level(#)	set confidence level; default is level(95)
artests(#)	use # as maximum order for AR tests; default is artests(2)
display_options	control spacing and line width
coeflegend	display legend instead of statistics

A panel variable and a time variable must be specified; use xtset; see [XT] **xtset**.

indepvars and all *varlists*, except pre(*varlist*[...]) and endogenous(*varlist*[...]), may contain time-series operators; see [U] **11.4.4 Time-series varlists**. The specification of *depvar* may not contain time-series operators.

by, collect, statsby, and xi 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

Model

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

inst(*varlist*) specifies a set of variables to be used as additional instruments. These instruments are not differenced by xtabond before including them in the instrument matrix.

lags(#) sets p, the number of lags of the dependent variable to be included in the model. The default is $p = 1$.

maxldep(#) sets the maximum number of lags of the dependent variable that can be used as instruments. The default is to use all $T_i - p - 2$ lags.

maxlags(#) sets the maximum number of lags of the predetermined and endogenous variables that can be used as instruments. For predetermined variables, the default is to use all $T_i - p - 1$ lags. For endogenous variables, the default is to use all $T_i - p - 2$ lags.

twostep specifies that the two-step estimator be calculated.

Predetermined

pre(*varlist*[, lagstruct(*prelags*, *premaxlags*)]) specifies that a set of predetermined variables be included in the model. Optionally, you may specify that *prelags* lags of the specified variables also be included. The default for *prelags* is 0. Specifying *premaxlags* sets the maximum number of further lags of the predetermined variables that can be used as instruments. The default is to include $T_i - p - 1$ lagged levels as instruments for predetermined variables. You may specify as many sets of predetermined variables as you need within the standard Stata limits on matrix size. Each set of predetermined variables may have its own number of *prelags* and *premaxlags*.

Endogenous

endogenous(*varlist*[, lagstruct(*endlags*, *endmaxlags*)]) specifies that a set of endogenous variables be included in the model. Optionally, you may specify that *endlags* lags of the specified variables also be included. The default for *endlags* is 0. Specifying *endmaxlags* sets the maximum number of further lags of the endogenous variables that can be used as instruments. The default is to include $T_i - p - 2$ lagged levels as instruments for endogenous variables. You may specify as many sets of endogenous variables as you need within the standard Stata limits on matrix size. Each set of endogenous variables may have its own number of *endlags* and *endmaxlags*.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that are robust to some kinds of misspecification; see *Remarks and examples* below.

vce(gmm), the default, uses the conventionally derived variance estimator for generalized method of moments estimation.

vce(robust) uses the robust estimator. After one-step estimation, this is the Arellano–Bond robust VCE estimator. After two-step estimation, this is the Windmeijer (2005) WC-robust estimator.

Reporting

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

artests(#) specifies the maximum order of the autocorrelation test to be calculated. The tests are reported by estat abond; see [XT] **xtabond postestimation**. Specifying the order of the highest test at estimation time is more efficient than specifying it to estat abond, because estat abond must refit the model to obtain the test statistics. The maximum order must be less than or equal to the number of periods in the longest panel. The default is artests(2).

display_options: vsquish and nolstretch; see [R] **Estimation options**.

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

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

Remarks and examples

Linear dynamic panel-data models include p lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random. By construction, the unobserved panel-level effects are correlated with the lagged dependent variables, making standard estimators inconsistent. Arellano and Bond (1991) derived a consistent generalized method of moments (GMM) estimator for the parameters of this model; `xtabond` implements this estimator.

Anderson and Hsiao (1981, 1982) propose using further lags of the level or the difference of the dependent variable to instrument the lagged dependent variables that are included in a dynamic panel-data model after the panel-level effects have been removed by first-differencing. A version of this estimator can be obtained from `xtivreg` (see [XT] **xtivreg**). Arellano and Bond (1991) build upon this idea by noting that, in general, there are many more instruments available. Building on Holtz-Eakin, Newey, and Rosen (1988) and using the GMM framework developed by Hansen (1982), they identify how many lags of the dependent variable, the predetermined variables, and the endogenous variables are valid instruments and how to combine these lagged levels with first differences of the strictly exogenous variables into a potentially large instrument matrix. Using this instrument matrix, Arellano and Bond (1991) derive the corresponding one-step and two-step GMM estimators, as well as the robust VCE estimator for the one-step model. They also found that the robust two-step VCE was seriously biased. Windmeijer (2005) worked out a bias-corrected (WC) robust estimator for VCEs of two-step GMM estimators, which is implemented in `xtabond`. The test of autocorrelation of order m and the Sargan test of overidentifying restrictions derived by Arellano and Bond (1991) can be obtained with `estat abond` and `estat sargan`, respectively; see [XT] **xtabond postestimation**.

The Arellano–Bond estimator is designed for datasets with many panels and few periods, and it requires that there be no autocorrelation in the idiosyncratic errors. For a related estimator that uses additional moment conditions, but still requires no autocorrelation in the idiosyncratic errors, see [XT] **xtdpdsys**. For estimators that allow for some autocorrelation in the idiosyncratic errors, at the cost of a more complicated syntax, see [XT] **xtdpd**.

▷ Example 1: One-step estimator

Arellano and Bond (1991) apply their new estimators and test statistics to a model of dynamic labor demand that had previously been considered by Layard and Nickell (1986) using data from an unbalanced panel of firms from the United Kingdom. All variables are indexed over the firm i and time t. In this dataset, $\mathtt{n}_{it}$ is the log of employment in firm i at time t, $\mathtt{w}_{it}$ is the natural log of the real product wage, $\mathtt{k}_{it}$ is the natural log of the gross capital stock, and $\mathtt{ys}_{it}$ is the natural log of industry output. The model also includes time dummies `yr1980`, `yr1981`, `yr1982`, `yr1983`, and `yr1984`. In table 4 of Arellano and Bond (1991), the authors present the results they obtained from several specifications.

In column a1 of table 4, Arellano and Bond report the coefficients and their standard errors from the robust one-step estimators of a dynamic model of labor demand in which $\mathtt{n}_{it}$ is the dependent variable and its first two lags are included as regressors. To clarify some important issues, we will begin with the homoskedastic one-step version of this model and then consider the robust case. Here is the command using `xtabond` and the subsequent output for the homoskedastic case:

```
. use https://www.stata-press.com/data/r19/abdata

. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) noconstant

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6

Number of instruments =     41                  Wald chi2(16)     =    1757.07
                                                Prob > chi2       =     0.0000
One-step results

            n   Coefficient  Std. err.      z    P>|z|     [95% conf. interval]

            n
          L1.     .6862261   .1486163     4.62   0.000     .3949435    .9775088
          L2.    -.0853582   .0444365    -1.92   0.055    -.1724523    .0017358

            w
          --.    -.6078208   .0657694    -9.24   0.000    -.7367265   -.4789151
          L1.     .3926237   .1092374     3.59   0.000     .1785222    .6067251

            k
          --.     .3568456   .0370314     9.64   0.000     .2842653    .4294259
          L1.    -.0580012   .0583051    -0.99   0.320     -.172277    .0562747
          L2.    -.0199475   .0416274    -0.48   0.632    -.1015357    .0616408

           ys
          --.     .6085073   .1345412     4.52   0.000     .3448115    .8722031
          L1.    -.7111651   .1844599    -3.86   0.000      -1.0727   -.3496304
          L2.     .1057969   .1428568     0.74   0.459    -.1741974    .3857912

       yr1980     .0029062   .0212705     0.14   0.891    -.0387832    .0445957
       yr1981    -.0404378   .0354707    -1.14   0.254    -.1099591    .0290836
       yr1982    -.0652767    .048209    -1.35   0.176    -.1597646    .0292111
       yr1983    -.0690928   .0627354    -1.10   0.271    -.1920521    .0538664
       yr1984    -.0650302   .0781322    -0.83   0.405    -.2181665    .0881061
         year     .0095545   .0142073     0.67   0.501    -.0182912    .0374002

Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980
                  D.yr1981 D.yr1982 D.yr1983 D.yr1984 D.year
```

The coefficients are identical to those reported in column a1 of table 4, as they should be. Of course, the standard errors are different because we are considering the homoskedastic case. Although the moment conditions use first-differenced errors, `xtabond` estimates the coefficients of the level model and reports them accordingly.

The footer in the output reports the instruments used. The first line indicates that `xtabond` used lags from 2 on back to create the GMM-type instruments described in Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988); also see *Methods and formulas* in [XT] **xtdpd**. The second and third lines indicate that the first difference of all the exogenous variables were used as standard instruments. GMM-type instruments use the lags of a variable to contribute multiple columns to the instrument matrix, whereas each standard instrument contributes one column to the instrument matrix. The notation `L(2/.).n` indicates that GMM-type instruments were created using lag 2 of `n` from on back. (`L(2/4).n` would indicate that GMM-type instruments were created using only lags 2, 3, and 4 of `n`.)

After `xtabond`, `estat sargan` reports the Sargan test of overidentifying restrictions.

```
. estat sargan
Sargan test of overidentifying restrictions
H0: Overidentifying restrictions are valid

        chi2(25)     =  65.81806
        Prob > chi2  =    0.0000
```

Only for a homoskedastic error term does the Sargan test have an asymptotic χ^2 distribution. In fact, Arellano and Bond (1991) show that the one-step Sargan test overrejects in the presence of heteroskedasticity. Because its asymptotic distribution is not known under the assumptions of the `vce(robust)` model, `xtabond` does not compute it when `vce(robust)` is specified. The Sargan test, reported by Arellano and Bond (1991, table 4, column a1), comes from the one-step homoskedastic estimator and is the same as the one reported here. The output above presents strong evidence against the null hypothesis that the overidentifying restrictions are valid. Rejecting this null hypothesis implies that we need to reconsider our model or our instruments, unless we attribute the rejection to heteroskedasticity in the data-generating process. Although performing the Sargan test after the two-step estimator is an alternative, Arellano and Bond (1991) found a tendency for this test to underreject in the presence of heteroskedasticity. (See [XT] **xtdpd** for an example indicating that this rejection may be due to misspecification.)

By default, `xtabond` calculates the Arellano–Bond test for first- and second-order autocorrelation in the first-differenced errors. (Use `artests()` to compute tests for higher orders.) There are versions of this test for both the homoskedastic and the robust cases, although their values are different. Use `estat abond` to report the test results.

```
. estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors
H0: No autocorrelation

Order          z    Prob > z
----------------------------
    1   -3.9394     0.0001
    2   -.54239     0.5876
----------------------------
```

When the idiosyncratic errors are independent and identically distributed (i.i.d.), the first-differenced errors are first-order serially correlated. So, as expected, the output above presents strong evidence against the null hypothesis of zero autocorrelation in the first-differenced errors at order 1. Serial correlation in the first-differenced errors at an order higher than 1 implies that the moment conditions used by `xtabond` are not valid; see [XT] **xtdpd** for an example of an alternative estimation method. The output above presents no significant evidence of serial correlation in the first-differenced errors at order 2.

◁

▷ Example 2: A one-step estimator with robust VCE

Consider the output from the one-step robust estimator of the same model:

```
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) vce(robust)
> noconstant

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6

Number of instruments =     41                  Wald chi2(16)     =    1727.45
                                                Prob > chi2       =     0.0000
One-step results
                                     (Std. err. adjusted for clustering on id)

                             Robust
           n   Coefficient  std. err.      z    P>|z|     [95% conf. interval]

           n
         L1.     .6862261   .1445943     4.75   0.000     .4028266    .9696257
         L2.    -.0853582   .0560155    -1.52   0.128    -.1951467    .0244302

           w
         --.    -.6078208   .1782055    -3.41   0.001    -.9570972   -.2585445
         L1.     .3926237   .1679931     2.34   0.019     .0633632    .7218842

           k
         --.     .3568456   .0590203     6.05   0.000      .241168    .4725233
         L1.    -.0580012   .0731797    -0.79   0.428    -.2014308    .0854284
         L2.    -.0199475   .0327126    -0.61   0.542    -.0840631    .0441681

          ys
         --.     .6085073   .1725313     3.53   0.000     .2703522    .9466624
         L1.    -.7111651   .2317163    -3.07   0.002    -1.165321   -.2570095
         L2.     .1057969   .1412021     0.75   0.454    -.1709542     .382548

      yr1980     .0029062   .0158028     0.18   0.854    -.0280667    .0338791
      yr1981    -.0404378   .0280582    -1.44   0.150    -.0954307    .0145552
      yr1982    -.0652767   .0365451    -1.79   0.074    -.1369038    .0063503
      yr1983    -.0690928    .047413    -1.46   0.145    -.1620205    .0238348
      yr1984    -.0650302   .0576305    -1.13   0.259    -.1779839    .0479235
        year     .0095545   .0102896     0.93   0.353    -.0106127    .0297217

Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980
                  D.yr1981 D.yr1982 D.yr1983 D.yr1984 D.year
```

The coefficients are the same, but now the standard errors match that reported in Arellano and Bond (1991, table 4, column a1). Most of the robust standard errors are higher than those that assume a homoskedastic error term.

The Sargan statistic cannot be calculated after requesting a robust VCE, but robust tests for serial correlation are available.

```
. estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors
H0: No autocorrelation

Order          z   Prob > z
───────────────────────────
    1    -3.5996     0.0003
    2    -.51603     0.6058
───────────────────────────
```

The value of the test for second-order autocorrelation matches those reported in Arellano and Bond (1991, table 4, column a1) and presents no evidence of model misspecification.

◁

▷ Example 3: The Wald model test

xtabond reports the Wald statistic of the null hypothesis that all the coefficients except the constant are zero. Here the null hypothesis is that all the coefficients are zero, because there is no constant in the model. In our previous example, the null hypothesis is soundly rejected. In column a1 of table 4, Arellano and Bond report a χ^2 test of the null hypothesis that all the coefficients are zero, except the time trend and the time dummies. Here is this test in Stata:

```
. test l.n l2.n w l.w k l.k l2.k ys l.ys l2.ys

 ( 1)  L.n = 0
 ( 2)  L2.n = 0
 ( 3)  w = 0
 ( 4)  L.w = 0
 ( 5)  k = 0
 ( 6)  L.k = 0
 ( 7)  L2.k = 0
 ( 8)  ys = 0
 ( 9)  L.ys = 0
 (10)  L2.ys = 0

           chi2( 10) =  408.29
         Prob > chi2 =    0.0000
```

◁

▷ Example 4: A two-step estimator with Windmeijer bias-corrected robust VCE

The two-step estimator with the Windmeijer bias-corrected robust VCE of the same model produces the following output:

```
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) twostep
> vce(robust) noconstant

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                             min =          4
                                                             avg =   4.364286
                                                             max =          6

Number of instruments =     41                  Wald chi2(16)     =    1104.72
                                                Prob > chi2       =     0.0000
Two-step results
                                     (Std. err. adjusted for clustering on id)

------------------------------------------------------------------------------
             |              WC-robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6287089   .1934138     3.25   0.001     .2496248    1.007793
         L2. |  -.0651882   .0450501    -1.45   0.148    -.1534847    .0231084
             |
           w |
         --. |  -.5257597   .1546107    -3.40   0.001     -.828791   -.2227284
         L1. |   .3112899   .2030006     1.53   0.125     -.086584    .7091638
             |
           k |
         --. |   .2783619   .0728019     3.82   0.000     .1356728    .4210511
         L1. |   .0140994   .0924575     0.15   0.879     -.167114    .1953129
         L2. |  -.0402484   .0432745    -0.93   0.352    -.1250649    .0445681
             |
          ys |
         --. |   .5919243   .1730916     3.42   0.001      .252671    .9311776
         L1. |  -.5659863   .2611008    -2.17   0.030    -1.077734   -.0542381
         L2. |   .1005433   .1610987     0.62   0.533    -.2152043    .4162908
             |
      yr1980 |   .0006378   .0168042     0.04   0.970    -.0322978    .0335734
      yr1981 |  -.0550044   .0313389    -1.76   0.079    -.1164275    .0064187
      yr1982 |   -.075978   .0419276    -1.81   0.070    -.1581545    .0061986
      yr1983 |  -.0740708   .0528381    -1.40   0.161    -.1776315      .02949
      yr1984 |  -.0906606   .0642615    -1.41   0.158    -.2166108    .0352896
        year |   .0112155   .0116783     0.96   0.337    -.0116735    .0341045
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980
                  D.yr1981 D.yr1982 D.yr1983 D.yr1984 D.year
```

Arellano and Bond recommend against using the two-step nonrobust results for inference on the coefficients because the standard errors tend to be biased downward (see Arellano and Bond 1991 for details). The output above uses the Windmeijer bias-corrected (WC) robust VCE, which Windmeijer (2005) showed to work well. The magnitudes of several of the coefficient estimates have changed, and one even switched its sign.

The test for autocorrelation presents no evidence of model misspecification:

```
. estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors
H0: No autocorrelation

Order          z   Prob > z
---------------------------
    1    -2.1255     0.0335
    2    -.35166     0.7251
---------------------------
```

◁

Manuel Arellano (1957–) was born in Elda in Alicante, Spain. He earned degrees in economics from the University of Barcelona and the London School of Economics. After various posts in Oxford and London, he returned to Spain as professor of econometrics at Madrid in 1991. He is a leading expert on panel-data econometrics.

Stephen Roy Bond (1963–) earned degrees in economics from Cambridge and Oxford. Following various posts at Oxford, he now works mainly at the Institute for Fiscal Studies in London. His research interests include company taxation, dividends, and the links between financial markets, corporate control, and investment.

▷ Example 5: Including an estimator for the constant

Thus far we have been specifying the `noconstant` option to keep to the standard Arellano–Bond estimator, which uses instruments only for the difference equation. The constant estimated by `xtabond` is a constant in the level equation, and it is estimated from the level errors. The output below illustrates that including a constant in the model does not affect the other parameter estimates.

```
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) twostep vce(robust)

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6

Number of instruments =     42                  Wald chi2(16)     =    1104.72
                                                Prob > chi2       =     0.0000

Two-step results
                                       (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6287089   .1934138     3.25   0.001     .2496248    1.007793
         L2. |  -.0651882   .0450501    -1.45   0.148    -.1534847    .0231084
             |
           w |
         --. |  -.5257597   .1546107    -3.40   0.001     -.828791   -.2227284
         L1. |   .3112899   .2030006     1.53   0.125     -.086584    .7091638
             |
           k |
         --. |   .2783619   .0728019     3.82   0.000     .1356728    .4210511
         L1. |   .0140994   .0924575     0.15   0.879     -.167114    .1953129
         L2. |  -.0402484   .0432745    -0.93   0.352    -.1250649    .0445681
             |
          ys |
         --. |   .5919243   .1730916     3.42   0.001      .252671    .9311776
         L1. |  -.5659863   .2611008    -2.17   0.030    -1.077734   -.0542381
         L2. |   .1005433   .1610987     0.62   0.533    -.2152043    .4162908
             |
      yr1980 |   .0006378   .0168042     0.04   0.970    -.0322978    .0335734
      yr1981 |  -.0550044   .0313389    -1.76   0.079    -.1164275    .0064187
      yr1982 |   -.075978   .0419276    -1.81   0.070    -.1581545    .0061986
      yr1983 |  -.0740708   .0528381    -1.40   0.161    -.1776315      .02949
      yr1984 |  -.0906606   .0642615    -1.41   0.158    -.2166108    .0352896
        year |   .0112155   .0116783     0.96   0.337    -.0116735    .0341045
       _cons |  -21.53725   23.23138    -0.93   0.354    -67.06992    23.99542
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1980
                  D.yr1981 D.yr1982 D.yr1983 D.yr1984 D.year
Instruments for level equation
        Standard: _cons
```

Including the constant does not affect the other parameter estimates because it is identified only by the level errors; see [XT] **xtdpd** for details.

◁

▷ Example 6: Including predetermined covariates

Sometimes we cannot assume strict exogeneity. Recall that a variable, x_{it}, is said to be strictly exogenous if $E[x_{it}\epsilon_{is}] = 0$ for all t and s. If $E[x_{it}\epsilon_{is}] \neq 0$ for $s < t$ but $E[x_{it}\epsilon_{is}] = 0$ for all $s \geq t$, the variable is said to be predetermined. Intuitively, if the error term at time t has some feedback on the subsequent realizations of x_{it}, x_{it} is a predetermined variable. Because unforecastable errors today might affect future changes in the real wage and in the capital stock, we might suspect that the log of the real product wage and the log of the gross capital stock are predetermined instead of strictly exogenous. Here we treat `w` and `k` as predetermined and use lagged levels as instruments.

```
. xtabond n l(0/1).ys yr1980-yr1984 year, lags(2) twostep pre(w, lag(1,.))
> pre(k, lag(2,.)) noconstant vce(robust)

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6

Number of instruments =     83                  Wald chi2(15)     =     958.30
                                                Prob > chi2       =     0.0000

Two-step results
                                     (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .8580958   .1265515     6.78   0.000     .6100594    1.106132
         L2. |   -.081207   .0760703    -1.07   0.286    -.2303022    .0678881
             |
           w |
         --. |  -.6910855   .1387684    -4.98   0.000    -.9630666   -.4191044
         L1. |   .5961712   .1497338     3.98   0.000     .3026982    .8896441
             |
           k |
         --. |   .4140654   .1382788     2.99   0.003     .1430439    .6850868
         L1. |  -.1537048   .1220244    -1.26   0.208    -.3928681    .0854586
         L2. |  -.1025833   .0710886    -1.44   0.149    -.2419143    .0367477
             |
          ys |
         --. |   .6936392   .1728623     4.01   0.000     .3548354    1.032443
         L1. |  -.8773678   .2183085    -4.02   0.000    -1.305245    -.449491
             |
      yr1980 |  -.0072451    .017163    -0.42   0.673    -.0408839    .0263938
      yr1981 |  -.0609608    .030207    -2.02   0.044    -.1201655   -.0017561
      yr1982 |  -.1130369   .0454826    -2.49   0.013    -.2021812   -.0238926
      yr1983 |  -.1335249   .0600213    -2.22   0.026    -.2511645   -.0158853
      yr1984 |  -.1623177   .0725434    -2.24   0.025    -.3045001   -.0201352
        year |   .0264501   .0119329     2.22   0.027      .003062    .0498381
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n L(1/.).L.w L(1/.).L2.k
        Standard: D.ys LD.ys D.yr1980 D.yr1981 D.yr1982 D.yr1983 D.yr1984
                  D.year
```

The footer informs us that we are now including GMM-type instruments from the first lag of `L.w` on back and from the first lag of `L2.k` on back.

◁

❑ Technical note

The above example illustrates that `xtabond` understands `pre(w, lag(1, .))` to mean that `L.w` is a predetermined variable and `pre(k, lag(2, .))` to mean that `L2.k` is a predetermined variable. This is a stricter definition than the alternative that `pre(w, lag(1, .))` means only that `w` is predetermined but includes a lag of `w` in the model and that `pre(k, lag(2, .))` means only that `k` is predetermined but includes first and second lags of `k` in the model. If you prefer the weaker definition, `xtabond` still gives you consistent estimates, but it is not using all possible instruments; see [XT] **xtdpd** for an example of how to include all possible instruments.

❑

▷ Example 7: Including endogenous covariates

We might instead suspect that `w` and `k` are endogenous in that $E[x_{it}\epsilon_{is}] \neq 0$ for $s \leq t$ but $E[x_{it}\epsilon_{is}] = 0$ for all $s > t$. By this definition, endogenous variables differ from predetermined variables only in that the former allow for correlation between the x_{it} and the ϵ_{it} at time t, whereas the latter do not. Endogenous variables are treated similarly to the lagged dependent variable. Levels of the endogenous variables lagged two or more periods can serve as instruments. In this example, we treat `w` and `k` as endogenous variables.

```
. xtabond n l(0/1).ys yr1980-yr1984 year, lags(2) twostep
> endogenous(w, lag(1,.)) endogenous(k, lag(2,.)) noconstant vce(robust)

Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year

                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6

Number of instruments =     71                  Wald chi2(15)     =     967.61
                                                Prob > chi2       =     0.0000

Two-step results
                                     (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6640937   .1278908     5.19   0.000     .4134323     .914755
         L2. |   -.041283    .081801    -0.50   0.614    -.2016101    .1190441
             |
           w |
         --. |  -.7143942     .13083    -5.46   0.000    -.9708162   -.4579721
         L1. |   .3644198    .184758     1.97   0.049     .0023008    .7265388
             |
           k |
         --. |   .5028874   .1205419     4.17   0.000     .2666296    .7391452
         L1. |  -.2160842   .0972855    -2.22   0.026    -.4067603    -.025408
         L2. |  -.0549654   .0793673    -0.69   0.489    -.2105225    .1005917
             |
          ys |
         --. |   .5989356   .1779731     3.37   0.001     .2501148    .9477564
         L1. |  -.6770367   .1961166    -3.45   0.001    -1.061418   -.2926553
             |
      yr1980 |  -.0061122   .0155287    -0.39   0.694    -.0365478    .0243235
      yr1981 |    -.04715   .0298348    -1.58   0.114    -.1056252    .0113251
      yr1982 |  -.0817646   .0486049    -1.68   0.093    -.1770285    .0134993
      yr1983 |  -.0939251   .0675804    -1.39   0.165    -.2263802    .0385299
      yr1984 |   -.117228   .0804716    -1.46   0.145    -.2749493    .0404934
        year |   .0208857   .0103485     2.02   0.044     .0006031    .0411684
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n L(2/.).L.w L(2/.).L2.k
        Standard: D.ys LD.ys D.yr1980 D.yr1981 D.yr1982 D.yr1983 D.yr1984
                  D.year
```

Although some estimated coefficients changed in magnitude, none changed in sign, and these results are similar to those obtained by treating `w` and `k` as predetermined.

◁

The Arellano–Bond estimator is for datasets with many panels and few periods. (Technically, the large-sample properties are derived with the number of panels going to infinity and the number of periods held fixed.) The number of instruments increases quadratically in the number of periods. If your dataset is better described by a framework in which both the number of panels and the number of periods is large, then you should consider other estimators such as those in [XT] **xtivreg** or `xtreg, fe` in [XT] **xtreg**; see Alvarez and Arellano (2003) for a discussion of this case.

▷ Example 8: Restricting the number of instruments

Treating variables as predetermined or endogenous quickly increases the size of the instrument matrix. (See *Methods and formulas* in [XT] **xtdpd** for a discussion of how this matrix is created and what determines its size.) GMM estimators with too many overidentifying restrictions may perform poorly in small samples. (See Kiviet 1995 for a discussion of the dynamic panel-data case.)

To handle these problems, you can set a maximum number of lagged levels to be included as instruments for lagged-dependent or the predetermined variables. Here is an example in which a maximum of three lagged levels of the predetermined variables are included as instruments:

```
. xtabond n l(0/1).ys yr1980-yr1984 year, lags(2) twostep
> pre(w, lag(1,3)) pre(k, lag(2,3)) noconstant vce(robust)
Arellano-Bond dynamic panel-data estimation     Number of obs     =        611
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          4
                                                              avg =   4.364286
                                                              max =          6
Number of instruments =     67                  Wald chi2(15)     =    1116.89
                                                Prob > chi2       =     0.0000
Two-step results
                                     (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .931121   .1456964     6.39   0.000     .6455612    1.216681
         L2. |  -.0759918   .0854356    -0.89   0.374    -.2434425    .0914589
             |
           w |
         --. |  -.6475372   .1687931    -3.84   0.000    -.9783656   -.3167089
         L1. |   .6906238   .1789698     3.86   0.000     .3398493    1.041398
             |
           k |
         --. |   .3788106   .1848137     2.05   0.040     .0165824    .7410389
         L1. |  -.2158533   .1446198    -1.49   0.136    -.4993028    .0675962
         L2. |  -.0914584   .0852267    -1.07   0.283    -.2584997    .0755829
             |
          ys |
         --. |   .7324964    .176748     4.14   0.000     .3860766    1.078916
         L1. |  -.9428141   .2735472    -3.45   0.001    -1.478957   -.4066715
             |
      yr1980 |  -.0102389   .0172473    -0.59   0.553    -.0440431    .0235652
      yr1981 |  -.0763495   .0296992    -2.57   0.010    -.1345589   -.0181402
      yr1982 |  -.1373829   .0441833    -3.11   0.002    -.2239806   -.0507853
      yr1983 |  -.1825149   .0613674    -2.97   0.003    -.3027928   -.0622369
      yr1984 |  -.2314023   .0753669    -3.07   0.002    -.3791186    -.083686
        year |   .0310012   .0119167     2.60   0.009     .0076448    .0543576
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n L(1/3).L.w L(1/3).L2.k
        Standard: D.ys LD.ys D.yr1980 D.yr1981 D.yr1982 D.yr1983 D.yr1984
                  D.year
```

◁

▷ Example 9: Missing observations in the middle of panels

xtabond handles data in which there are missing observations in the middle of the panels. In this example, we deliberately set the dependent variable to missing in the year 1980:

```
. replace n=. if year==1980
(140 real changes made, 140 to missing)

. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year, lags(2) noconstant
>  vce(robust)
note: yr1980 omitted from div() because of collinearity.
note: yr1981 omitted from div() because of collinearity.
note: yr1982 omitted from div() because of collinearity.
note: yr1980 omitted because of collinearity.
note: yr1981 omitted because of collinearity.
note: yr1982 omitted because of collinearity.

Arellano-Bond dynamic panel-data estimation     Number of obs     =        115
Group variable: id                              Number of groups  =        101
Time variable: year

                                                Obs per group:
                                                             min =          1
                                                             avg =   1.138614
                                                             max =          2

Number of instruments =     18                  Wald chi2(12)     =      44.48
                                                Prob > chi2       =     0.0000

One-step results
                                     (Std. err. adjusted for clustering on id)
```

n	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
n						
L1.	.1790577	.2204682	0.81	0.417	-.253052	.6111674
L2.	.0214253	.0488476	0.44	0.661	-.0743143	.1171649
w						
--.	-.2513405	.1402114	-1.79	0.073	-.5261498	.0234689
L1.	.1983952	.1445875	1.37	0.170	-.0849912	.4817815
k						
--.	.3983149	.0883352	4.51	0.000	.2251811	.5714488
L1.	-.025125	.0909236	-0.28	0.782	-.203332	.1530821
L2.	-.0359338	.0623382	-0.58	0.564	-.1581144	.0862468
ys						
--.	.3663201	.3824893	0.96	0.338	-.3833451	1.115985
L1.	-.6319976	.4823958	-1.31	0.190	-1.577476	.3134807
L2.	.5318404	.4105269	1.30	0.195	-.2727775	1.336458
yr1983	-.0047543	.024855	-0.19	0.848	-.0534692	.0439606
yr1984	0	(omitted)				
year	.0014465	.010355	0.14	0.889	-.0188489	.0217419

```
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1983
                  D.yr1984 D.year
```

There are two important aspects to this example. First, xtabond reports that variables have been omitted from the model and from the div() instrument list. For xtabond, the div() instrument list is the list of instruments created from the strictly exogenous variables; see [XT] **xtdpd** for more about the div()

instrument list. Second, because xtabond uses time-series operators in its computations, if statements and missing values are not equivalent. An if statement causes the false observations to be excluded from the sample, but it computes the time-series operators wherever possible. In contrast, missing data prevent evaluation of the time-series operators that involve missing observations. Thus the example above is not equivalent to the following one:

```
. use https://www.stata-press.com/data/r19/abdata, clear
. xtabond n l(0/1).w l(0/2).(k ys) yr1980-yr1984 year if year!=1980,
> lags(2) noconstant vce(robust)
note: yr1980 omitted from div() because of collinearity.
note: yr1980 omitted because of collinearity.
Arellano-Bond dynamic panel-data estimation     Number of obs     =        473
Group variable: id                              Number of groups  =        140
Time variable: year
                                                Obs per group:
                                                              min =          3
                                                              avg =   3.378571
                                                              max =          5
Number of instruments =     37                  Wald chi2(15)     =    1041.61
                                                Prob > chi2       =     0.0000
One-step results
                                     (Std. err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
           n | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .7210062   .1321214     5.46   0.000     .4620531    .9799593
         L2. |  -.0960646   .0570547    -1.68   0.092    -.2078898    .0157606
             |
           w |
         --. |  -.6684175   .1739484    -3.84   0.000     -1.00935   -.3274849
         L1. |    .482322   .1647185     2.93   0.003     .1594797    .8051642
             |
           k |
         --. |   .3802777   .0728546     5.22   0.000     .2374853    .5230701
         L1. |   -.104598    .088597    -1.18   0.238     -.278245     .069049
         L2. |  -.0272055   .0379994    -0.72   0.474     -.101683    .0472721
             |
          ys |
         --. |   .4655989   .1864368     2.50   0.013     .1001895    .8310082
         L1. |  -.8562492   .2187886    -3.91   0.000    -1.285067   -.4274315
         L2. |   .0896556   .1440035     0.62   0.534     -.192586    .3718972
             |
      yr1981 |  -.0711626   .0205299    -3.47   0.001    -.1114005   -.0309247
      yr1982 |  -.1212749   .0334659    -3.62   0.000    -.1868669   -.0556829
      yr1983 |  -.1470248   .0461714    -3.18   0.001    -.2375191   -.0565305
      yr1984 |  -.1519021   .0543904    -2.79   0.005    -.2585054   -.0452988
        year |   .0203277   .0108732     1.87   0.062    -.0009833    .0416387
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1981
                  D.yr1982 D.yr1983 D.yr1984 D.year
```

The year 1980 is omitted from the sample, but when the value of a variable from 1980 is required because a lag or difference is required, the 1980 value is used.

◁

Stored results

xtabond 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(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(t_min)	minimum time in sample
e(t_max)	maximum time in sample
e(chi2)	χ^2
e(arm#)	test for autocorrelation of order #
e(artests)	number of AR tests computed
e(sig2)	estimate of σ_ϵ^2
e(rss)	sum of squared differenced residuals
e(sargan)	Sargan test statistic
e(rank)	rank of e(V)
e(zrank)	rank of instrument matrix
Macros	
e(cmd)	xtabond
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(twostep)	twostep, if specified
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(system)	system, if system estimator
e(transform)	specified transform
e(datasignature)	checksum from datasignature
e(datasignaturevars)	variables used in calculation of checksum
e(properties)	b V
e(estat_cmd)	program used to implement estat
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
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

A dynamic panel-data model has the form

$$y_{it} = \sum_{j=1}^{p} \alpha_j y_{i,t-j} + \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{w}_{it}\boldsymbol{\beta}_2 + \nu_i + \epsilon_{it} \quad i = 1,\ldots,N \quad t = 1,\ldots,T_i \tag{1}$$

where

the α_j are p parameters to be estimated,
$\mathbf{x}_{it}$ is a $1 \times k_1$ vector of strictly exogenous covariates,
$\boldsymbol{\beta}_1$ is a $k_1 \times 1$ vector of parameters to be estimated,
$\mathbf{w}_{it}$ is a $1 \times k_2$ vector of predetermined and endogenous covariates,
$\boldsymbol{\beta}_2$ is a $k_2 \times 1$ vector of parameters to be estimated,
ν_i are the panel-level effects (which may be correlated with the covariates), and
ϵ_{it} are i.i.d. over the whole sample with variance σ_ϵ^2.

The ν_i and the ϵ_{it} are assumed to be independent for each i over all t.

By construction, the lagged dependent variables are correlated with the unobserved panel-level effects, making standard estimators inconsistent. With many panels and few periods, estimators are constructed by first-differencing to remove the panel-level effects and using instruments to form moment conditions.

`xtabond` uses a GMM estimator to estimate $\alpha_1, \ldots, \alpha_p, \boldsymbol{\beta}_1$, and β_2. The moment conditions are formed from the first-differenced errors from (1) and instruments. Lagged levels of the dependent variable, the predetermined variables, and the endogenous variables are used to form GMM-type instruments. See Arellano and Bond (1991) and Holtz-Eakin, Newey, and Rosen (1988) for discussions of GMM-type instruments. First differences of the strictly exogenous variables are used as standard instruments.

`xtabond` uses `xtdpd` to perform its computations, so the formulas are given in *Methods and formulas* of [XT] **xtdpd**.

References

Alvarez, J., and M. Arellano. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71: 1121–1159. https://doi.org/10.1111/1468-0262.00441.

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.

———. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47–82. https://doi.org/10.1016/0304-4076(82)90095-1.

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.

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

Blackburne, E. F., III, and M. W. Frank. 2007. Estimation of nonstationary heterogeneous panels. *Stata Journal* 7: 197–208.

Bruno, G. S. F. 2005. Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *Stata Journal* 5: 473–500.

De Vos, I., G. Everaert, and I. Ruyssen. 2015. Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal* 15: 986–1018.

Grotti, R., and G. Cutuli. 2018. xtpdyn: A community-contributed command for fitting dynamic random-effects probit models with unobserved heterogeneity. *Stata Journal* 18: 844–862.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054. https://doi.org/10.2307/1912775.

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–1395. https://doi.org/10.2307/1913103.

Kiviet, J. F. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68: 53–78. https://doi.org/10.1016/0304-4076(94)01643-E.

Kripfganz, S. 2016. Quasi–maximum likelihood estimation of linear dynamic short-T panel-data models. *Stata Journal* 16: 1013–1038.

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

Neal, T. 2014. Panel cointegration analysis with xtpedroni. *Stata Journal* 14: 684–692.

Pinzon, E. 2015. xtabond cheat sheet. *The Stata Blog: Not Elsewhere Classified.* https://blog.stata.com/2015/11/12/xtabond-cheat-sheet/.

Seo, M. H., S. Kim, and Y.-J. Kim. 2019. Estimation of dynamic panel threshold model using Stata. *Stata Journal* 19: 685–697.

Williams, R., P. D. Allison, and E. Moral-Benito. 2018. Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling. *Stata Journal* 18: 293–326.

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51. https://doi.org/10.1016/j.jeconom.2004.02.005.

Also see

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

[XT] **xtdpd** — Linear dynamic panel-data estimation

[XT] **xtdpdsys** — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

[XT] **xtivreg** — Instrumental variables and two-stage least squares for panel-data models

[XT] **xtreg** — Linear models for panel data

[XT] **xtregar** — Fixed- and random-effects linear models with an AR(1) disturbance

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

[U] 20 Estimation and postestimation commands

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