Title

stata.com

xtdpdsys — Arellano–Bover/Blundell–Bond linear dynamic panel-data estimation

Syntax

`xtdpdsys` *depvar* [*indepvars*] [*if*] [*in*] [, *options*]

options	Description
Model	
`noconstant`	suppress constant term
`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 **[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`, `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.

Menu

Statistics > Longitudinal/panel data > Dynamic panel data (DPD) > Arellano-Bover/Blundell-Bond estimation

Description

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 this model. The Arellano and Bond estimator can perform poorly if the autoregressive parameters are too large or the ratio of the variance of the panel-level effect to the variance of idiosyncratic error is too large. Building on the work of Arellano and Bover (1995), Blundell and Bond (1998) developed a system estimator that uses additional moment conditions; `xtdpdsys` implements this estimator.

This estimator is designed for datasets with many panels and few periods. This method assumes that there is no autocorrelation in the idiosyncratic errors and requires the initial condition that the panel-level effects be uncorrelated with the first difference of the first observation of the dependent variable.

Options

Model

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

`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 *Methods and formulas* in [XT] **xtdpd**.

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

`vce(robust)` uses the robust estimator. For the one-step estimator, this is the Arellano–Bond robust VCE estimator. For the two-step estimator, 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] **xtdpdsys 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 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 `xtdpdsys` but is not shown in the dialog box:

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

Remarks and examples

stata.com

If you have not read [XT] **xtabond**, you may want to do so before continuing.

Consider the dynamic panel-data model

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

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 or 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^2_ϵ.

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, the Arellano–Bond estimator is constructed by first-differencing to remove the panel-level effects and using instruments to form moment conditions.

Blundell and Bond (1998) show that the lagged-level instruments in the Arellano–Bond estimator become weak as the autoregressive process becomes too persistent or the ratio of the variance of the panel-level effects ν_i to the variance of the idiosyncratic error ϵ_{it} becomes too large. Building on the work of Arellano and Bover (1995), Blundell and Bond (1998) proposed a system estimator that uses moment conditions in which lagged differences are used as instruments for the level equation in addition to the moment conditions of lagged levels as instruments for the differenced equation. The additional moment conditions are valid only if the initial condition $E[\nu_i \Delta y_{i2}] = 0$ holds for all i; see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).

`xtdpdsys` fits dynamic panel-data estimators with the Arellano–Bover/Blundell–Bond system estimator. Because `xtdpdsys` extends `xtabond`, [XT] **xtabond** provides useful background.

▷ Example 1: A dynamic panel model

In their article, Arellano and Bond (1991) apply their 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, $\texttt{n}_{it}$ is the log of employment in firm i 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`.

For comparison, we begin by using `xtabond` to fit a model to these data.

```
. use http://www.stata-press.com/data/r13/abdata
. xtabond n L(0/2).(w k) yr1980-yr1984 year, 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 =     40               Wald chi2(13)        =   1318.68
                                             Prob > chi2          =    0.0000
One-step results
                                   (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6286618   .1161942     5.41   0.000     .4009254    .8563983
             |
           w |
         --. |  -.5104249   .1904292    -2.68   0.007    -.8836592   -.1371906
         L1. |   .2891446    .140946     2.05   0.040     .0128954    .5653937
         L2. |  -.0443653   .0768135    -0.58   0.564     -.194917    .1061865
             |
           k |
         --. |   .3556923   .0603274     5.90   0.000     .2374528    .4739318
         L1. |  -.0457102   .0699732    -0.65   0.514    -.1828552    .0914348
         L2. |  -.0619721   .0328589    -1.89   0.059    -.1263743    .0024301
             |
      yr1980 |  -.0282422   .0166363    -1.70   0.090    -.0608488    .0043643
      yr1981 |  -.0694052    .028961    -2.40   0.017    -.1261677   -.0126426
      yr1982 |  -.0523678   .0423433    -1.24   0.216    -.1353591    .0306235
      yr1983 |  -.0256599   .0533747    -0.48   0.631    -.1302723    .0789525
      yr1984 |  -.0093229   .0696241    -0.13   0.893    -.1457837    .1271379
        year |   .0019575   .0119481     0.16   0.870    -.0214604    .0253754
       _cons |  -2.543221   23.97919    -0.11   0.916    -49.54158    44.45514
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w L2D.w D.k LD.k L2D.k D.yr1980 D.yr1981 D.yr1982
                  D.yr1983 D.yr1984 D.year
Instruments for level equation
        Standard: _cons
```

Now we fit the same model by using `xtdpdsys`:

```
. xtdpdsys n L(0/2).(w k) yr1980-yr1984 year, vce(robust)

System dynamic panel-data estimation        Number of obs          =       751
Group variable: id                          Number of groups       =       140
Time variable: year
                                            Obs per group:     min =         5
                                                               avg =  5.364286
                                                               max =         7

Number of instruments =     47              Wald chi2(13)          =   2579.96
                                            Prob > chi2            =    0.0000
One-step results
```

n	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
n						
L1.	.8221535	.093387	8.80	0.000	.6391184	1.005189
w						
--.	-.5427935	.1881721	-2.88	0.004	-.911604	-.1739831
L1.	.3703602	.1656364	2.24	0.025	.0457189	.6950015
L2.	-.0726314	.0907148	-0.80	0.423	-.2504292	.1051664
k						
--.	.3638069	.0657524	5.53	0.000	.2349346	.4926792
L1.	-.1222996	.0701521	-1.74	0.081	-.2597951	.015196
L2.	-.0901355	.0344142	-2.62	0.009	-.1575862	-.0226849
yr1980	-.0308622	.016946	-1.82	0.069	-.0640757	.0023512
yr1981	-.0718417	.0293223	-2.45	0.014	-.1293123	-.014371
yr1982	-.0384806	.0373631	-1.03	0.303	-.1117111	.0347498
yr1983	-.0121768	.0498519	-0.24	0.807	-.1098847	.0855311
yr1984	-.0050903	.0655011	-0.08	0.938	-.1334701	.1232895
year	.0058631	.0119867	0.49	0.625	-.0176304	.0293566
_cons	-10.59198	23.92087	-0.44	0.658	-57.47602	36.29207

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

If you are unfamiliar with the `L().()` notation, see [U] **13.9 Time-series operators**. That the system estimator produces a much higher estimate of the coefficient on lagged employment agrees with the results in Blundell and Bond (1998), who show that the system estimator does not have the downward bias that the Arellano–Bond estimator has when the true value is high.

Comparing the footers illustrates the difference between the two estimators; `xtdpdsys` includes lagged differences of `n` as instruments for the level equation, whereas `xtabond` does not. Comparing the headers shows that `xtdpdsys` has seven more instruments than `xtabond`. (As it should; there are 7 observations on `LD.n` available in the complete panels that run from 1976–1984, after accounting for the first two years that are lost because the model has two lags.) Only the first lags of the variables are used because the moment conditions using higher lags are redundant; see Blundell and Bond (1998) and Blundell, Bond, and Windmeijer (2000).

`estat abond` reports the Arellano–Bond test for serial correlation in the first-differenced errors. The moment conditions are valid only if there is no serial correlation in the idiosyncratic errors.

Because the first difference of independently and identically distributed idiosyncratic errors will be autocorrelated, rejecting the null hypothesis of no serial correlation at order one in the first-differenced errors does not imply that the model is misspecified. Rejecting the null hypothesis at higher orders implies that the moment conditions are not valid. See [XT] **xtdpd** for an alternative estimator in this case.

```
. estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors
```

Order	z	Prob > z
1	-4.6414	0.0000
2	-1.0572	0.2904

```
H0: no autocorrelation
```

The above output does not present evidence that the model is misspecified.

◁

▷ Example 2: 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.

```
. xtdpdsys n yr1980-yr1984 year, pre(w k, lag(2, .)) vce(robust)
System dynamic panel-data estimation          Number of obs         =       751
Group variable: id                            Number of groups      =       140
Time variable: year
                                              Obs per group:    min =         5
                                                                avg =  5.364286
                                                                max =         7

Number of instruments =     95                Wald chi2(13)         =   7562.80
                                              Prob > chi2           =    0.0000
One-step results
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .913278   .0460602    19.83   0.000     .8230017    1.003554
             |
           w |
         --. |   -.728159   .1019044    -7.15   0.000     -.927888   -.5284301
         L1. |   .5602737   .1939617     2.89   0.004     .1801156    .9404317
         L2. |  -.0523028   .1487653    -0.35   0.725    -.3438775    .2392718
             |
           k |
         --. |   .4820097   .0760787     6.34   0.000     .3328983    .6311212
         L1. |  -.2846944   .0831902    -3.42   0.001    -.4477442   -.1216446
         L2. |  -.1394181   .0405709    -3.44   0.001    -.2189356   -.0599006
             |
      yr1980 |  -.0325146   .0216371    -1.50   0.133    -.0749226    .0098935
      yr1981 |  -.0726116   .0346482    -2.10   0.036    -.1405207   -.0047024
      yr1982 |  -.0477038   .0451914    -1.06   0.291    -.1362772    .0408696
      yr1983 |  -.0396264   .0558734    -0.71   0.478    -.1491362    .0698835
      yr1984 |  -.0810383   .0736648    -1.10   0.271    -.2254186     .063342
        year |   .0192741   .0145326     1.33   0.185    -.0092092    .0477574
       _cons |  -37.34972   28.77747    -1.30   0.194    -93.75253    19.05308
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n L(1/.).L2.w L(1/.).L2.k
        Standard: D.yr1980 D.yr1981 D.yr1982 D.yr1983 D.yr1984 D.year
Instruments for level equation
        GMM-type: LD.n L2D.w L2D.k
        Standard: _cons
```

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 for the differenced errors and the second lags of the differences of `w` and `k` as instruments for the level errors.

◁

❑ Technical note

The above example illustrates that `xtdpdsys` understands `pre(w k, lag(2, .))` to mean that `L2.w` and `L2.k` are predetermined variables. This is a stricter definition than the alternative that `pre(w k, lag(2, .))` means only that `w k` are predetermined but to include two lags of `w` and two lags of `k` in the model. If you prefer the weaker definition, `xtdpdsys` 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.

❑

Stored results

xtdpdsys 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)	xtdpdsys
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(hascons)	hascons, if specified
e(transform)	specified transform
e(datasignature)	checksum from datasignature
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

Methods and formulas

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

Acknowledgment

We thank David Roodman of the Center for Global Development, who wrote xtabond2.

References

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

——. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47–82.

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.

Arellano, M., and O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29–51.

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.

Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.

Blundell, R., S. Bond, and F. Windmeijer. 2000. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In *Nonstationary Panels, Cointegrating Panels and Dynamic Panels*, ed. B. H. Baltagi, 53–92. New York: Elsevier.

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.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

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

Layard, R., and S. J. Nickell. 1986. Unemployment in Britain. *Economica* 53: S121–S169.

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25–51.

Also see

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

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

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

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

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

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

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

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