Description

`xtdpdsys` fits a linear dynamic panel-data model where the unobserved panel-level effects are correlated with the lags of the dependent variable. This model is an extension of the Arellano–Bond estimator that accommodates large autoregressive parameters and a large ratio of the variance of the panel-level effect to the variance of idiosyncratic error. This is known as the Arellano–Bover/Blundell–Bond system 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 that the panel-level effects be uncorrelated with the first difference of the first observation of the dependent variable.

Quick start

Dynamic panel-data regression of y on x with default Arellano–Bond instruments and lagged difference of y

```
xtdpdsys y x
```

Add the lagged difference of x as an instrument

```
xtdpdsys y x, pre(x)
```

Set the maximum number of lags of the dependent variable used as instruments to 2

```
xtdpdsys y x, maxldep(2)
```
**Syntax**

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

<table>
<thead>
<tr>
<th>options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>lags(#)</code></td>
<td>use # lags of dependent variable as covariates; default is <code>lags(1)</code></td>
</tr>
<tr>
<td><code>maxldep(#)</code></td>
<td>maximum lags of dependent variable for use as instruments</td>
</tr>
<tr>
<td><code>maxlags(#)</code></td>
<td>maximum lags of predetermined and endogenous variables for use as instruments</td>
</tr>
<tr>
<td><code>twostep</code></td>
<td>compute the two-step estimator instead of the one-step estimator</td>
</tr>
</tbody>
</table>

**Predetermined**

`pre(varlist[...])`  
predetermined variables; can be specified more than once

**Endogenous**

`endogenous(varlist[...])`  
enogenous 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.

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

**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**

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

Linear dynamic panel-data models include $p$ lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random
Consider the dynamic panel-data model

\[ y_{it} = \sum_{j=1}^{p} \alpha_j y_{i,t-j} + x_{it} \beta_1 + w_{it} \beta_2 + \nu_i + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \]  

(1)

where the \( \alpha_j \) are \( p \) parameters to be estimated, 
\( x_{it} \) is a \( 1 \times k_1 \) vector of strictly exogenous covariates, 
\( \beta_1 \) is a \( k_1 \times 1 \) vector of parameters to be estimated, 
\( w_{it} \) is a \( 1 \times k_2 \) vector of predetermined or endogenous covariates, 
\( \beta_2 \) is a \( k_2 \times 1 \) vector of parameters to be estimated, 
\( \nu_i \) are the panel-level effects (which may be correlated with the covariates), and 
\( \epsilon_{it} \) are i.i.d. over the whole sample with variance \( \sigma^2_\epsilon \).

The \( \nu_i \) and the \( \epsilon_{it} \) 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. Arellano and Bond (1991) derived a consistent generalized method of moments (GMM) estimator for this model. 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 \( \nu_i \) to the variance of the idiosyncratic error \( \epsilon_{it} \) 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 difference 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).

\texttt{xtdpdsys} fits dynamic panel-data estimators with the Arellano–Bover/Blundell–Bond system 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. Because \texttt{xtdpdsys} extends \texttt{xtabond}, [\texttt{XT}] \texttt{xtabond} provides useful background.

\textbf{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, \( n_{it} \) is the log of employment in firm \( i \) at time \( t \), \( w_{it} \) is the natural log of the real product wage, \( k_{it} \) is the natural log of the gross capital stock, and \( y_{it} \) is the natural log of industry output. The model also includes time dummies \texttt{yr1980}, \texttt{yr1981}, \texttt{yr1982}, \texttt{yr1983}, and \texttt{yr1984}.
For comparison, we begin by using xtabond to fit a model to these data.

```bash
use https://www.stata-press.com/data/r16/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)

|    | 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 | .069732  | -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 | -.1355391 .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
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)
```

<table>
<thead>
<tr>
<th>System dynamic panel-data estimation</th>
<th>Number of obs    = 751</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
<td>Number of groups      = 140</td>
</tr>
<tr>
<td>Time variable: year</td>
<td>Obs per group:</td>
</tr>
<tr>
<td></td>
<td>min = 5</td>
</tr>
<tr>
<td></td>
<td>avg = 5.364286</td>
</tr>
<tr>
<td></td>
<td>max = 7</td>
</tr>
<tr>
<td>Number of instruments = 47</td>
<td>Wald chi2(13)         = 2579.96</td>
</tr>
<tr>
<td></td>
<td>Prob &gt; chi2           = 0.0000</td>
</tr>
<tr>
<td>One-step results</td>
<td></td>
</tr>
</tbody>
</table>

| n         | Coef.     | Std. Err. | z      | P>|z|  | [95% Conf. Interval] |
|-----------|-----------|-----------|--------|------|---------------------|
| 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| -.1875862 -.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 .034798  |
| yr1983    | -.0121768 | .0498519  | -0.24  | 0.807| -1.098847 .0855311 |
| yr1984    | -.0050903 | .0655011  | -0.08  | 0.938| -.1334701 .1232895 |
| year      | .0058631  | .0119276  | 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`

Instruments for level equation

- `GMM-type: LD.n`
- `Standard: _cons`

If you are unfamiliar with the `L().()` notation, see [U] 13.10 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 independent 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 \texttt{XT xtdpd} for an alternative estimator in this case.

\begin{verbatim}
. estat abond
Arellano-Bond test for zero autocorrelation in first-differenced errors

<table>
<thead>
<tr>
<th>Order</th>
<th>z</th>
<th>Prob &gt; z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.6414</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>-1.0572</td>
<td>0.2904</td>
</tr>
</tbody>
</table>

H0: no autocorrelation
\end{verbatim}

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

\section*{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

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.  Std. Err.  z  P&gt;</td>
</tr>
<tr>
<td>n</td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.913278 .0460602 19.83 0.000 .8230017 1.003554</td>
</tr>
<tr>
<td>w</td>
<td>-.728159 .1019044 -7.15 0.000 -.9278879 -.5284301</td>
</tr>
<tr>
<td>L1.</td>
<td>.5602737 .1939617 2.89 0.004 .1801156 .9404317</td>
</tr>
<tr>
<td>L2.</td>
<td>-.0523028 .1487653 -0.35 0.725 -.3438775 .2392718</td>
</tr>
<tr>
<td>k</td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>-.2846944 .0831902 -3.42 0.001 -.4477442 -.1216446</td>
</tr>
<tr>
<td>L2.</td>
<td>-.1394181 .0405709 -3.44 0.001 -.2189356 -.0599006</td>
</tr>
<tr>
<td>yr1980</td>
<td>-.0326146 .0216371 -1.50 0.133 -.0749226 .0098935</td>
</tr>
<tr>
<td>yr1981</td>
<td>-.0726116 .0346482 -2.10 0.036 -.1405207 -.0047024</td>
</tr>
<tr>
<td>yr1982</td>
<td>-.0477038 .0451914 -1.06 0.291 -.1362772 .0408696</td>
</tr>
<tr>
<td>yr1983</td>
<td>-.0396264 .0558734 -0.71 0.478 -.1491362 .0698835</td>
</tr>
<tr>
<td>yr1984</td>
<td>-.0810383 .0736648 -1.10 0.271 -.2254186 .0633424</td>
</tr>
<tr>
<td>year</td>
<td>.0192741 .0145326 1.33 0.185 -.0092092 .0477574</td>
</tr>
<tr>
<td>_cons</td>
<td>-37.34972 28.77747 -1.30 0.194 -93.75253 19.05308</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/.) n L(1/.) L2.w L(1/.) L2.k

Instruments for level equation
GMM-type: L.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)` \( \chi^2 \)
- `e(arm#)` test for autocorrelation of order \#
- `e(artests)` number of AR tests computed
- `e(sig2)` estimate of \( \sigma^2 \)
- `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(transform)` specified transform
- `e(diffvars)` already-differenced exogenous variables
- `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

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 Open Philanthropy Project, who wrote `xtabond2`.
References


Also see

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

[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

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

[U] 20 Estimation and postestimation commands