

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

[Description](#)

[Quick start](#)

[Menu](#)

[Syntax](#)

[Options](#)

[Remarks and examples](#)

[Stored results](#)

[Methods and formulas](#)

[Acknowledgment](#)

[References](#)

[Also see](#)

## Description

`xtregar` fits cross-sectional time-series regression models when the disturbance term is first-order autoregressive. `xtregar` offers a within estimator for fixed-effects models and a GLS estimator for random-effects models. `xtregar` can accommodate unbalanced panels whose observations are unequally spaced over time.

## Quick start

Random-effects model of `y` on `x1` with an AR(1) disturbance using `xtset` data

```
xtregar y x1
```

Add `x2` and `x3` as covariates and perform Baltagi–Wu LBI test

```
xtregar y x1 x2 x3, lbi
```

Fixed-effects model using the within estimator and observations where `tvar` is greater than 2,000

```
xtregar y x1 x2 x3 if tvar > 2000, fe
```

## Menu

Statistics > Longitudinal/panel data > Linear models > Linear regression with AR(1) disturbance (FE, RE)

## Syntax

GLS random-effects (RE) model

```
xtregar depvar [indepvars] [if] [in] [, re options]
```

Fixed-effects (FE) model

```
xtregar depvar [indepvars] [if] [in] [weight] , fe [options]
```

*options*

Description

---

Model	
<code>re</code>	use random-effects estimator; the default
<code>fe</code>	use fixed-effects estimator
<code>rhotype(<i>rhomethod</i>)</code>	specify method to compute autocorrelation; seldom used
<code>rhof(#)</code>	use # for $\rho$ and do not estimate $\rho$
<code>twostep</code>	perform two-step estimate of correlation
Reporting	
<code>llevel(#)</code>	set confidence level; default is <code>llevel(95)</code>
<code>lbi</code>	perform Baltagi–Wu LBI test
<code>display_options</code>	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
<code>coeflegend</code>	display legend instead of statistics

---

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

`indepvars` may contain factor variables; see [U] 11.4.3 **Factor variables**.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 **Time-series varlists**.

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

`fweights` and `awweights` are allowed for the fixed-effects model with `rhotype(regress)` or `rhotype(freg)`, or with a fixed rho; see [U] 11.1.6 **weight**. Weights must be constant within panel.

`coeflegend` does not appear in the dialog box.

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

## Options

Model

`re` requests the GLS estimator of the random-effects model, which is the default.

`fe` requests the within estimator of the fixed-effects model.

`rhotype(rhmethod)` allows the user to specify any of the following estimators of  $\rho$ :

<code>dw</code>	$\rho_{dw} = 1 - d/2$ , where $d$ is the Durbin–Watson $d$ statistic
<code>regress</code>	$\rho_{reg} = \beta$ from the residual regression $\epsilon_t = \beta\epsilon_{t-1}$
<code>freg</code>	$\rho_{freg} = \beta$ from the residual regression $\epsilon_t = \beta\epsilon_{t+1}$
<code>tscorr</code>	$\rho_{tscorr} = \epsilon'\epsilon_{t-1}/\epsilon'\epsilon$ , where $\epsilon$ is the vector of residuals and $\epsilon_{t-1}$ is the vector of lagged residuals
<code>theil</code>	$\rho_{theil} = \rho_{tscorr}(N - k)/N$
<code>nagar</code>	$\rho_{nagar} = (\rho_{dw}N^2 + k^2)/(N^2 - k^2)$
<code>onestep</code>	$\rho_{onestep} = (n/m_c)(\epsilon'\epsilon_{t-1}/\epsilon'\epsilon)$ , where $\epsilon$ is the vector of residuals, $n$ is the number of observations, and $m_c$ is the number of consecutive pairs of residuals

`dw` is the default method. Except for `onestep`, the details of these methods are given in [TS] [prais](#). `prais` handles unequally spaced data. `onestep` is the one-step method proposed by Baltagi and Wu (1999). More details on this method are available below in [Methods and formulas](#).

`rhof(#)` specifies that the given number be used for  $\rho$  and that  $\rho$  not be estimated.

`twostep` requests that a two-step implementation of the *rhmethod* estimator of  $\rho$  be used. Unless a fixed value of  $\rho$  is specified (with the `rhof()` option),  $\rho$  is estimated by running `prais` on the de-meanded data. When `twostep` is specified, `prais` will stop on the first iteration after the equation is transformed by  $\rho$ —the two-step efficient estimator. Although it is customary to iterate these estimators to convergence, they are efficient at each step. When `twostep` is not specified, the FGLS process iterates to convergence as described in the [Methods and formulas](#) of [TS] [prais](#).

#### Reporting

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

`lbi` requests that the Baltagi–Wu (1999) locally best invariant (LBI) test statistic that  $\rho = 0$  and a modified version of the Bhargava, Franzini, and Narendranathan (1982) Durbin–Watson statistic be calculated and reported. The default is not to report them.  $p$ -values are not reported for either statistic. Although Bhargava, Franzini, and Narendranathan (1982) published critical values for their statistic, no tables are currently available for the Baltagi–Wu LBI. Baltagi and Wu (1999) derive a normalized version of their statistic, but this statistic cannot be computed for datasets of moderate size. You can also specify these options upon replay.

*display\_options*: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] [Estimation options](#).

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

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

## Remarks and examples

Remarks are presented under the following headings:

[Introduction](#)  
[The fixed-effects model](#)  
[The random-effects model](#)

### Introduction

If you have not read [\[XT\] xt](#), please do so.

**xtregar** fits cross-sectional time-series regression models when the disturbance term is first-order autoregressive. The models of interest are described by

$$y_{it} = \alpha + \mathbf{x}_{it}\beta + \nu_i + \epsilon_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T_i \quad (1)$$

where

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + \eta_{it} \quad (2)$$

and where  $|\rho| < 1$  and  $\eta_{it}$  is independent and identically distributed (i.i.d.) with mean 0 and variance  $\sigma_\eta^2$ .

In the fixed-effects model, the  $\nu_i$  are assumed to be correlated with the covariates  $\mathbf{x}_{it}$ , whereas in the random-effects model they are assumed to follow an i.i.d. process with mean 0 and variance  $\sigma_\eta^2$  and to be uncorrelated with the  $\mathbf{x}_{it}$ .

Similar to other linear panel-data models, any  $\mathbf{x}_{it}$  that do not vary over  $t$  are collinear with the  $\nu_i$  and will be omitted from the fixed-effects model. In contrast, the random-effects model can accommodate covariates that are constant over time.

**xtregar** offers a within estimator for the fixed-effect model and the Baltagi–Wu (1999) GLS estimator of the random-effects model. Both of these estimators offer several estimators of  $\rho$ .

The Baltagi–Wu (1999) GLS estimator extends the balanced panel estimator in [Baltagi and Li \(1991\)](#) to a case of exogenously unbalanced panels with unequally spaced observations. Specifically, the dataset contains observations on individual  $i$  at times  $t_{ij}$  for  $j = 1, \dots, n_i$ . The difference  $t_{ij} - t_{i,j-1}$  plays an integral role in the estimation techniques used by **xtregar**.

For this reason, you must specify the `delta()` option when you `xtset panelvar timevar` if, for example, you have quarterly data with a monthly `timevar` recorded every three months instead of a quarterly `timevar`; see [\[XT\] xtset](#).

### The fixed-effects model

Let's examine the fixed-effect model first. The basic approach is common to all fixed-effects models. The  $\nu_i$  are treated as nuisance parameters. We use a transformation of the model that removes the nuisance parameters and leaves behind the parameters of interest in an estimable form. Subtracting the group means from (1) removes the  $\nu_i$  from the model

$$y_{it_{ij}} - \bar{y}_i = (\bar{\mathbf{x}}_{it_{ij}} - \bar{\mathbf{x}}_i) \beta + \epsilon_{it_{ij}} - \bar{\epsilon}_i \quad (3)$$

where

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{it_{ij}} \quad \bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{it_{ij}} \quad \bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{it_{ij}}$$

After the transformation, (3) is a linear AR(1) model, potentially with unequally spaced observations. (3) can be used to estimate  $\rho$ . Given an estimate of  $\rho$ , we must do a Cochrane–Orcutt transformation on each panel and then remove the within-panel means and add back the overall mean for each variable. OLS on the transformed data will produce the within estimates of  $\alpha$  and  $\beta$ .

### ► Example 1: Fixed-effects model

Let's use the Grunfeld investment dataset to illustrate how `xtregar` can be used to fit the fixed-effects model. This dataset contains information on 10 firms' investment, market value, and the value of their capital stocks. The data were collected annually between 1935 and 1954. The following output shows that we have `xtset` our data and gives the results of running a fixed-effects model with investment as a function of market value and the capital stock.

```
. use https://www.stata-press.com/data/r17/grunfeld
. xtset
Panel variable: company (strongly balanced)
Time variable: year, 1935 to 1954
Delta: 1 year
. xtregar invest mvalue kstock, fe
FE (within) regression with AR(1) disturbances   Number of obs   =   190
Group variable: company                         Number of groups =   10
R-squared:                                     Obs per group:
    Within = 0.5927                               min =   19
    Between = 0.7989                             avg =  19.0
    Overall = 0.7904                             max =   19
corr(u_i, Xb) = -0.0454                         F(2,178)       =  129.49
                                                Prob > F       =   0.0000
```

invest	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
mvalue	.0949999	.0091377	10.40	0.000	.0769677	.113032
kstock	.350161	.0293747	11.92	0.000	.2921935	.4081286
_cons	-63.22022	5.648271	-11.19	0.000	-74.36641	-52.07402
rho_ar	.67210608					
sigma_u	91.507609					
sigma_e	40.992469					
rho_fov	.8328647	(fraction of variance because of u_i)				

F test that all u\_i=0: F(9,178) = 11.53 Prob > F = 0.0000

Because there are 10 groups, the panel-by-panel Cochrane–Orcutt method decreases the number of available observations from 200 to 190. The above example used the default `dw` estimator of  $\rho$ . Using the `tscorr` estimator of  $\rho$  yields

```

. xtregar invest mvalue kstock, fe rhotype(tscorr)
FE (within) regression with AR(1) disturbances  Number of obs    =    190
Group variable: company                        Number of groups   =    10
R-squared:                                     Obs per group:
  Within = 0.6583                               min =           19
  Between = 0.8024                              avg =          19.0
  Overall = 0.7933                               max =           19
                                                F(2,178)         =    171.47
corr(u_i, Xb) = -0.0709                        Prob > F         =    0.0000

```

	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
invest						
mvalue	.0978364	.0096786	10.11	0.000	.0787369	.1169359
kstock	.346097	.0242248	14.29	0.000	.2982922	.3939018
_cons	-61.84403	6.621354	-9.34	0.000	-74.91049	-48.77758
rho_ar	.54131231					
sigma_u	90.893572					
sigma_e	41.592151					
rho_fov	.82686297	(fraction of variance because of u_i)				

F test that all u\_i=0: F(9,178) = 19.73 Prob > F = 0.0000



## □ Technical note

The `tscorr` estimator of  $\rho$  is bounded in  $[-1, 1]$ . The other estimators of  $\rho$  are not. In samples with short panels, the estimates of  $\rho$  produced by the other estimators of  $\rho$  may be outside  $[-1, 1]$ . If this happens, use the `tscorr` estimator. However, simulations have shown that the `tscorr` estimator is biased toward zero. `dw` is the default because it performs well in Monte Carlo simulations. In the example above, the estimate of  $\rho$  produced by `tscorr` is much smaller than the one produced by `dw`.



## ▷ Example 2: Using `xtset`

`xtregar` will complain if you try to run `xtregar` on a dataset that has not been `xtset`:

```

. xtset, clear
. xtregar invest mvalue kstock, fe
must specify panelvar and timevar; use xtset
r(459);

```

You must `xtset` your data to ensure that `xtregar` understands the nature of your time variable. Suppose that our observations were taken quarterly instead of annually. We will get the same results with the quarterly variable `t2` that we did with the annual variable `year`.

```
. generate t = year - 1934
. generate t2 = tq(1934q4) + t
. format t2 %tq
. list year t2 in 1/5
```

	year	t2
1.	1935	1935q1
2.	1936	1935q2
3.	1937	1935q3
4.	1938	1935q4
5.	1939	1936q1

```
. xtset company t2
```

```
Panel variable: company (strongly balanced)
```

```
Time variable: t2, 1935q1 to 1939q4
```

```
Delta: 1 quarter
```

```
. xtregar invest mvalue kstock, fe
```

```
FE (within) regression with AR(1) disturbances Number of obs = 190
```

```
Group variable: company Number of groups = 10
```

```
R-squared:
```

```
Within = 0.5927
```

```
Between = 0.7989
```

```
Overall = 0.7904
```

```
Obs per group:
```

```
min = 19
```

```
avg = 19.0
```

```
max = 19
```

```
F(2,178) = 129.49
```

```
Prob > F = 0.0000
```

```
corr(u_i, Xb) = -0.0454
```

invest	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
mvalue	.0949999	.0091377	10.40	0.000	.0769677	.113032
kstock	.350161	.0293747	11.92	0.000	.2921935	.4081286
_cons	-63.22022	5.648271	-11.19	0.000	-74.36641	-52.07402
rho_ar	.67210608					
sigma_u	91.507609					
sigma_e	40.992469					
rho_fov	.8328647	(fraction of variance because of u_i)				

```
F test that all u_i=0: F(9,178) = 11.53
```

```
Prob > F = 0.0000
```

◀

In all the examples thus far, we have assumed that  $\epsilon_{it}$  is first-order autoregressive. Testing the hypothesis of  $\rho = 0$  in a first-order autoregressive process produces test statistics with extremely complicated distributions. [Bhargava, Franzini, and Narendranathan \(1982\)](#) extended the Durbin–Watson statistic to the case of balanced, equally spaced panel datasets. [Baltagi and Wu \(1999\)](#) modify their statistic to account for unbalanced panels with unequally spaced data. In the same article, [Baltagi and Wu \(1999\)](#) derive the locally best invariant test statistic of  $\rho = 0$ . Both these test statistics have extremely complicated distributions, although [Bhargava, Franzini, and Narendranathan \(1982\)](#) did publish some critical values in their article. Specifying the `lbi` option to `xtregar` causes Stata to calculate and report the modified Bhargava et al. Durbin–Watson and the Baltagi–Wu LBI.

### ► Example 3: Testing for autocorrelation

In this example, we calculate the modified Bhargava et al. Durbin–Watson statistic and the Baltagi–Wu LBI. We exclude periods 9 and 10 from the sample, thereby reproducing the results of [Baltagi](#)

and Wu (1999, 822).  $p$ -values are not reported for either statistic. Although Bhargava, Franzini, and Narendranathan (1982) published critical values for their statistic, no tables are currently available for the Baltagi–Wu (LBI). Baltagi and Wu (1999) did derive a normalized version of their statistic, but this statistic cannot be computed for datasets of moderate size.

```
. xtregar invest mvalue kstock if year !=1934 & year !=1944, fe lbi
FE (within) regression with AR(1) disturbances   Number of obs   =       180
Group variable: company                         Number of groups =        10
R-squared:                                       Obs per group:
  Within   = 0.5954                               min       =        18
  Between  = 0.7952                               avg       =       18.0
  Overall  = 0.7889                               max       =        18
                                                F(2,168)       =       123.63
corr(u_i, Xb) = -0.0516                          Prob > F       =        0.0000
```

	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
invest						
mvalue	.0941122	.0090926	10.35	0.000	.0761617	.1120627
kstock	.3535872	.0303562	11.65	0.000	.2936584	.4135161
_cons	-64.82534	5.946885	-10.90	0.000	-76.56559	-53.08509
rho_ar	.6697198					
sigma_u	93.320452					
sigma_e	41.580712					
rho_fov	.83435413	(fraction of variance because of u_i)				

```
F test that all u_i=0: F(9,168) = 11.55                               Prob > F = 0.0000
Modified Bhargava et al. Durbin-Watson = .71380994
Baltagi-Wu LBI = 1.0134522
```

4

## The random-effects model

In the random-effects model, the  $\nu_i$  are assumed to be realizations of an i.i.d. process with mean 0 and variance  $\sigma_\nu^2$ . Furthermore, the  $\nu_i$  are assumed to be independent of both the  $\epsilon_{it}$  and the covariates  $\mathbf{x}_{it}$ . The latter of these assumptions can be strong, but inference is not conditional on the particular realizations of the  $\nu_i$  in the sample. See Mundlak (1978) for a discussion of this point.

### ► Example 4: Random-effects model

By specifying the `re` option, we obtain the Baltagi–Wu GLS estimator of the random-effects model. This estimator can accommodate unbalanced panels and unequally spaced data. We run this model on the Grunfeld dataset:



```

. xtregar invest mvalue kstock if year !=1934 & year !=1944, re lbi
RE GLS regression with AR(1) disturbances      Number of obs      =      190
Group variable: company                       Number of groups   =       10
R-squared:                                    Obs per group:
    Within = 0.7707                            min =              19
    Between = 0.8039                            avg =             19.0
    Overall = 0.7958                            max =              19
                                                Wald chi2(3)       =     351.37
corr(u_i, Xb) = 0 (assumed)                   Prob > chi2        =      0.0000

```

invest	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
mvalue	.0947714	.0083691	11.32	0.000	.0783683	.1111746
kstock	.3223932	.0263226	12.25	0.000	.2708019	.3739845
_cons	-45.21427	27.12492	-1.67	0.096	-98.37814	7.949603
rho_ar	.6697198	(estimated autocorrelation coefficient)				
sigma_u	74.662876					
sigma_e	42.253042					
rho_fov	.75742494	(fraction of variance due to u_i)				
theta	.66973313					

```

Modified Bhargava et al. Durbin-Watson = .71380994
Baltagi-Wu LBI = 1.0134522

```

The modified Bhargava et al. Durbin–Watson and the Baltagi–Wu LBI are the same as those reported for the fixed-effects model because the formulas for these statistics do not depend on fitting the fixed-effects model or the random-effects model.

◀

## Stored results

`xtregar`, `re` stores the following in `e()`:

### Scalars

<code>e(N)</code>	number of observations
<code>e(N_g)</code>	number of groups
<code>e(df_m)</code>	model degrees of freedom
<code>e(g_min)</code>	smallest group size
<code>e(g_avg)</code>	average group size
<code>e(g_max)</code>	largest group size
<code>e(d1)</code>	Bhargava et al. Durbin–Watson
<code>e(LBI)</code>	Baltagi–Wu LBI statistic
<code>e(N_LBI)</code>	number of obs used in <code>e(LBI)</code>
<code>e(Tcon)</code>	1 if $T$ is constant
<code>e(sigma_u)</code>	panel-level standard deviation
<code>e(sigma_e)</code>	standard deviation of $\eta_{it}$
<code>e(r2_w)</code>	$R^2$ for within model
<code>e(r2_o)</code>	$R^2$ for overall model
<code>e(r2_b)</code>	$R^2$ for between model
<code>e(chi2)</code>	$\chi^2$
<code>e(rho_ar)</code>	autocorrelation coefficient
<code>e(rho_fov)</code>	$u_i$ fraction of variance
<code>e(thta_min)</code>	minimum $\theta$
<code>e(thta_5)</code>	$\theta$ , 5th percentile
<code>e(thta_50)</code>	$\theta$ , 50th percentile
<code>e(thta_95)</code>	$\theta$ , 95th percentile
<code>e(thta_max)</code>	maximum $\theta$
<code>e(Tbar)</code>	harmonic mean of group sizes
<code>e(rank)</code>	rank of <code>e(V)</code>

### Macros

<code>e(cmd)</code>	<code>xtregar</code>
<code>e(cmdline)</code>	command as typed
<code>e(depvar)</code>	name of dependent variable
<code>e(ivar)</code>	variable denoting groups
<code>e(tvar)</code>	variable denoting time within groups
<code>e(model)</code>	<code>re</code>
<code>e(rhotype)</code>	method of estimating $\rho_{ar}$
<code>e(dw)</code>	<code>lbi</code> , if <code>lbi</code> specified
<code>e(chi2type)</code>	Wald; type of model $\chi^2$ test
<code>e(properties)</code>	<code>b V</code>
<code>e(predict)</code>	program used to implement <code>predict</code>
<code>e(marginsok)</code>	predictions allowed by <code>margins</code>
<code>e(marginsnotok)</code>	predictions disallowed by <code>margins</code>
<code>e(asbalanced)</code>	factor variables <code>fvset</code> as <code>asbalanced</code>
<code>e(asobserved)</code>	factor variables <code>fvset</code> as <code>asobserved</code>

### Matrices

<code>e(b)</code>	coefficient vector
<code>e(V)</code>	VCE for random-effects model

### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

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

### Matrices

<code>r(table)</code>	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.

xtregar, fe stores the following in e():

#### Scalars

e(N)	number of observations
e(N_g)	number of groups
e(df_m)	model degrees of freedom
e(mss)	model sum of squares
e(rss)	residual sum of squares
e(g_min)	smallest group size
e(g_avg)	average group size
e(g_max)	largest group size
e(d1)	Bhargava et al. Durbin–Watson
e(LBI)	Baltagi–Wu LBI statistic
e(N_LBI)	number of obs used in e(LBI)
e(Tcon)	1 if $T$ is constant
e(corr)	$\text{corr}(u_i, Xb)$
e(sigma_u)	panel-level standard deviation
e(sigma_e)	standard deviation of $\epsilon_{it}$
e(r2_a)	adjusted $R^2$
e(r2_w)	$R^2$ for within model
e(r2_o)	$R^2$ for overall model
e(r2_b)	$R^2$ for between model
e(ll)	log likelihood
e(ll_0)	log likelihood, constant-only model
e(rho_ar)	autocorrelation coefficient
e(rho_fov)	$u_i$ fraction of variance
e(F)	$F$ statistic
e(F_f)	$F$ for $u_i=0$
e(df_r)	residual degrees of freedom
e(df_a)	degrees of freedom for absorbed effect
e(df_b)	numerator degrees of freedom for $F$ statistic
e(rmse)	root mean squared error
e(Tbar)	harmonic mean of group sizes
e(rank)	rank of e(V)

#### Macros

e(cmd)	xtregar
e(cmdline)	command as typed
e(depvar)	name of dependent variable
e(ivar)	variable denoting groups
e(tvar)	variable denoting time within groups
e(wtype)	weight type
e(wexp)	weight expression
e(model)	fe
e(rhotype)	method of estimating $\rho_{ar}$
e(dw)	lbi, if lbi specified
e(properties)	b V
e(predict)	program used to implement predict
e(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed by margins
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved

#### Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators

#### Functions

e(sample)	marks estimation sample
-----------	-------------------------

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

Matrices	
<code>r(table)</code>	matrix containing the coefficients with their standard errors, test statistics, $p$ -values, and confidence intervals

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

## Methods and formulas

Consider a linear panel-data model described by (1) and (2). The data can be unbalanced and unequally spaced. Specifically, the dataset contains observations on individual  $i$  at times  $t_{ij}$  for  $j = 1, \dots, n_i$ .

Methods and formulas are presented under the following headings:

*Estimating  $\rho$*   
*Transforming the data to remove the AR(1) component*  
*The within estimator of the fixed-effects model*  
*The Baltagi–Wu GLS estimator*  
*The test statistics*

### Estimating $\rho$

The estimate of  $\rho$  is always obtained after removing the group means. Let  $\tilde{y}_{it} = y_{it} - \bar{y}_i$ , let  $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$ , and let  $\tilde{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i$ .

Then, except for the `onestep` method, all the estimates of  $\rho$  are obtained by running Stata's `prais` on

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \tilde{\epsilon}_{it}$$

See [TS] `prais` for the formulas for each of the methods.

When `onestep` is specified, a regression is run on the above equation, and the residuals are obtained. Let  $e_{it_{ij}}$  be the residual used to estimate the error  $\tilde{\epsilon}_{it_{ij}}$ . If  $t_{ij} - t_{i,j-1} > 1$ ,  $e_{it_{ij}}$  is set to zero. Given this series of residuals

$$\hat{\rho}_{\text{onestep}} = \frac{n}{m_c} \frac{\sum_{i=1}^N \sum_{t=2}^T e_{it} e_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2}$$

where  $n$  is the number of nonzero elements in  $e$  and  $m_c$  is the number of consecutive pairs of nonzero  $e_{it}$ s.

### Transforming the data to remove the AR(1) component

After estimating  $\rho$ , Baltagi and Wu (1999) derive a transformation of the data that removes the AR(1) component. Their  $C_i(\rho)$  can be written as

$$y_{it_{ij}}^* = \begin{cases} (1 - \rho^2)^{1/2} y_{it_{ij}} & \text{if } t_{ij} = 1 \\ (1 - \rho^2)^{1/2} \left\{ y_{i,t_{ij}} \frac{1}{(1 - \rho^{2(t_{ij} - t_{i,j-1})})^{1/2}} - y_{i,t_{i,j-1}} \frac{\rho^{(t_{ij} - t_{i,j-1})}}{(1 - \rho^{2(t_{i,j} - t_{i,j-1})})^{1/2}} \right\} & \text{if } t_{ij} > 1 \end{cases}$$

Using the analogous transform on the independent variables generates transformed data without the AR(1) component. Performing simple OLS on the transformed data leaves behind the residuals  $\mu^*$ .

## The within estimator of the fixed-effects model

To obtain the within estimator, we must transform the data that come from the AR(1) transform. For the within transform to remove the fixed effects, the first observation of each panel must be omitted. Specifically, let

$$\begin{aligned}\check{y}_{it_{ij}} &= y_{it_{ij}}^* - \bar{y}_i^* + \bar{\bar{y}}^* & \forall j > 1 \\ \check{\mathbf{x}}_{it_{ij}} &= \mathbf{x}_{it_{ij}}^* - \bar{\mathbf{x}}_i^* + \bar{\bar{\mathbf{x}}}^* & \forall j > 1 \\ \check{\epsilon}_{it_{ij}} &= \epsilon_{it_{ij}}^* - \bar{\epsilon}_i^* + \bar{\bar{\epsilon}}^* & \forall j > 1\end{aligned}$$

where

$$\begin{aligned}\bar{y}_i^* &= \frac{\sum_{j=2}^{n_i-1} y_{it_{ij}}^*}{n_i - 1} \\ \bar{\bar{y}}^* &= \frac{\sum_{i=1}^N \sum_{j=2}^{n_i-1} y_{it_{ij}}^*}{\sum_{i=1}^N n_i - 1} \\ \bar{\mathbf{x}}_i^* &= \frac{\sum_{j=2}^{n_i-1} \mathbf{x}_{it_{ij}}^*}{n_i - 1} \\ \bar{\bar{\mathbf{x}}}^* &= \frac{\sum_{i=1}^N \sum_{j=2}^{n_i-1} \mathbf{x}_{it_{ij}}^*}{\sum_{i=1}^N n_i - 1} \\ \bar{\epsilon}_i^* &= \frac{\sum_{j=2}^{n_i-1} \epsilon_{it_{ij}}^*}{n_i - 1} \\ \bar{\bar{\epsilon}}^* &= \frac{\sum_{i=1}^N \sum_{j=2}^{n_i-1} \epsilon_{it_{ij}}^*}{\sum_{i=1}^N n_i - 1}\end{aligned}$$

The within estimator of the fixed-effects model is then obtained by running OLS on

$$\check{y}_{it_{ij}} = \alpha + \check{\mathbf{x}}_{it_{ij}} \boldsymbol{\beta} + \check{\epsilon}_{it_{ij}}$$

Reported as  $R^2$  within is the  $R^2$  from the above regression.

Reported as  $R^2$  between is  $\left\{ \text{corr}(\bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}}, \bar{y}_i) \right\}^2$ .

Reported as  $R^2$  overall is  $\left\{ \text{corr}(\mathbf{x}_{it} \hat{\boldsymbol{\beta}}, y_{it}) \right\}^2$ .

## The Baltagi–Wu GLS estimator

The residuals  $\mu^*$  can be used to estimate the variance components. Translating the matrix formulas given in Baltagi and Wu (1999) into summations yields the following variance-components estimators:

$$\begin{aligned}\widehat{\sigma}_\omega^2 &= \sum_{i=1}^N \frac{(\mu_i^{*'} g_i)^2}{(g_i' g_i)} \\ \widehat{\sigma}_\epsilon^2 &= \frac{\left[ \sum_{i=1}^N (\mu_i^{*'} \mu_i^*) - \sum_{i=1}^N \left\{ \frac{(\mu_i^{*'} g_i)^2}{(g_i' g_i)} \right\} \right]}{\sum_{i=1}^N (n_i - 1)} \\ \widehat{\sigma}_\mu^2 &= \frac{\left[ \sum_{i=1}^N \left\{ \frac{(\mu_i^{*'} g_i)^2}{(g_i' g_i)} \right\} - N \widehat{\sigma}_\epsilon^2 \right]}{\sum_{i=1}^N (g_i' g_i)}\end{aligned}$$

where

$$g_i = \left[ 1, \frac{\{1 - \rho^{(t_{i,2} - t_{i,1})}\}}{\{1 - \rho^{2(t_{i,2} - t_{i,1})}\}^{\frac{1}{2}}}, \dots, \frac{\{1 - \rho^{(t_{i,n_i} - t_{i,n_i-1})}\}}{\{1 - \rho^{2(t_{i,n_i} - t_{i,n_i-1})}\}^{\frac{1}{2}}} \right]'$$

and  $\mu_i^*$  is the  $n_i \times 1$  vector of residuals from  $\mu^*$  that correspond to person  $i$ .

Then

$$\widehat{\theta}_i = 1 - \left( \frac{\widehat{\sigma}_\mu}{\widehat{\omega}_i} \right)$$

where

$$\widehat{\omega}_i^2 = g_i' g_i \widehat{\sigma}_\mu^2 + \widehat{\sigma}_\epsilon^2$$

With these estimates in hand, we can transform the data via

$$z_{it_{ij}}^{**} = z_{it_{ij}}^* - \widehat{\theta}_i g_{ij} \frac{\sum_{s=1}^{n_i} g_{is} z_{it_{is}}^*}{\sum_{s=1}^{n_i} g_{is}^2}$$

for  $z \in \{y, \mathbf{x}\}$ .

Running OLS on the transformed data  $y^{**}, \mathbf{x}^{**}$  yields the feasible GLS estimator of  $\alpha$  and  $\beta$ .

Reported as  $R^2$  between is  $\left\{ \text{corr}(\bar{\mathbf{x}}_i \widehat{\beta}, \bar{y}_i) \right\}^2$ .

Reported as  $R^2$  within is  $\left\{ \text{corr}\{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \widehat{\beta}, y_{it} - \bar{y}_i\} \right\}^2$ .

Reported as  $R^2$  overall is  $\left\{ \text{corr}(\mathbf{x}_{it} \widehat{\beta}, y_{it}) \right\}^2$ .

## The test statistics

The Baltagi–Wu LBI is the sum of terms

$$d_* = d_1 + d_2 + d_3 + d_4$$

where

$$d_1 = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} \{\tilde{z}_{it_{ij}} - \tilde{z}_{it_{i,j-1}} I(t_{ij} - t_{i,j-1} = 1)\}^2}{\sum_{i=1}^N \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_2 = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i-1} \tilde{z}_{it_{ij}}^2 \{1 - I(t_{i,j+1} - t_{ij} = 1)\}}{\sum_{i=1}^N \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_3 = \frac{\sum_{i=1}^N \tilde{z}_{it_{i1}}^2}{\sum_{i=1}^N \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_4 = \frac{\sum_{i=1}^N \tilde{z}_{it_{in_i}}^2}{\sum_{i=1}^N \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$I()$  is the indicator function that takes the value of 1 if the condition is true and 0 otherwise. The  $\tilde{z}_{it_{i,j-1}}$  are residuals from the within estimator.

Baltagi and Wu (1999) also show that  $d_1$  is the Bhargava et al. Durbin–Watson statistic modified to handle cases of unbalanced panels and unequally spaced data.

## Acknowledgment

We thank Badi Baltagi of the Department of Economics at Syracuse University for his helpful comments.

## References

- Baltagi, B. H. 2009. *A Companion to Econometric Analysis of Panel Data*. Chichester, UK: Wiley.
- . 2013. *Econometric Analysis of Panel Data*. 5th ed. Chichester, UK: Wiley.
- Baltagi, B. H., and Q. Li. 1991. A transformation that will circumvent the problem of autocorrelation in an error-component model. *Journal of Econometrics* 48: 385–393. [https://doi.org/10.1016/0304-4076\(91\)90070-T](https://doi.org/10.1016/0304-4076(91)90070-T).
- Baltagi, B. H., and P. X. Wu. 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823. <https://doi.org/10.1017/S0266466699156020>.
- Bhargava, A., L. Franzini, and W. Narendranathan. 1982. Serial correlation and the fixed effects model. *Review of Economic Studies* 49: 533–549. <https://doi.org/10.2307/2297285>.
- Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3: 168–178.
- Hoechle, D. 2007. Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal* 7: 281–312.
- Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46: 69–85. <https://doi.org/10.2307/1913646>.

Sosa-Escudero, W., and A. K. Bera. 2008. Tests for unbalanced error-components models under local misspecification. *Stata Journal* 8: 68–78.

Wursten, J. 2018. Testing for serial correlation in fixed-effects panel models. *Stata Journal* 18: 76–100.

## Also see

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

[XT] **xtgee** — Fit population-averaged panel-data models by using GEE

[XT] **xtgls** — Fit panel-data models by using GLS

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

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

[TS] **newey** — Regression with Newey–West standard errors

[TS] **prais** — Prais–Winsten and Cochrane–Orcutt regression

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