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

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

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

fweights and aweights 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(*rhomethod*) allows the user to specify any of the following estimators of ρ:

dw	$\rho_{\text{dw}} = 1 - d/2$, where d is the Durbin–Watson d statistic
regress	$\rho_{\text{reg}} = \beta$ from the residual regression $\epsilon_t = \beta\epsilon_{t-1}$
freg	$\rho_{\text{freg}} = \beta$ from the residual regression $\epsilon_t = \beta\epsilon_{t+1}$
tscorr	$\rho_{\text{tscorr}} = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}_{t-1}/\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the vector of residuals and $\boldsymbol{\epsilon}_{t-1}$ is the vector of lagged residuals
theil	$\rho_{\text{theil}} = \rho_{\text{tscorr}}(N-k)/N$
nagar	$\rho_{\text{nagar}} = (\rho_{\text{dw}}N^2 + k^2)/(N^2 - k^2)$
onestep	$\rho_{\text{onestep}} = (n/m_c)(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}_{t-1}/\boldsymbol{\epsilon}'\boldsymbol{\epsilon})$, where $\boldsymbol{\epsilon}$ 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 ρ and that ρ not be estimated.

twostep requests that a two-step implementation of the *rhomethod* estimator of ρ be used. Unless a fixed value of ρ is specified (with the rhof() option), ρ is estimated by running prais on the de-meaned data. When twostep is specified, prais will stop on the first iteration after the equation is transformed by ρ—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

stata.com

Remarks are presented under the following headings:

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

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

where

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

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

In the fixed-effects model, the ν_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 σ_η^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 ν_i and will be dropped 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 ρ.

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 ν_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 ν_i from the model

$$y_{it_{ij}} - \overline{y}_i = \left(\overline{\mathbf{x}}_{it_{ij}} - \overline{\mathbf{x}}_i\right)\boldsymbol{\beta} + \epsilon_{it_{ij}} - \overline{\epsilon}_i \tag{3}$$

where

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

After the transformation, (3) is a linear AR(1) model, potentially with unequally spaced observations. (3) can be used to estimate ρ. Given an estimate of ρ, 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 α and $\boldsymbol{\beta}$.

▷ 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 http://www.stata-press.com/data/r14/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-sq:                                           Obs per group:
     within  = 0.5927                                         min =         19
     between = 0.7989                                         avg =       19.0
     overall = 0.7904                                         max =         19
                                                F(2,178)          =     129.49
corr(u_i, Xb)  = -0.0454                        Prob > F          =     0.0000

      invest       Coef.   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 ρ. Using the `tscorr` estimator of ρ 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-sq:                                           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

      invest        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

      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 ρ is bounded in $[-1, 1]$. The other estimators of ρ are not. In samples with short panels, the estimates of ρ produced by the other estimators of ρ 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 ρ 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-sq:                                           Obs per group:
     within  = 0.5927                                         min =        19
     between = 0.7989                                         avg =      19.0
     overall = 0.7904                                         max =        19

                                                F(2,178)           =    129.49
corr(u_i, Xb)  = -0.0454                        Prob > F           =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   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 ϵ_{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-sq:                                           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

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      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
```

◁

The random-effects model

In the random-effects model, the ν_i are assumed to be realizations of an i.i.d. process with mean 0 and variance σ^2_ν. Furthermore, the ν_i are assumed to be independent of both the ϵ_{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 ν_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-sq:                                           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 |      Coef.   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

`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(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(sigma_u)`	panel-level standard deviation
`e(sigma_e)`	standard deviation of η_{it}
`e(r2_w)`	R-squared for within model
`e(r2_o)`	R-squared for overall model
`e(r2_b)`	R-squared for between model
`e(chi2)`	χ^2
`e(rho_ar)`	autocorrelation coefficient
`e(rho_fov)`	u_i fraction of variance
`e(thta_min)`	minimum θ
`e(thta_5)`	θ, 5th percentile
`e(thta_50)`	θ, 50th percentile
`e(thta_95)`	θ, 95th percentile
`e(thta_max)`	maximum θ
`e(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(model)`	`re`
`e(rhotype)`	method of estimating ρ_{ar}
`e(dw)`	LBI, if requested
`e(chi2type)`	`Wald`; type of model χ^2 test
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
`e(asbalanced)`	factor variables `fvset` as `asbalanced`
`e(asobserved)`	factor variables `fvset` as `asobserved`

Matrices

`e(b)`	coefficient vector
`e(V)`	VCE for random-effects model

Functions

`e(sample)`	marks estimation sample

`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)`	corr(u_i, `Xb`)
`e(sigma_u)`	panel-level standard deviation
`e(sigma_e)`	standard deviation of ϵ_{it}
`e(r2_a)`	adjusted R-squared
`e(r2_w)`	R-squared for within model
`e(r2_o)`	R-squared for overall model
`e(r2_b)`	R-squared 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 ρ_{ar}
`e(dw)`	LBI, if requested
`e(properties)`	`b V`
`e(predict)`	program used to implement `predict`
`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

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 ρ

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

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

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

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 $\widetilde{\epsilon}_{it_{ij}}$. If $t_{ij} - t_{i,j-1} > 1$, $e_{it_{ij}}$ is set to zero. Given this series of residuals

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

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

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

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 μ^*.

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 dropped. Specifically, let

$$\begin{aligned}
\breve{y}_{it_{ij}} &= y^*_{it_{ij}} - \overline{y}^*_i + \overline{\overline{y}}^* \qquad \forall j > 1 \\
\breve{\mathbf{x}}_{it_{ij}} &= \mathbf{x}^*_{it_{ij}} - \overline{\mathbf{x}}^*_i + \overline{\overline{\mathbf{x}}}^* \qquad \forall j > 1 \\
\breve{\epsilon}_{it_{ij}} &= \epsilon^*_{it_{ij}} - \overline{\epsilon}^*_i + \overline{\overline{\epsilon}}^* \qquad \forall j > 1
\end{aligned}$$

where

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

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

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

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

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

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

The Baltagi–Wu GLS estimator

The residuals μ^* 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:

$$\widehat{\sigma}^2_\omega = \sum_{i=1}^{N} \frac{(\mu_i^{*\prime} g_i)^2}{(g_i' g_i)}$$

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

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

where

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

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

Then

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

where

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

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 α and $\boldsymbol{\beta}$.

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

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

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

The test statistics

The Baltagi–Wu LBI is the sum of terms

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

where

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

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

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

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

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

Baltagi, B. H., and P. X. Wu. 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823.

Bhargava, A., L. Franzini, and W. Narendranathan. 1982. Serial correlation and the fixed effects model. *Review of Economic Studies* 49: 533–549.

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.

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

Also see

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

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

[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

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

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

[U] 20 Estimation and postestimation commands