Title

xtreg — Fixed-, between-, and random-effects, and population-averaged linear models

Syntax

GLS random-effects (RE) model

`xtreg` *depvar* [*indepvars*] [*if*] [, `re` *RE_options*]

Between-effects (BE) model

`xtreg` *depvar* [*indepvars*] [*if*] , `be` [*BE_options*]

Fixed-effects (FE) model

`xtreg` *depvar* [*indepvars*] [*if*] [*weight*] , `fe` [*FE_options*]

ML random-effects (MLE) model

`xtreg` *depvar* [*indepvars*] [*if*] [*weight*] , `mle` [*MLE_options*]

Population-averaged (PA) model

`xtreg` *depvar* [*indepvars*] [*if*] [*weight*] , `pa` [*PA_options*]

RE_options	description
Model	
`re`	use random-effects estimator; the default
`sa`	use Swamy–Arora estimator of the variance components
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `conventional`, `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
`nonest`	do not check that panels are nested within clusters
Reporting	
`level(`*#*`)`	set confidence level; default is `level(95)`
`theta`	report θ

BE_options	description
Model	
be	use between-effects estimator
wls	use weighted least squares
SE	
vce(*vcetype*)	*vcetype* may be conventional, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)

FE_options	description
Model	
fe	use fixed-effects estimator
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, robust, cluster *clustvar*, bootstrap, or jackknife
nonest	do not check that panels are nested within clusters
dfadj	adjust the cluster–robust–VCE estimator for the within transform; seldom used
Reporting	
level(#)	set confidence level; default is level(95)

MLE_options	description
Model	
noconstant	suppress constant term
mle	use ML random-effects estimator
SE	
vce(*vcetype*)	*vcetype* may be oim, bootstrap, or jackknife
Reporting	
level(#)	set confidence level; default is level(95)
Max options	
maximize_options	control the maximization process; seldom used

PA_options	description
Model	
noconstant	suppress constant term
pa	use population-averaged estimator
offset(*varname*)	include *varname* in model with coefficient constrained to 1
Correlation	
corr(*correlation*)	within-group correlation structure
force	estimate even if observations unequally spaced in time
SE/Robust	
vce(*vcetype*)	*vcetype* may be conventional, robust, bootstrap, or jackknife
nmp	use divisor $N-P$ instead of the default N
rgf	multiply the robust variance estimate by $(N-1)/(N-P)$
scale(*parm*)	overrides the default scale parameter; *parm* may be x2, dev, phi, or #
Reporting	
level(#)	set confidence level; default is level(95)
Opt options	
optimize_options	control the optimization process; seldom used

correlation	description
exchangeable	exchangeable
independent	independent
unstructured	unstructured
fixed *matname*	user-specified
ar #	autoregressive of order #
stationary #	stationary of order #
nonstationary #	nonstationary of order #

A panel variable must be specified. For xtreg, pa, correlation structures other than exchangeable and independent require that a time variable also be specified. Use xtset; see [XT] **xtset**.

depvar and *indepvars* may contain time-series operators; see [U] **11.4.3 Time-series varlists**.

by, statsby, xi are allowed; see [U] **11.1.10 Prefix commands**.

aweights, fweights, and pweights are allowed for the fixed-effects model. iweights, fweights, and pweights are allowed for the population-averaged model. iweights are allowed for the maximum-likelihood random-effects (MLE) model. See [U] **11.1.6 weight**. Weights must be constant within panel.

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

Description

xtreg fits regression models to panel data. In particular, xtreg with the be option fits random-effects models by using the between regression estimator; with the fe option, it fits fixed-effects models (by using the within regression estimator); and with the re option, it fits random-effects models by using the GLS estimator (producing a matrix-weighted average of the between and within results). See [XT] **xtdata** for a faster way to fit fixed- and random-effects models.

Options for RE model

Model

`re`, the default, requests the GLS random-effects estimator.

`sa` specifies that the small-sample Swamy–Arora estimator individual-level variance component be used instead of the default consistent estimator. See the *Methods and Formulas* section for details.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory, that are robust to some kinds of misspecification, that allow for intragroup correlation, and that use bootstrap or jackknife methods; see [XT] ***vce_options***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

`nonest` removes the check that the panels are nested within clusters.

Reporting

`level(`#`)`; see [XT] **estimation options**.

`theta`, used with `xtreg, re` only, specifies that the output include the estimated value of θ used in combining the between and fixed estimators. For balanced data, this is a constant, and for unbalanced data, a summary of the values is presented in the header of the output.

Options for BE model

Model

`be` requests the between regression estimator.

`wls` specifies that, for unbalanced data, weighted least squares be used rather than the default OLS. Both methods produce consistent estimates. The true variance of the between-effects residual is $\sigma^2_\nu + T_i\sigma^2_\epsilon$ (see *Methods and Formulas* below). WLS produces a "stabilized" variance of $\sigma^2_\nu/T_i + \sigma^2_\epsilon$, which is also not constant. Thus the choice between OLS and WLS amounts to which is more stable.

Comment: `xtreg, be` is rarely used anyway, but between estimates are an ingredient in the random-effects estimate. Our implementation of `xtreg, re` uses the OLS estimates for this ingredient, based on our judgment that σ^2_ν is large relative to σ^2_ϵ in most models. Formally, only a consistent estimate of the between estimates is required.

SE

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that use bootstrap or jackknife methods; see [XT] ***vce_options***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

Reporting

`level(`#`)`; see [XT] **estimation options**.

Options for FE model

Model

fe requests the fixed-effects (within) regression estimator.

SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory, that are robust to some kinds of misspecification, that allow for intragroup correlation, and that use bootstrap or jackknife methods; see [XT] ***vce_options***.

vce(conventional), the default, uses the conventionally derived variance estimator for generalized least-squares regression.

nonest removes the check that the panels are nested within clusters. The cluster–robust–VCE estimator generally assumes that the panels are nested within the clusters or that there are many observations per panel.

dfadj adjusts the cluster–robust–VCE estimator for the within transform. dfadj will produce a conservative VCE when panels are not nested within clusters, even when there are only a few observations per panel.

Reporting

level(*#*); see [XT] **estimation options**.

Options for MLE model

Model

noconstant; see [XT] **estimation options**.

mle requests the maximum-likelihood random-effects estimator.

SE

vce(*vcetype*) specifies the type of standard error reported, which includes types that are derived from asymptotic theory and that use bootstrap or jackknife methods; see [XT] ***vce_options***.

Reporting

level(*#*); see [XT] **estimation options**.

Max options

maximize_options: iterate(*#*), [no]log, trace, tolerance(*#*), ltolerance(*#*), from(*init_specs*); see [R] **maximize**. These options are seldom used.

Options for PA model

Model

noconstant; see [XT] **estimation options**.

`pa` requests the population-averaged estimator. For linear regression, this is the same as a random-effects estimator (both interpretations hold).

`xtreg, pa` is equivalent to `xtgee, family(gaussian) link(id) corr(exchangeable)`, which are the defaults for the `xtgee` command. `xtreg, pa` allows all the relevant `xtgee` options such as `vce(robust)`. Whether you use `xtreg, pa` or `xtgee` makes no difference. See [XT] **xtgee**.

`offset(`*varname*`)`; see [XT] **estimation options**.

Correlation

`corr(`*correlation*`)`, `force`; see [XT] **estimation options**.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory, that are robust to some kinds of misspecification, and that use bootstrap or jackknife methods; see [XT] ***vce_options***.

`vce(conventional)`, the default, uses the conventionally derived variance estimator for generalized least-squares regression.

`nmp`; see [XT] ***vce_options***.

`rgf` specifies that the robust variance estimate is multiplied by $(N-1)/(N-P)$, where N is the total number of observations and P is the number of coefficients estimated. This option can be used with `family(gaussian)` only when `vce(robust)` is either specified or implied by the use of `pweights`. Using this option implies that the robust variance estimate is not invariant to the scale of any weights used.

`scale(x2 | dev | phi | `*#*`)`; see [XT] ***vce_options***.

Reporting

`level(`*#*`)`; see [XT] **estimation options**.

Opt options

optimize_options control the iterative optimization process. These options are seldom used.

`iterate(`*#*`)` specifies the maximum number of iterations. When the number of iterations equals #, the optimization stops and presents the current results, even if convergence has not been reached. The default is `iterate(100)`.

`tolerance(`*#*`)` specifies the tolerance for the coefficient vector. When the relative change in the coefficient vector from one iteration to the next is less than or equal to #, the optimization process is stopped. `tolerance(1e-6)` is the default.

`nolog` suppresses display of the iteration log.

`trace` specifies that the current estimates be printed at each iteration.

Remarks

If you have not read [XT] **xt**, please do so.

See Baltagi (2005, chap. 2) and Wooldridge (2002, chap. 10) for good overviews of fixed-effects and random-effects models.

Consider fitting models of the form

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \nu_i + \epsilon_{it} \tag{1}$$

In this model, $\nu_i + \epsilon_{it}$ is the residual in that we have little interest in it; we want estimates of $\boldsymbol{\beta}$. ν_i is the unit-specific residual; it differs between units, but for any particular unit, its value is constant. In the pulmonary data of [XT] **xt**, a person who exercises less would presumably have a lower FEV year after year and so would have a negative ν_i.

ϵ_{it} is the "usual" residual with the usual properties (mean 0, uncorrelated with itself, uncorrelated with $\mathbf{x}$, uncorrelated with ν, and homoskedastic), although in a more thorough development, we could decompose $\epsilon_{it} = \upsilon_t + \omega_{it}$, assume that ω_{it} is a standard residual, and better describe υ_t.

Before making the assumptions necessary for estimation, let us perform some useful algebra on (1). Whatever the properties of ν_i and ϵ_{it}, if (1) is true, it must also be true that

$$\overline{y}_i = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta} + \nu_i + \overline{\epsilon}_i \tag{2}$$

where $\overline{y}_i = \sum_t y_{it}/T_i$, $\overline{\mathbf{x}}_i = \sum_t \mathbf{x}_{it}/T_i$, and $\overline{\epsilon}_i = \sum_t \epsilon_{it}/T_i$. Subtracting (2) from (1), it must be equally true that

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

These three equations provide the basis for estimating $\boldsymbol{\beta}$. In particular, `xtreg, fe` provides what is known as the fixed-effects estimator—also known as the within estimator—and amounts to using OLS to perform the estimation of (3). `xtreg, be` provides what is known as the between estimator and amounts to using OLS to perform the estimation of (2). `xtreg, re` provides the random-effects estimator and is a (matrix) weighted average of the estimates produced by the between and within estimators. In particular, the random-effects estimator turns out to be equivalent to estimation of

$$(y_{it} - \theta\overline{y}_i) = (1-\theta)\alpha + (\mathbf{x}_{it} - \theta\overline{\mathbf{x}}_i)\boldsymbol{\beta} + \{(1-\theta)\nu_i + (\epsilon_{it} - \theta\overline{\epsilon}_i)\} \tag{4}$$

where θ is a function of σ^2_ν and σ^2_ϵ. If $\sigma^2_\nu = 0$, meaning that ν_i is always 0, $\theta = 0$ and (1) can be estimated by OLS directly. Alternatively, if $\sigma^2_\epsilon = 0$, meaning that ϵ_{it} is 0, $\theta = 1$ and the within estimator returns all the information available (which will, in fact, be a regression with an R^2 of 1).

For more reasonable cases, few assumptions are required to justify the fixed-effects estimator of (3). The estimates are, however, conditional on the sample in that the ν_i are not assumed to have a distribution but are instead treated as fixed and estimable. This statistical fine point can lead to difficulty when making out-of-sample predictions, but that aside, the fixed-effects estimator has much to recommend it.

More is required to justify the between estimator of (2), but the conditioning on the sample is not assumed since $\nu_i + \overline{\epsilon}_i$ is treated as a residual. Newly required is that we assume that ν_i and $\overline{\mathbf{x}}_i$ are uncorrelated. This follows from the assumptions of the OLS estimator but is also transparent: were ν_i and $\overline{\mathbf{x}}_i$ correlated, the estimator could not determine how much of the change in $\overline{y}_i$, associated with an increase in $\overline{\mathbf{x}}_i$, to assign to $\boldsymbol{\beta}$ versus how much to attribute to the unknown correlation. (This, of course, suggests the use of an instrumental-variable estimator, $\overline{\mathbf{z}}_i$, which is correlated with $\overline{\mathbf{x}}_i$ but uncorrelated with ν_i, though that approach is not implemented here.)

The random-effects estimator of (4) requires the same no-correlation assumption. In comparison with the between estimator, the random-effects estimator produces more efficient results, albeit ones with unknown small-sample properties. The between estimator is less efficient because it discards the over-time information in the data in favor of simple means; the random-effects estimator uses both the within and the between information.

All this would seem to leave the between estimator of (2) with no role (except for a minor, technical part it plays in helping to estimate σ_ν^2 and σ_ϵ^2, which are used in the calculation of θ, on which the random-effects estimates depend). Let us, however, consider a variation on (1):

$$y_{it} = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta}_1 + (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta}_2 + \nu_i + \epsilon_{it} \tag{1$'$}$$

In this model, we postulate that changes in the average value of $\mathbf{x}$ for an individual have a different effect from temporary departures from the average. In an economic situation, y might be purchases of some item and $\mathbf{x}$ income; a change in average income should have more effect than a transitory change. In a clinical situation, y might be a physical response and $\mathbf{x}$ the level of a chemical in the brain; the model allows a different response to permanent rather than transitory changes.

The variations of (2) and (3) corresponding to (1$'$) are

$$\overline{y}_i = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta}_1 + \nu_i + \overline{\epsilon}_i \tag{2$'$}$$

$$(y_{it} - \overline{y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta}_2 + (\epsilon_{it} - \overline{\epsilon}_i) \tag{3$'$}$$

That is, the between estimator estimates $\boldsymbol{\beta}_1$ and the within $\boldsymbol{\beta}_2$, and neither estimates the other. Thus even when estimating equations like (1), it is worth comparing the within and between estimators. Differences in results can suggest models like (1$'$), or at the least some other specification error.

Finally, it is worth understanding the role of the between and within estimators with regressors that are constant over time or constant over units. Consider the model

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{s}_i\boldsymbol{\beta}_2 + \mathbf{z}_t\boldsymbol{\beta}_3 + \nu_i + \epsilon_{it} \tag{1$''$}$$

This model is the same as (1), except that we explicitly identify the variables that vary over both time and i ($\mathbf{x}_{it}$, such as output or FEV); variables that are constant over time ($\mathbf{s}_i$, such as race or sex); and variables that vary solely over time ($\mathbf{z}_t$, such as the consumer price index or age in a cohort study). The corresponding between and within equations are

$$\overline{y}_i = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta}_1 + \mathbf{s}_i\boldsymbol{\beta}_2 + \overline{\mathbf{z}}\boldsymbol{\beta}_3 + \nu_i + \overline{\epsilon}_i \tag{2$''$}$$

$$(y_{it} - \overline{y}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta}_1 + (\mathbf{z}_t - \overline{\mathbf{z}})\boldsymbol{\beta}_3 + (\epsilon_{it} - \overline{\epsilon}_i) \tag{3$''$}$$

In the between estimator of (2$''$), no estimate of $\boldsymbol{\beta}_3$ is possible because $\overline{\mathbf{z}}$ is a constant across the i observations; the regression-estimated intercept will be an estimate of $\alpha + \overline{\mathbf{z}}\boldsymbol{\beta}_3$. On the other hand, it can provide estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. It can estimate effects of factors that are constant over time, such as race and sex, but to do so it must assume that ν_i is uncorrelated with those factors.

The within estimator of (3$''$), like the between estimator, provides an estimate of $\boldsymbol{\beta}_1$ but provides no estimate of $\boldsymbol{\beta}_2$ for time-invariant factors. Instead, it provides an estimate of $\boldsymbol{\beta}_3$, the effects of the time-varying factors. The between estimator can also provide estimates u_i for ν_i. More correctly, the estimator u_i is an estimator of $\nu_i + \mathbf{s}_i\boldsymbol{\beta}_2$. Thus u_i is an estimator of ν_i only if there are no time-invariant variables in the model. If there are time-invariant variables, u_i is an estimate of ν_i plus the effects of the time-invariant variables.

Assessing goodness of fit

R^2 is a popular measure of goodness of fit in ordinary regression. In our case, given $\widehat{\alpha}$ and $\widehat{\boldsymbol{\beta}}$ estimates of α and $\boldsymbol{\beta}$, we can assess the goodness of fit with respect to (1), (2), or (3). The prediction equations are, respectively,

$$\widehat{y}_{it} = \widehat{\alpha} + \mathbf{x}_{it}\widehat{\boldsymbol{\beta}} \qquad (1''')$$

$$\widehat{\overline{y}}_i = \widehat{\alpha} + \overline{\mathbf{x}}_i\widehat{\boldsymbol{\beta}} \qquad (2''')$$

$$\widehat{\widetilde{y}}_{it} = (\widehat{y}_{it} - \widehat{\overline{y}}_i) = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\widehat{\boldsymbol{\beta}} \qquad (3''')$$

`xtreg` reports "R-squares" corresponding to these three equations. R-squares is in quotes because the R-squares reported do not have all the properties of the OLS R^2.

The ordinary properties of R^2 include being equal to the squared correlation between $\widehat{y}$ and y and being equal to the fraction of the variation in y explained by $\widehat{y}$—formally defined as $\mathrm{Var}(\widehat{y})/\mathrm{Var}(y)$. The identity of the definitions is from a special property of the OLS estimates; in general, given a prediction $\widehat{y}$ for y, the squared correlation is not equal to the ratio of the variances, and the ratio of the variances is not required to be less than 1.

`xtreg` reports R^2 values calculated as correlations squared, calling them R^2 overall, corresponding to (1'''); R^2 between, corresponding to (2'''); and R^2 within, corresponding to (3'''). In fact, you can think of each of these three numbers as having all the properties of ordinary R^2s if you bear in mind that the prediction being judged is not $\widehat{y}_{it}$, $\widehat{\overline{y}}_i$, and $\widehat{\widetilde{y}}_{it}$, but $\gamma_1\widehat{y}_{it}$ from the regression $y_{it} = \gamma_1\widehat{y}_{it}$; $\gamma_2\widehat{\overline{y}}_i$ from the regression $\overline{y}_i = \gamma_2\widehat{\overline{y}}_i$; and $\gamma_3\widehat{\widetilde{y}}_{it}$ from $\widetilde{y}_{it} = \gamma_3\widehat{\widetilde{y}}_{it}$.

In particular, `xtreg, be` obtains its estimates by performing OLS on (2), and therefore its reported R^2 between is an ordinary R^2. The other two reported R^2s are merely correlations squared, or, if you prefer, R^2s from the second-round regressions $y_{it} = \gamma_{11}\widehat{y}_{it}$ and $\widetilde{y}_{it} = \gamma_{13}\widehat{\widetilde{y}}_{it}$.

`xtreg, fe` obtains its estimates by performing OLS on (3), so its reported R^2 within is an ordinary R^2. As with `be`, the other R^2s are correlations squared, or, if you prefer, R^2s from the second-round regressions $\overline{y_i} = \gamma_{22}\widehat{\overline{y}}_i$ and, as with `be`, $\widetilde{y}_{it} = \gamma_{23}\widehat{\widetilde{y}}_{it}$.

`xtreg, re` obtains its estimates by performing OLS on (4); none of the R^2s corresponding to (1'''), (2'''), or (3''') correspond directly to this estimator (the "relevant" R^2 is the one corresponding to (4)). All three reported R^2s are correlations squared, or, if you prefer, from second-round regressions.

xtreg and associated commands

▷ Example 1: Between-effects model

Using `nlswork.dta` described in [XT] **xt**, we will model `ln_wage` in terms of completed years of schooling (`grade`), current age and age squared, current years worked (experience) and experience squared, current years of tenure on the current job and tenure squared, whether black, whether residing in an area not designated an SMSA (standard metropolitan statistical area), and whether residing in the South. Most of these variables are in the data, but we need to construct a few:

```
. use http://www.stata-press.com/data/r10/nlswork
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)
. generate age2 = age^2
(24 missing values generated)
. generate ttl_exp2 = ttl_exp^2
```

```
. generate tenure2 = tenure^2
(433 missing values generated)
. generate byte black = race==2
```

To obtain the between-effects estimates, we use `xtreg, be`. `nlswork.dta` has previously been `xtset idcode year` because that is what is true of the data, but for running `xtreg`, it would have been sufficient to have `xtset idcode` by itself.

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, be

Between regression (regression on group means)  Number of obs      =     28091
Group variable: idcode                          Number of groups   =      4697

R-sq:  within  = 0.1591                         Obs per group: min =         1
       between = 0.4900                                        avg =       6.0
       overall = 0.3695                                        max =        15

                                                F(10,4686)         =    450.23
sd(u_i + avg(e_i.))=  .3036114                  Prob > F           =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0607602   .0020006    30.37   0.000     .0568382    .0646822
         age |   .0323158   .0087251     3.70   0.000     .0152105    .0494211
        age2 |  -.0005997   .0001429    -4.20   0.000    -.0008799   -.0003194
     ttl_exp |   .0138853   .0056749     2.45   0.014     .0027598    .0250108
    ttl_exp2 |   .0007342   .0003267     2.25   0.025     .0000936    .0013747
      tenure |   .0698419   .0060729    11.50   0.000     .0579361    .0817476
     tenure2 |  -.0028756   .0004098    -7.02   0.000    -.0036789   -.0020722
       black |  -.0564167   .0105131    -5.37   0.000    -.0770272   -.0358061
    not_smsa |  -.1860406   .0112495   -16.54   0.000    -.2080949   -.1639862
       south |  -.0993378    .010136    -9.80   0.000    -.1192091   -.0794665
       _cons |   .3339113   .1210434     2.76   0.006     .0966093    .5712133
------------------------------------------------------------------------------
```

The between-effects regression is estimated on person-averages, so the "`n = 4697`" result is relevant. `xtreg, be` reports the "number of observations" and group-size information: `describe` in [XT] **xt** showed that we have 28,534 "observations"—person-years, really—of data. If we take the subsample that has no missing values in `ln_wage`, `grade`, ..., `south` leaves us with 28,091 observations on person-years, reflecting 4,697 persons, each observed for an average of 5.98 years.

For goodness of fit, the R^2 between is directly relevant; our R^2 is .4900. If, however, we use these estimates to predict the within model, we have an R^2 of .1591. If we use these estimates to fit the overall data, our R^2 is .3695.

The F statistic tests that the coefficients on the regressors `grade`, `age`, ..., `south` are all jointly zero. Our model is significant.

The root mean squared error of the fitted regression, which is an estimate of the standard deviation of $\nu_i + \overline{\epsilon}_i$, is .3036.

For our coefficients, each year of schooling increases hourly wages by 6.1%; age increases wages up to age 26.9 and thereafter decreases them (because the quadratic $ax^2 + bx + c$ turns over at $x = -b/2a$, which for our `age` and `age2` coefficients is $.0323158/(2 \times .0005997) \approx 26.9$); total experience increases wages at an increasing rate (which is surprising and bothersome); tenure on the current job increases wages up to a tenure of 12.1 years and thereafter decreases them; wages of blacks are, these things held constant, (approximately) 5.6% below that of nonblacks (approximately because `black` is an indicator variable); residing in a non-SMSA (rural area) reduces wages by 18.6%; and residing in the South reduces wages by 9.9%.

◁

▷ Example 2: Fixed-effects model

To fit the same model with the fixed-effects estimator, we specify the `fe` option.

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, fe
Fixed-effects (within) regression               Number of obs      =     28091
Group variable: idcode                          Number of groups   =      4697

R-sq:  within  = 0.1727                         Obs per group: min =         1
       between = 0.3505                                        avg =       6.0
       overall = 0.2625                                        max =        15

                                                F(8,23386)         =    610.12
corr(u_i, Xb)  = 0.1936                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |  (dropped)
         age |   .0359987   .0033864    10.63   0.000     .0293611    .0426362
        age2 |  -.000723    .0000533   -13.58   0.000    -.0008274   -.0006186
     ttl_exp |   .0334668   .0029653    11.29   0.000     .0276545     .039279
    ttl_exp2 |   .0002163   .0001277     1.69   0.090    -.0000341    .0004666
      tenure |   .0357539   .0018487    19.34   0.000     .0321303    .0393775
     tenure2 |  -.0019701    .000125   -15.76   0.000    -.0022151   -.0017251
       black |  (dropped)
    not_smsa |  -.0890108   .0095316    -9.34   0.000    -.1076933   -.0703282
       south |  -.0606309   .0109319    -5.55   0.000    -.0820582   -.0392036
       _cons |    1.03732   .0485546    21.36   0.000     .9421497     1.13249
-------------+----------------------------------------------------------------
     sigma_u |  .35562203
     sigma_e |  .29068923
         rho |  .59946283   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(4696,23386) =     5.13          Prob > F = 0.0000
```

The observation summary at the top is the same as for the between-effects model, although this time it is the “`Number of obs`” that is relevant.

Our three R^2s are not too different from those reported previously; the R^2 within is slightly higher (.1727 versus .1591), and the R^2 between is a little lower (.3505 versus .4900), as expected, since the between estimator maximizes R^2 between and the within estimator R^2 within. In terms of overall fit, these estimates are somewhat worse (.2625 versus .3695).

`xtreg, fe` can estimate σ_ν and σ_ϵ, although how you interpret these estimates depends on whether you are using `xtreg` to fit a fixed-effects model or random-effects model. To clarify this fine point, in the fixed-effects model, ν_i are formally fixed—they have no distribution. If you subscribe to this view, think of the reported $\widehat{\sigma}_\nu$ as merely an arithmetic way to describe the range of the estimated but fixed ν_i. If, however, you are using the fixed-effects estimator of the random-effects model, .355622 is an estimate of σ_ν or would be if there were no dropped variables.

Here both `grade` and `black` were dropped from the model because they do not vary over time. Since `grade` and `race` are time invariant, our estimate u_i is an estimate of ν_i plus the effects of `grade` and `race`, so our estimate of the standard deviation is based on the variation in ν_i, `grade`, and `race`. On the other hand, had `race` and `grade` been dropped merely because they were collinear with the other regressors in our model, u_i would be an estimate of ν_i, and .3556 would be an estimate of σ_ν. (`xtsum` and `xttab` allow you to determine whether a variable is time invariant; see [XT] **xtsum** and [XT] **xttab**.)

Regardless of the status of u_i, our estimate of the standard deviation of ϵ_{it} is valid (and, in fact, is the estimate that would be used by the random-effects estimator to produce its results).

Our estimate of the correlation of u_i with $\mathbf{x}_{it}$ suffers from the problem of what u_i measures. We find correlation but cannot say whether this is correlation of ν_i with $\mathbf{x}_{it}$ or merely correlation of `grade` and `race` with $\mathbf{x}_{it}$. In any case, the fixed-effects estimator is robust to such a correlation, and the other estimates it produces are unbiased.

So, although this estimator produces no estimates of the effects of `grade` and `race`, it does predict that age has a positive effect on wages up to age 24.9 years (compared with 26.9 years estimated by the between estimator); that total experience still increases wages at an increasing rate (which is still bothersome); that tenure increases wages up to 9.1 years (compared with 12.1); that living in a non-SMSA reduces wages by 8.9% (compared with a more drastic 18.6%); and that living in the South reduces wages by 6.1% (as compared with 9.9%).

◁

▷ Example 3: Fixed-effects models with robust standard errors

If we suspect that there is heteroskedasticity in the idiosyncratic error term ϵ_{it}, we could specify the `vce(robust)` option:

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, fe vce(robust)

Fixed-effects (within) regression               Number of obs      =     28091
Group variable: idcode                          Number of groups   =      4697

R-sq:  within  = 0.1727                         Obs per group: min =         1
       between = 0.3505                                        avg =       6.0
       overall = 0.2625                                        max =        15

                                                F(8,23386)         =    553.03
corr(u_i, Xb)  = 0.1936                         Prob > F           =    0.0000

------------------------------------------------------------------------------
             |               Robust
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |  (dropped)
         age |   .0359987   .0039755     9.06   0.000     .0282064    .0437909
        age2 |   -.000723   .0000634   -11.40   0.000    -.0008473   -.0005987
     ttl_exp |   .0334668    .003215    10.41   0.000     .0271652    .0397684
    ttl_exp2 |   .0002163    .000141     1.53   0.125    -.0000601    .0004926
      tenure |   .0357539   .0019756    18.10   0.000     .0318817    .0396261
     tenure2 |  -.0019701   .0001362   -14.47   0.000     -.002237   -.0017032
       black |  (dropped)
    not_smsa |  -.0890108   .0113004    -7.88   0.000    -.1111603   -.0668613
       south |  -.0606309    .013096    -4.63   0.000       -.0863   -.0349618
       _cons |    1.03732   .0564117    18.39   0.000     .9267494     1.14789
-------------+----------------------------------------------------------------
     sigma_u |  .35562203
     sigma_e |  .29068923
         rho |  .59946283   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Although the estimated coefficients are the same with and without the `vce(robust)` option, the robust estimator produced larger standard errors and a p-value for `ttl_exp2` above the conventional 10%. The F test of $\nu_i = 0$ is suppressed because it is too difficult to compute the robust form of the statistic when there are more than a few panels.

◁

❑ Technical Note

Clustering on the panel variable produces an estimator of the VCE that is robust to cross-sectional heteroskedasticity and within-panel (serial) correlation that is asymptotically equivalent to that proposed by Arellano (1987). Although the example above applies the fixed-effects estimator, the robust and cluster–robust VCE estimators are also available for the random-effects estimator. Wooldridge (2002) and Arellano (2003) discuss these robust and cluster–robust VCE estimators for the fixed-effects and random-effects estimators. More details are available in *Methods and Formulas*.

❑

▷ Example 4: Random-effects model

Refitting our log-wage model with the random-effects estimator, we obtain

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, re theta

Random-effects GLS regression                   Number of obs      =     28091
Group variable: idcode                          Number of groups   =      4697

R-sq:  within  = 0.1715                         Obs per group: min =         1
       between = 0.4784                                        avg =       6.0
       overall = 0.3708                                        max =        15

Random effects u_i ~ Gaussian                   Wald chi2(10)      =   9244.87
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------- theta --------------------
  min      5%       median        95%      max
0.2520   0.2520     0.5499     0.7016   0.7206

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0646499   .0017811    36.30   0.000     .0611589    .0681408
         age |    .036806   .0031195    11.80   0.000     .0306918    .0429201
        age2 |  -.0007133     .00005   -14.27   0.000    -.0008113   -.0006153
     ttl_exp |   .0290207   .0024219    11.98   0.000     .0242737    .0337676
    ttl_exp2 |   .0003049   .0001162     2.62   0.009      .000077    .0005327
      tenure |    .039252   .0017555    22.36   0.000     .0358114    .0426927
     tenure2 |  -.0020035   .0001193   -16.80   0.000    -.0022373   -.0017697
       black |  -.0530532   .0099924    -5.31   0.000    -.0726379   -.0334685
    not_smsa |  -.1308263   .0071751   -18.23   0.000    -.1448891   -.1167634
       south |  -.0868927   .0073031   -11.90   0.000    -.1012066   -.0725788
       _cons |   .2387209   .0494688     4.83   0.000     .1417639     .335678
-------------+----------------------------------------------------------------
     sigma_u |  .25790313
     sigma_e |  .29069544
         rho |  .44043812   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

According to the R^2s, this estimator performs worse within than the within fixed-effects estimator and worse between than the between estimator, as it must, and slightly better overall.

We estimate that σ_ν is .2579 and σ_ϵ is .2907 and, by assertion, assume that the correlation of ν and **x** is zero.

All that is known about the random-effects estimator is its asymptotic properties, so rather than reporting an F statistic for overall significance, `xtreg, re` reports a χ^2. Taken jointly, our coefficients are significant.

`xtreg, re` also reports a summary of the distribution of θ_i, an ingredient in the estimation of (4). θ is not a constant here because we observe women for unequal periods.

We estimate that schooling has a rate of return of 6.5% (compared with 6.1% between and no estimate within); that the increase of wages with age turns around at 25.8 years (compared with 26.9 between and 24.9 within); that total experience yet again increases wages increasingly; that the effect of job tenure turns around at 9.8 years (compared with 12.1 between and 9.1 within); that being black reduces wages by 5.3% (compared with 5.6% between and no estimate within); that living in a non-SMSA reduces wages 13.1% (compared with 18.6% between and 8.9% within); and that living in the South reduces wages 8.7% (compared with 9.9% between and 6.1% within).
◁

▷ Example 5: Random-effects model fitted using ML

We could also have fitted this random-effects model with the maximum likelihood estimator:

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, mle
Fitting constant-only model:
Iteration 0:   log likelihood = -13690.161
Iteration 1:   log likelihood = -12819.317
Iteration 2:   log likelihood = -12662.039
Iteration 3:   log likelihood = -12649.744
Iteration 4:   log likelihood = -12649.614
Fitting full model:
Iteration 0:   log likelihood =  -8922.145
Iteration 1:   log likelihood = -8853.6409
Iteration 2:   log likelihood = -8853.4255
Iteration 3:   log likelihood = -8853.4254
Random-effects ML regression                    Number of obs      =     28091
Group variable: idcode                          Number of groups   =      4697
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       6.0
                                                               max =        15
                                                LR chi2(10)        =   7592.38
Log likelihood  = -8853.4254                    Prob > chi2        =    0.0000
```

ln_wage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
grade	.0646093	.0017372	37.19	0.000	.0612044	.0680142
age	.0368531	.0031226	11.80	0.000	.030733	.0429732
age2	-.0007132	.0000501	-14.24	0.000	-.0008113	-.000615
ttl_exp	.0288196	.0024143	11.94	0.000	.0240877	.0335515
ttl_exp2	.000309	.0001163	2.66	0.008	.0000811	.0005369
tenure	.0394371	.0017604	22.40	0.000	.0359868	.0428875
tenure2	-.0020052	.0001195	-16.77	0.000	-.0022395	-.0017709
black	-.0533394	.0097338	-5.48	0.000	-.0724172	-.0342615
not_smsa	-.1323433	.0071322	-18.56	0.000	-.1463221	-.1183644
south	-.0875599	.0072143	-12.14	0.000	-.1016998	-.0734201
_cons	.2390837	.0491902	4.86	0.000	.1426727	.3354947
/sigma_u	.2485556	.0035017			.2417863	.2555144
/sigma_e	.2918458	.001352			.289208	.2945076
rho	.4204033	.0074828			.4057959	.4351212

```
Likelihood-ratio test of sigma_u=0: chibar2(01)= 7339.84 Prob>=chibar2 = 0.000
```

The estimates are nearly the same as those produced by `xtreg, re`—the GLS estimator. For instance, `xtreg, re` estimated the coefficient on `grade` to be .0646499, `xtreg, mle` estimated .0646093, and the ratio is $.0646499/.0646093 = 1.001$ to three decimal places. Similarly, the standard errors are nearly equal: $.0017812/.0017372 = 1.025$. Below we compare all 11 coefficients:

Estimator	Coefficient ratio mean	min.	max.	SE ratio mean	min.	max.
xtreg, mle (ML)	1.	1.	1.	1.	1.	1.
xtreg, re (GLS)	.997	.987	1.007	1.006	.997	1.027

◁

▷ Example 6: Population-averaged model

We could also have fitted this model with the population-averaged estimator:

```
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, pa
Iteration 1: tolerance = .0310561
Iteration 2: tolerance = .00074898
Iteration 3: tolerance = .0000147
Iteration 4: tolerance = 2.880e-07

GEE population-averaged model                   Number of obs      =     28091
Group variable:                     idcode      Number of groups   =      4697
Link:                             identity      Obs per group: min =         1
Family:                           Gaussian                     avg =       6.0
Correlation:                  exchangeable                     max =        15
                                                Wald chi2(10)      =   9598.89
Scale parameter:                  .1436709      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0645427   .0016829    38.35   0.000     .0612442    .0678412
         age |    .036932   .0031509    11.72   0.000     .0307564    .0431076
        age2 |  -.0007129   .0000506   -14.10   0.000    -.0008121   -.0006138
     ttl_exp |   .0284878   .0024169    11.79   0.000     .0237508    .0332248
    ttl_exp2 |   .0003158   .0001172     2.69   0.007      .000086    .0005456
      tenure |   .0397468   .0017779    22.36   0.000     .0362621    .0432315
     tenure2 |   -.002008   .0001209   -16.61   0.000    -.0022449   -.0017711
       black |  -.0538314   .0094086    -5.72   0.000     -.072272   -.0353909
    not_smsa |  -.1347788   .0070543   -19.11   0.000    -.1486049   -.1209526
       south |  -.0885969   .0071132   -12.46   0.000    -.1025386   -.0746552
       _cons |   .2396286   .0491465     4.88   0.000     .1433034    .3359539
------------------------------------------------------------------------------
```

These results differ from those produced by `xtreg, re` and `xtreg, mle`. Coefficients are larger and standard errors smaller. `xtreg, pa` is simply another way to run the `xtgee` command. That is, we would have obtained the same output had we typed

```
. xtgee ln_w grade age* ttl_exp* tenure* black not_smsa south
```
(*output omitted because it is the same as above*)

See [XT] **xtgee**. In the language of `xtgee`, the random-effects model corresponds to an `exchangeable` correlation structure and `identity` link, and `xtgee` also allows other correlation structures. Let us stay with the random-effects model, however. `xtgee` will also produce robust estimates of variance, and we refit this model that way by typing

```
. xtgee ln_w grade age* ttl_exp* tenure* black not_smsa south, vce(robust)
```
(*output omitted, coefficients the same, standard errors different*)

In the previous example, we presented a table comparing `xtreg, re` with `xtreg, mle`. Below we add the results from the estimates shown and the ones we did with `xtgee, vce(robust)`:

Estimator		Coefficient ratio mean	min.	max.	SE ratio mean	min.	max.
`xtreg, mle`	(ML)	1.	1.	1.	1.	1.	1.
`xtreg, re`	(GLS)	.997	.987	1.007	1.006	.997	1.027
`xtreg, pa`	(PA)	1.060	.847	1.317	.853	.626	.986
`xtgee, vce(robust)`	(PA)	1.060	.847	1.317	1.306	.957	1.545

So, which are right? This is a real dataset, and we do not know. However, in example 2 in [XT] **xtreg postestimation**, we will present evidence that the assumptions underlying the `xtreg, re` and `xtreg, mle` results are not met.

◁

Acknowledgments

We thank Richard Goldstein, who wrote the first draft of the routine that fits random-effects regressions, and Badi Baltagi of Syracuse University and Manuelita Ureta of Texas A&M University, who assisted us in working our way through the literature.

Saved Results

`xtreg, re` saves the following in `e()`:

Scalars

`e(N)`	number of observations	`e(r2_b)`	R-squared for between model
`e(N_g)`	number of groups	`e(sigma)`	ancillary parameter (`gamma`, `lnormal`)
`e(df_m)`	model degrees of freedom	`e(sigma_u)`	panel-level standard deviation
`e(g_max)`	largest group size	`e(sigma_e)`	standard deviation of ϵ_{it}
`e(g_min)`	smallest group size	`e(rmse)`	root mean squared error of GLS regression
`e(g_avg)`	average group size		
`e(chi2)`	χ^2	`e(thta_min)`	minimum θ
`e(rho)`	ρ	`e(thta_5)`	θ, 5th percentile
`e(Tbar)`	harmonic mean of group sizes	`e(thta_50)`	θ, 50th percentile
`e(Tcon)`	`1` if T is constant	`e(thta_95)`	θ, 95th percentile
`e(r2_w)`	R-squared for within model	`e(thta_max)`	maximum θ
`e(r2_o)`	R-squared for overall model	`e(N_clust)`	number of clusters

Macros

`e(cmd)`	`xtreg`	`e(vcetype)`	title used to label Std. Err.
`e(cmdline)`	command as typed	`e(chi2type)`	`Wald`; type of model χ^2 test
`e(depvar)`	name of dependent variable	`e(sa)`	Swamy–Arora estimator of the variance components (`sa` only)
`e(model)`	`re`		
`e(ivar)`	variable denoting groups	`e(properties)`	`b V`
`e(clustvar)`	name of cluster variable	`e(predict)`	program used to implement `predict`
`e(vce)`	*vcetype* specified in `vce()`		

Matrices

`e(b)`	coefficient vector	`e(Vf)`	VCE for fixed-effects model
`e(theta)`	θ	`e(bf)`	coefficient vector for fixed-effects model
`e(V)`	variance–covariance matrix of the estimators		

Functions

`e(sample)`	marks estimation sample

`xtreg, be` saves the following in `e()`:

Scalars

e(N)	number of observations	e(ll)	log likelihood
e(N_g)	number of groups	e(ll_0)	log likelihood, constant-only model
e(mss)	model sum of squares	e(g_max)	largest group size
e(df_m)	model degrees of freedom	e(g_min)	smallest group size
e(rss)	residual sum of squares	e(g_avg)	average group size
e(df_r)	residual degrees of freedom	e(Tbar)	harmonic mean of group sizes
e(r2)	R-squared	e(Tcon)	1 if T is constant
e(r2_a)	adjusted R-squared	e(r2_w)	R-squared for within model
e(F)	F statistic	e(r2_o)	R-squared for overall model
e(rmse)	root mean squared error	e(r2_b)	R-squared for between model

Macros

e(cmd)	xtreg	e(ivar)	variable denoting groups
e(cmdline)	command as typed	e(vce)	*vcetype* specified in vce()
e(depvar)	name of dependent variable	e(vcetype)	title used to label Std. Err.
e(title)	title in estimation output	e(properties)	b V
e(model)	be	e(predict)	program used to implement predict

Matrices

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

Functions

e(sample)	marks estimation sample

`xtreg, fe` saves the following in `e()`:

Scalars

`e(N)`	number of observations	`e(F)`	F statistic
`e(N_g)`	number of groups	`e(g_max)`	largest group size
`e(mss)`	model sum of squares	`e(g_min)`	smallest group size
`e(tss)`	total sum of squares	`e(g_avg)`	average group size
`e(df_m)`	model degrees of freedom	`e(rho)`	ρ
`e(rss)`	residual sum of squares	`e(Tbar)`	harmonic mean of group sizes
`e(df_r)`	residual degrees of freedom	`e(Tcon)`	`1` if T is constant
`e(r2)`	R-squared	`e(r2_w)`	R-squared for within model
`e(r2_a)`	adjusted R-squared	`e(r2_o)`	R-squared for overall model
`e(sigma)`	ancillary parameter (`gamma`, `lnormal`)	`e(r2_b)`	R-squared for between model
`e(rmse)`	root mean squared error	`e(N_clust)`	number of clusters
`e(ll)`	log likelihood	`e(corr)`	corr(u_i, `Xb`)
`e(ll_0)`	log likelihood, constant-only model	`e(sigma_u)`	panel-level standard deviation
`e(df_a)`	degrees of freedom for absorbed effect	`e(sigma_e)`	standard deviation of ϵ_{it}
`e(df_b)`	numerator degrees of freedom for F statistic	`e(F_f)`	F for u_i=0

Macros

`e(cmd)`	`xtreg`	`e(wexp)`	weight expression
`e(cmdline)`	command as typed	`e(vce)`	*vcetype* specified in `vce()`
`e(depvar)`	name of dependent variable	`e(vcetype)`	title used to label Std. Err.
`e(model)`	`fe`	`e(properties)`	`b V`
`e(ivar)`	variable denoting groups	`e(predict)`	program used to implement `predict`
`e(clustvar)`	name of cluster variable		
`e(wtype)`	weight type		

Matrices

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

Functions

`e(sample)`	marks estimation sample

(*Continued on next page*)

xtreg, mle saves the following in e():

Scalars

e(N)	number of observations	e(g_min)	smallest group size
e(N_g)	number of groups	e(g_avg)	average group size
e(df_m)	model degrees of freedom	e(chi2)	χ^2
e(ll)	log likelihood	e(chi2_c)	χ^2 for comparison test
e(ll_0)	log likelihood, constant-only model	e(rho)	ρ
e(ll_c)	log likelihood, comparison model	e(sigma_u)	panel-level standard deviation
e(g_max)	largest group size	e(sigma_e)	standard deviation of ϵ_{it}

Macros

e(cmd)	xtreg	e(vcetype)	title used to label Std. Err.
e(cmdline)	command as typed	e(chi2type)	Wald or LR; type of model χ^2 test
e(depvar)	name of dependent variable	e(chi2_ct)	Wald or LR; type of model χ^2 test corresponding to e(chi2_c)
e(model)	ml		
e(ivar)	variable denoting groups	e(distrib)	Gaussian; the distribution of the RE
e(wtype)	weight type	e(crittype)	optimization criterion
e(wexp)	weight expression	e(properties)	b V
e(title)	title in estimation output	e(predict)	program used to implement predict
e(vce)	*vcetype* specified in vce()		

Matrices

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

Functions

e(sample)	marks estimation sample

xtreg, pa saves the following in e():

Scalars

e(N)	number of observations	e(df_pear)	degrees of freedom for Pearson χ^2
e(N_g)	number of groups	e(deviance)	deviance
e(df_m)	model degrees of freedom	e(chi2_dev)	χ^2 test of deviance
e(g_max)	largest group size	e(dispers)	deviance dispersion
e(g_min)	smallest group size	e(chi2_dis)	χ^2 test of deviance dispersion
e(g_avg)	average group size	e(tol)	target tolerance
e(rc)	return code	e(dif)	achieved tolerance
e(chi2)	χ^2	e(phi)	scale parameter

Macros

e(cmd)	xtgee	e(scale)	x2, dev, phi, or #; scale parameter
e(cmd2)	xtreg	e(ivar)	variable denoting groups
e(cmdline)	command as typed	e(vce)	*vcetype* specified in vce()
e(depvar)	name of dependent variable	e(vcetype)	title used to label Std. Err.
e(model)	pa	e(chi2type)	Wald; type of model χ^2 test
e(wtype)	weight type	e(disp)	deviance dispersion
e(wexp)	weight expression	e(offset)	offset
e(family)	Gaussian	e(crittype)	optimization criterion
e(link)	identity; link function	e(properties)	b V
e(corr)	correlation structure	e(predict)	program used to implement predict

Matrices

e(b)	coefficient vector	e(V)	variance–covariance matrix of the estimators
e(R)	estimated working correlation matrix		

Functions

e(sample)	marks estimation sample

Methods and Formulas

The model to be fitted is

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \nu_i + \epsilon_{it}$$

for $i = 1, \ldots, n$ and, for each i, $t = 1, \ldots, T$, of which T_i periods are actually observed.

xtreg, fe

xtreg, fe produces estimates by running OLS on

$$(y_{it} - \overline{y}_i + \overline{\overline{y}}) = \alpha + (\mathbf{x}_{it} - \overline{\mathbf{x}}_i + \overline{\overline{\mathbf{x}}})\boldsymbol{\beta} + (\epsilon_{it} - \overline{\epsilon}_i + \overline{\nu}) + \overline{\overline{\epsilon}}$$

where $\overline{y}_i = \sum_{t=1}^{T_i} y_{it}/T_i$, and similarly, $\overline{\overline{y}} = \sum_i \sum_t y_{it}/(nT_i)$. The conventional covariance matrix of the estimators is adjusted for the extra $n-1$ estimated means, so results are the same as using OLS on (1) to estimate ν_i directly. Specifying vce(robust) or vce(cluster *clustvar*) causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [U] **20.15 Obtaining robust variance estimates** and [P] **_robust** for details. Wooldridge (2002) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator.

Clustering on the panel variable produces a consistent VCE estimator when the disturbances are correlated within panels and not identically distributed over the panels. The most common example is within-panel serial-correlation and cross-panel heteroskedasticity.

The cluster–robust–VCE estimator requires that there are many clusters and the disturbances are uncorrelated across the clusters. Usually, the panel variable must be nested within the cluster variable because of the within-panel correlation induced by the within transform.

Care should be taken in using `nonest`, which removes the restriction that the panel variable be nested within the clusters. When using this option with `xtreg, fe`, you will need many observations per panel so that the within-panel correlation caused by the within transform becomes negligible.

The `dfadj` option adjusts the cluster–robust–VCE estimator for the n degrees-of-freedom used in the within transform. Although this option usually causes the standard errors to be too large, it may be useful when the panels are not nested within the clusters and there are few observations per panel.

From the estimates $\widehat{\alpha}$ and $\widehat{\boldsymbol{\beta}}$, estimates u_i of ν_i are obtained as $u_i = \overline{y}_i - \widehat{\alpha} - \overline{\mathbf{x}}_i\widehat{\boldsymbol{\beta}}$. Reported from the calculated u_i are its standard deviation and its correlation with $\overline{\mathbf{x}}_i\widehat{\boldsymbol{\beta}}$. Reported as the standard deviation of e_{it} is the regression's estimated root mean squared error, s, which is adjusted (as previously stated) for the $n-1$ estimated means.

Reported as R^2 within is the R^2 from the mean-deviated regression.

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

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

xtreg, be

`xtreg, be` fits the following model:

$$\overline{y}_i = \alpha + \overline{\mathbf{x}}_i\boldsymbol{\beta} + \nu_i + \overline{\epsilon}_i$$

Estimation is via OLS unless T_i is not constant and the `wls` option is specified. Otherwise, the estimation is performed via WLS. The estimates and conventional VCE are obtained from `regress` for both cases, but for WLS, `[aweight=`T_i`]` is specified.

Reported as R^2 between is the R^2 from the fitted regression.

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

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

xtreg, re

The key to the random-effects estimator is the GLS transform. Given estimates of the idiosyncratic component, $\widehat{\sigma}_e^2$, and the individual component, $\widehat{\sigma}_u^2$, the GLS transform of a variable z for the random-effects model is

$$z_{it}^* = z_{it} - \widehat{\theta}_i\overline{z}_i$$

where $\overline{z}_i = \frac{1}{T_i}\sum_t^{T_i} z_{it}$ and

$$\widehat{\theta}_i = 1 - \sqrt{\frac{\widehat{\sigma}_e^2}{T_i\widehat{\sigma}_u^2 + \widehat{\sigma}_e^2}}$$

Given an estimate of $\widehat{\theta}_i$, one transforms the dependent and independent variables, and then the coefficient estimates and the conventional variance–covariance matrix come from an OLS regression of y_{it}^* on $\mathbf{x}_{it}^*$ and the transformed constant $1-\widehat{\theta}_i$. Specifying `vce(robust)` or `vce(cluster` *clustvar*`)` causes the Huber/White/sandwich VCE estimator to be calculated for the coefficients estimated in this regression. See [U] **20.15 Obtaining robust variance estimates** and [P] **_robust** for details. Wooldridge (2002) and Arellano (2003) discuss this application of the Huber/White/sandwich VCE estimator.

Stata has two implementations of the Swamy–Arora method for estimating the variance components. They produce the same results in balanced panels and share the same estimator of σ_e^2. However, the two methods differ in their estimator of σ_u^2 in unbalanced panels. We call the first $\widehat{\sigma}^2_{u\overline{T}}$ and the second $\widehat{\sigma}^2_{\mathrm{uSA}}$. Both estimators are consistent; however, $\widehat{\sigma}^2_{\mathrm{uSA}}$ has a more elaborate adjustment for small samples than $\widehat{\sigma}^2_{u\overline{T}}$. (See Baltagi [2005], Baltagi and Chang [1994], and Swamy and Arora [1972] for derivations of these methods.)

Both methods use the same function of within residuals to estimate the idiosyncratic error component σ_e. Specifically,

$$\widehat{\sigma}_e^2 = \frac{\sum_i^n \sum_t^{T_i} e_{it}^2}{N - n - K + 1}$$

where

$$e_{it} = (y_{it} - \overline{y}_i + \overline{\overline{y}}) - \widehat{\alpha}_w - (\mathbf{x}_{it} - \overline{\mathbf{x}}_i + \overline{\overline{\mathbf{x}}})\widehat{\boldsymbol{\beta}}_w$$

and $\widehat{\alpha}_w$ and $\widehat{\boldsymbol{\beta}}_w$ are the within estimates of the coefficients and $N = \sum_i^n T_i$. After passing the within residuals through the within transform, only the idiosyncratic errors are left.

The default method for estimating σ_u^2 is

$$\widehat{\sigma}^2_{u\overline{T}} = \max\left\{0, \frac{SSR_b}{n-K} - \frac{\widehat{\sigma}_e^2}{\overline{T}}\right\}$$

where

$$SSR_b = \sum_i^n T_i \left(\overline{y}_i - \widehat{\alpha}_b - \overline{\mathbf{x}}_i\widehat{\boldsymbol{\beta}}_b\right)^2$$

$\widehat{\alpha}_b$ and $\widehat{\boldsymbol{\beta}}_b$ are coefficient estimates from the between regression and $\overline{T}$ is the harmonic mean of T_i:

$$\overline{T} = \frac{n}{\sum_i^n \frac{1}{T_i}}$$

This estimator is consistent for σ_u^2 and is computationally less expensive than the second method. The sum of squared residuals from the between model estimate a function of both the idiosyncratic component and the individual component. Using our estimator of σ_e^2, we can remove the idiosyncratic component, leaving only the desired individual component.

The second method is the Swamy–Arora method for unbalanced panels derived by Baltagi and Chang (1994), which has a more precise small-sample adjustment. Using this method,

$$\widehat{\sigma}^2_{\mathrm{uSA}} = \max\left\{0, \frac{SSR_b - (n-K)\widehat{\sigma}_e^2}{N - tr}\right\}$$

where

$$tr = \text{trace}\left\{(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\right\}$$

$$\mathbf{P} = \text{diag}\left\{\left(\frac{1}{T_i}\right)\boldsymbol{\iota}_{T_i}\boldsymbol{\iota}'_{T_i}\right\}$$

$$\mathbf{Z} = \text{diag}\left[\boldsymbol{\iota}_{T_i}\right]$$

$\mathbf{X}$ is the $N \times K$ matrix of covariates, including the constant, and $\boldsymbol{\iota}_{T_i}$ is a $T_i \times 1$ vector of ones.

The estimated coefficients $(\widehat{\alpha}_r, \widehat{\boldsymbol{\beta}}_r)$ and their covariance matrix $\mathbf{V}_r$ are reported together with the previously calculated quantities $\widehat{\sigma}_e$ and $\widehat{\sigma}_u$. The standard deviation of $\nu_i + e_{it}$ is calculated as $\sqrt{\widehat{\sigma}_e^2 + \widehat{\sigma}_u^2}$.

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

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

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

xtreg, mle

The log likelihood for the ith unit is

$$l_i = -\frac{1}{2}\left(\frac{1}{\sigma_e^2}\left[\sum_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}\boldsymbol{\beta})^2 - \frac{{\sigma_u}^2}{T_i\sigma_u^2 + \sigma_e^2}\left\{\sum_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}\boldsymbol{\beta})\right\}^2\right] + \ln\left(T_i\frac{\sigma_u^2}{\sigma_e^2} + 1\right) + T_i\ln(2\pi\sigma_e^2)\right)$$

The `mle` and `re` options yield essentially the same results, except when total $N = \sum_i T_i$ is small (200 or less) and the data are unbalanced.

xtreg, pa

See [XT] **xtgee** for details on the methods and formulas used to calculate the population-averaged model using a generalized estimating equations approach.

References

Andrews, M., T. Schank, and R. Upward. 2006. Practical fixed-effects estimation methods for the three-way error-components model. *Stata Journal* 6: 461–481.

Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49: 431–434.

———. 2003. *Panel Data Econometrics*. New York: Oxford University Press.

Baltagi, B. H. 1985. Pooling cross-sections with unequal time-series lengths. *Economics Letters* 18: 133–136.

———. 2005. *Econometric Analysis of Panel Data*. 3rd ed. New York: Wiley.

Baltagi, B. H., and Y. Chang. 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics* 62: 67–89.

Baum, C. F. 2001. Residual diagnostics for cross-section time series regression models. *Stata Journal* 1: 101–104.

Blackwell, J. L., III. 2005. Estimation and testing of fixed-effect panel-data systems. *Stata Journal* 5: 202–207.

Bottai, M., and N. Orsini. 2004. Confidence intervals for the variance component of random-effects linear models. *Stata Journal* 4: 429–435.

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.

Dwyer, J., and M. Feinleib. 1992. Introduction to statistical models for longitudinal observation. In *Statistical Models for Longitudinal Studies of Health*, ed. J. Dwyer, M. Feinleib, P. Lippert, and H. Hoffmeister, 3–48. New York: Oxford University Press.

Greene, W. H. 1983. Simultaneous estimation of factor substitution, economies of scale, and non-neutral technical change. In *Econometric Analyses of Productivity*, ed. A. Dogramaci. Boston: Kluwer.

——. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hoyos, R. E. De, and V. Sarafidis. 2006. Testing for cross-sectional dependence in panel-data models. *Stata Journal* 6: 482–496.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: Wiley.

Lee, L., and W. Griffiths. 1979. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. University of New England Working Papers in Econometrics and Applied Statistics No. 1, Armidale, Australia.

Rabe-Hesketh, S., A. Pickles, and C. Taylor. 2000. sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin* 53: 47–57. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 293–307.

Sosa-Escudero, W., and A. K. Bera. 2001. sg164: Specification tests for linear panel data models. *Stata Technical Bulletin* 61: 18–21. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 307–311.

Swamy, P. A. V. B., and S. S. Arora. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40: 643–657.

Taub, A. J. 1979. Prediction in the context of the variance-components model. *Journal of Econometrics* 10: 103–108.

Twisk, J. W. R. 2003. *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge: Cambridge University Press.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Also See

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

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

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

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

[XT] **xtmixed** — Multilevel mixed-effects linear regression

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

[R] **areg** — Linear regression with a large dummy-variable set

[R] **regress** — Linear regression

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

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