```
------------------------------------------------------------------------------
      name:  <unnamed>
       log:  H:\Documents and Settings\Lee\My Documents\Document\stata\panel.log
  log type:  text
 opened on:   9 Feb 2010, 09:34:23

. use http://www.stata-press.com/data/r11/nlswork, clear
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)

. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#
> c.tenure 2.race 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
             |
 c.age#c.age |  -.0005997   .0001429    -4.20   0.000    -.0008799   -.0003194
             |
     ttl_exp |   .0138853   .0056749     2.45   0.014     .0027598    .0250108
             |
   c.ttl_exp#|
   c.ttl_exp |   .0007342   .0003267     2.25   0.025     .0000936    .0013747
             |
      tenure |   .0698419   .0060729    11.50   0.000     .0579361    .0817476
             |
    c.tenure#|
    c.tenure |  -.0028756   .0004098    -7.02   0.000    -.0036789   -.0020722
             |
      2.race |  -.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
------------------------------------------------------------------------------
```

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#
> c.tenure 2.race not_smsa south, fe
note: grade omitted because of collinearity
note: 2.race omitted because of collinearity

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 |  (omitted)
         age |   .0359987   .0033864    10.63   0.000     .0293611    .0426362
             |
 c.age#c.age |   -.000723   .0000533   -13.58   0.000    -.0008274   -.0006186
             |
     ttl_exp |   .0334668   .0029653    11.29   0.000     .0276545     .039279
             |
   c.ttl_exp#|
   c.ttl_exp |   .0002163   .0001277     1.69   0.090    -.0000341    .0004666
             |
      tenure |   .0357539   .0018487    19.34   0.000     .0321303    .0393775
             |
    c.tenure#|
    c.tenure |  -.0019701    .000125   -15.76   0.000    -.0022151   -.0017251
             |
      2.race |  (omitted)
    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     .9421496     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
```

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#
> c.tenure 2.race not_smsa south, fe vce(robust)
note: grade omitted because of collinearity
note: 2.race omitted because of collinearity

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(8,4696)          =    273.86
corr(u_i, Xb)  = 0.1936                         Prob > F           =    0.0000

                              (Std. Err. adjusted for 4697 clusters in idcode)
------------------------------------------------------------------------------
             |               Robust
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |  (omitted)
         age |   .0359987   .0052407     6.87   0.000     .0257243     .046273
             |
 c.age#c.age |   -.000723   .0000845    -8.56   0.000    -.0008887   -.0005573
             |
     ttl_exp |   .0334668    .004069     8.22   0.000     .0254896    .0414439
             |
   c.ttl_exp#|
   c.ttl_exp |   .0002163   .0001763     1.23   0.220    -.0001294    .0005619
             |
      tenure |   .0357539   .0024683    14.49   0.000     .0309148     .040593
             |
    c.tenure#|
    c.tenure |  -.0019701   .0001696   -11.62   0.000    -.0023026   -.0016376
             |
      2.race |  (omitted)
    not_smsa |  -.0890108   .0137629    -6.47   0.000    -.1159926    -.062029
       south |  -.0606309   .0163366    -3.71   0.000    -.0926583   -.0286035
       _cons |    1.03732   .0739644    14.02   0.000     .8923149    1.182325
-------------+----------------------------------------------------------------
     sigma_u |  .35562203
     sigma_e |  .29068923
         rho |  .59946283   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#
> c.tenure 2.race 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)      =   9244.74
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   .0017812    36.30   0.000     .0611589    .0681409
         age |   .0368059   .0031195    11.80   0.000     .0306918    .0429201
             |
 c.age#c.age |  -.0007133     .00005   -14.27   0.000    -.0008113   -.0006153
             |
     ttl_exp |   .0290208    .002422    11.98   0.000     .0242739    .0337678
             |
   c.ttl_exp#|
   c.ttl_exp |   .0003049   .0001162     2.62   0.009      .000077    .0005327
             |
      tenure |   .0392519   .0017554    22.36   0.000     .0358113    .0426925
             |
    c.tenure#|
    c.tenure |  -.0020035   .0001193   -16.80   0.000    -.0022373   -.0017697
             |
      2.race |   -.053053   .0099926    -5.31   0.000    -.0726381   -.0334679
    not_smsa |  -.1308252   .0071751   -18.23   0.000    -.1448881   -.1167622
       south |  -.0868922   .0073032   -11.90   0.000    -.1012062   -.0725781
       _cons |   .2387207    .049469     4.83   0.000     .1417633    .3356781
-------------+----------------------------------------------------------------
     sigma_u |  .25790526
     sigma_e |  .29068923
         rho |  .44045273   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. xtreg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#
> c.tenure 2.race 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
Iteration 5:   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
             |
 c.age#c.age |  -.0007132   .0000501   -14.24   0.000    -.0008113    -.000615
             |
     ttl_exp |   .0288196   .0024143    11.94   0.000     .0240877    .0335515
             |
   c.ttl_exp#|
   c.ttl_exp |    .000309   .0001163     2.66   0.008     .0000811    .0005369
             |
      tenure |   .0394371   .0017604    22.40   0.000     .0359868    .0428875
             |
    c.tenure#|
    c.tenure |  -.0020052   .0001195   -16.77   0.000    -.0022395   -.0017709
             |
      2.race |  -.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
```

```
. reg ln_w grade age c.age#c.age ttl_exp c.ttl_exp#c.ttl_exp  tenure c.tenure#c.
> tenure 2.race not_smsa south, cluster(idcode)

Linear regression                                      Number of obs =   28091
                                                       F( 10,  4696) =  560.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3745
                                                       Root MSE      =  .37797

                               (Std. Err. adjusted for 4697 clusters in idcode)
------------------------------------------------------------------------------
             |               Robust
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0629238    .002076    30.31   0.000     .0588539    .0669937
         age |    .038598   .0049782     7.75   0.000     .0288385    .0483575
             |
 c.age#c.age |  -.0007082   .0000839    -8.44   0.000    -.0008726   -.0005437
             |
     ttl_exp |   .0211279   .0036516     5.79   0.000     .0139691    .0282867
             |
   c.ttl_exp#|
   c.ttl_exp |   .0004473   .0002113     2.12   0.034     .0000332    .0008615
             |
      tenure |   .0473687   .0028918    16.38   0.000     .0416995     .053038
             |
    c.tenure#|
    c.tenure |   -.002027   .0001984   -10.22   0.000    -.0024161    -.001638
             |
      2.race |  -.0699386   .0098805    -7.08   0.000     -.089309   -.0505681
    not_smsa |  -.1720455    .009528   -18.06   0.000    -.1907248   -.1533662
       south |  -.1003387   .0088154   -11.38   0.000     -.117621   -.0830565
       _cons |   .2472833   .0704249     3.51   0.000     .1092175     .385349
------------------------------------------------------------------------------

. log close
      name:  <unnamed>
       log:  H:\Documents and Settings\Lee\My Documents\Document\stata\panel.log
  log type:  text
 closed on:   9 Feb 2010, 09:38:20
```

Assessing goodness of fit

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

$$\widehat{y}_{it} = \widehat{\alpha} + \mathbf{x}_{it}\widehat{\beta} \tag{1'''}$$

$$\widehat{\overline{y}}_i = \widehat{\alpha} + \overline{\mathbf{x}}_i\widehat{\beta} \tag{2'''}$$

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

`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.