STATA Commands for Unobserved Effects Panel Data

John C Frain

November 19, 2007

Contents

1 Introduction

Panel data or cross-sectional timeseries are observations on a panel of i units or cases over t time periods. Most panel data commands start with **`xt`** For an overview type **`help xt`**. These notes present the annotated log of a STATA session demonstrating the use of many of these commands. The data sets used are those used in the STATA cross-sectional time series reference manual. This note should be regarded as an introduction to that manual and to the STATA on-line help files which give comprehensive descriptions of the facilities in STATA for cross-sectional time series analysis.

To obtain the optimum benefit from these notes I would recommend that one should work through the STATA session with a copy of Wooldridge available for reference. The emphasis here is on the implementation of the methods described in Chapter 10 of Wooldridge and no attempt is made to explain the theory set out there. Note the different fonts used for comments (this font), instructions in these comments (**`help xt`**) and for computer input/output (`help xt`).

```
. help xt

-------------------------------------------------------------------------------
help for xt, iis, tis                                          manual:  [XT] xt
                                                              dialogs:  iis  tsset
-------------------------------------------------------------------------------
```

Cross-sectional time-series analysis

```
xt ... [, i(varname) t(varname) ... ]

iis [varname] [, clear]

tis [varname] [, clear]
```

Description

The xt series of commands provide tools for analyzing cross-sectional time-series (panel) datasets:

help xtdes	Describe pattern of xt data
help xtsum	Summarize xt data
help xttab	Tabulate xt data
help xtdata	Faster specification searches with xt data
help xtline	Line plots with xt data
help xtreg	Fixed-, between- and random-effects, and population-averaged linear models
help xtregar	Fixed- and random-effects linear models with an AR(1) disturbance
help xtgls	Panel-data models using GLS
help xtpcse	OLS or Prais-Winsten models with panel-corrected standard errors
help xtrchh	Hildreth-Houck random coefficients models
help xtivreg	Instrumental variables and two-stage least squares for panel-data models
help xtabond	Arellano-Bond linear, dynamic panel data estimator
help xttobit	Random-effects tobit models
help xtintreg	Random-effects interval data regression models
help xtlogit	Fixed-effects, random-effects, & population-averaged logit models
help xtprobit	Random-effects and population-averaged probit models
help xtcloglog	Random-effects and population-averaged cloglog models
help xtpoisson	Fixed-effects, random-effects, & population-averaged Poisson models
help xtnbreg	Fixed-effects, random-effects, & population-averaged negative binomial models
help xtgee	Population-averaged panel-data models using GEE

Each observation in a cross-sectional time-series (xt) dataset is an observation on x for unit i at time t.

iis is related to the i() option of the other xt commands. Command iis or option i() sets the name of the variable corresponding to index i.

tis is similarly related to the t() option. Command tis or option t() sets the name of the variable corresponding to index t.

Some xt commands use time-series operators in their internal calculations and thus require that your data be tsset; see help tsset. For instance, since xtabond uses time-series operators in its internal calculations, you must tsset your data before using it. The particular help file will indicate if tsset is required for the command.

Options

i(varname) specifies the variable name corresponding to index i. This must be a single, numeric variable, although whether it takes on the values 1, 2, 3 or 1, 7, 9, etc., is irrelevant. (If the identifying variable is a string, use egen's group() function to make a numeric variable; see help egen.)

t(varname) specifies the variable name corresponding to index t. This must be a single, numeric variable.

clear removes the definition of i or t. For instance, typing "tis, clear" makes Stata forget the identity of the t() variable.

Remarks

Once i() and t() have been specified, either by option or by the iis and tis commands, they need not be specified again except to change the variable's identity.

iis and tis, without arguments, list the current name of the index variable.

Example

An xt dataset:

pid	yr_visit	fev	age	sex	height	smokes
1071	1991	1.21	25	1	69	0
1071	1992	1.52	26	1	69	0
1071	1993	1.32	28	1	68	0
1072	1991	1.33	18	1	71	1
1072	1992	1.18	20	1	71	1

```
        1072    1993   1.19   21   1      71       0

    The other xt commands need to know the identities of the variables
    identifying patient and time.  You could type

        . iis pid
        . tis yr_visit


Also see

    Manual:  [XT] intro,
             [XT] xt

    Online:  help for xtabond, xtcloglog, xtdata, xtdes, xtgee, xtgls,
             xtintreg, xtivreg, xtline, xtlogit, xtnbreg, xtpcse, xtpoisson,
             xtprobit, xtrchh, xtreg, xtregar, xtsum, xttab, xttobit; tsset
```

Load the data set nlswork.dta

```
. use nlswork, clear
. describe

Contains data                                     National Longitudinal Survey.
                                                    Young Women 14-26 years of age
                                                    in 1968
  obs:        28,534                              18 Feb 2005 22:17
 vars:            21
 size:     1,055,758
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
idcode          int    %8.0g                  NLS id
year            byte   %8.0g                  interview year
birth_yr        byte   %8.0g                  birth year
age             byte   %8.0g                  age in current year
race            byte   %8.0g                  1=white, 2=black, 3=other
msp             byte   %8.0g                  1 if married, spouse present
nev_mar         byte   %8.0g                  1 if never yet married
grade           byte   %8.0g                  current grade completed
collgrad        byte   %8.0g                  1 if college graduate
not_smsa        byte   %8.0g                  1 if not SMSA
c_city          byte   %8.0g                  1 if central city
south           byte   %8.0g                  1 if south
ind_code        byte   %8.0g                  industry of employment
occ_code        byte   %8.0g                  occupation
union           byte   %8.0g                  1 if union
wks_ue          byte   %8.0g                  weeks unemployed last year
```

```
ttl_exp         float  %9.0g                 total work experience
tenure          float  %9.0g                 job tenure, in years
hours           int    %8.0g                 usual hours worked
wks_work        int    %8.0g                 weeks worked last year
ln_wage         float  %9.0g                 ln(wage/GNP deflator)
-------------------------------------------------------------------------------
Sorted by:  idcode  year
```

To start one must set the indices i (units) and t (time). As already described this can be done using the **iis tis** commands, **i() t()** options or the **tsset** command. Examples of the commands follow.

```
. iis idcode

. tis year

.
. iis
i() is idcode

. tis
t() is year

. iis, clear

. iis
(i() has not been defined)

.
. tsset idcode year
       panel variable:  idcode, 1 to 5159
        time variable:  year, 68 to 88, but with gaps

. tsset
       panel variable:  idcode, 1 to 5159
        time variable:  year, 68 to 88, but with gaps
```

xtdes describes the participation pattern of panel data. We have 4711 women in the survey. The maximum number of years over which any women is observed is 15. the most common pattern is participation in only the first year (136 or 2.89% are observed in this pattern). The bottom line of the table give the totals for participation patterns not observed.

```
. xtdes

  idcode:  1, 2, ..., 5159                                n =       4711
    year:  68, 69, ..., 88                                T =         15
           Delta(year) = 1; (88-68)+1 = 21
           (idcode*year uniquely identifies each observation)
```

```
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       3         5         9      13      15

     Freq.  Percent    Cum. |  Pattern
 ----------------------------+----------------------
       136      2.89    2.89 |  1...................
       114      2.42    5.31 |  .....................1
        89      1.89    7.20 |  ...................1.11
        87      1.85    9.04 |  .....................11
        86      1.83   10.87 |  111111.1.11.1.11.1.11
        61      1.29   12.16 |  ..............11.1.11
        56      1.19   13.35 |  11...................
        54      1.15   14.50 |  ..............1.1.11
        54      1.15   15.64 |  .......1.11.1.11.1.11
      3974     84.36  100.00 | (other patterns)
 ----------------------------+----------------------
      4711    100.00         |  XXXXXX.X.XX.X.XX.X.XX

. xtdes, pattern(20)

  idcode:  1, 2, ..., 5159                                n =       4711
    year:  68, 69, ..., 88                                T =         15
           Delta(year) = 1; (88-68)+1 = 21
           (idcode*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       3         5         9      13      15

     Freq.  Percent    Cum. |  Pattern
 ----------------------------+----------------------
       136      2.89    2.89 |  1...................
       114      2.42    5.31 |  .....................1
        89      1.89    7.20 |  ...................1.11
        87      1.85    9.04 |  .....................11
        86      1.83   10.87 |  111111.1.11.1.11.1.11
        61      1.29   12.16 |  ..............11.1.11
        56      1.19   13.35 |  11...................
        54      1.15   14.50 |  ..............1.1.11
        54      1.15   15.64 |  .......1.11.1.11.1.11
        49      1.04   16.68 |  .........11.1.11.1.11
        45      0.96   17.64 |  ............1.11.1.11
        43      0.91   18.55 |  1111................
        42      0.89   19.44 |  ...1................
        40      0.85   20.29 |  .....1.1.11.1.11.1.11
        38      0.81   21.10 |  ....11.1.11.1.11.1.11
        38      0.81   21.91 |  111..................
        34      0.72   22.63 |  ..1111.1.11.1.11.1.11
        31      0.66   23.29 |  ................1...
        30      0.64   23.92 |  ..........1.1.11.1.11
        29      0.62   24.54 |  ...111.1.11.1.11.1.11
```

```
     3555     75.46  100.00 | (other patterns)
 ---------------------------+----------------------
     4711    100.00         |  XXXXXX.X.XX.X.XX.X.XX
```

xtsum generalizes summarize by reporting means and standard for panel data. It differs from summarize in that it decomposes the standard deviation into between and within components.

```
. summ hours

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       hours |     28467    36.55956    9.869623          1        168

. xtsum hours

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+-----------------------------------------+----------------
hours    overall |  36.55956   9.869623          1        168 |     N =   28467
         between |             7.846585          1       83.5 |     n =    4710
         within  |             7.520712  -2.154726   130.0596 | T-bar = 6.04395

. xtsum birth_yr  /* Time invariant variable */

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+-----------------------------------------+----------------
birth_yr overall |  48.08509   3.012837         41         54 |     N =   28534
         between |             3.051795         41         54 |     n =    4711
         within  |                   0   48.08509   48.08509 | T-bar = 6.05689
```

xttab generalizes tabulate by performing one-way tabulations and by decomposing counts into between and within components in panel data.

```
. summ msp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         msp |     28518    .6029175    .4893019          0          1

. tab msp

      1 if |
   married, |
     spouse |
    present |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |     11,324       39.71       39.71
          1 |     17,194       60.29      100.00
------------+-----------------------------------
```

```
      Total |     28,518      100.00

. xttab msp

                  Overall             Between            Within
      msp |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |   11324     39.71      3113     66.08          55.06
        1 |   17194     60.29      3643     77.33          71.90
----------+-----------------------------------------------------
    Total |   28518    100.00      6756    143.41          64.14
                             (n = 4711)
```

xttrans is another generalization of tabulate. It reports changes in a single categorical variable over time.

```
. xttrans msp

      1 if |
  married, | 1 if married, spouse
    spouse |        present
   present |         0          1 |     Total
-----------+----------------------+----------
         0 |     80.49      19.51 |    100.00
         1 |      7.96      92.04 |    100.00
-----------+----------------------+----------
     Total |     37.11      62.89 |    100.00


. xttrans msp, freq /* Does not normalize for missing time periods */

      1 if |
  married, | 1 if married, spouse
    spouse |        present
   present |         0          1 |     Total
-----------+----------------------+----------
         0 |     7,697      1,866 |     9,563
           |     80.49      19.51 |    100.00
-----------+----------------------+----------
         1 |     1,133     13,100 |    14,233
           |      7.96      92.04 |    100.00
-----------+----------------------+----------
     Total |     8,830     14,966 |    23,796
           |     37.11      62.89 |    100.00


.
. * Rectangularize the data
. fillin idcode year
```

```
. xttrans msp, freq

     1 if |
  married, | 1 if married, spouse
    spouse |        present
   present |         0          1 |     Total
-----------+----------------------+----------
         0 |     6,792      1,446 |     8,238
           |     82.45      17.55 |    100.00
-----------+----------------------+----------
         1 |       813     10,954 |    11,767
           |      6.91      93.09 |    100.00
-----------+----------------------+----------
     Total |     7,605     12,400 |    20,005
           |     38.02      61.98 |    100.00
```

xtline draws line plots for panel data.

2 Estimation using xtreg

The basic linear unobserved effects panel data model may is

$$y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + c_i + u_{it} \tag{1}$$

(For a full explanation of the symbols see Wooldridge page 251, etc.). In equation 1 u_i is the unit specific residual and differs *between* units but not across time *within* units. Averaging equation 1 over time we get

$$\bar{y_i} = \bar{\mathbf{X}}_i\boldsymbol{\beta} + c_i + \bar{u_i} \tag{2}$$

Subtracting equation 2 from equation 1 gives equation 3

$$(y_{it} - \bar{y_i}) = (\mathbf{X}_{it} - \bar{\mathbf{X}_i})\boldsymbol{\beta} + (u_{it} - \bar{u_i}) \tag{3}$$

These three equations form the basis for the various ways of estimating $\boldsymbol{\beta}$.

xtreg ...,fe gives the fixed effects or within estimator of $\boldsymbol{\beta}$ and is derived from equation 3. It is equivalent to performing *OLS* on equation pd3.

xtreg ...,be gives the between effects and corresponds to *OLS* estimation of equation 2.

xtreg ...,re gives the random effects estimator and is a weighted average of the within and between effects estimator. The random effects estimator is equivalent to estimating

$$(y_{it} - \theta\bar{y_i}) = (\mathbf{X}_{it} - \theta\bar{\mathbf{X}_i})\boldsymbol{\beta} + (1-\theta)c_i + (u_{it} - \theta\bar{u_i}) \tag{4}$$

where θ is a function of σ_c^2 and σ_u^2.

xtreg ...,mle produces maximum liklihood estimates of the random effects estimator.

For other options available with the **xtreg** command see the on-line help files or the STATA manuals.

```
. tsset idcode year
       panel variable:  idcode, 1 to 5159
        time variable:  year, 68 to 88, but with gaps

.
. qui gen age2 = age^2

. qui gen ttl_exp2 = ttl_exp^2

. qui gen tenure2 = tenure^2

. gen byte black = race==2

.
. * OLS
. regress ln_w grade age* ttl_exp* tenure* black not_smsa south

      Source |       SS       df       MS              Number of obs =   28091
-------------+------------------------------           F( 10, 28080) = 1681.47
       Model |  2402.22796    10  240.222796           Prob > F      =  0.0000
    Residual |  4011.63592 28080  .142864527           R-squared     =  0.3745
-------------+------------------------------           Adj R-squared =  0.3743
       Total |  6413.86388 28090  .228332641           Root MSE      =  .37797

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0629238   .0010313    61.01   0.000     .0609024    .0649452
         age |    .038598    .003467    11.13   0.000     .0318025    .0453935
        age2 |  -.0007082   .0000563   -12.57   0.000    -.0008186   -.0005978
     ttl_exp |   .0211279    .002335     9.05   0.000     .0165511    .0257046
    ttl_exp2 |   .0004473   .0001246     3.59   0.000     .0002031    .0006916
      tenure |   .0473687   .0019626    24.14   0.000     .0435219    .0512156
     tenure2 |   -.002027   .0001338   -15.15   0.000    -.0022893   -.0017648
       black |  -.0699386   .0053207   -13.14   0.000    -.0803673   -.0595098
    not_smsa |  -.1720455   .0051675   -33.29   0.000     -.182174    -.161917
       south |  -.1003387   .0048938   -20.50   0.000    -.1099308   -.0907467
       _cons |   .2472833   .0493319     5.01   0.000     .1505903    .3439762
------------------------------------------------------------------------------

.
. * Fixed-effects model (within-group estimator)
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, fe

Fixed-effects (within) regression               Number of obs      =     28091
Group variable (i): 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
```

Version 8 of STATA does not have a robust option on the `xtreg` command. (This option is available in version 9. In version 8 use the `areg` command.

```
. areg  ln_w grade age* ttl_exp* tenure* black not_smsa south, ///
        absorb(idcode) cluster(idcode)

Regression with robust standard errors                Number of obs =   28091
                                                      F(  9,  4696) =  202.74
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.6919
                                                      Adj R-squared =  0.6299
                                                      Root MSE      =   .2907

                          (standard errors adjusted for clustering on idcode)
------------------------------------------------------------------------------
             |               Robust
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |  -.0147384   237927.9    -0.00   1.000    -466450.3    466450.3
         age |   .0359987    .005743     6.27   0.000     .0247397    .0472576
        age2 |   -.000723   .0000926    -7.81   0.000    -.0009045   -.0005414
     ttl_exp |   .0334668   .0044589     7.51   0.000     .0247251    .0422084
    ttl_exp2 |   .0002163   .0001932     1.12   0.263    -.0001625    .0005951
      tenure |   .0357539   .0027049    13.22   0.000     .0304511    .0410567
```

```
     tenure2 |  -.0019701   .0001859   -10.60   0.000    -.0023345   -.0016057
       black |  (dropped)
    not_smsa |  -.0890108   .0150819    -5.90   0.000    -.1185784   -.0594432
       south |  -.0606309   .0179022    -3.39   0.001    -.0957276   -.0255341
       _cons |   1.222086    2982750     0.00   1.000     -5847589     5847592
-------------+----------------------------------------------------------------
      idcode |   absorbed                                     (4697 categories)
```

Note that STATA has no direct command for two way fixed effects. If you wish to also introduce a second set of fixed effects for, say, time periods create a set of appropriate dummy variables for inclusion in your regressions and use a one way estimator.

```
. * Between-group estimator
. 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 (i): 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
------------------------------------------------------------------------------


.
. * Random-effects model (GLS estimator)
. xtreg ln_w grade age* ttl_exp* tenure* black not_smsa south, re

Random-effects GLS regression                   Number of obs      =     28091
Group variable (i): 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

------------------------------------------------------------------------------
     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)
------------------------------------------------------------------------------

. estimates hold re

.
. * Random-effects model (Gaussian ML or fully iterated GLS 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 (i): 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    70.98   0.000     .2416925    .2554187
    /sigma_e |   .2918458    .001352   215.87   0.000      .289196    .2944956
-------------+----------------------------------------------------------------
         rho |   .4204033   .0074828                     .4057959    .4351212
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)= 7339.84 Prob>=chibar2 = 0.000
```

3 Testing after `xtreg`

```
*

. /* After xtreg, re */
.
. estimates unhold re

.
. * Breusch & Pagan score test for random effects
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

        ln_wage[idcode,t] = Xb + u[idcode] + e[idcode,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                 ---------+-----------------------------
                  ln_wage |   .2283326       .4778416
                        e |   .0845038       .2906954
                        u |    .066514       .2579031
```

```
        Test:   Var(u) = 0
                                chi2(1) = 14779.98
                              Prob > chi2 =     0.0000

.
. * Hausman specification test (compares fe and re)

.qui xtreg ln_wage  grade age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south,fe

    F(4696, 23386) =     5.19         Prob > F = 0.0000

. estimates store fe

. qui xtreg ln_wage  grade age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south, re

. estimates store re

. hausman fe re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
         age |    .0359987     .0363062       -.0003075        .0013183
        age2 |    -.000723     -.000705        -.000018        .0000184
     ttl_exp |    .0334668     .0292321        .0042347        .0017085
    ttl_exp2 |    .0002163     .0002946       -.0000783        .0000529
      tenure |    .0357539     .0390983       -.0033444        .0005789
     tenure2 |   -.0019701    -.0020014        .0000313        .0000372
    not_smsa |   -.0890108    -.1268961        .0378853        .0063038
       south |   -.0606309     -.094716        .0340851         .008259
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      142.53
                Prob>chi2 =      0.0000


.
```

4 Prediction after `xtreg`

```
.
. * Syntax:
```

```
. * predict [type] newvarname [if exp] [in range] [, statistic nooffset]
. * where statistic is:
. *    xb     fitted values (the default)
. *    stdp   standard error of the fitted values
. *    ue     the combined residuals
. *    xbu    prediction, including effect
. *    u      the fixed effect component
. *    e      the random error component
.
. predict xb  /* computes the linear predictor (the default) */
(option xb assumed; fitted values)
(443 missing values generated)

. predict stdp, stdp
(443 missing values generated)

.
```

5 Faster estimation of alternative models using `xtdata`

`xtdata varlist ...` produces a converted data set of the variables specified or, if varlist is not specified, all the variables in the data. Once converted, Stata's ordinary regress command may be used to perform various panel data regressions more quickly than use **`xtreg`**. Before using **`xdata`** you must eliminate any variables that you do not intend to use and that have missing values. After converting the data, with **`xdata`** you may form linear transformations of the regressors. All nonlinear transformations of the data must be done before conversion. .

```
. xtdata ln_w grade age* ttl_exp* tenure* black not_smsa south, fe clear

. regress ln_w grade age ttl_exp tenure black not_smsa south

      Source |       SS       df       MS              Number of obs =   28091
-------------+------------------------------           F(  6, 28084) =  820.44
       Model |  356.233455     6  59.3722424           Prob > F      =  0.0000
    Residual |  2032.33275 28084  .072366214           R-squared     =  0.1491
-------------+------------------------------           Adj R-squared =  0.1490
       Total |   2388.5662 28090  .085032617           Root MSE      =  .26901

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |   .0375399   5.04e+08     0.00   1.000    -9.88e+08    9.88e+08
         age |  -.0026787   .0007876    -3.40   0.001    -.0042224   -.0011349
     ttl_exp |   .0287709   .0013209    21.78   0.000     .0261819    .0313599
      tenure |   .0114355   .0008422    13.58   0.000     .0097847    .0130863
       black |  (dropped)
    not_smsa |  -.0921689   .0088194   -10.45   0.000    -.1094553   -.0748825
       south |  -.0633396   .0101132    -6.26   0.000     -.083162   -.0435172
```

```
       _cons |   1.121064   6.32e+09     0.00   1.000    -1.24e+10    1.24e+10
------------------------------------------------------------------------------

. regress ln_w grade age* ttl_exp* tenure* black not_smsa south

      Source |       SS       df       MS              Number of obs =   28091
-------------+------------------------------           F(  9, 28081) =  651.21
       Model |  412.443881     9  45.8270979           Prob > F      =  0.0000
    Residual |  1976.12232 28081   .07037222           R-squared     =  0.1727
-------------+------------------------------           Adj R-squared =  0.1724
       Total |   2388.5662 28090  .085032617           Root MSE      =  .26528

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grade |  -.0147384   4.97e+08    -0.00   1.000    -9.75e+08    9.75e+08
         age |   .0359987   .0030904    11.65   0.000     .0299414    .0420559
        age2 |   -.000723   .0000486   -14.88   0.000    -.0008183   -.0006277
     ttl_exp |   .0334668   .0027061    12.37   0.000     .0281626    .0387709
    ttl_exp2 |   .0002163   .0001166     1.86   0.064    -.0000122    .0004448
      tenure |   .0357539   .0016871    21.19   0.000     .0324471    .0390607
     tenure2 |  -.0019701   .0001141   -17.27   0.000    -.0021937   -.0017465
       black |  (dropped)
    not_smsa |  -.0890108   .0086984   -10.23   0.000      -.10606   -.0719616
       south |  -.0606309   .0099763    -6.08   0.000    -.0801849   -.0410769
       _cons |   1.222086   6.23e+09     0.00   1.000    -1.22e+10    1.22e+10
------------------------------------------------------------------------------

.
. xtdata ln_w grade age* ttl_exp* tenure* black not_smsa south, re ratio(.95) clear

------------------- theta --------------------
  min      5%       median        95%      max
0.2750   0.2750     0.5741     0.7198   0.7377

. * (ratio is the ratio of the std. dev. of the individual effect and the
. * random error)
. regress ln_w constant grade age ttl_exp tenure black not_smsa south, noconstant

      Source |       SS       df       MS              Number of obs =   28091
-------------+------------------------------           F(  6, 28085) =27121.45
       Model |  11775.6413     6  1962.60688           Prob > F      =  0.0000
    Residual |  2032.33275 28085  .072363637           R-squared     =  0.8528
-------------+------------------------------           Adj R-squared =  0.8528
       Total |   13807.974 28091  .491544411           Root MSE      =    .269

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    constant |  (dropped)
```

```
       grade |   .1269649   .0013954    90.99   0.000     .1242299    .1296999
         age |  -.0026787   .0007876    -3.40   0.001    -.0042224    -.001135
     ttl_exp |   .0287709   .0013209    21.78   0.000      .026182    .0313599
      tenure |   .0114355   .0008422    13.58   0.000     .0097847    .0130863
       black |  (dropped)
    not_smsa |  -.0921689   .0088192   -10.45   0.000     -.109455   -.0748828
       south |  -.0633396    .010113    -6.26   0.000    -.0831616   -.0435175
------------------------------------------------------------------------------

. regress ln_w constant grade age* ttl_exp* tenure* black not_smsa south, noconstant

      Source |       SS       df       MS              Number of obs =   28091
-------------+------------------------------           F(  9, 28082) =18682.05
       Model |  11831.8517     9  1314.65019           Prob > F      =  0.0000
    Residual |  1976.12232 28082  .070369714           R-squared     =  0.8569
-------------+------------------------------           Adj R-squared =  0.8568
       Total |   13807.974 28091  .491544411           Root MSE      =  .26527

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    constant |  (dropped)
       grade |   .0827449   .0035478    23.32   0.000     .0757911    .0896987
         age |   .0359987   .0030903    11.65   0.000     .0299415    .0420558
        age2 |   -.000723   .0000486   -14.88   0.000    -.0008183   -.0006277
     ttl_exp |   .0334668   .0027061    12.37   0.000     .0281627    .0387708
    ttl_exp2 |   .0002163   .0001166     1.86   0.064    -.0000122    .0004447
      tenure |   .0357539   .0016871    21.19   0.000     .0324472    .0390606
     tenure2 |  -.0019701   .0001141   -17.27   0.000    -.0021937   -.0017465
       black |  (dropped)
    not_smsa |  -.0890108   .0086982   -10.23   0.000    -.1060597   -.0719619
       south |  -.0606309   .0099761    -6.08   0.000    -.0801845   -.0410772
------------------------------------------------------------------------------
```

6 More general error structures

xtpcse and **xtgls** estimate linear panel data models using feasible GLS. **xtpcse** computes *OLS* estimates with panel-corrected standard errors, while **xtgls** computes feasible *GLS* estimates. These commands allow estimation in the presence of AR(1) autocorrelation within panels, as well as heteroscedasticity or cross-sectional correlation across panels. In the case of cross-sectional correlation, **xtgls** requires $T > n$.

```
. use invest2, clear

. tsset company time
       panel variable:  company, 1 to 5
        time variable:  time, 1 to 20
```

```
.
. * OLS with panel-corrected standard errors
. xtpcse invest market stock  /* Heterosk. and contemp. correlation (the default) */

Linear regression, correlated panels corrected standard errors (PCSEs)

Group variable:   company                       Number of obs      =       100
Time variable:    time                          Number of groups   =         5
Panels:           correlated (balanced)         Obs per group: min =        20
Autocorrelation:  no autocorrelation                           avg =        20
                                                               max =        20
Estimated covariances      =        15          R-squared          =    0.7789
Estimated autocorrelations =         0          Wald chi2(2)       =    755.43
Estimated coefficients     =         3          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .1050854   .0083183    12.63   0.000     .0887818    .1213891
       stock |   .3053655   .0330427     9.24   0.000      .240603    .3701281
       _cons |  -48.02974   10.81437    -4.44   0.000     -69.2255   -26.83397
------------------------------------------------------------------------------

. xtpcse invest market stock, corr(ar1) /* Heterosk., contemp. correlation and AR(1) autoc
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable:   company                       Number of obs      =       100
Time variable:    time                          Number of groups   =         5
Panels:           correlated (balanced)         Obs per group: min =        20
Autocorrelation:  common AR(1)                                 avg =        20
                                                               max =        20
Estimated covariances      =        15          R-squared          =    0.5909
Estimated autocorrelations =         1          Wald chi2(2)       =    124.32
Estimated coefficients     =         3          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |    .093367   .0125705     7.43   0.000     .0687294    .1180046
       stock |    .354706   .0571221     6.21   0.000     .2427486    .4666633
       _cons |  -39.39866   40.22722    -0.98   0.327    -118.2426    39.44524
-------------+----------------------------------------------------------------
         rho |   .8530976
------------------------------------------------------------------------------

. xtpcse invest market stock, corr(psar1) rhotype(tscorr)
```

```
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable:   company                       Number of obs      =       100
Time variable:    time                          Number of groups   =         5
Panels:           correlated (balanced)         Obs per group: min =        20
Autocorrelation:  panel-specific AR(1)                         avg =        20
                                                               max =        20
Estimated covariances      =        15          R-squared          =    0.8734
Estimated autocorrelations =         5          Wald chi2(2)       =    483.87
Estimated coefficients     =         3          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .0976686    .009442    10.34   0.000     .0791627    .1161746
       stock |   .3726526   .0384121     9.70   0.000     .2973662     .447939
       _cons |  -46.95183   16.78803    -2.80   0.005    -79.85576    -14.0479
------------------------------------------------------------------------------
        rhos =  .4735903   .704354  .8977688  .5249498  .8558518
------------------------------------------------------------------------------

. xtpcse invest market stock, hetonly  /* Heterosk., no contemp. correlation */

Linear regression, heteroskedastic panels corrected standard errors

Group variable:   company                       Number of obs      =       100
Time variable:    time                          Number of groups   =         5
Panels:           heteroskedastic (balanced)    Obs per group: min =        20
Autocorrelation:  no autocorrelation                           avg =        20
                                                               max =        20
Estimated covariances      =         5          R-squared          =    0.7789
Estimated autocorrelations =         0          Wald chi2(2)       =    720.01
Estimated coefficients     =         3          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |            Het-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .1050854   .0090625    11.60   0.000     .0873232    .1228476
       stock |   .3053655   .0409468     7.46   0.000     .2251113    .3856198
       _cons |  -48.02974   14.20367    -3.38   0.001    -75.86841   -20.19106
------------------------------------------------------------------------------

. xtpcse invest market stock, hetonly corr(ar1)  /* Heterosk. and AR(1) autocorr., no cont
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, heteroskedastic panels corrected standard errors
```

```
Group variable:   company                         Number of obs      =       100
Time variable:    time                            Number of groups   =         5
Panels:           heteroskedastic (balanced)      Obs per group: min =        20
Autocorrelation:  common AR(1)                                   avg =        20
                                                                 max =        20
Estimated covariances      =         5          R-squared          =    0.5909
Estimated autocorrelations =         1          Wald chi2(2)       =    120.57
Estimated coefficients     =         3          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |            Het-corrected
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |    .093367   .0128727     7.25   0.000     .0681369    .1185971
       stock |    .354706   .0587917     6.03   0.000     .2394763    .4699357
       _cons |  -39.39866   37.19875    -1.06   0.290    -112.3069    33.50954
-------------+----------------------------------------------------------------
         rho |   .8530976
------------------------------------------------------------------------------


. 
. * Feasible GLS
. xtgls invest market stock, panel(iid) corr(indep) nmk

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        homoskedastic
Correlation:   no autocorrelation

Estimated covariances      =         1          Number of obs      =       100
Estimated autocorrelations =         0          Number of groups   =         5
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    341.63
Log likelihood             = -624.9928          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .1050854   .0113778     9.24   0.000     .0827853    .1273855
       stock |   .3053655   .0435078     7.02   0.000     .2200918    .3906393
       _cons |  -48.02974   21.48016    -2.24   0.025    -90.13009   -5.929387
------------------------------------------------------------------------------

. * (same as regress ..., nmk uses n - k to normalize the RSS)
. 
. xtgls invest market stock, i(company) panel(hetero) /* Heterosk., no contemp. correlatio

Cross-sectional time-series FGLS regression
```

```
Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   no autocorrelation

Estimated covariances      =         5          Number of obs      =       100
Estimated autocorrelations =         0          Number of groups   =         5
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    865.38
Log likelihood             = -570.1305          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .0949905    .007409    12.82   0.000     .0804692    .1095118
       stock |   .3378129   .0302254    11.18   0.000     .2785722    .3970535
       _cons |   -36.2537   6.124363    -5.92   0.000    -48.25723   -24.25017
------------------------------------------------------------------------------

. xtgls invest market stock, panel(corr) corr(ar1) /* Heterosk., contemp. correlation and

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   common AR(1) coefficient for all panels  (0.8651)

Estimated covariances      =        15          Number of obs      =       100
Estimated autocorrelations =         1          Number of groups   =         5
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    153.66
Log likelihood             = -491.3974          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .0745101   .0091391     8.15   0.000     .0565978    .0924225
       stock |   .3150971   .0447361     7.04   0.000     .2274158    .4027783
       _cons |  -2.770019   13.78308    -0.20   0.841    -29.78435    24.24431
------------------------------------------------------------------------------

. matrix list e(Sigma)

symmetric e(Sigma)[5,5]
             _ee        _ee2        _ee3        _ee4        _ee5
 _ee   5223.2164
_ee2  -101.18031   302.56293
_ee3   37.474924   146.75692   2578.9016
_ee4  -173.62446   57.848228   619.37254   262.40269
_ee5  -1093.8519    111.5931   537.76577   704.40596     8835.32
```

```
. xtgls invest market stock, panel(corr) corr(ar1) rhotype(dw)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   common AR(1) coefficient for all panels  (0.8179)

Estimated covariances      =        15          Number of obs      =       100
Estimated autocorrelations =         1          Number of groups   =         5
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    203.26
Log likelihood             = -495.6259          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .0758752   .0090285     8.40   0.000     .0581796    .0935708
       stock |   .3289528   .0409971     8.02   0.000     .2485999    .4093056
       _cons |  -10.08235    11.9502    -0.84   0.399    -33.50432    13.33961
------------------------------------------------------------------------------

. xtgls invest market stock, panel(corr) corr(psar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   panel-specific AR(1)

Estimated covariances      =        15          Number of obs      =       100
Estimated autocorrelations =         5          Number of groups   =         5
Estimated coefficients     =         3          Time periods       =        20
                                                Wald chi2(2)       =    331.55
Log likelihood             = -484.6178          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      invest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      market |   .0820264   .0081381    10.08   0.000      .066076    .0979767
       stock |   .3800689   .0313874    12.11   0.000     .3185508     .441587
       _cons |  -11.51848   12.69055    -0.91   0.364    -36.39151    13.35455
------------------------------------------------------------------------------

.
.
end of do-file

. exit, clear
```