STATA Commands for Unobserved Effects Panel Data

John C Frain

November 24, 2008

Contents

1 Introduction

Panel data, cross-sectional timeseries or longitudinal data are observations on a panel of i units or cases over t time periods. Most panel data commands start with **`xt`** For an overview of panel data type **`help xt`**. A typical panel data might record data on the income and expenditure of a group of individuals repeated over a number of years.

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 (2002) (or Cameron and Trivedi

(2005) or your favourite text) available for reference. The emphasis here is on the implementation of the methods described in Chapter 10 of Wooldridge (2002) Wooldridge and no attempt is made to explain the theory set out there[1] . Note the different fonts used for comments (this font), instructions in these comments (**`help xt`**) and for computer input/output (`help xt`).

```
help xt                                                      dialogs:  xtset
-------------------------------------------------------------------------------

Title

    [XT] xt -- Introduction to xt commands


Syntax

        xtcmd ... [, i(varname_i) t(varname_t) ...]


Description

    The xt series of commands provide tools for analyzing panel data (also known
    as longitudinal data or in some disciplines as cross-sectional time series
    when there is an explicit time component):

        xtset        Declare a dataset to be panel data
        xtdescribe   Describe pattern of xt data
        xtsum        Summarize xt data
        xttab        Tabulate xt data
        xtdata       Faster specification searches with xt data

        xtline       Line plots with xt data

        xtreg        Fixed-, between- and random-effects, and
                          population-averaged linear models
        xtregar      Fixed- and random-effects linear models with an AR(1)
                          disturbance
        xtmixed      Multilevel mixed-effects linear regression
        xtgls        Panel-data models using GLS
        xtpcse       OLS or Prais-Winsten models with panel-corrected standard
                          errors
        xtrc         Random coefficients models
        xtivreg      Instrumental variables and two-stage least squares for
                          panel-data models

        xtabond      Arellano-Bond linear dynamic panel-data estimator
        xtdpdsys     Arellano-Bond/Blundell-Bond estimation
        xtdpd        Linear dynamic panel-data estimation

        xttobit      Random-effects tobit models
        xtintreg     Random-effects interval data regression models

        xtlogit      Fixed-effects, random-effects, & population-averaged logit
                          models
        xtprobit     Random-effects and population-averaged probit models
        xtcloglog    Random-effects and population-averaged cloglog models
```

[1] The references at the end of this note are to books on panel data analysis or on the use of Stata in econometrics. These panel data books are not always easy going and are are suitable for those undertaking quantitative research using panel data. If you have problems you might consult one of the more general texts recommended for your course as an introduction to the topic.

```
    xtpoisson    Fixed-effects, random-effects, & population-averaged
                      Poisson models
    xtnbreg      Fixed-effects, random-effects, & population-averaged
                      negative binomial models

    xtmelogit    Multilevel mixed-effects logistic regression
    xtmepoisson  Multilevel mixed-effects Poisson regression

    xtgee        Population-averaged panel-data models using GEE

Panel datasets have the form x_it, where x_it is a vector of observations
for unit i and time t.  The particular commands (such as xtdescribe, xtsum,
and xtreg) are documented in their own help file entries.  This entry deals
with concepts common across commands.

The xtset command sets the panel variable and the time variable.  Most xt
commands require that the panel variable be specified, and some require that
the time variable also be specified. Once you xtset your data, you need not
do it again.  The xtset information is stored with your data.

If you have previously tsset your data by using both a panel and a time
variable, these settings will be recognized by xtset, and you need not xtset
your data.


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

        . xtset pid yr_visit


Also see

    Manual:  [XT] xt

    Online:  [XT] xtset
```

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 in Stata 10 using the **xtset** command.[2] Examples of the commands follow.

```
. 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
```

2 Examining Panel Data

The instructions in this section may be used to extract various panel properties of a panel data set. The temptation may be to skip this material and move to the estimation instructions in the next section. You should, at least, have a look at it as these instrustions may be very useful in examining the kind of panel data found in real examples.

xtdescribe describes the participation pattern in 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. The **pattern(#)** option allows one to increase the number of patterns shown.

[2] In earlier versions of Stata one used the **iis tis** commands, the **i() t()** options or the **tsset** command. While these can still be used it is recommended that one use the newer version 10 methods.

```
. 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. The between figures refers to the standard deviation, minimum and maximum of the averages for each individual ($\bar{x}_i$) (4710 individuals). The within figure calculates the statistics for the deviations of each individual from his own average ($x_{ij} - \bar{x}_i + \bar{\bar{x}}$) 28467 observations. Thus is for each person you calculate the average hours worked and then calculate the standard deviation of these averages you will get 7.846585. If, for each time period and each person you calculate his deviation from his own average and then calculate the standard deviation of all these 28467 adjusted observations yo get 7.520712. (The addition of an overall average to the individual deviations does not change the variance but does impinge on the explanation of the maximum and minimum. If you find the meaning of between and within a bit difficult leave it and return to it later)

```
. 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 generalises tabulate by performing one-way tabulations and giving details of between and within frequencies. 3113 in the between category is the number of women who showed a 0 in some year in the sample while 3643 is the number who showed a 1. As there were only 4711 in the sample we can see that many changed status during the sample period. The 55.06% is the fraction of time, on average, a women recorded 0 under `msp` given that she recorded 0 in some period. Similarly 71.90 is the fraction of time, on average, that a woman recorded 1 given that she recorded 1 in some period. For comparison purposes **xttab** is repeated for race, a characteristic which does not change over the period.

```
. 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 looks at transitions from on state to another over time. Thus 80.49% of the women who were 0 in one year were also zero in the next recorded year while 19.51% changed to 1. As we have seen above this data set is not balanced (i.e. There are not observations for all persons for all years. Thus some of the transitions may be over a period of years. To solve this problem we need to fill in the missing observations with NAs. This is accomplished by the **fillin** command. Rerunning the **xttrans** command gives appropriate estimates of the transition probabilities.

```
. 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
```

At this stage we might mention the **reshape** command. To illustrate this we load a small artificial data set which is listed in the computer output below. This is in what is known as long format. In long format each record has two identifiers. Here the identifiers are the `idcode` and the `year`. There are two variables, wage and tax, and 12 observations on each variable. In wide format as shown in the second list command there are three records indexed by `idcode`. There are four wage (wage2001, wage2002, wage2003 and wage2004) and four tax variables. This allows one to summarise data for each year if required. On some occasions one may find panel data on an excel spreadsheet in this format and may need to transform it to long format for analysis. Transfer from long to wide format is accomplished by **reshape wide** and the refverse transfer by **reshape long** In Stata 10 details of the **reshape** command are given in the Data manual.

xtline draws line plots for panel data. detail of wide and long

```
//
insheet using long.csv, comma clear
xtset idcode year
describe
summarize
list
// Switch to wide format
reshape wide wage tax , i(idcode) j(year)
describe
summarize
list
// Switch back to narrow format
reshape long wage tax , i(idcode) j(year)
describe
summarize
list
//
```

3 Estimation using xtreg

3.1 Introduction

The basic linear unobserved effects panel data model is

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

(For a full explanation of the symbols see Wooldridge page 251, etc.). In equation (1) c_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) which does not include the unit specific effect.

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

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 (3).

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

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

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

```
. //
. // Using xtreg
. //
. use nlswork, clear
(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
```

3.2 Pooled of Stacked OLS

For comparison purposes, we first estimate the model using OLS. Some might refer to this model as a stacked or pooled OLS estimate.

```
. //
. // OLS
. //
. regress ln_wage 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
------------------------------------------------------------------------------

. estimates store m_ols
```

3.3 Fixed Effects Estiamator

Next we have a fixed effects estimator first assuming heteroscedasticity and then with a robust estimator of the variance covariance matrix. Note at the bottom of the the first table the F-test that all the fixed effects are zero. This test is not given in the table containing the results of the robust estimation procedure as the equivalent robust statistic is difficult to calculate. Finally there is a table summarising the results of the OLS and fixed effects estimator. There is also a Stata command **areg** which may be used for OLS estimation when one has a large number of dummy variables. **areg** is designed for datasets with many groups, but not a number of groups that increases with the sample size. **xtreg, fe** can handle the case in which the number of groups increases with the sample size. With the addition of the vce() options to the **xtreg, fe** command access to the **areg** command is not now required for that purpose.

```
. //
. //Fixed Effects
. //
. xtreg ln_wage 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     .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

. estimates store m_fe

. //
. qui xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, fe vce(ro
> bust)

. estimates store m_fe_ro


. estimates table m_ols m_fe m_fe_ro m_fe_cl, b(%7.3g) se(%7.3g)

--------------------------------------------
    Variable |  m_ols     m_fe     m_fe_ro
-------------+------------------------------
       grade |   .0629         0         0
             |  .00103         0         0
         age |   .0386      .036      .036
             |  .00347    .00339    .00524
        age2 | -.00071   -.00072   -.00072
             | 5.6e-05   5.3e-05   8.5e-05
     ttl_exp |   .0211     .0335     .0335
             |  .00234    .00297    .00407
    ttl_exp2 |  .00045    .00022    .00022
             |  .00012    .00013    .00018
      tenure |   .0474     .0358     .0358
             |  .00196    .00185    .00247
     tenure2 | -.00203   -.00197   -.00197
             |  .00013    .00012    .00017
       black |  -.0699         0         0
             |  .00532         0         0
    not_smsa |   -.172     -.089     -.089
             |  .00517    .00953     .0138
       south |     -.1    -.0606    -.0606
             |  .00489     .0109     .0163
       _cons |    .247      1.04      1.04
             |   .0493     .0486      .074
--------------------------------------------
                                legend: b/se
```

3.4 Between Effects Estimator

We now calculate a between effects estimator.

```
. //
. // Betweeen Effects Estimator
. //
. xtreg ln_wage 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
------------------------------------------------------------------------------
```

3.5 Random Effects Estimator

Four examples of the random effects estimator follow — random effects, random effects with robust errors, Maximum likelihood. Note that the maximum likelihood estimator give a test for zero variance of the random effects.

```
. //
. // Random Effects Estimators
. //
. xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, re

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

------------------------------------------------------------------------------
     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 store m_re

. qui xtreg ln_wage grade age* ttl_exp* tenure* black not_smsa south, re vce(ro
> bust)

. estimates store m_re_ro

. xtreg ln_wage 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
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
        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


. estimates store m_mle

. estimates table m_re m_re_ro m_mle, b(%7.3g) se(%7.3g)

--------------------------------------------
    Variable |  m_re     m_re_ro    m_mle
-------------+------------------------------
_            |
       grade |   .0646     .0646
             |  .00178    .00193
         age |   .0368     .0368
             |  .00312    .00354
        age2 | -.00071   -.00071
             | 5.0e-05   5.8e-05
     ttl_exp |    .029      .029
             |  .00242    .00261
    ttl_exp2 |   .0003     .0003
             |  .00012    .00013
      tenure |   .0393     .0393
             |  .00176    .00185
     tenure2 |   -.002     -.002
             |  .00012    .00013
       black |  -.0531    -.0531
             |  .00999    .00983
    not_smsa |   -.131     -.131
             |  .00718    .00772
       south |  -.0869    -.0869
             |   .0073    .00772
       _cons |    .239      .239
             |   .0495     .0549
-------------+------------------------------
ln_wage      |
       grade |                       .0646
             |                      .00174
         age |                       .0369
             |                      .00312
        age2 |                     -.00071
             |                     5.0e-05
     ttl_exp |                       .0288
             |                      .00241
    ttl_exp2 |                      .00031
             |                      .00012
      tenure |                       .0394
             |                      .00176
     tenure2 |                     -.00201
             |                      .00012
       black |                      -.0533
             |                      .00973
    not_smsa |                       -.132
             |                      .00713
       south |                      -.0876
             |                      .00721
       _cons |                        .239
             |                       .0492
-------------+------------------------------
sigma_u      |
       _cons |                        .249
```

```
             |                       .0035
-------------+-----------------------------
sigma_e      |
       _cons |                        .292
             |                      .00135
-------------------------------------------
                               legend: b/se
```

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.

4 Testing after `xtreg`

*

Note that the post estimation commands **`test`**, **`testnl`**, **`estimates`**, **`lincom`**, **`lrtest`**, **`mfx`**, **`nlcom`**, **`predict`**, **`predictnl`** and **`hausman`** are also available after **`xtreg`**. The command **`xttest0`** is the Breusch and Pagan LM test for random effects. We can also do a Hausman test. When the common effects are orthogonal to the regressors both fixed and random effect estimators are consistent (but fixed effects are not efficient) under the alternative fixed effects are consistent whereas random effects are not. The Hausman test looks at the difference between the coefficients estimated using a random effects estimator and a fixed effects estimator. Roughly speaking in both estimates are similar we can use a random effects estimator. If they are different one may use the fixed effects estimator. The format of the **`hausman`** instruction is

`hausman` *m_consistent m_efficient*

where ***m_consistent*** and ***m_efficient*** are estimates of two models that have been estimated and saved (**`estimates store`**). ***m_consistent*** is the estimate that is consistent in both cases but not efficient and ***m_efficient*** is efficient where the first model is not efficient. In this case the estimates are significantly different and we reject the random effects estimator.

```
. /* After xtreg, re */
// After the random effects estimate

.
. * 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


.
```

5 Prediction after **xtreg**

After **xtreg** the **predict** command has the following options

xb $\boldsymbol{x}_j\boldsymbol{b}$, fitted values

stdp standard error of the fitted values

ue $\hat{c}_i + \hat{u}_{it}$, combined residual

xbu $\boldsymbol{x}_j\boldsymbol{b} + c_i$,prediction including effect

u c_i, The fixed or random effect component

e u_{it}, the error component

The last three of these are only available within sample

6 Other Stata Panel estimators

Here we mention some more detail regarding other Stata panel data commands. A complete list has already been given in Section (1). Details are given in the manuals ot the help files. Before using any of these commands one should be familiar with the theory involved.

6.1 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 using **`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. The gain in this instruction is where one needs to do a large specification search. The following stata commands illustrate this procedure. Output is suppressed. .

```
. qui xtdata ln\_wage grade age* ttl_exp* tenure* black not_smsa south, fe clear
. qui regress ln\_wage grade age ttl_exp tenure black not_smsa south
. qui regress ln\_wage grade age* ttl_exp* tenure* black not_smsa south
. qui xtdata ln\_wage grade age* ttl_exp* tenure* black not_smsa south, re ratio(.95) clear
. qui regress ln\_wage constant grade age ttl_exp tenure black not_smsa south, noconstant
. qui regress ln\_wage constant grade age* ttl_exp* tenure* black not_smsa south, noconstant
. qui xtdata ln_wage grade age* ttl_exp* tenure* black not_smsa south, fe clear
. qui regress ln_w grade age ttl_exp tenure black not_smsa south
```

6.2 More general error structures

`xtregar` fits fixed affects and random effects models where the disturbance follows an AR(1) process i.e.

$$
\begin{aligned}
y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta} + c_i + u_{it} \\
u_{it} &= \rho u_{i,t-1} + \eta_{it}
\end{aligned}
$$

and η_{it} are independent $N(0, \sigma_\eta^2)$.

`xtpcse` and **`xtgls`** estimate panel models under various assumptions about heterogeneity of variances across panels and possible serial correlation.

6.3 Dynamic panel data

The commands of interest here are **`xtabond`**, **`xtdpsys`** and **`xtdpd`**. There are considerable changes here relative to earlier versions of Stata.

6.4 Limited Dependent Variables in Panel Data

There are a variety of LDV estimation commands corresponding to the standard methods. These include **`xttobit`**, **`xtintreg`**, **`xtlogit`**, **`xtprobit`**, **`xtclogit`**, **`xtpoisson`**, **`xtnbreg`**, **`xtmelogit`** and **`xtmepoisson`**.

References

Adkins, L. C. and R. C. Hill (2008). *Using Stata for Principles of Econometrics Third Edition*. Wiley.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*. Wiley.

Baum, C. F. (2006). *An Introduction to Modern Econometrics using Stata*. Stata Press.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Hill, R. C., W. E. Griffiths, and G. C. Lim (2008). *Principles of Econometrics* (Third ed.). Wiley.

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge University Press.

Lee, M. (2002). *Panel Data Econometrics: Method of Moments and Limited Dependent Variables*. Academic Press.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press.