```
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------
      name:  <unnamed>
       log:  C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex2.log
  log type:  text
 opened on:   8 Nov 2011, 09:29:34

.
. ************************************************************
. * Load the panel data and tell Stata it is panel data         *
. ************************************************************
.
. use "`working'\longperson_unbal.dta"

. *change the path*
.
. * covert a string variable to the numberic variable *
. cap drop id

. destring xwaveid, gen (id)
xwaveid has all characters numeric; id generated as long

. tsset id wave
       panel variable:  id (unbalanced)
        time variable:  wave, 1 to 9, but with gaps
                delta:  1 unit


.
. ****************************************************************
. * To start with, let's focus on a very simple model to see how   *
. * the command work, and to look at age, cohort(yr of birth and   *
. * period effects. First, select a sample of employees of working *
. * age and drop observations with missing value                    *
. ****************************************************************
. gen year=2000+wave

. gen cohort=year-hgage //derive year of birth from age

. replace jbhruc=. if jbhruc<0
(0 real changes made)

. * use imputed income *
. gen w_hr=wscei/jbhruc if esbrd==1 //hourly wage and salary for all jobs
(42493 missing values generated)

. gen lwage_hr=ln(w_hr)
(50018 missing values generated)

. gen age=hgage/10        /* <-- better not /10 here  */

.
. keep if esdtl==1|esdtl==2 //employed
(42395 observations deleted)

. drop if hgage>63 & sex==2 //state retirement age for female
(691 observations deleted)
```

```
. drop if hgage>65 & sex==1 //state retirement age for male
(904 observations deleted)

. drop if lwage_hr==.
(6961 observations deleted)

. drop if hgage==.
(0 observations deleted)

. browse id wave year cohort lwage age esdtl w_hr esbrd

.
. ****************************************************************
. * As always, first examine the sample, e.g. using xtsum          *
. * Any problems?                                                 *
. ****************************************************************
. xtdes

      id:  100001, 100002, ..., 901101                       n =
14390
    wave:  1, 2, ..., 9                                      T =
9
           Delta(wave) = 1 unit
           Span(wave)  = 9 periods
           (id*wave uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%
max
                         1       1       2         4         8       9
9

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
     2421     16.82   16.82 |  111111111
      975      6.78   23.60 |  1........
      689      4.79   28.39 |  .........1
      480      3.34   31.72 |  11.......
      379      2.63   34.36 |  .......11
      355      2.47   36.82 |  ......111
      321      2.23   39.05 |  111......
      304      2.11   41.17 |  .1.......
      288      2.00   43.17 |  .....1111
     8178     56.83  100.00 | (other patterns)
 ---------------------------+-----------
    14390    100.00         |  XXXXXXXXX

. xtsum lwage_hr age cohort year

Variable         |      Mean   Std. Dev.       Min        Max |
Observations
-----------------+--------------------------------------------+----------
------
lwage_hr overall |  2.971187   .5621937  -3.688879   7.378696 |     N =
65734
         between |             .5188068  -1.473306   6.982181 |     n =
14390
```

```
         within  |             .3303074  -1.557092   6.429971 | T-bar =
4.56803
                 |                                           |
age      overall |  3.724642   1.267182        1.5        6.5 |     N =
65734
         between |             1.336906        1.5        6.5 |     n =
14390
         within  |             .2193802   3.199642   4.249642 | T-bar =
4.56803
                 |                                           |
cohort   overall |   1967.78   12.91942       1936       1994 |     N =
65734
         between |             14.11535       1936       1994 |     n =
14390
         within  |                   0    1967.78    1967.78 | T-bar =
4.56803
                 |                                           |
year     overall |  2005.026   2.605328       2001       2009 |     N =
65734
         between |             2.127319       2001       2009 |     n =
14390
         within  |             2.193802   1999.776   2010.276 | T-bar =
4.56803

. 
. ******************************************************************
. * Estimate a FE and BE model with the log hourly wage and         *
. * salary regressed on age, cohort and year                        *
. ******************************************************************
. xtreg lwage_hr /*age cohort*/ year, fe

Fixed-effects (within) regression               Number of obs      =
65734
Group variable: id                              Number of groups   =
14390

R-sq:  within  = 0.1789                         Obs per group: min =
1
       between = 0.0023                                        avg =
4.6
       overall = 0.0435                                        max =
9

                                                F(1,51343)         =
11183.77
corr(u_i, Xb)  = -0.1049                        Prob > F           =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
        year |   .0636769   .0006021   105.75   0.000     .0624967
.0648571
       _cons |  -124.7027    1.20728  -103.29   0.000     -127.069   -
122.3364
```

```
-------------+----------------------------------------------------------------
     sigma_u |  .52993552
     sigma_e |  .33867026
         rho |  .71001497   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(14389, 51343) =     8.47        Prob > F =
0.0000

. xtreg lwage_hr /*age*/ cohort year, be

Between regression (regression on group means)  Number of obs      =
65734
Group variable: id                              Number of groups   =
14390

R-sq:  within  = 0.1789                         Obs per group: min =
1
       between = 0.0904                                        avg =
4.6
       overall = 0.1111                                        max =
9

                                                F(2,14387)         =
715.08
sd(u_i + avg(e_i.))=  .4948309                  Prob > F           =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      cohort |  -.0120063   .0003215   -37.34   0.000    -.0126365    -
.011376
        year |   .0447999   .0021334    21.00   0.000     .0406182
.0489815
       _cons |  -63.29595   4.054339   -15.61   0.000    -71.24298   -
55.34892
------------------------------------------------------------------------------

.
.
. ****************************************************************
. * Now we are going to estimate the FE and between models manually,*
. * for comparison with the STATA estimates. Recall that the FE     *
. * model can be estimated using deviations from individual means,  *
. * while the between model uses the individual means. Create the   *
. * means and deviations and examine them                           *
. ****************************************************************
. * individual means
. cap drop mlwage_hr mage mcohort myear

. egen mlwage_hr = mean(lwage_hr), by (id)
```

```
. sort id wave

. browse id wave lwage_hr mlwage_hr

. egen mage=mean(age), by (id)

. egen mcohort=mean(cohort), by (id)

. egen myear=mean(year), by (id)

.
. * within deviations
. gen devlwage_hr=lwage_hr-mlwage_hr

. gen devage=age-mage

. gen devcohort=cohort-mcohort

. gen devyear=year-myear

.
. xtsum mlwage_hr mage mcohort myear devlwage_hr devage devcohort devyear

Variable         |      Mean   Std. Dev.       Min        Max |
Observations
-----------------+--------------------------------------------+----------
------
mlwage~r overall |  2.971187   .4549272  -1.473306   6.982181 |     N =
65734
         between |             .5188068  -1.473306   6.982181 |     n =
14390
         within  |                   0   2.971187   2.971187 | T-bar =
4.56803
                 |                                          |
mage     overall |  3.724642   1.248047        1.5        6.5 |     N =
65734
         between |             1.336906        1.5        6.5 |     n =
14390
         within  |                   0   3.724642   3.724642 | T-bar =
4.56803
                 |                                          |
mcohort  overall |   1967.78   12.91942       1936       1994 |     N =
65734
         between |             14.11535       1936       1994 |     n =
14390
         within  |                   0    1967.78    1967.78 | T-bar =
4.56803
                 |                                          |
myear    overall |  2005.026   1.405335       2001       2009 |     N =
65734
         between |             2.127319       2001       2009 |     n =
14390
         within  |                   0   2005.026   2005.026 | T-bar =
4.56803
                 |                                          |
devlwa~r overall |  3.82e-10   .3303074   -4.52828   3.458783 |     N =
65734
```

```
         between |             6.31e-08  -2.38e-07   2.38e-07 |     n =
14390
         within  |             .3303074   -4.52828   3.458783 | T-bar =
4.56803
                 |                                           |
devage   overall | -1.37e-09   .2193802  -.5250001   .5250001 |     N =
65734
         between |             9.52e-08  -2.38e-07   2.38e-07 |     n =
14390
         within  |             .2193802  -.5250002   .5250002 | T-bar =
4.56803
                 |                                           |
devcoh~t overall |         0          0          0          0 |     N =
65734
         between |                    0          0          0 |     n =
14390
         within  |                    0          0          0 | T-bar =
4.56803
                 |                                           |
devyear  overall |  3.27e-07   2.193802      -5.25       5.25 |     N =
65734
         between |             .0000139  -.0000523   .0000523 |     n =
14390
         within  |             2.193802      -5.25       5.25 | T-bar =
4.56803

.
. ******************************************************************
. * Perform an OLS regression of the derivation of log wage on the *
. * Deviations of year, age and cohort.                            *
. * Run the regression without a constant (why?)                   *
. * Are there any differences compared to the Stata FE estimates   *
. * Stata FE estimates (coefficient and SEs)?                      *
. ******************************************************************
.
. regress devlwage_hr /*devage devcohort*/ devyear, nocons

      Source |       SS       df       MS              Number of obs =
65734
-------------+------------------------------           F(  1, 65733)
=14318.27
       Model |  1282.75109     1  1282.75109           Prob > F      =
0.0000
    Residual |  5888.91615 65733  .089588428           R-squared     =
0.1789
-------------+------------------------------           Adj R-squared =
0.1789
       Total |  7171.66724 65734  .109101336           Root MSE      =
.29931

------------------------------------------------------------------------
-----
 devlwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
     devyear |   .0636769   .0005322   119.66   0.000     .0626339
.0647199
```

```
------------------------------------------------------------------------------

. * force to estimate age coeff *
. regress devlwage_hr devage /*devcohort*/, nocons

      Source |       SS       df       MS              Number of obs =
65734
-------------+------------------------------           F(  1, 65733)
=14318.27
       Model |  1282.75109     1  1282.75109           Prob > F      =
0.0000
    Residual |  5888.91616 65733  .089588428           R-squared     =
0.1789
-------------+------------------------------           Adj R-squared =
0.1789
       Total |  7171.66724 65734  .109101336           Root MSE      =
.29931

------------------------------------------------------------------------------
 devlwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      devage |   .6367691   .0053215   119.66   0.000     .6263389
.6471993
------------------------------------------------------------------------------

.
. * Compare with FE *
. xtreg lwage age /* cohort year*/, fe

Fixed-effects (within) regression               Number of obs      =
65734
Group variable: id                              Number of groups   =
14390

R-sq:  within  = 0.1789                         Obs per group: min =
1
       between = 0.0738                                        avg =
4.6
       overall = 0.0709                                        max =
9

                                                F(1,51343)         =
11183.77
corr(u_i, Xb)  = -0.8241                        Prob > F           =
0.0000

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

```
         age |   .6367691   .0060213   105.75   0.000     .6249674
.6485709
       _cons |   .5994504   .0224659    26.68   0.000      .555417
.6434839
-------------+----------------------------------------------------------
-----
     sigma_u |  .86831217
     sigma_e |  .33867026
         rho |  .86796091   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----
F test that all u_i=0:     F(14389, 51343) =     8.13        Prob > F =
0.0000

.
. ****************************************************************
. * Now do a similar regression using the individual mean.          *
. * By default, Stata does not take account of the fact that some   *
. * individuals contribute more observations than others. To
*
. * replicate this, do a simple (unweighted) OLS regression, using  *
. * one observation per individual                                   *
. ****************************************************************
. by id: gen firstobs=_n==1 //each person's first observation

. reg mlwage_hr mage mcohort /*myear*/ if firstobs==1

      Source |       SS       df       MS              Number of obs =
14390
-------------+------------------------------           F(  2, 14387) =
715.08
       Model |  350.183329     2  175.091665           Prob > F      =
0.0000
    Residual |  3522.76671 14387  .244857629           R-squared     =
0.0904
-------------+------------------------------           Adj R-squared =
0.0903
       Total |  3872.95004 14389  .269160472           Root MSE      =
.49483

------------------------------------------------------------------------
-----
   mlwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
        mage |   .4479986   .0213337    21.00   0.000     .4061817
.4898154
     mcohort |   .0327936   .0020206    16.23   0.000      .028833
.0367542
       _cons |  -63.29596   4.054339   -15.61   0.000    -71.24298   -
55.34893
------------------------------------------------------------------------
-----

.
. * to compare*
. xtreg lwage_hr age cohort /*year*/ , be
```

```
Between regression (regression on group means)  Number of obs      =
65734
Group variable: id                              Number of groups   =
14390

R-sq:  within  = 0.1789                         Obs per group: min =
1
       between = 0.0904                                        avg =
4.6
       overall = 0.1111                                        max =
9

                                                F(2,14387)         =
715.08
sd(u_i + avg(e_i.))=  .4948309                  Prob > F           =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
         age |   .4479985   .0213337    21.00   0.000     .4061817
.4898154
      cohort |   .0327936   .0020206    16.23   0.000      .028833
.0367542
       _cons |  -63.29595   4.054339   -15.61   0.000    -71.24298   -
55.34892
------------------------------------------------------------------------------

. * drop year to force give coefficient on age *
.
.
. *********************************************************************
. * Now we estimate a more complete model of hourly  wage and         *
. * salary: as a function of age, age square, birth cohort, martial   *
. * status, sex, job tenure, possession of degree or further education*
. * trade union coverage, job contract type and born in Australia     *
. *********************************************************************
.
. *********************************************************************
. * Assume we don't care why the values are missing set the negative  *
. * values to missing, drop the missing value and derive binary       *
. * variables for the model                                           *
. *********************************************************************
. * creat marriage categories *
. replace mrcurr=. if mrcurr<0  //change negative values to missing
(18 real changes made, 18 to missing)

. drop if mrcurr==.
(18 observations deleted)

. recode mrcurr(1=1) (else=0), gen (married)
(33991 differences between mrcurr and married)
```

```
. recode mrcurr(2=1) (else=0), gen (defacto)
(65716 differences between mrcurr and defacto)

. recode mrcurr(3=1) (else=0), gen (separated)
(65716 differences between mrcurr and separated)

. recode mrcurr(4=1) (else=0), gen (divorced)
(65716 differences between mrcurr and divorced)

. recode mrcurr(5=1) (else=0), gen (widowed)
(65716 differences between mrcurr and widowed)

. recode mrcurr(6=1) (else=0), gen (single)
(65716 differences between mrcurr and single)

. tab wave

       wave |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      7,608       11.58       11.58
          2 |      7,190       10.94       22.52
          3 |      7,030       10.70       33.22
          4 |      6,885       10.48       43.69
          5 |      7,241       11.02       54.71
          6 |      7,382       11.23       65.94
          7 |      7,409       11.27       77.22
          8 |      7,434       11.31       88.53
          9 |      7,537       11.47      100.00
------------+-----------------------------------
      Total |     65,716      100.00

.
. * contract of current work *
. replace jbmcnt=. if jbmcnt<0
(5475 real changes made, 5475 to missing)

. drop if jbmcnt==.
(5475 observations deleted)

. recode jbmcnt(1=1) (else=0), gen (fixedterm)
(54639 differences between jbmcnt and fixedterm)

. recode jbmcnt(2=1) (else=0), gen (casual)
(60241 differences between jbmcnt and casual)

. recode jbmcnt(3=1) (else=0), gen (permanent)
(60241 differences between jbmcnt and permanent)

. replace fixedterm=1 if jbmcnt==8 /*collapse 14 other into fixed term*/
(206 real changes made)

. tab wave

       wave |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      6,749       11.20       11.20
          2 |      6,445       10.70       21.90
          3 |      6,434       10.68       32.58
```

```
          4 |      6,314       10.48       43.06
          5 |      6,675       11.08       54.14
          6 |      6,829       11.34       65.48
          7 |      6,860       11.39       76.87
          8 |      6,918       11.48       88.35
          9 |      7,017       11.65      100.00
------------+-----------------------------------
      Total |     60,241      100.00

.
. * highest education achieved *
. replace edhigh =. if edhigh<0
(0 real changes made)

. drop if edhigh==.
(0 observations deleted)

. recode edhigh(1/3=1) (else=0), gen (degree) //have a degree or higher
qualification
(57911 differences between edhigh and degree)

. recode edhigh(1/7=1) (else=0), gen (further) //have done some further
studies after school
(57911 differences between edhigh and further)

. recode edhigh(8=1) (else=0), gen (yr12)
(60241 differences between edhigh and yr12)

. recode edhigh(9=1) (else=0), gen (yr11)
(60241 differences between edhigh and yr11)

. recode edhigh(10=1) (else=0), gen (edunk)
(60241 differences between edhigh and edunk)

. tab wave

       wave |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      6,749       11.20       11.20
          2 |      6,445       10.70       21.90
          3 |      6,434       10.68       32.58
          4 |      6,314       10.48       43.06
          5 |      6,675       11.08       54.14
          6 |      6,829       11.34       65.48
          7 |      6,860       11.39       76.87
          8 |      6,918       11.48       88.35
          9 |      7,017       11.65      100.00
------------+-----------------------------------
      Total |     60,241      100.00

. browse xwaveid wave edhigh degree further yr12

.
.
. * trade union *
. cap drop tucov

. *label list ajbmunio
```

```
. replace jbmunio=. if jbmunio<0
(437 real changes made, 437 to missing)

. drop if jbmunio==.
(437 observations deleted)

. recode jbmunio(1=1) (else=0), gen(tucov)
(42982 differences between jbmunio and tucov)

.
.
. * job tenure *
. * label list ajbempt
. replace jbempt=. if jbempt<0
(45 real changes made, 45 to missing)

. drop if jbempt==.
(45 observations deleted)

.
. * Details on the state of living*
. *label list ahhstate
. replace hhstate=. if hhstate<0
(0 real changes made)

. drop if hhstate<0
(0 observations deleted)

. recode hhstate(1=1) (else=0), gen(nsw)
(42227 differences between hhstate and nsw)

. recode hhstate(2=1) (else=0), gen(vic)
(59759 differences between hhstate and vic)

. recode hhstate(3=1) (else=0), gen(qld)
(59759 differences between hhstate and qld)

. recode hhstate(4=1) (else=0), gen(sa)
(59759 differences between hhstate and sa)

. recode hhstate(5=1) (else=0), gen(wa)
(59759 differences between hhstate and wa)

. recode hhstate(6=1) (else=0), gen(tas)
(59759 differences between hhstate and tas)

. recode hhstate(7=1) (else=0), gen(nt)
(59759 differences between hhstate and nt)

. recode hhstate(8=1) (else=0), gen(act)
(59759 differences between hhstate and act)

.
. * convert sex to a binary variable *
. * label list sex
. replace sex=. if sex<0
(0 real changes made)
```

```
. drop if sex==.
(0 observations deleted)

. recode sex(2=1) (else=0), gen(female)
(59759 differences between sex and female)

.
. gen agesq=age*age

. ********************************************************
. * check variable jbempt for any implausible value       *
. ********************************************************
. sum jbempt, detail

          DV: Tenure with current employer (years)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .0384615       .0192308
 5%     .1538462       .0192308
10%     .3076923       .0192308       Obs                59759
25%            1       .0192308       Sum of Wgt.        59759

50%            3                      Mean           5.833331
                        Largest       Std. Dev.      7.319051
75%            8             47
90%           16             48       Variance        53.5685
95%           22             50       Skewness        1.94705
99%           32             50       Kurtosis        6.789296

. gen prob=1 if hgage<jbemp
(59755 missing values generated)

. tab prob

       prob |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          4      100.00      100.00
------------+-----------------------------------
      Total |          4      100.00

. drop if prob==1   // 4 observations deleted
(4 observations deleted)

. drop prob

.
. xtsum jbempt

Variable         |      Mean   Std. Dev.       Min        Max |
Observations
-----------------+--------------------------------------------+----------
------
jbempt   overall |  5.832181    7.31787   .0192308         50 |     N =
59755
         between |             6.288826   .0192308         50 |     n =
13367
         within  |             2.955178  -24.10628   38.83218 | T-bar =
4.47034
```

```
.
. ************************************************************
. * Estimate a FE model and interpret the coefficients. How     *
. * important are the individual effects? Estimate the between  *
. * group model. How do the coefficients compare to the FE model*
. * ? How do you interpret any differences?                     *
. * drop year from all the models from now on                   *
. ************************************************************
. sort id wave

. browse id wave lwage_hr age agesq married female jbemp degree further
///
>            tucov permanent nsw

.
. xtdes

      id:  100001, 100002, ..., 901101                       n =
13367
    wave:  1, 2, ..., 9                                      T =
9
           Delta(wave) = 1 unit
           Span(wave)  = 9 periods
           (id*wave uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%
max
                         1       1       2         4         7       9
9

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------
     2142     16.02   16.02 |  111111111
      868      6.49   22.52 |  1........
      697      5.21   27.73 |  ........1
      418      3.13   30.86 |  11.......
      387      2.90   33.75 |  .......11
      344      2.57   36.33 |  ......111
      313      2.34   38.67 |  111......
      292      2.18   40.85 |  .1.......
      277      2.07   42.93 |  .....1111
     7629     57.07  100.00 | (other patterns)
 ---------------------------+-----------
    13367    100.00         |  XXXXXXXXX

.
. xtreg lwage_hr age agesq married female jbemp degree further ///
>           tucov permanent nsw, fe
note: female omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  = 0.2412                         Obs per group: min =
1
```

```
       between = 0.1494                                        avg =
4.5
       overall = 0.1330                                        max =
9

                                                    F(9,46379)         =
1638.44
corr(u_i, Xb)  = -0.8513                            Prob > F           =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
         age |   1.345099   .0212097    63.42   0.000     1.303528
1.38667
       agesq |  -.0921478    .002585   -35.65   0.000    -.0972144   -
.0870812
     married |  -.0018379   .0069057    -0.27   0.790    -.0153731
.0116973
      female |  (omitted)
      jbempt |   .0012362   .0004347     2.84   0.004     .0003842
.0020883
      degree |   .0152046    .014657     1.04   0.300    -.0135233
.0439325
     further |    .062598   .0102969     6.08   0.000      .042416
.08278
       tucov |   .0275785   .0052327     5.27   0.000     .0173223
.0378348
   permanent |  -.0341137    .004087    -8.35   0.000    -.0421243   -
.0261032
         nsw |   .0252218   .0132444     1.90   0.057    -.0007374
.0511809
       _cons |   -.606643    .040114   -15.12   0.000    -.6852671    -
.528019
-------------+----------------------------------------------------------------
     sigma_u |  .85089062
     sigma_e |  .30106476
         rho |  .88873817   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(13366, 46379) =     6.38        Prob > F =
0.0000

.
. xtreg lwage_hr age agesq married female jbemp degree further ///
>           tucov permanent nsw, be

Between regression (regression on group means)  Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  = 0.0733                         Obs per group: min =
1
```

```
       between = 0.3203                                        avg =
4.5
       overall = 0.2694                                        max =
9

                                                    F(10,13356)       =
629.42
sd(u_i + avg(e_i.))=  .4031659                      Prob > F          =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
         age |   .5235305   .0181119    28.91   0.000     .4880285
.5590324
       agesq |  -.0618675   .0023544   -26.28   0.000    -.0664824   -
.0572525
     married |    .044588   .0090958     4.90   0.000     .0267591
.062417
      female |  -.1028382   .0070736   -14.54   0.000    -.1167036   -
.0889729
      jbempt |   .0055679   .0006912     8.06   0.000     .0042131
.0069227
      degree |   .2751084   .0101278    27.16   0.000     .2552565
.2949603
     further |   .0963877   .0088601    10.88   0.000     .0790207
.1137546
       tucov |   .0669973   .0101034     6.63   0.000     .0471932
.0868014
   permanent |   .1038804   .0096285    10.79   0.000     .0850072
.1227535
         nsw |   .0350543   .0077902     4.50   0.000     .0197845
.0503242
       _cons |   1.740149   .0288373    60.34   0.000     1.683624
1.796674
------------------------------------------------------------------------------

.
.
.
. ****************************************************************
. * Estimate the individual effects u(i), For help with          *
. * prediction options, type help xtreg post-estimation. Perform*
. * a second step regression to estimate the coeffs associated  *
. * with time invariant characteristics, female and cohort.      *
. * First ignore the different numbers of observations           *
. * contributed by individuals. Second, weight the observations *
. * accordingly. Hint: see wls option for xtreg, be              *
. ****************************************************************
.
. * first need to run FE again - do it "quietly" *
.
. quietly xtreg lwage_hr hgage agesq married female jbemp degree  ///
>                     further tucov permanent nsw, fe
```

```
. predict ui, u

. xtreg ui cohort female, be

Between regression (regression on group means)  Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  =      .                         Obs per group: min =
1
       between = 0.7544                                        avg =
4.5
       overall = 0.7799                                        max =
9

                                                F(2,13364)         =
20521.04
sd(u_i + avg(e_i.))=  .4217458                  Prob > F           =
0.0000

------------------------------------------------------------------------------
          ui |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      cohort |   .0527708   .0002612   202.01   0.000     .0522588
.0532829
      female |  -.1145501   .0072958   -15.70   0.000    -.1288509   -
.1002493
       _cons |   -103.846   .5147516  -201.74   0.000    -104.8549    -
102.837
------------------------------------------------------------------------------

. xtreg u cohort female, be wls

Between regression (regression on group means)  Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  =      .                         Obs per group: min =
1
       between = 0.7799                                        avg =
4.5
       overall = 0.7799                                        max =
9

                                                F(2,13364)         =
23670.49
sd(u_i + avg(e_i.))=  .3668426                  Prob > F           =
0.0000

------------------------------------------------------------------------------
```

```
          ui |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      cohort |   .0532641   .0002458   216.67   0.000     .0527822
.0537459
      female |   -.114513   .0063461   -18.04   0.000    -.1269524   -
.1020737
       _cons |  -104.7853   .4839319  -216.53   0.000    -105.7339   -
103.8367
------------------------------------------------------------------------------

.
.
. ****************************************************************
. *                       Additional Exercise
*
. ****************************************************************
. ********************************************************************
. * Allows for more general distributions of e(it).                  *
. * Correcting the standard errors for arbitrary heteroscedasticity *
. ********************************************************************
.
. xtreg lwage_hr hgage agesq married female jbemp degree further ///
>         tucov permanent nsw, fe robust
note: female omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  = 0.2412                         Obs per group: min =
1
       between = 0.1494                                        avg =
4.5
       overall = 0.1330                                        max =
9

                                                F(9,13366)         =
970.51
corr(u_i, Xb)  = -0.8513                        Prob > F           =
0.0000

                                  (Std. Err. adjusted for 13367 clusters
in id)
------------------------------------------------------------------------------
             |               Robust
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
       hgage |   .1345099   .0029513    45.58   0.000      .128725
.1402949
       agesq |  -.0921478   .0035552   -25.92   0.000    -.0991165   -
.0851791
```

```
     married |  -.0018379   .0074132    -0.25   0.804    -.0163688
.012693
      female |  (omitted)
      jbempt |   .0012362   .0005344     2.31   0.021     .0001888
.0022836
      degree |   .0152046   .0196401     0.77   0.439    -.0232927
.0537019
     further |    .062598   .0140489     4.46   0.000     .0350602
.0901358
       tucov |   .0275785   .0061439     4.49   0.000     .0155355
.0396215
   permanent |  -.0341137   .0049542    -6.89   0.000    -.0438247   -
.0244027
         nsw |   .0252218   .0188212     1.34   0.180    -.0116705
.0621141
       _cons |  -.6066431   .0564119   -10.75   0.000    -.7172184   -
.4960678
-------------+----------------------------------------------------------
-----
     sigma_u |  .85089062
     sigma_e |  .30106476
         rho |  .88873817   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----

.
. ****************************************************************
. * testing for no serial correlation of e(it)                     *
. * Need to install xtserial
*
. ****************************************************************
. findit xtserial

. net sj 3-2 st0039

------------------------------------------------------------------------
------------------------------------------------------------------------
------------
package st0039 from http://www.stata-journal.com/software/sj3-2
------------------------------------------------------------------------
------------------------------------------------------------------------
------------

TITLE
      SJ3-2 st0039.  Testing for serial correlation in linear ...

DESCRIPTION/AUTHOR(S)
      Testing for serial correlation in linear panel-data models
      by David M. Drukker, Stata Corporation
      Support:  ddrukker@stata.com
      After installation, type help xtserial

INSTALLATION FILES                                  (type net install
st0039)
      st0039/xtserial.ado
      st0039/xtserial.hlp

ANCILLARY FILES                                     (type net get st0039)
```

```
      st0039/xtserial.do
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
------------

. net install st0039
checking st0039 consistency and verifying not already installed...
all files already exist and are up to date.

.
. xtserial lwage_hr hgage agesq married female jbemp degree further ///
>              tucov permanent nsw

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,    7711) =    152.589
           Prob > F =      0.0000

.
. ****************************************************************
. * Correcting standard errors for arbitrary serial correlation *
. ****************************************************************
.
. xtreg lwage_hr hgage agesq married /*female*/ jbemp degree further ///
>           tucov permanent nsw,  fe robust cluster(id)

Fixed-effects (within) regression               Number of obs      =
59755
Group variable: id                              Number of groups   =
13367

R-sq:  within  = 0.2412                         Obs per group: min =
1
       between = 0.1494                                        avg =
4.5
       overall = 0.1330                                        max =
9

                                                F(9,13366)         =
970.51
corr(u_i, Xb)  = -0.8513                        Prob > F           =
0.0000

                                  (Std. Err. adjusted for 13367 clusters
in id)
------------------------------------------------------------------------------
             |               Robust
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
       hgage |   .1345099   .0029513    45.58   0.000      .128725
.1402949
       agesq |  -.0921478   .0035552   -25.92   0.000    -.0991165   -
.0851791
     married |  -.0018379   .0074132    -0.25   0.804    -.0163688
.012693
```

```
      jbempt |   .0012362   .0005344     2.31   0.021     .0001888
.0022836
      degree |   .0152046   .0196401     0.77   0.439    -.0232927
.0537019
     further |    .062598   .0140489     4.46   0.000     .0350602
.0901358
       tucov |   .0275785   .0061439     4.49   0.000     .0155355
.0396215
   permanent |  -.0341137   .0049542    -6.89   0.000    -.0438247   -
.0244027
         nsw |   .0252218   .0188212     1.34   0.180    -.0116705
.0621141
       _cons |  -.6066431   .0564119   -10.75   0.000    -.7172184   -
.4960678
-------------+----------------------------------------------------------
-----
     sigma_u |  .85089062
     sigma_e |  .30106476
         rho |  .88873817   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----

.
. ****************************************************************
. * Estimate model assuming serial correlation is AR(1)         *
. ****************************************************************
.
. xtregar lwage_hr hgage agesq married female jbemp degree further ///
>             tucov permanent nsw, fe
note: female dropped because of collinearity

FE (within) regression with AR(1) disturbances  Number of obs      =
46388
Group variable: id                              Number of groups   =
10208

R-sq:  within  = 0.2802                         Obs per group: min =
1
       between = 0.1184                                        avg =
4.5
       overall = 0.1047                                        max =
8

                                                F(9,36171)         =
1564.16
corr(u_i, Xb)  = -0.8218                        Prob > F           =
0.0000

------------------------------------------------------------------------
-----
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
       hgage |   .1126726   .0017013    66.23   0.000     .1093379
.1160073
       agesq |  -.0730688   .0026333   -27.75   0.000    -.0782302   -
.0679074
```

```
     married |  -.0015793   .0088161    -0.18   0.858    -.0188591
.0157005
      female |  (omitted)
      jbempt |   .0012677   .0004865     2.61   0.009     .0003141
.0022213
      degree |   .0325187   .0193611     1.68   0.093    -.0054296
.070467
     further |   .0741285   .0137562     5.39   0.000     .0471659
.1010911
       tucov |   .0175519   .0060042     2.92   0.003     .0057836
.0293203
   permanent |  -.0355731   .0045486    -7.82   0.000    -.0444885   -
.0266577
         nsw |   .0118447   .0160096     0.74   0.459    -.0195345
.043224
       _cons |  -.1057544   .0207072    -5.11   0.000     -.146341   -
.0651678
-------------+----------------------------------------------------------
-----
      rho_ar |   .3019444
     sigma_u |  .73757376
     sigma_e |    .291292
     rho_fov |  .86507298   (fraction of variance because of u_i)
------------------------------------------------------------------------
-----
F test that all u_i=0:     F(10207,36171) =     3.76         Prob > F =
0.0000

.
.
. ****************************************************************
. *               End of additional exercise                     *
. ****************************************************************
.
. ****************************************************************
. * Save the new data set for ex.3
*
. ****************************************************************
.
. save "`working'\panelex3.dta", replace
file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\panelex3.dta saved

.
. set more on

.
. log c
      name:  <unnamed>
       log:  C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex2.log
  log type:  text
 closed on:   8 Nov 2011, 09:29:49
------------------------------------------------------------------------
------------------------------------------------------------------------
------------
```