```
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
------------
      name:  <unnamed>
       log:  C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex4.log
  log type:  text
 opened on:   8 Nov 2011, 15:10:25

. 
. ************************************************************
. * Load the panel data and tell Stata it is panel data      *
. ************************************************************
. 
. use "`working'\panelex4.dta" ,clear

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

. 
. ************************************************************
. * We continue with the wage equation. We suspect that the  *
. * error terms may be correlated with job tenure.           *
. * Various sources of bias are possible:                    *
. * - people with higher earning ability tend to hold jobs   *
. *   for longer(?). Implies positive correlation between u(i) *
. *   and jbemp. Note this is controlled for in the FE model.  *
. * - people stay longer in high-paid jobs (high lwage_hr)).  *
. *   Implies positive correlation between high lwage_hr and  *
. *   jbemp.                                                  *
. * - people move to good (high paid) jobs. Implies negative  *
. *   correlation between high lwage_hr and jbemp.            *
. * - Any others?                                            *
. * To control for a possible correlation of jbemp with      *
. * lwage_hr. we will use instrumental variables. We need    *
. * variables which help explain jbemp, but are uncorrelated  *
. * with wages (conditional on regressors). Suggested        *
. * instruments are hours worked by spouse (spousehr) [zero if *
. * not employed employed or no spouse], preference to stay in *
. * current home (wantstay), and whether renting home (tenant).*
. * To allow for the effects to vary over men and women,     *
. * interact the instruments with a gender dummy variable.   *
. * Based on Ex3, the appropriate model is FE.Before estimating*
. * the model it is necessary to check that the instruments  *
. * help explain jbemp (using FE (why?) equation).           *
. ************************************************************
. 
. * Derive extra variables for the model *
. * preference to stay in current home *
. sort id wave

. browse id wave losathl

. tab losathl
```

```
   K6 Satisfaction - The home in which |
                              you live |      Freq.     Percent
Cum.
-------------------------------------+-----------------------------------
--
              [0] Totally dissatisfied |        187        0.31
0.31
                                     1 |        228        0.38
0.69
                                     2 |        575        0.96
1.66
                                     3 |      1,051        1.76
3.42
                                     4 |      1,331        2.23
5.65
[5] Neither satisfied nor dissatisfied |      3,504        5.87
11.51
                                     6 |      4,434        7.42
18.94
                                     7 |      9,761       16.34
35.28
                                     8 |     15,596       26.11
61.40
                                     9 |     11,709       19.61
81.00
                [10] Totally satisfied |     11,346       19.00
100.00
-------------------------------------+-----------------------------------
--
                                 Total |     59,722      100.00

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

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

. capture drop wantstay

. recode losathl(6/10=1) (else=0), gen (wantstay)
(59535 differences between losathl and wantstay)

. tab wantstay

  RECODE of |
losathl (K6 |
Satisfactio |
    n - The |
    home in |
  which you |
      live) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      6,876       11.51       11.51
          1 |     52,846       88.49      100.00
------------+-----------------------------------
      Total |     59,722      100.00

.
```

```
.
. *whether renting home*
. capture drop tenant

. recode hstenur(2=1) (else=0), gen(tenant)
(59722 differences between hstenur and tenant)

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

.
. save "`working'\panelex4.dta",replace
file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\panelex4.dta saved

. clear

.
. * spouse information *
. use "`working'\longperson_unbal.dta"

. keep hhpxid wave hgage hgsex jbhruc edhigh esdtl

.  tab wave if hhpxid =""

       wave |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      5,188       11.49       11.49
          2 |      4,927       10.91       22.40
          3 |      4,877       10.80       33.19
          4 |      4,813       10.66       43.85
          5 |      5,035       11.15       55.00
          6 |      5,077       11.24       66.24
          7 |      4,978       11.02       77.26
          8 |      4,983       11.03       88.29
          9 |      5,288       11.71      100.00
------------+-----------------------------------
      Total |     45,166      100.00

. drop if hhpxid=""
(45166 observations deleted)

. rename hhpxid xwaveid

. rename hgage spouseage

. rename hgsex spousesex

. rename jbhruc spousehr

. rename edhigh spouseedhigh

. rename esdtl spouseesdtl

. sort xwaveid wave

. save "`working'\partner.dta", replace
```

```
file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\partner.dta saved

. clear

.
. use "`working'\panelex4.dta"

. sort xwaveid wave

. merge 1:1 xwaveid wave using "`working'\partner.dta"
(label AJBHRUC already defined)
(label AHGSEX already defined)
(label AHGAGE already defined)
(label AESDTL already defined)
(label AEDHIGH already defined)

    Result                           # of obs.
    -----------------------------------------
    not matched                         61,685
        from master                     24,944  (_merge==1)
        from using                      36,741  (_merge==2)

    matched                             34,778  (_merge==3)
    -----------------------------------------

. drop if _merge==2
(36741 observations deleted)

. * if no spouse assign value zero *
. replace spouseage=0 if _merge==1
(24944 real changes made)

. replace spousesex=0 if _merge==1
(24944 real changes made)

. replace spouseesdtl=0 if _merge==1
(24944 real changes made)

. replace spousehr=0 if _merge==1
(24944 real changes made)

. replace spouseedhigh=0 if _merge==1
(24944 real changes made)

. drop _merge

. browse

.
. * set spousehr=0 if not employed *
. replace spousehr=0 if spouseesdtl!=1 & spouseesdtl!=2 & spouseesdtl!=7
(6104 real changes made)

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

. drop if spousehr==. //drop the non-responding records from the analysis
```

```
(25 observations deleted)

.
. gen femten=female*tenant

. gen femstay=female*wantstay

. gen femsphr=female*spousehr

. browse

.
. xtreg jbemp hgage age cohort married female degree further ///
>       /*tucov*/ permanent nsw tenant femten wantstay  ///
>           femstay spousehr femsphr, fe
note: age omitted because of collinearity
note: cohort omitted because of collinearity
note: female omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

R-sq:  within  = 0.0745                         Obs per group: min =
1
       between = 0.2547                                        avg =
4.5
       overall = 0.2222                                        max =
9

                                                F(12,46325)        =
310.53
corr(u_i, Xb)  = -0.1797                        Prob > F           =
0.0000

------------------------------------------------------------------------
-----
      jbempt |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
       hgage |   .3509702   .0064365    54.53   0.000     .3383547
.3635858
         age |  (omitted)
      cohort |  (omitted)
     married |  -.0701032    .076043    -0.92   0.357    -.2191487
.0789423
      female |  (omitted)
      degree |  -1.101143   .1564094    -7.04   0.000    -1.407708   -
.7945785
     further |  -.3674437   .1089724    -3.37   0.001    -.5810312   -
.1538562
   permanent |   .7554432   .0433483    17.43   0.000     .6704799
.8404065
         nsw |   .8499127   .1417408     6.00   0.000     .5720987
1.127727
```

```
      tenant |  -.2838275   .0777469    -3.65   0.000    -.4362127   -
.1314423
      femten |   .0051369    .111543     0.05   0.963    -.2134891
.2237629
    wantstay |   .0535554   .0784402     0.68   0.495    -.1001887
.2072994
     femstay |   .1360356   .1096383     1.24   0.215    -.0788571
.3509283
    spousehr |  -.0008721   .0018952    -0.46   0.645    -.0045866
.0028425
     femsphr |  -.0055362   .0024926    -2.22   0.026    -.0104218   -
.0006506
       _cons |  -7.228886   .2352341   -30.73   0.000    -7.689949   -
6.767824
-------------+----------------------------------------------------------
-----
     sigma_u |   5.640542
     sigma_e |  3.2223168
         rho |  .75394403   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----
F test that all u_i=0:     F(13359, 46325) =    13.60        Prob > F =
0.0000

.
. test tenant femten wantstay femstay spousehr femsphr

 ( 1)  tenant = 0
 ( 2)  femten = 0
 ( 3)  wantstay = 0
 ( 4)  femstay = 0
 ( 5)  spousehr = 0
 ( 6)  femsphr = 0

       F(  6, 46325) =    8.19
            Prob > F =    0.0000

. test wantstay femstay

 ( 1)  wantstay = 0
 ( 2)  femstay = 0

       F(  2, 46325) =    3.29
            Prob > F =    0.0372

.
. ************************************************************
. * The "canned" Stata routine for panel IV is xtivreg. We are *
. * going to download a more advanced command which provides   *
. * more diagnostic statistics (but is for FE models only).    *
. ************************************************************
. ssc install ranktest, replace
checking ranktest consistency and verifying not already installed...
all files already exist and are up to date.

. ssc install ivreg2, replace
checking ivreg2 consistency and verifying not already installed...
all files already exist and are up to date.
```

```
. ssc install xtivreg2, replace
checking xtivreg2 consistency and verifying not already installed...
all files already exist and are up to date.

. ssc describe ivreg2

------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------
package ivreg2 from http://fmwww.bc.edu/repec/bocode/i
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------

TITLE
      'IVREG2': module for extended instrumental variables/2SLS and GMM
estimation

DESCRIPTION/AUTHOR(S)

      ivreg2 provides extensions to Stata's official ivregress and
      newey. Its main capabilities: two-step feasible GMM estimation;
      continuously updated GMM estimation (CUE); LIML and k-class
      estimation; automatic output of the Hansen-Sargan or
      Anderson-Rubin statistic for overidentifying restrictions; C
      statistic test of exogeneity of subsets of instruments (orthog()
      option); kernel-based autocorrelation-consistent (AC) and
      heteroskedastic and autocorrelation-consistent (HAC) estimation,
      with user-specified choice of kernel; Cragg's "heteroskedastic
      OLS" (HOLS) estimator; default reporting of large-sample
      statistics (z and chi-squared rather than t and F); small option
      to report small-sample statistics; first-stage regression
      reported with F-test of excluded instruments and R-squared with
      included instruments "partialled-out"; enhanced Kleibergen-Paap
      and Cragg-Donald tests for weak instruments, redundancy of
      instruments, significance of endogenous regressors; two-way
      clustering of standard errors;  Kiefer and Driscoll-Kraay
      standard errors. ivreg2 can also be used for ordinary least
      squares (OLS) estimation using the same command syntax as Stata's
      official regress and newey. This is version 3.0.08 of ivreg2,
      updated from that published in Stata Journal, 5(4), requiring
      Stata 10.1 or better. Stata 9.2 users may use ivreg29 (q.v.).
      Stata 8 users may use ivreg28 (q.v.) Stata 7 users may use the
      Stata Journal version of ivreg2, accessible via net search
      ivreg2.

      KW: instrumental variables
      KW: Sargan test
      KW: robust estimation
      KW: orthogonality
      KW: GMM
      KW: Hansen's J
      KW: heteroskedastic OLS,

      Requires: Stata version 10.1 and ranktest from SSC
```

```
      Author: Christopher F Baum, Boston College
      Support: email baum@bc.edu

      Author: Mark E Schaffer, Heriot-Watt University
      Support: email m.e.schaffer@hw.ac.uk

      Author: Steven Stillman, Motu Economic and Public Policy Research
      Support: email  stillman@motu.org.nz

      Distribution-Date: 20111006


INSTALLATION FILES                                  (type net install
ivreg2)
      ivreg2.ado
      ivreg2.sthlp
      ivreg2_p.ado

ANCILLARY FILES                                     (type net get ivreg2)
      ../c/cs_ivreg2_3.0.8.do
------------------------------------------------------------------------------
------------------------------------------------------------------------------
------------
(type -ssc install ivreg2- to install)

.
. xtivreg2 lwage_hr hgage agesq cohort married female degree ///
>         further /*tucov*/ permanent nsw ///
>         (jbemp=tenant femten wantstay femstay spousehr femsphr), fe
Warning - singleton groups detected.  3156 observation(s) not used.
Warning - collinearities detected
Vars dropped:  cohort female

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =     10204                    Obs per group: min =
2
                                                               avg =
5.5
                                                               max =
9

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =
56541
                                                      F(  8, 46329) =
1620.00
                                                      Prob > F      =
0.0000
Total (centered) SS     =  5533.303524                Centered R2   =
0.1403
Total (uncentered) SS   =  5533.303524                Uncentered R2 =
0.1403
```

```
Residual SS             =  4757.196366                Root MSE      =
.3204

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      jbempt |  -.0325755   .0166258    -1.96   0.050    -.0651613
.0000104
       hgage |   .1376245   .0027534    49.98   0.000      .132228
.143021
       agesq |  -.0810061   .0061008   -13.28   0.000    -.0929634   -
.0690487
     married |   -.000207   .0073927    -0.03   0.978    -.0146965
.0142825
      degree |  -.0162425   .0226944    -0.72   0.474    -.0607228
.0282378
     further |    .056042   .0114037     4.91   0.000     .0336911
.0783929
   permanent |  -.0060852   .0139931    -0.43   0.664    -.0335112
.0213409
         nsw |   .0533724   .0195754     2.73   0.006     .0150053
.0917395
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):
35.850
                                                      Chi-sq(6) P-val =
0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):
5.978
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias
19.28
                                         10% maximal IV relative bias
11.12
                                         20% maximal IV relative bias
6.76
                                         30% maximal IV relative bias
5.15
                                         10% maximal IV size
29.18
                                         15% maximal IV size
16.23
                                         20% maximal IV size
11.72
                                         25% maximal IV size
9.38
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):
11.703
                                                      Chi-sq(5) P-val =
0.0391
```

```
------------------------------------------------------------------------------
Instrumented:         jbempt
Included instruments: hgage agesq married degree further permanent nsw
Excluded instruments: tenant femten wantstay femstay spousehr femsphr
Dropped collinear:    cohort female
------------------------------------------------------------------------------

. 
. ************************************************************
. * The Sargan test of instruments overidentification suggests *
. * that instruments are valid instruments,i.e. iv not         *
. * correlated with error terms, and the excluded instruments  *
. * are correctly excluded from the estimated equation         *
. ************************************************************
. 
. ************************************************************
. * Try dropping the tenancy instruments.                     *
. * what do you conclude ?
* 
. ************************************************************
. xtivreg2 lwage_hr hgage agesq cohort married female degree ///
>         further /*tucov*/ permanent nsw ///
>         (jbemp= spousehr femsphr), fe
Warning - singleton groups detected.  3156 observation(s) not used.
Warning - collinearities detected
Vars dropped:  cohort female

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =     10204                     Obs per group: min =
2
                                                                avg =
5.5
                                                                max =
9

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =
56541
                                                      F(  8, 46329) =
1826.71
                                                      Prob > F      =
0.0000
Total (centered) SS     =  5533.303524                Centered R2   =
0.2378
Total (uncentered) SS   =  5533.303524                Uncentered R2 =
0.2378
Residual SS             =  4217.627672                Root MSE      =
.3017
```

```
------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      jbempt |  -.0042827    .034358    -0.12   0.901     -.071623    .0630577
       hgage |   .1349434   .0038886    34.70   0.000     .1273219    .1425649
       agesq |  -.0902746   .0115492    -7.82   0.000    -.1129107   -.0676386
     married |  -.0014906   .0070978    -0.21   0.834    -.0154021    .0124209
      degree |   .0118204   .0371064     0.32   0.750    -.0609069    .0845477
     further |   .0613765   .0121881     5.04   0.000     .0374883    .0852647
   permanent |  -.0287288   .0277985    -1.03   0.301    -.0832128    .0257552
         nsw |   .0302528   .0310539     0.97   0.330    -.0306117    .0911173
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):           7.442
                                                      Chi-sq(2) P-val =    0.0242
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                3.721
Stock-Yogo weak ID test critical values: 10% maximal IV size            19.93
                                         15% maximal IV size            11.59
                                         20% maximal IV size             8.75
                                         25% maximal IV size             7.25
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):           9.384
                                                      Chi-sq(1) P-val =    0.0022
------------------------------------------------------------------------------
Instrumented:         jbempt
Included instruments: hgage agesq married degree further permanent nsw
Excluded instruments: spousehr femsphr
Dropped collinear:    cohort female
------------------------------------------------------------------------------

.
. ************************************************************
```

```
. * Now test whether job tenure is, in fact, exogenous in the  *
. * FE model. We can use a Hausman test for this, assuming our *
. * instruments are valid. Recall we need an estimator that is *
. * consistent under both H0 (tenure is exogenous) and H1      *
. * (tenure is endogenous); and we need a second estimator     *
. * that is efficient under H0, but inconsistent under H1.     *
. ************************************************************
.
. * consistent under H0 and H1 *
. xtivreg2 lwage_hr hgage agesq cohort married female degree ///
>         further /*tucov*/ permanent nsw ///
>         (jbemp= tenant femten wantstay femstay spousehr femsphr), fe
Warning - singleton groups detected.  3156 observation(s) not used.
Warning - collinearities detected
Vars dropped:  cohort female

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =     10204                     Obs per group: min =
2
                                                                 avg =
5.5
                                                                 max =
9

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =
56541
                                                      F(  8, 46329) =
1620.00
                                                      Prob > F      =
0.0000
Total (centered) SS     =  5533.303524                Centered R2   =
0.1403
Total (uncentered) SS   =  5533.303524                Uncentered R2 =
0.1403
Residual SS             =  4757.196366                Root MSE      =
.3204

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
      jbempt |  -.0325755   .0166258    -1.96   0.050    -.0651613
.0000104
       hgage |   .1376245   .0027534    49.98   0.000      .132228
.143021
       agesq |  -.0810061   .0061008   -13.28   0.000    -.0929634   -
.0690487
     married |   -.000207   .0073927    -0.03   0.978    -.0146965
.0142825
```

```
      degree |  -.0162425   .0226944    -0.72   0.474    -.0607228
.0282378
     further |    .056042   .0114037     4.91   0.000     .0336911
.0783929
   permanent |  -.0060852   .0139931    -0.43   0.664    -.0335112
.0213409
         nsw |   .0533724   .0195754     2.73   0.006     .0150053
.0917395
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):
35.850
                                                      Chi-sq(6) P-val =
0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):
5.978
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias
19.28
                                         10% maximal IV relative bias
11.12
                                         20% maximal IV relative bias
6.76
                                         30% maximal IV relative bias
5.15
                                         10% maximal IV size
29.18
                                         15% maximal IV size
16.23
                                         20% maximal IV size
11.72
                                         25% maximal IV size
9.38
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):
11.703
                                                      Chi-sq(5) P-val =
0.0391
------------------------------------------------------------------------------
Instrumented:         jbempt
Included instruments: hgage agesq married degree further permanent nsw
Excluded instruments: tenant femten wantstay femstay spousehr femsphr
Dropped collinear:    cohort female
------------------------------------------------------------------------------

.
. estimates store ivfe

.
. * efficient under H0; inconsistent under H1 *
. xtreg lwage_hr hgage agesq cohort married female degree ///
>         further /*tucov*/ permanent nsw, fe
note: cohort omitted because of collinearity
```

```
note: female omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

R-sq:  within  = 0.2404                         Obs per group: min =
1
       between = 0.1475                                        avg =
4.5
       overall = 0.1313                                        max =
9

                                                F(7,46330)         =
2095.02
corr(u_i, Xb)  = -0.8508                        Prob > F           =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
       hgage |   .1345376   .0021226    63.38   0.000     .1303773
.1386979
       agesq |  -.0916776   .0025838   -35.48   0.000    -.0967419   -
.0866134
      cohort |  (omitted)
     married |  -.0016849    .006913    -0.24   0.807    -.0152345
.0118647
      female |  (omitted)
      degree |   .0160683   .0146562     1.10   0.273    -.0126581
.0447947
     further |    .062184   .0103067     6.03   0.000     .0419827
.0823853
   permanent |  -.0321563   .0040709    -7.90   0.000    -.0401354   -
.0241773
         nsw |   .0267532   .0132479     2.02   0.043     .0007871
.0527192
       _cons |  -.6016509   .0401058   -15.00   0.000    -.6802589   -
.5230429
-------------+----------------------------------------------------------------
     sigma_u |  .85131941
     sigma_e |  .30119291
         rho |  .88875364   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(13359, 46330) =     5.80        Prob > F =
0.0000

.
. estimates store fe

.
. hausman ivfe fe
```

```
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-
V_B))
             |      ivfe          fe         Difference          S.E.
-------------+------------------------------------------------------------
-----
       hgage |    .1376245     .1345376        .0030869        .0017537
       agesq |   -.0810061    -.0916776        .0106716        .0055266
     married |    -.000207    -.0016849        .0014779        .0026197
      degree |   -.0162425     .0160683       -.0323108        .0173272
     further |     .056042      .062184        -.006142        .0048803
   permanent |   -.0060852    -.0321563        .0260712        .0133879
         nsw |    .0533724     .0267532        .0266192        .0144115
------------------------------------------------------------------------
-----
                       b = consistent under Ho and Ha; obtained from
xtivreg2
            B = inconsistent under Ha, efficient under Ho; obtained from
xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        3.83
                Prob>chi2 =      0.7986


. 
. 
. ************************************************************
. * What do you conclude? What is your best estimates of the   *
. * tenure effect (with confidence interval)?                      *
. ************************************************************
. 
. 
. ************************************************************
. * Now we are going to illustrate the Hausman Taylor Method.  *
. * Its main attraction is to allow some time-invariant        *
. * characteristics to be correlated with u(i). To identify        *
. * their coefs, we required at least as many time-varying         *
. * characteristics which are uncorrelated with u(i). We are   *
. * going to modify the wage equation to include some highest  *
. * education someone ever attained, ant try to estimate the       *
. * returns to education based on this variation across        *
. * individuals. Estimate returns using RE model for comparison*
. ************************************************************
. egen everdeg=max(degree), by(id) //ever got degree

. egen everfur=max(further), by(id) //ever got further edu

. replace everfur=0 if everdeg   //replace with highest
(16896 real changes made)

. assert everfur+everdeg<2   if everdeg != . //check only got the highest

. browse xwaveid wave edhigh degree everdeg further everfur

. 
```

```
. xtreg lwage_hr hgage agesq cohort married female ///
>         everdeg everfur jbemp /*tucov*/ permanent nsw, re

Random-effects GLS regression                   Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

R-sq:  within  = 0.2376                         Obs per group: min =
1
       between = 0.3214                                        avg =
4.5
       overall = 0.3153                                        max =
9

                                                Wald chi2(10)      =
21024.75
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
       hgage |   .1255369    .001419    88.47   0.000     .1227558
.128318
       agesq |  -.0835477     .00168   -49.73   0.000    -.0868404    -
.080255
      cohort |   .0522863    .000574    91.09   0.000     .0511613
.0534112
     married |   .0266724   .0054191     4.92   0.000     .0160513
.0372936
      female |  -.1228024   .0067963   -18.07   0.000     -.136123   -
.1094818
     everdeg |   .3524897   .0087361    40.35   0.000     .3353673
.3696121
     everfur |   .0805616   .0080655     9.99   0.000     .0647534
.0963697
      jbempt |   .0040673    .000356    11.43   0.000     .0033696
.004765
   permanent |  -.0050001   .0037427    -1.34   0.182    -.0123357
.0023355
         nsw |   .0378613   .0065569     5.77   0.000     .0250099
.0507126
       _cons |  -103.4095   1.149843   -89.93   0.000    -105.6632   -
101.1559
-------------+----------------------------------------------------------------
     sigma_u |  .34603298
     sigma_e |  .30134495
         rho |  .56870191   (fraction of variance due to u_i)
------------------------------------------------------------------------------

.
. ************************************************************
```

```
. * Assume that age is uncorrelated with u(i), but that all    *
. * other time-varying characteristics are correlated with     *
. * u(i). Does this model satisfy the identification condition?*
. * Check how strongly age and age squared are correlated      *
. * with education.                                            *
. ************************************************************
. correlate hgage agesq further degree
(obs=59697)

             |    hgage    agesq  further   degree
-------------+------------------------------------
       hgage |   1.0000
       agesq |   0.9864   1.0000
     further |   0.1998   0.1577   1.0000
      degree |   0.1108   0.0829   0.5164   1.0000


.
. ************************************************************
. * Do these correlations suggest that age and age squared are *
. * good instruments?
*
. * Estimate HT model
*
. ************************************************************
.
. xthtaylor lwage_hr hgage agesq cohort married female   ///
>         everdeg everfur jbemp /*tucov*/ permanent nsw, ///
> endog(married everdeg everfur jbemp /*tucov*/ permanent nsw)

Hausman-Taylor estimation                       Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

                                                Obs per group: min =
1
                                                               avg =
4.5
                                                               max =
9

Random effects u_i ~ i.i.d.                     Wald chi2(10)      =
17583.17
                                                Prob > chi2        =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
       hgage |   .1392252   .0018511    75.21   0.000     .1355971
.1428533
       agesq |  -.0967671   .0022981   -42.11   0.000    -.1012713    -
.092263
```

```
TVendogenous |
     married |  -.0029956    .006318    -0.47   0.635    -.0153786
.0093874
      jbempt |   .0013358   .0003971     3.36   0.001     .0005575
.0021142
   permanent |  -.0312916   .0037245    -8.40   0.000    -.0385915   -
.0239917
         nsw |    .026316   .0121152     2.17   0.030     .0025707
.0500613
TIexogenous  |
      cohort |    .050224   .0032291    15.55   0.000     .0438951
.0565528
      female |  -.4284428   .1590214    -2.69   0.007     -.740119   -
.1167667
TIendogenous |
     everdeg |   1.565475   .6697541     2.34   0.019     .2527806
2.878169
     everfur |  -2.429372   1.266983    -1.92   0.055    -4.912613
.0538686
             |
       _cons |  -98.94146   6.638212   -14.90   0.000    -111.9521   -
85.93081
-------------+----------------------------------------------------------
-----
     sigma_u |  2.9042385
     sigma_e |  .30132544
         rho |  .98934982   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----
Note:  TV refers to time varying; TI refers to time invariant.

.
. ************************************************************
. * What do you conclude?
*
. * Other time-varying characteristics can be included in the  *
. * instrument set, but only if we are satisfied they are not  *
. * correlated with the individual effect. Otherwise estimates *
. * are biased. For example, assume that the living in NSW is  *
. * not correlated with u(i) (is this plausible?):                *
. ************************************************************
.
. correlate hgage agesq nsw further degree
(obs=59697)

             |    hgage    agesq      nsw  further   degree
-------------+---------------------------------------------
       hgage |   1.0000
       agesq |   0.9864   1.0000
         nsw |   0.0235   0.0236   1.0000
     further |   0.1998   0.1577   0.0376   1.0000
      degree |   0.1108   0.0829   0.0342   0.5164   1.0000



.
. xthtaylor lwage_hr hgage agesq cohort married female ///
>             everdeg everfur jbemp /*tucov*/ permanent nsw, ///
>             endog(married everdeg everfur jbemp /*tucov*/ permanent)
```

```
Hausman-Taylor estimation                       Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

                                                Obs per group: min =
1
                                                               avg =
4.5
                                                               max =
9

Random effects u_i ~ i.i.d.                     Wald chi2(10)      =
16966.47
                                                Prob > chi2        =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
       hgage |   .1392201   .0018875    73.76   0.000     .1355207
.1429196
       agesq |   -.096762   .0023433   -41.29   0.000    -.1013549   -
.0921692
         nsw |   .0257103   .0119314     2.15   0.031     .0023253
.0490953
TVendogenous |
     married |  -.0029939   .0064428    -0.46   0.642    -.0156216
.0096339
      jbempt |   .0013366   .0004049     3.30   0.001     .0005429
.0021302
   permanent |  -.0312893   .0037981    -8.24   0.000    -.0387335   -
.0238451
TIexogenous  |
      cohort |   .0502983   .0025739    19.54   0.000     .0452536
.055343
      female |  -.4222835   .1228836    -3.44   0.001    -.6631308   -
.1814361
TIendogenous |
     everdeg |   1.535095   .5107617     3.01   0.003     .5340204
2.536169
     everfur |   -2.38077   .9807347    -2.43   0.015    -4.302975   -
.4585657
             |
       _cons |  -99.09877   5.284907   -18.75   0.000     -109.457   -
88.74055
-------------+----------------------------------------------------------------
     sigma_u |  2.2717992
     sigma_e |  .30132544
         rho |  .98271148   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
Note:  TV refers to time varying; TI refers to time invariant.

.
. ************************************************************
. * A difficulty with HT is finding good instruments from          *
. * within the model which are also uncorrelated with u(i).       *
. * You might like to consider other instruments.                  *
. ************************************************************
.
. ************************************************************
. * An alternative strategy would be use external instruments  *
. * and a conventional RE model. But all regressors that are   *
. * not instrumented must be uncorrelated with u(i).               *
. * Parental background measures are sometimes used to             *
. * instrument educational attainment (assumed correlated with *
. * children's education but not with their wage). Examine the *
. * dummy variables for father's one digit occupation. What do *
. * you notice?
*
. ************************************************************
.
. *label list afmfo62
.
. * derive the dummy variables for father's occupation *
. replace fmfo62 =. if fmfo62<0
(4650 real changes made, 4650 to missing)

. recode fmfo62(10/19=1) (else=0), gen(pamanager)
(59697 differences between fmfo62 and pamanager)

. recode fmfo62(20/29=1) (else=0), gen(paprof)
(59697 differences between fmfo62 and paprof)

. recode fmfo62(30/39=1) (else=0), gen (paassoprof)
(59697 differences between fmfo62 and paassoprof)

. recode fmfo62(40/49=1) (else=0), gen (patradeperson)
(59697 differences between fmfo62 and patradeperson)

. recode fmfo62(50/59=1) (else=0), gen (paadvclerical)
(59697 differences between fmfo62 and paadvclerical)

. recode fmfo62(60/69=1) (else=0), gen (paintclerical)
(59697 differences between fmfo62 and paintclerical)

. recode fmfo62(70/79=1) (else=0), gen (paintprodworker)
(59697 differences between fmfo62 and paintprodworker)

. recode fmfo62(80/89=1) (else=0), gen (paeleclerical)
(59697 differences between fmfo62 and paeleclerical)

. recode fmfo62(90/99=1) (else=0), gen (palabour)
(59697 differences between fmfo62 and palabour)

.
. ************************************************************
. * To deal with missing values, create a dummy variable to    *
. * indicate father's occupation missing.                           *
```

```
. ************************************************************
.
. gen pamiss=1 if pamanager==0 & paprof==0 & paassoprof==0 &
patradeperson==0 ///
>          & paadvclerical==0 & paintclerical==0 & paintprodworker==0 ///
>          & paeleclerical==0 & palabour==0
(55047 missing values generated)

. replace pamiss=0 if pamiss==.
(55047 real changes made)

.
.
. browse xwaveid wave fmfo62 pamanager paprof paassoprof ///
>         patradeperson paadvclerical paintclerical paintprodworker ///
>         paeleclerical palabour pamiss

.
. xtsum pamanager paprof paassoprof ///
>         patradeperson paadvclerical paintclerical paintprodworker ///
>         paeleclerical palabour

Variable         |      Mean   Std. Dev.       Min        Max |
Observations
-----------------+--------------------------------------------+----------
------
pamana~r overall |  .2185034   .4132343          0          1 |     N =
59697
         between |             .4075561          0          1 |     n =
13360
         within  |             .0362297  -.6703854   .7185034 | T-bar =
4.46834
                 |                                           |
paprof   overall |   .152872   .3598671          0          1 |     N =
59697
         between |             .3584852          0          1 |     n =
13360
         within  |             .0365209  -.7360169    .652872 | T-bar =
4.46834
                 |                                           |
paasso~f overall |  .2257735   .4180942          0          1 |     N =
59697
         between |             .4137833          0          1 |     n =
13360
         within  |             .0350174  -.6631154   .7257735 | T-bar =
4.46834
                 |                                           |
patrad~n overall |  .0324137   .1770976          0          1 |     N =
59697
         between |             .1812939          0          1 |     n =
13360
         within  |             .0108244  -.8564752   .5324137 | T-bar =
4.46834
                 |                                           |
paadvc~l overall |   .053604   .2252365          0          1 |     N =
59697
         between |             .2167983          0          1 |     n =
13360
```

```
         within  |             .0201606  -.8352849    .553604 | T-bar =
4.46834
                 |                                            |
paintc~l overall |  .0426152   .2019897          0          1 |     N =
59697
         between |             .2034548          0          1 |     n =
13360
         within  |             .0181169  -.8462737   .5426152 | T-bar =
4.46834
                 |                                            |
paintp~r overall |  .1051309   .3067246          0          1 |     N =
59697
         between |             .3072686          0          1 |     n =
13360
         within  |             .0263084   -.783758   .6051309 | T-bar =
4.46834
                 |                                            |
paelec~l overall |  .0911939    .287887          0          1 |     N =
59697
         between |             .2892532          0          1 |     n =
13360
         within  |             .0174062   -.797695   .5911939 | T-bar =
4.46834
                 |                                            |
palabour overall |         0          0          0          0 |     N =
59697
         between |                    0          0          0 |     n =
13360
         within  |                    0          0          0 | T-bar =
4.46834

. 
. ************************************************************
. * check correlations of occupational dummies and education   *
. ************************************************************
. corr everfur everdeg pamanager paprof paassoprof ///
>         patradeperson paadvclerical paintclerical paintprodworker ///
>         paeleclerical /*palabour*/ pamiss
(obs=59697)

             |  everfur  everdeg pamana~r   paprof paasso~f patrad~n
paadvc~l paintc~l paintp~r paelec~l   pamiss
-------------+------------------------------------------------------------
-------------------------------------
     everfur |   1.0000
     everdeg |  -0.4442   1.0000
   pamanager |  -0.0256   0.0701   1.0000
      paprof |  -0.0912   0.1871  -0.2246   1.0000
  paassoprof |   0.0644  -0.0724  -0.2855  -0.2294   1.0000
patradeper~n |   0.0084  -0.0226  -0.0968  -0.0778  -0.0988   1.0000
paadvcleri~l |  -0.0051   0.0134  -0.1258  -0.1011  -0.1285  -0.0436
1.0000
paintcleri~l |   0.0044  -0.0105  -0.1116  -0.0896  -0.1139  -0.0386  -
0.0502   1.0000
paintprodw~r |   0.0353  -0.1013  -0.1812  -0.1456  -0.1851  -0.0627  -
0.0816  -0.0723   1.0000
paelecleri~l |   0.0266  -0.0838  -0.1675  -0.1346  -0.1711  -0.0580  -
0.0754  -0.0668  -0.1086   1.0000
```

```
      pamiss |  -0.0122  -0.0290  -0.1537  -0.1235  -0.1570  -0.0532  -
0.0692  -0.0613  -0.0996  -0.0921   1.0000


.
.
. * estimate the model *
. xtivreg lwage_hr hgage agesq cohort married female   ///
>           jbemp /*tucov*/ permanent ///
> (everdeg everfur=pamanager paprof paassoprof ///
>         patradeperson paadvclerical paintclerical paintprodworker ///
>         paeleclerical palabour pamiss), re
note: palabour omitted because of collinearity
note: pamiss omitted because of collinearity

G2SLS random-effects IV regression              Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

R-sq:  within  = 0.2371                         Obs per group: min =
1
       between = 0.2413                                        avg =
4.5
       overall = 0.2281                                        max =
9

                                                Wald chi2(9)       =
18025.44
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
     everdeg |   .8228528   .1120266     7.35   0.000     .6032846
1.042421
     everfur |   .6369579   .2416616     2.64   0.008     .1633099
1.110606
       hgage |   .1140456   .0047394    24.06   0.000     .1047564
.1233347
       agesq |  -.0677138   .0060517   -11.19   0.000    -.0795748   -
.0558528
      cohort |   .0553492   .0010936    50.61   0.000     .0532058
.0574927
     married |   .0202976   .0058626     3.46   0.001     .0088071
.0317881
      female |  -.1025759   .0168636    -6.08   0.000     -.135628   -
.0695238
      jbempt |   .0033917   .0003791     8.95   0.000     .0026488
.0041346
   permanent |  -.0184497   .0048268    -3.82   0.000    -.0279101   -
.0089892
       _cons |  -109.5419   2.192259   -49.97   0.000    -113.8387   -
105.2452
```

```
-------------+----------------------------------------------------------------
     sigma_u |  .41916117
     sigma_e |  .30136427
         rho |  .65923154   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   everdeg everfur
Instruments:    hgage agesq cohort married female jbempt permanent
pamanager paprof paassoprof patradeperson paadvclerical paintclerical
paintprodworker
                paeleclerical
------------------------------------------------------------------------------

. 
. ************************************************************
. * Test for exogeneity of education. What is your overall     *
. * conclusion?
* 
. ************************************************************
. 
. est store ivre

. 
. xtreg lwage_hr hgage agesq cohort married female   ///
>       everdeg everfur jbemp /*tucov*/ permanent, re

Random-effects GLS regression                   Number of obs      =
59697
Group variable: id                              Number of groups   =
13360

R-sq:  within  = 0.2376                         Obs per group: min =
1
       between = 0.3200                                        avg =
4.5
       overall = 0.3138                                        max =
9

                                                Wald chi2(9)       =
20976.19
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =
0.0000

------------------------------------------------------------------------------
    lwage_hr |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
       hgage |   .1253424   .0014192    88.32   0.000     .1225608
.128124
       agesq |  -.0833992   .0016807   -49.62   0.000    -.0866934    -
.080105
      cohort |   .0522023    .000574    90.95   0.000     .0510773
.0533272
```

```
     married |    .026797   .0054214     4.94   0.000     .0161712
.0374228
      female |  -.1227188   .0068033   -18.04   0.000     -.136053   -
.1093846
     everdeg |   .3542426   .0087398    40.53   0.000      .337113
.3713723
     everfur |   .0814384   .0080723    10.09   0.000      .065617
.0972598
      jbempt |   .0041058    .000356    11.53   0.000      .003408
.0048036
   permanent |  -.0049332   .0037435    -1.32   0.188    -.0122703
.0024039
       _cons |  -103.2293   1.149735   -89.79   0.000    -105.4827   -
100.9759
-------------+----------------------------------------------------------
-----
     sigma_u |    .346451
     sigma_e |  .30135451
         rho |  .56927851   (fraction of variance due to u_i)
------------------------------------------------------------------------
-----

.
. est store re

.
. hausman ivre re

                 ---- Coefficients ----
              |      (b)          (B)            (b-B)     sqrt(diag(V_b-
V_B))
              |      ivre          re         Difference          S.E.
-------------+----------------------------------------------------------
-----
     everdeg |    .8228528     .3542426        .4686101        .1116852
     everfur |    .6369579     .0814384        .5555195        .2415267
       hgage |    .1140456     .1253424       -.0112968         .004522
       agesq |   -.0677138    -.0833992        .0156854        .0058136
      cohort |    .0553492     .0522023         .003147        .0009309
     married |    .0202976      .026797       -.0064994        .0022313
      female |   -.1025759    -.1227188        .0201429        .0154304
      jbempt |    .0033917     .0041058       -.0007141        .0001301
   permanent |   -.0184497    -.0049332       -.0135165        .0030471
------------------------------------------------------------------------
-----
                           b = consistent under Ho and Ha; obtained from
xtivreg
            B = inconsistent under Ha, efficient under Ho; obtained from
xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      127.79
                Prob>chi2 =      0.0000

.
. ************************************************************
```

```
. * Save data for Exercise 5.
*
. ************************************************************
. save "`working'\panelex5.dta", replace
file C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\panelex5.dta saved

.
. set more on

.
. log c
      name:  <unnamed>
       log:  C:\Users\nw\Documents\HILDA Project\Training\Melbourne Nov
2011\Exercises\ex4.log
  log type:  text
 closed on:   8 Nov 2011, 15:11:12
-------------------------------------------------------------------------
-------------------------------------------------------------------------
------------
```