Answers to Computer Exercise 8

Summarise the data set.

```
. su

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        doim |     10016     9.70008    1.320034        -9         12
        dobm |     10016    6.322484    3.426489         1         12
        doby |     10016    1955.523    9.603142      1925       1975
         sex |     10016    1.471246    .4991974         1          2
        race |     10016    -6.86272    3.019825        -8          9
-------------+--------------------------------------------------------
      jbsize |     10016    5.298722    2.376668        -9         11
       jbhrs |     10016    34.89277    9.753727         2         97
         pid |     10016    1.47e+07     2531708   1.00e+07   1.91e+07
         age |     10016    38.76697    9.872366        16         72
      region |     10016    8.232428    5.441074        -9         18
-------------+--------------------------------------------------------
        year |     10016        94.5    2.291402        91         98
     hourpay |     10016    8.326613    4.867058   .5034965   92.30769
      tenure |     10016    5.274661    5.801229         0         38
       edage |     10016    17.78275    3.172386        10         33
  experience |     10016    20.98423    10.66915         0         58
-------------+--------------------------------------------------------
      london |     10016    .0873602    .2823763         0          1
         rpi |     10015    .8575959    .0545567   .7766879    .949221
     yearsed |     10016    12.78275    3.172386         5         28
```

In particular work out the average hourly wage, age of respondents and the proportion living in London. Then compare these average values in 1991 and 1998. How have these averages changed over time?

```
. su hourpay age london if year==91

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     hourpay |      1252    6.903067    4.368866   .5566802   92.30769
         age |      1252    35.30272    9.612154         16         65
      london |      1252    .0950479    .2933983          0          1

since London is a proportion, standard errors here are wrong (need standard
error of a proportion given by)

. ci london if year==91

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
      london |       1252    .0950479    .0082919        .0787803    .1113155

. su hourpay age london if year==98

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     hourpay |      1252    9.765613    5.337395   .6813187   44.17582
         age |      1252     42.2492    9.604249         23         72
      london |      1252    .0822684     .274883          0          1
```

```
. ci london if year==98

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
      london |       1252    .0822684    .0077687        .0670273    .0975094
```

The (nominal) hourly wage has risen by 42% (9/77/6.90) over the seven year sample period. The average age of the panel has risen by 7 years (not surprising given that this is panel data) – this may also help explain the rise in the average wage (wages tend to rise with age and inflation) and the proportion living in London has fallen over the period (though mot significantly). Since this is a panel suggests net out-migration from London.

Test whether the returns to education have changed over the period.

1st estimate wage equation pooled over all 8 years of the sample by OLS. The estimated coefficients give average effect over the sample period.

```
. g exper2=experience^2

. g lhw=log(hourpay)


. reg lhw experience exper2 yearsed london

      Source |       SS       df       MS              Number of obs =   10016
-------------+------------------------------           F(  4, 10011) =  611.33
       Model |   528.22826     4  132.057065           Prob > F      =  0.0000
    Residual |  2162.53125 10011  .216015508           R-squared     =  0.1963
-------------+------------------------------           Adj R-squared =  0.1960
       Total |  2690.75951 10015  .268672942           Root MSE      =  .46477
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0437931   .0016754    26.14   0.000     .0405089    .0470773
      exper2 |  -.0008205   .0000357   -22.96   0.000    -.0008905   -.0007504
     yearsed |   .0641491   .0016064    39.93   0.000     .0610003    .0672978
      london |   .1918926   .0166147    11.55   0.000     .1593243    .2244608
       _cons |   .6817042   .0296455    23.00   0.000     .6235931    .7398153
------------------------------------------------------------------------------
```

So the return to 1 extra year of education dLnhourlyWage/dYears_education = .064
1 year extra education adds 6.4% to hourly pay

Now generate interaction terms to pick up how return to education changes over the sample period

To generate dummy variables (one for each year)
tab year, gen(d)

Then create the interaction terms

```
. g ed91=yearsed*d91
. g ed92=yearsed*d92
. g ed93=yearsed*d93
. g ed94=yearsed*d94
. g ed95=yearsed*d95
. g ed96=yearsed*d96
. g ed97=yearsed*d97
. g ed98=yearsed*d98
```

```
. reg lhw experience exper2 yearsed london ed92-ed98

      Source |       SS       df       MS              Number of obs =   10016
-------------+------------------------------           F( 11, 10004) =  264.80
       Model |  606.784057    11   55.162187           Prob > F      =  0.0000
    Residual |  2083.97546 10004   .20831422           R-squared     =  0.2255
-------------+------------------------------           Adj R-squared =  0.2247
       Total |  2690.75951 10015  .268672942           Root MSE      =  .45641
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0389767   .0016648    23.41   0.000     .0357133    .0422401
      exper2 |    -.00076   .0000352   -21.56   0.000    -.0008291   -.0006909
     yearsed |   .0496928   .0018779    26.46   0.000     .0460118    .0533738
      london |   .2006257   .0163222    12.29   0.000     .1686308    .2326205
        ed92 |   .0056759   .0013858     4.10   0.000     .0029594    .0083923
        ed93 |   .0068445   .0013878     4.93   0.000      .004124    .0095649
        ed94 |   .0101975   .0013909     7.33   0.000      .007471    .0129239
        ed95 |   .0134783   .0013949     9.66   0.000     .0107441    .0162126
        ed96 |    .015498   .0013994    11.08   0.000      .012755    .0182411
        ed97 |   .0193321   .0014044    13.77   0.000     .0165792     .022085
        ed98 |   .0221553   .0014101    15.71   0.000     .0193913    .0249193
       _cons |   .7843714   .0295964    26.50   0.000     .7263565    .8423862
------------------------------------------------------------------------------
```

Coefficients on interaction terms give additional impact of an extra year of education relative to the base year effect which is captured by yearsed

Eg. dLnWage/dYears_education = .050 in 1991 (base year)
dLnWage/dYears_education = .050 + .006 in 1992
(so additional return is coefficient on interaction term)
dLnWage/dYears_education = .050 + .022 in 1998

So (nominal) returns to education appear to have risen through the 1990s (though this may be because the model does not include separate year dummies, so part of the effect is the nominal rather than the real effect)

Dependent variable is the nominal hourly wage. Since nominal wages tend to grow over time just to keep up with inflation, we are more interested in the effect of education on real (inflation adjusted) wages.

There are two ways of allowing for price changes with pooled data.

```
. g rhwage=hourpay/rpi
```

/* create a real wage variable by dividing the nominal wage by the price index for the relevant year*/

Then regress log of the real wage on rhs variables

```
. g lrhwage=log(rhwage)

. reg lrhw experience exper2 yearsed london d92-d98

      Source |       SS       df       MS              Number of obs =   10016
-------------+------------------------------           F( 11, 10004) =  226.25
       Model |  516.111598    11  46.9192362           Prob > F      =  0.0000
    Residual |  2074.63057 10004  .207380105           R-squared     =  0.1992
-------------+------------------------------           Adj R-squared =  0.1983
       Total |  2590.74217 10015  .258686188           Root MSE      =  .45539
------------------------------------------------------------------------------
```

```
        lrhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0397353   .0016543    24.02   0.000     .0364926    .0429779
      exper2 |  -.0007812   .0000351   -22.27   0.000      -.00085   -.0007124
     yearsed |   .0610771    .001581    38.63   0.000     .0579779    .0641763
      london |   .2010158   .0162851    12.34   0.000     .1690938    .2329377
         d92 |   .0430992   .0182096     2.37   0.018     .0074048    .0787936
         d93 |   .0483373   .0182344     2.65   0.008     .0125942    .0840804
         d94 |   .0707984   .0182719     3.87   0.000     .0349818    .1066149
         d95 |   .0804617   .0183212     4.39   0.000     .0445485    .1163749
         d96 |   .0868127   .0183781     4.72   0.000      .050788    .1228374
         d97 |   .1029824   .0184439     5.58   0.000     .0668286    .1391361
         d98 |   .1122749   .0185197     6.06   0.000     .0759726    .1485772
       _cons |   .8711056   .0306492    28.42   0.000      .811027    .9311841
------------------------------------------------------------------------------
```

Compare this with a regression of the log of nominal hourly wages

```
. reg lhw experience exper2 yearsed london d92-d98

      Source |       SS       df       MS              Number of obs =   10016
-------------+------------------------------           F( 11, 10004) =  269.97
       Model |  615.909544    11  55.9917768           Prob > F      =  0.0000
    Residual |  2074.84997 10004  .207402036           R-squared     =  0.2289
-------------+------------------------------           Adj R-squared =  0.2281
       Total |  2690.75951 10015  .268672942           Root MSE      =  .45541
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0397511   .0016543    24.03   0.000     .0365082    .0429939
      exper2 |  -.0007813   .0000351   -22.27   0.000    -.0008501   -.0007125
     yearsed |   .0610812   .0015811    38.63   0.000     .0579819    .0641805
      london |   .2007183   .0162859    12.32   0.000     .1687946    .2326419
         d92 |   .0756177   .0182105     4.15   0.000     .0399214     .111314
         d93 |    .095826   .0182354     5.25   0.000      .060081    .1315711
         d94 |   .1416405   .0182729     7.75   0.000     .1058221     .177459
         d95 |   .1836828   .0183222    10.03   0.000     .1477677    .2195979
         d96 |   .2158445    .018379    11.74   0.000     .1798179    .2518711
         d97 |   .2674807   .0184449    14.50   0.000      .231325    .3036363
         d98 |   .3078367   .0185207    16.62   0.000     .2715325     .344141
       _cons |   .6222571   .0306508    20.30   0.000     .5621753    .6823388
------------------------------------------------------------------------------
```

Can see the coefficients on the original rhs variables are the same but the coefficients on the year dummies are different

The reason is that log(hourpay/rpi) = log(hourpay) – log(rpi)

Since log(rpi) is approximately constant (for each year) – it varies a little because individuals are interviewd in 3 different months of each year - the effect is to subsume it into the constant for each year (ie the year-specific dummy variables) Hence the remaining variance is the same whether log(hourpay) or log(real wage) is used. Hence the experience and education coefficients are unchanged, but the year dummies differ across the specification.

Hence when studying the effects of age or education on wages you do not need to worry about whether real or nominal wages are used as a dependent variable as long as the 1) the dependent variable is in logs and 2) year dummies are included in the list of rhs variables.

Now estimate (1) – without the year interaction dummies – by within groups and by random effects.

By pooled OLS

```
reg lhw experience exper2 yearsed london
```

```
      Source |       SS       df       MS              Number of obs =   10016
-------------+------------------------------           F(  4, 10011) =  611.33
       Model |   528.22826     4  132.057065           Prob > F      =  0.0000
    Residual |  2162.53125 10011  .216015508           R-squared     =  0.1963
-------------+------------------------------           Adj R-squared =  0.1960
       Total |  2690.75951 10015  .268672942           Root MSE      =  .46477
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0437931   .0016754    26.14   0.000     .0405089    .0470773
      exper2 |  -.0008205   .0000357   -22.96   0.000    -.0008905   -.0007504
     yearsed |   .0641491   .0016064    39.93   0.000     .0610003    .0672978
      london |   .1918926   .0166147    11.55   0.000     .1593243    .2244608
       _cons |   .6817042   .0296455    23.00   0.000     .6235931    .7398153
------------------------------------------------------------------------------
```

with panel data estimates

1) Within Groups

```
xtreg lhw experience exper2 yearsed london, fe i(pid)
```

```
Fixed-effects (within) regression               Number of obs      =     10016
Group variable (i): pid                         Number of groups   =      1252

R-sq:  within  = 0.2468                         Obs per group: min =         8
       between = 0.0011                                        avg =       8.0
       overall = 0.0005                                        max =         8

                                                F(3,8761)          =    956.82
corr(u_i, Xb)  = -0.7154                        Prob > F           =    0.0000
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |   .0755017   .0021044    35.88   0.000     .0713766    .0796269
      exper2 |  -.0006486   .0000448   -14.47   0.000    -.0007364   -.0005607
     yearsed |  (dropped)
      london |   .0027659    .035285     0.08   0.938     -.066401    .0719328
       _cons |   .7575866   .0251147    30.17   0.000     .7083558    .8068173
-------------+----------------------------------------------------------------
     sigma_u |  .68421578
     sigma_e |  .21418521
         rho |  .91075304   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1251, 8761) =    30.68          Prob > F = 0.0000
```

2) Random Effects

```
. xtreg lhw experience exper2 yearsed london, re i(pid)

Random-effects GLS regression                   Number of obs      =     10016
Group variable (i): pid                         Number of groups   =      1252

R-sq:  within  = 0.2386                         Obs per group: min =         8
       between = 0.1170                                        avg =       8.0
       overall = 0.1209                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(4)       =   2312.83
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         lhw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  experience |    .065283   .0019663    33.20   0.000     .0614292    .0691369
      exper2 |    -.00077   .0000419   -18.36   0.000    -.0008521   -.0006878
     yearsed |   .0970992   .0039283    24.72   0.000     .0893999    .1047985
      london |   .0452169   .0280717     1.61   0.107    -.0098027    .1002365
       _cons |  -.2056268   .0590953    -3.48   0.001    -.3214515   -.0898022
-------------+----------------------------------------------------------------
     sigma_u |   .4024581
     sigma_e |  .21418521
         rho |  .77928417   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Note that variables which are constant over time are now estimated in the random effects estimation (in fixed effects the effect of constant variables can not be separated from that of the unobserved fixed effects)

Also coefficients on both experience and education are higher when account for unobserveables. Suggests OLS estimates were biased down by omitted variable bias (and hence that correlation between unobserveables and education is positive – see lecture notes on omitted variable bias)

To run Hausman test , do following commands

```
xtreg lwage exper exp2 yearsed married nonwhite, re i(id)
est store fixed
xtreg lhw experience exper2 yearsed london, re i(pid)
. hausman fixed

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
  experience |    .0755017      .065283        .0102187        .0007498
      exper2 |   -.0006486      -.00077        .0001214        .0000158
      london |    .0027659     .0452169        -.042451        .0213778
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      856.50
                Prob>chi2 =      0.0000
```

So Chi-squared test value exceeds critical value at 5% level. Reject null hypothesis that coefficients are the same in the two estimation techniques. Conclude can reject (not accept) hypothesis that unobserveables are uncorrelated with the right hand side variables ie in this case a fixed effect estimation strategy is the better choice.

(note that this test can't estimate the difference in the yearsed variable – or any variable that is constant)

Why is the years of education variable dropped but the dummy variable "London" is retained in the fixed effect estimates?

Variables which are constant over time are dropped from the within-groups (and first differenced) estimates. This explains why years of education is dropped – it does not vary over the sample period. However the London dummy variable is not dropped. This can only be because some individuals move house over the sample period and hence either move in or out of London which has the effect of making the London dummy not constant across individuals over time.