BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Ms. Yu, Fall 2003

Problem Set 6 Solutions

Problem sets should be your own work. You may work together with classmates, but if you're not figuring this out on your own, you will eventually regret it.

1. (**7.13**)

```
 . use http://fmwww.bc.edu/ec-p/data/wooldridge/CEOSAL1

. gen rosneg=(ros<0)

. browse

. regress  lsalary lsales roe rosneg

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  3,   205) =   28.81
       Model |  19.7902034     3  6.59673446           Prob > F      =  0.0000
    Residual |  46.9319665   205  .228936422           R-squared     =  0.2966
-------------+------------------------------           Adj R-squared =  0.2863
       Total |  66.7221699   208  .320779663           Root MSE      =  .47847

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lsales |   .2883868   .0336172     8.58   0.000      .222107    .3546666
         roe |   .0166571   .0039681     4.20   0.000     .0088336    .0244806
      rosneg |  -.2256748    .109338    -2.06   0.040     -.441246   -.0101036
       _cons |   4.297602   .2932526    14.65   0.000     3.719424     4.87578
------------------------------------------------------------------------------
```

The estimated equation is

$$\begin{aligned} \log(\widehat{salary}) &= \underset{(.29)}{4.30} + \underset{(.034)}{.228}\log(sales) + \underset{(.0040)}{.167}\, roe - \underset{(.109)}{.226}\, rosneg \\ n &= 209, R^2 = .297, \overline{R}^2 = .286. \end{aligned}$$

The coefficient on *rosneg* implies that if the CEO's firm had a negative return on its stock over the 1988 to 1990 period, the CEO salary was predicted to be about 22.6% lower, for given levels of *sales* and *roe*. The t statistic is about -2.07, which is significant at the 5% level against a two-sided alternative.

2. (**7.14**)

(i)
```
. by male, sort: regress  sleep totwrk educ age agesq yngkid

------------------------------------------------------------------------------
-> male = 0
```

```
      Source |       SS       df       MS              Number of obs =     306
-------------+------------------------------           F(  5,   300) =    6.50
       Model |  6201576.18     5  1240315.24           Prob > F      =  0.0000
    Residual |  57288575.9   300   190961.92           R-squared     =  0.0977
-------------+------------------------------           Adj R-squared =  0.0826
       Total |  63490152.1   305  208164.433           Root MSE      =  436.99

------------------------------------------------------------------------------
       sleep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      totwrk |  -.1399495   .0276594    -5.06   0.000    -.1943806   -.0855184
        educ |  -10.20514   9.588848    -1.06   0.288    -29.07506    8.664786
         age |  -30.35657   18.53091    -1.64   0.102    -66.82361    6.110463
       agesq |   .3679406   .2233398     1.65   0.101    -.0715705    .8074516
      yngkid |  -118.2826   93.18757    -1.27   0.205    -301.6666    65.10153
       _cons |   4238.729   384.8923    11.01   0.000     3481.299     4996.16
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> male = 1

      Source |       SS       df       MS              Number of obs =     400
-------------+------------------------------           F(  5,   394) =   14.59
       Model |  11806161.6     5  2361232.32           Prob > F      =  0.0000
    Residual |  63763979.0   394   161837.51           R-squared     =  0.1562
-------------+------------------------------           Adj R-squared =  0.1455
       Total |  75570140.6   399  189398.849           Root MSE      =  402.29

------------------------------------------------------------------------------
       sleep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      totwrk |  -.1821232   .0244855    -7.44   0.000    -.2302618   -.1339846
        educ |  -13.05238   7.414218    -1.76   0.079    -27.62876    1.523995
         age |   7.156591   14.32037     0.50   0.618    -20.99731    35.31049
       agesq |  -.0447674   .1684053    -0.27   0.791    -.3758528     .286318
      yngkid |   60.38021   59.02278     1.02   0.307    -55.65877    176.4192
       _cons |   3648.208   310.0393    11.77   0.000      3038.67    4257.747
------------------------------------------------------------------------------
```

The estimated equation for men is

$$\begin{aligned}\widehat{sleep} &= \underset{(310.0)}{3,648.2} - \underset{(.024)}{.182}\ totwrk - \underset{(7.41)}{13.05}\ educ + \underset{(14.32)}{7.16}\ age - \underset{(.1684)}{.0448}\ age^2 + \underset{(59.02)}{60.38}\ yngkid \\ n &= 400, R^2 = .156.\end{aligned}$$

The estimated equation for women is

$$\begin{aligned}\widehat{sleep} &= \underset{(384.9)}{4,238.7} - \underset{(.028)}{.140}\ totwrk - \underset{(9.59)}{10.21}\ educ - \underset{(18.53)}{30.36}\ age - \underset{(.223)}{.368}\ age^2 - \underset{(93.19)}{118.28}\ yngkid \\ n &= 306, R^2 = .098.\end{aligned}$$

There are certainly notable differences in the point estimates. For example, having a young child in the household leads to less lseep for women (about two hours a week) while men are estimated to sleep about an hour more. The quadratic in *age* is a hump-shape for men but a U-shape for women. The intercepts for men and women are also notably different.

(ii)
```
. gen  maletotwrk= male* totwrk

. gen  maleeduc= male* educ

. gen maleage= male* age

. gen maleagesq=male*agesq

. gen maleyugkid=male*yngkid

. regress  sleep totwrk educ age agesq yngkid  male maletotwrk
maleeduc maleage maleagesq maleyugkid

      Source |       SS       df       MS              Number of obs =     706
-------------+------------------------------           F( 11,   694) =    9.48
       Model |  18187280.8    11  1653389.17           Prob > F      =  0.0000
    Residual |   121052555   694  174427.313           R-squared     =  0.1306
-------------+------------------------------           Adj R-squared =  0.1168
       Total |   139239836   705  197503.313           Root MSE      =  417.64

------------------------------------------------------------------------------
       sleep |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      totwrk |  -.1399495   .0264349    -5.29   0.000    -.1918514   -.0880476
        educ |  -10.20514   9.164321    -1.11   0.266    -28.19826    7.787983
         age |  -30.35657   17.71049    -1.71   0.087    -65.12914    4.415998
       agesq |   .3679406   .2134519     1.72   0.085    -.0511483    .7870294
      yngkid |  -118.2826   89.06187    -1.33   0.185    -293.1456    56.58046
        male |  -590.5211   488.7916    -1.21   0.227    -1550.209    369.1665
  maletotwrk |  -.0421737    .036674    -1.15   0.251     -.114179    .0298317
    maleeduc |  -2.847243   11.96795    -0.24   0.812    -26.34497    20.65048
     maleage |   37.51316   23.12332     1.62   0.105    -7.886887    82.91321
   maleagesq |  -.4127079   .2759136    -1.50   0.135    -.9544333    .1290175
  maleyugkid |   178.6628   108.1051     1.65   0.099     -33.5895     390.915
       _cons |   4238.729   367.8519    11.52   0.000     3516.493    4960.965
------------------------------------------------------------------------------

. test  male maletotwrk maleeduc maleage maleagesq maleyugkid

 ( 1)  male = 0.0
 ( 2)  maletotwrk = 0.0
 ( 3)  maleeduc = 0.0
 ( 4)  maleage = 0.0
 ( 5)  maleagesq = 0.0
 ( 6)  maleyugkid = 0.0
```

```
        F(  6,   694) =    2.12
             Prob > F =    0.0495
```

The F statistic (with 6 and 694 *df*) is about 2.12 with p-value $\approx .05$, and so we reject the null that sleep equations are the same at the 5% level.

(iii)
```
. test  maletotwrk maleeduc maleage maleagesq maleyugkid

 ( 1)  maletotwrk = 0.0
 ( 2)  maleeduc = 0.0
 ( 3)  maleage = 0.0
 ( 4)  maleagesq = 0.0
 ( 5)  maleyugkid = 0.0

       F(  5,   694) =    1.26
            Prob > F =    0.2814
```

If we leave the coefficient on *male* unspecified under H_0, and test only the five interaction terms, $male \cdot totwrk$, $male \cdot educ$, $male \cdot age$, $male \cdot age^2$, and $male \cdot yngkid$, the F statistic (with 5 and 694 *df*) is about 1.26 and p-value $\approx .28$.

(iv) The outcome of the test in part (iii) shows that, once an intercept difference is allowed, there is not strong evidence of slope differences between men and women. this is one of those cases where the practically important differences in estimates for women and men in part (i) do not translate into statistically significant differences. We apparently need a larger sample size to determine whether there are differences in slopes. For the purposes of studying the sleep-work tradeoff, the original model with *male* added as an explanatory variable seems sufficient.

3. (7.15)

(i) When $educ = 12.5$, the approximate proportionate difference in estimated *wage* between women and men is $-.227 - .0056(12.5) = -.297$. When $educ = 0$, the difference is $-.227$. So the differential at 12.5 years of education is about 7 percentage points greater.

(ii)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/WAGE1

. gen femaleeduc1=female*(educ-12.5)

. regress  lwage female educ femaleeduc1 exper expersq tenure
tenursq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  7,   518) =   58.37
       Model |  65.4081526     7  9.3440218            Prob > F      =  0.0000
    Residual |  82.9216091   518  .160080326           R-squared     =  0.4410
-------------+------------------------------           Adj R-squared =  0.4334
       Total |  148.329762   525  .28253288            Root MSE      =   .4001
```

```
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.296345   .0358358    -8.27   0.000    -.3667465   -.2259436
        educ |   .0823692   .0084699     9.72   0.000     .0657296    .0990088
 femaleeduc1 |  -.0055645   .0130618    -0.43   0.670    -.0312252    .0200962
       exper |   .0293366   .0049842     5.89   0.000      .019545    .0391283
     expersq |  -.0005804   .0001075    -5.40   0.000    -.0007916   -.0003691
      tenure |   .0318967    .006864     4.65   0.000      .018412    .0453814
     tenursq |    -.00059   .0002352    -2.51   0.012    -.001052    -.000128
       _cons |    .388806   .1186871     3.28   0.001     .1556388    .6219733
------------------------------------------------------------------------------
```

We can write the model underlying (7.18) as

$$\begin{aligned}\log(wage) &= \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female \cdot educ + otherfactors \\ &= \beta_0 + (\delta_0 + 12.5\delta_1) female + \beta_1 educ + \delta_1 female \cdot (educ - 12.5) + otherfactors \\ &= \beta_0 + \theta_0 female + \beta_1 educ + \delta_1 female \cdot (educ - 12.5) + otherfactors,\end{aligned}$$

where $\theta_0 = \delta_0 + 12.5\delta_1$ is the gender differential at 12.5 years of education. When we run this regression we obtain about $-.294$ as the coefficient on $female$ (which differs from $-.297$ due to rounding error). Its standard error is about .036.

(iii) The t statistic on $female$ from part (ii) is about -8.17, which is very significant. This is because we are estimating the gender differential at a reasonable number of years of education, 12.5 which is close to the average. In equation (7.18), the coefficient on $female$ is the gender differential when $educ = 0$. There are no people of either gender with close to zero years of education, and so we cannot hope – nor do we want to – to estimate the gender differential at $educ = 0$.

4. (7.19)

(i)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge2k/401KSUBS-10

. summ nettfa

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      nettfa |       928    21.18766    74.44089   -121.472   1462.115
```

The average is 21.188 , the minimum is -1221.472 , the maximum is 1462.115 .

(ii)
```
. regress nettfa e401k

      Source |       SS       df       MS              Number of obs =     928
-------------+------------------------------           F(  1,   926) =   28.89
       Model |  155419.609     1  155419.609           Prob > F      =  0.0000
```

```
    Residual |  4981501.04   926  5379.59076           R-squared     =  0.0303
-------------+------------------------------           Adj R-squared =  0.0292
       Total |  5136920.65   927  5541.44622           Root MSE      =  73.346

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       e401k |   26.21824   4.877813     5.37   0.000     16.64538    35.79109
       _cons |   10.16922   3.162157     3.22   0.001     3.963395    16.37505
------------------------------------------------------------------------------
```

This can be easily done by regressing $nettfa$ on $e401k$ and doing a t test on $\hat{\beta_{ec401k}}$; the estimate is the average difference in $nettfa$ for those eligible for a 401(k) and those not eligible. Using the 928 observation gives $\hat{\beta_{ec401k}} = 26.218$ and $t_{e401k} = 4.878$. Therefore, we strongly reject the null hypothesis that there is no difference in the average. The coefficient implies that, on average, a family eligible for a 401(k) plan has 26,218 more on net total financial assets.

(iii)

```
. regress nettfa e401k inc incsq age agesq male

      Source |       SS       df       MS              Number of obs =     928
-------------+------------------------------           F(  6,   921) =   47.36
       Model |  1211139.92     6  201856.653           Prob > F      =  0.0000
    Residual |  3925780.74   921   4262.5198           R-squared     =  0.2358
-------------+------------------------------           Adj R-squared =  0.2308
       Total |  5136920.65   927  5541.44622           Root MSE      =  65.288

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       e401k |   14.21904   4.590444     3.10   0.002       5.2101    23.22799
         inc |  -.5482641    .253173    -2.17   0.031    -1.045127   -.0514011
       incsq |   .0140768   .0019759     7.12   0.000     .0101989    .0179546
         age |  -2.567236   1.818878    -1.41   0.158    -6.136862    1.002391
       agesq |   .0428191   .0209215     2.05   0.041     .0017597    .0838786
        male |    .201791   5.470784     0.04   0.971    -10.53486    10.93844
       _cons |   34.81393   37.44084     0.93   0.353    -38.66533    108.2932
------------------------------------------------------------------------------
```

The equation estimated by OLS is

$$
\begin{aligned}
\widehat{nettfa} &= \underset{(37.44)}{34.814} + \underset{(4.59)}{14.219}\, e401k - \underset{(.253)}{.548}\, inc + \underset{(.0020)}{.014}\, inc^2 - \underset{(1.819)}{2.567}\, age \\
&\quad + \underset{(.021)}{.0428}\, age^2 + \underset{(5.47)}{.202}\, male \\
n &= 928, R^2 = .236.
\end{aligned}
$$

Now holding income and age fixed, a 401(k)-eligible family is estimated to have \$14,219 more in wealth than a non-eligible family.

(iv)

```
. gen e401kage1= e401k*(age-41)

. gen e401kage2= e401k*(age-41)^2

. regress nettfa e401k inc incsq age agesq male e401kage1
e401kage2

      Source |       SS       df       MS              Number of obs =     928
-------------+------------------------------           F(  8,   919) =   37.25
       Model |  1257734.26     8  157216.782           Prob > F      =  0.0000
    Residual |  3879186.39   919   4221.0951           R-squared     =  0.2448
-------------+------------------------------           Adj R-squared =  0.2383
       Total |  5136920.65   927  5541.44622           Root MSE      =   64.97

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       e401k |   8.357268   6.187644     1.35   0.177    -3.786285    20.50082
         inc |  -.4700326   .2532375    -1.86   0.064    -.9670235    .0269583
       incsq |   .0133709   .0019785     6.76   0.000      .009488    .0172538
         age |  -1.791962   2.264044    -0.79   0.429    -6.235259    2.651334
       agesq |    .028537   .0258394     1.10   0.270    -.0221741    .0792481
        male |   .4487733   5.445848     0.08   0.934    -10.23897    11.13651
   e401kage1 |    1.14543   .4725547     2.42   0.016      .218019    2.072842
   e401kage2 |   .0595252   .0434693     1.37   0.171    -.0257854    .1448358
       _cons |   27.12249   47.16079     0.58   0.565    -65.43285    119.6778
------------------------------------------------------------------------------
```

Only the interaction $e401k \cdot (age - 41)$ is significant. Its coefficient is $1.145 (t = 2.42)$. It shows that the effect of 401(k) eligibility on financial wealth increases with age. The coefficient on $e401k \cdot (age - 41)^2$ is .060 (t statistic = 1.37), so it is not significant.

(v) The effect of e401k in part (iii) is the same for all ages, 14.219. For the regression in part (iv), the coefficient on $e401k$ from part (iv) is about 8.357, which is the effect at the average age, $age = 41$.

(vi)

```
. tab fsize, gen(fsize)

family size |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        203       21.88       21.88
          2 |        217       23.38       45.26
          3 |        198       21.34       66.59
          4 |        188       20.26       86.85
          5 |         74        7.97       94.83
          6 |         31        3.34       98.17
          7 |         11        1.19       99.35
          8 |          5        0.54       99.89
```

```
         13 |          1        0.11      100.00
------------+-----------------------------------
      Total |        928      100.00

. drop fsize5 fsize6 fsize7 fsize8 fsize9

. regress nettfa e401k inc incsq age agesq male fsize1 fsize2
fsize3 fsize4

      Source |       SS       df       MS              Number of obs =     928
-------------+------------------------------           F( 10,   917) =   29.47
       Model |  1249291.04    10  124929.104           Prob > F      =  0.0000
    Residual |  3887629.61   917  4239.50884           R-squared     =  0.2432
-------------+------------------------------           Adj R-squared =  0.2349
       Total |  5136920.65   927  5541.44622           Root MSE      =  65.112

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       e401k |   13.42462   4.595985     2.92   0.004     4.404754    22.44449
         inc |  -.5637908   .2564669    -2.20   0.028    -1.067121   -.0604606
       incsq |   .0142597    .001986     7.18   0.000     .0103621    .0181573
         age |  -1.732811   1.869153    -0.93   0.354    -5.401126    1.935504
       agesq |   .0321586   .0216034     1.49   0.137    -.0102393    .0745564
        male |  -1.783906   6.270077    -0.28   0.776    -14.08927    10.52146
      fsize1 |     9.1958   8.194099     1.12   0.262    -6.885564    25.27716
      fsize2 |   17.87712    7.54224     2.37   0.018     3.075066    32.67918
      fsize3 |   .5817076   7.547443     0.08   0.939    -14.23056    15.39397
      fsize4 |   6.537835   7.612689     0.86   0.391    -8.402482    21.47815
       _cons |   12.91241   39.44122     0.33   0.743    -64.49313    90.31795
------------------------------------------------------------------------------

. test fsize1 fsize2 fsize3 fsize4

 ( 1)  fsize1 = 0
 ( 2)  fsize2 = 0
 ( 3)  fsize3 = 0
 ( 4)  fsize4 = 0

       F(  4,   917) =    2.25
            Prob > F =    0.0620
```

I chose $fsize5$ as the base group. The estimated equation is

$$\begin{aligned}
\widehat{nettfa} &= \underset{(39.44)}{12.912} + \underset{(4.60)}{13.425}\, e401k - \underset{(.256)}{.564}\, inc + \underset{(.0020)}{.014}\, inc^2 - \underset{(1.869)}{1.733}\, age + \underset{(.022)}{.032}\, age^2 \\
&\quad - \underset{(6.27)}{1.784}\, male + \underset{(8.19)}{9.196}\, fsize1 + \underset{(7.54)}{17.877}\, fsize2 + \underset{(7.55)}{.582}\, fsize3 + \underset{(7.61)}{6.538}\, fsize4 \\
n &= 928, R^2 = .243.
\end{aligned}$$

The F statistic for joint significance of the four family size dummies is about 2.25. With 4 and 917 df, this gives p-value $= .062$, so they are not jointly significant.

5. (**8.9**)

(i)

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/VOTE1

. regress voteA prtystrA democA lexpendA lexpendB

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  4,   168) =  169.23
       Model |  38822.1768     4   9705.5442           Prob > F      =  0.0000
    Residual |  9635.07174   168  57.3516175           R-squared     =  0.8012
-------------+------------------------------           Adj R-squared =  0.7964
       Total |  48457.2486   172  281.728189           Root MSE      =  7.5731

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    prtystrA |   .2519175   .0712925     3.53   0.001     .1111729    .3926622
      democA |   3.792944    1.40652     2.70   0.008     1.016213    6.569674
    lexpendA |   5.779294   .3918197    14.75   0.000      5.00577    6.552819
    lexpendB |  -6.237836   .3974596   -15.69   0.000    -7.022495   -5.453178
       _cons |   37.66142   4.736036     7.95   0.000      28.3116    47.01123
------------------------------------------------------------------------------

. predict e if e(sample),resid

. regress e prtystrA democA lexpendA lexpendB

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  4,   168) =    0.00
       Model |           0     4           0           Prob > F      =  1.0000
    Residual |  9635.07169   168  57.3516172           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0238
       Total |  9635.07169   172  56.0178587           Root MSE      =  7.5731

------------------------------------------------------------------------------
           e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    prtystrA |   7.22e-11   .0712925     0.00   1.000    -.1407447    .1407447
      democA |  -2.62e-08    1.40652    -0.00   1.000    -2.776731     2.77673
    lexpendA |   2.44e-08   .3918197     0.00   1.000    -.7735247    .7735247
    lexpendB |   1.00e-09   .3974596     0.00   1.000    -.7846588    .7846589
       _cons |  -1.27e-07   4.736036    -0.00   1.000    -9.349812    9.349811
------------------------------------------------------------------------------
```

The estimated equation is

$$\begin{aligned}
\widehat{voteA} &= \underset{(4.74)}{37.66} + \underset{(.071)}{.252}\, prtystrA + \underset{(1.407)}{3.793}\, democA + \underset{(.392)}{5.779}\, \log(expendA) \\
&\quad - \underset{(0.397)}{6.238}\, \log(expendB) + \hat{u} \\
n &= 173, R^2 = .801, \overline{R}^2 = .796.
\end{aligned}$$

You can convince yourself that regressing the $\hat{u}_i$ on all of the explanatory variables yields an R-squared of zero, although it might not be exactly zero in your computer output due to rounding error. Remember, this is how OLS works: the estimates $\hat{\beta}_j$ are chosen to make the residuals be uncorrelated in the sample with each independent variable (as well as have zero sample average).

(ii) Use the F statistic version

```
. gen esq=e^2

. regress esq prtystrA democA lexpendA lexpendB

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  4,   168) =    2.33
       Model |  61537.0938     4  15384.2735           Prob > F      =  0.0581
    Residual |  1109198.47   168  6602.37183           R-squared     =  0.0526
-------------+------------------------------           Adj R-squared =  0.0300
       Total |  1170735.56   172   6806.6021           Root MSE      =  81.255

------------------------------------------------------------------------------
         esq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    prtystrA |  -.2992641   .7649293    -0.39   0.696    -1.809376    1.210848
      democA |   15.61921   15.09117     1.03   0.302    -14.17356    45.41198
    lexpendA |  -10.30573   4.204007    -2.45   0.015    -18.60522   -2.006238
    lexpendB |  -.0514033    4.26452    -0.01   0.990    -8.470355    8.367549
       _cons |   113.9635   50.81502     2.24   0.026      13.6452    214.2817
------------------------------------------------------------------------------

. test prtystrA democA lexpendA lexpendB

 ( 1)  prtystrA = 0
 ( 2)  democA = 0
 ( 3)  lexpendA = 0
 ( 4)  lexpendB = 0

       F(  4,   168) =    2.33
            Prob > F =    0.0581
```

use the bpagan test

```
. regress voteA prtystrA democA lexpendA lexpendB

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  4,   168) =  169.23
       Model |  38822.1768     4   9705.5442           Prob > F      =  0.0000
    Residual |  9635.07174   168  57.3516175           R-squared     =  0.8012
-------------+------------------------------           Adj R-squared =  0.7964
       Total |  48457.2486   172  281.728189           Root MSE      =  7.5731

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

```
-------------+----------------------------------------------------------------
    prtystrA |   .2519175   .0712925     3.53   0.001     .1111729    .3926622
      democA |   3.792944    1.40652     2.70   0.008     1.016213    6.569674
    lexpendA |   5.779294   .3918197    14.75   0.000      5.00577    6.552819
    lexpendB |  -6.237836   .3974596   -15.69   0.000    -7.022495   -5.453178
       _cons |   37.66142   4.736036     7.95   0.000      28.3116    47.01123
------------------------------------------------------------------------------

. bpagan prtystrA democA lexpendA lexpendB

Breusch-Pagan LM statistic:  9.919488  Chi-sq( 4)  P-value =
.0418
```

The B-P test entails regressing the $\hat{u}_i{}^2$ on the independent variables in part (i). The F statistic for joint significant (with 4 and 168 *df*) is about 2.33 with p-value $\approx .058$. Therefore, there is some evidence of heteroskedasticity, but not quite at the 5% level.

(iii)
```
use white test

. whitetst, fitted

White's special test statistic :  5.490049  Chi-sq( 2)  P-value =  .0642

use the F-statistic version

. predict voteA1 (option xb assumed; fitted values)

. gen voteA1sq= voteA1^2

. regress esq voteA1 voteA1sq

      Source |       SS       df       MS              Number of obs =     173
-------------+------------------------------           F(  2,   170) =    2.79
       Model |  37152.5749     2  18576.2875           Prob > F      =  0.0645
    Residual |  1133582.99   170  6668.13521           R-squared     =  0.0317
-------------+------------------------------           Adj R-squared =  0.0203
       Total |  1170735.56   172   6806.6021           Root MSE      =  81.659

------------------------------------------------------------------------------
         esq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      voteA1 |  -4.263682   2.166534    -1.97   0.051    -8.540455    .0130912
    voteA1sq |   .0357354   .0212419     1.68   0.094    -.0061964    .0776672
       _cons |   171.8584   53.14213     3.23   0.001     66.95499    276.7619
------------------------------------------------------------------------------

. test voteA1 voteA1sq

 ( 1)  voteA1 = 0
 ( 2)  voteA1sq = 0

       F(  2,   170) =    2.79
```

```
          Prob > F =    0.0645
```

Now we regress $\hat{u}_i{}^2$ on $\widehat{voteA}_i$ and $(\widehat{voteA}_i)^2$, where the $\widehat{voteA}_i$ are the OLS fitted values from part (i). The F test, with 2 and 170 *df*, is about 2.79 with p-value $\approx .065$. This is slightly less evidence of heteroskedasticity than provided by the B-P test, but the conclusion is very similar.

6. (**9.7**)

(i)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/WAGE2

. regress lwage educ exper tenure married south urban black KWW

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  8,   926) =   40.39
       Model |  42.8510762     8  5.35638452           Prob > F      =  0.0000
    Residual |  122.805218   926  .132619026           R-squared     =  0.2587
-------------+------------------------------           Adj R-squared =  0.2523
       Total |  165.656294   934  .177362199           Root MSE      =  .36417

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0576277    .006838     8.43   0.000     .0442079    .0710475
       exper |   .0122284    .003241     3.77   0.000     .0058678     .018589
      tenure |    .011072   .0024564     4.51   0.000     .0062512    .0158927
     married |   .1894612   .0390774     4.85   0.000     .1127707    .2661517
       south |  -.0916006   .0261562    -3.50   0.000     -.142933   -.0402683
       urban |   .1755452   .0270323     6.49   0.000     .1224936    .2285969
       black |  -.1642666   .0385304    -4.26   0.000    -.2398837   -.0886495
         KWW |   .0050275   .0018188     2.76   0.006     .0014581     .008597
       _cons |   5.358797   .1136002    47.17   0.000     5.135853    5.581741
------------------------------------------------------------------------------
```

We estimate the model from column (2) but with KWW in place of IQ. The coefficient on *educ* becomes about .058 (se $\approx$ .006), so this is similar to the estimate obtained with IQ, although slightly larger and more precisely estimated.

(ii)
```
. regress lwage educ exper tenure married south urban black KWW
IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  9,   925) =   37.28
       Model |  44.0968017     9  4.89964463           Prob > F      =  0.0000
    Residual |  121.559493   925  .131415668           R-squared     =  0.2662
-------------+------------------------------           Adj R-squared =  0.2591
       Total |  165.656294   934  .177362199           Root MSE      =  .36251

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

```
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0498375    .007262     6.86   0.000     .0355856    .0640893
       exper |   .0127522   .0032308     3.95   0.000     .0064117    .0190927
      tenure |   .0109248   .0024457     4.47   0.000      .006125    .0157246
     married |   .1921449   .0389094     4.94   0.000     .1157839    .2685059
       south |  -.0820295   .0262222    -3.13   0.002    -.1334913   -.0305676
       urban |   .1758226   .0269095     6.53   0.000     .1230118    .2286334
       black |  -.1303995   .0399014    -3.27   0.001    -.2087073   -.0520917
         KWW |    .003826   .0018521     2.07   0.039     .0001911    .0074608
          IQ |   .0031183   .0010128     3.08   0.002     .0011306    .0051059
       _cons |   5.175643    .127776    40.51   0.000     4.924879    5.426408
------------------------------------------------------------------------------
```

When KWW and IQ are both used as proxies, the coefficient on *educ* becomes about .049 (se $\approx$ .007). Compared with the estimate when only KWW is used as a proxy, the return to education has fallen by almost a full percentage point.

(iii)
```
. test  KWW IQ

 ( 1)  KWW = 0.0
 ( 2)  IQ = 0.0

       F(  2,   925) =    8.59
            Prob > F =    0.0002
```

The t statistic on IQ is about 3.08 while that on KWW is about 2.07, so each is significant at the 5% level against a two-sided alternative. They are jointly very significant, with $F_{2,925} \approx 8.59$ and p-value $\approx .0002$.