# Homework #7

A. Shawn Bandy

April 22, 2013

## 1 Lab Problems

L1 STATA log:

```
-------------------------------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/shawn/src/econ485/lab7/homework7.log
  log type:  text
 opened on:  23 Apr 2013, 09:10:19

. /*
> a. Use OLS regression to estimate the following model. First, run OLS assuming
> homoskedasticity. After you estimate that model, use the Breusch-Pagan test to test
> for heteroskedasticity. If you find that the errors are heteroskedastic, use robust
> standard errors to correct for heteroskedasticity by re-estimating that model. Also, test
> for multicollinearity after your 2nd model (if you need to account for
> heteroskedasticity).
> i. Dependent variable: lnwage; explanatory variables are educ exper
> black smsa
> */
>
> regress lnwage educ exper black smsa;

      Source |       SS       df       MS              Number of obs =    1184
-------------+------------------------------           F(  4,  1179) =   61.40
       Model |  34.9845998     4  8.74614994           Prob > F      =  0.0000
    Residual |   167.94877  1179  .142450187           R-squared     =  0.1724
-------------+------------------------------           Adj R-squared =  0.1696
       Total |   202.93337  1183   .17154131           Root MSE      =  .37743

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0714313   .0059652    11.97   0.000     .0597277    .0831349
       exper |   .0446856   .0037508    11.91   0.000     .0373267    .0520446
       black |  -.1896232    .037656    -5.04   0.000    -.2635035   -.1157428
        smsa |   .1431147   .0271522     5.27   0.000     .0898427    .1963868
```

```
       _cons |   4.898209   .1061538    46.14   0.000     4.689937     5.10648
------------------------------------------------------------------------------

. //rvfplot;
> estat hettest;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =     0.45
         Prob > chi2  =   0.5023

. vif;

    Variable |       VIF       1/VIF
-------------+----------------------
        educ |      1.55    0.646982
       exper |      1.52    0.656554
       black |      1.03    0.967542
        smsa |      1.03    0.974066
-------------+----------------------
    Mean VIF |      1.28

. /*
> b. Use outreg2 to create a nice formatted table named wageapril with your
> regression results from part a). Have the output be sent to BOTH word and excel.
> Make sure you title the table Wage Regression and that you title this column
> OLS. Also include the Adjusted R-squared e(r2_a) in your results.
> */
>
> outreg2 using wageapril, replace word title("Wage Regression") ctitle("OLS") e(r2_a);
wageapril.rtf
dir : seeout

. outreg2 using wageapril, replace excel title("Wage Regression") ctitle("OLS") e(r2_a);
wageapril.xml
dir : seeout

. outreg2 using wageapril, replace tex title("Wage Regression") ctitle("OLS") e(r2_a);
wageapril.tex
dir : seeout

. /*
> c. You decide that you need more explanatory variables. So, you decide to include the
> ID for each individual in your model. Use OLS regression to estimate the following
> model. If you found you needed to use robust standard errors in part a), do so again
> here.
> i. Dependent variable: lnwage; explanatory variables are educ exper
```

```
> black smsa id
> */
>
> regress lnwage educ exper black smsa id;

      Source |       SS       df       MS              Number of obs =    1184
-------------+------------------------------           F(  5,  1178) =   49.29
       Model |  35.1130527     5  7.02261054           Prob > F      =  0.0000
    Residual |  167.820317  1178  .142462069           R-squared     =  0.1730
-------------+------------------------------           Adj R-squared =  0.1695
       Total |   202.93337  1183   .17154131           Root MSE      =  .37744

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0713493   .0059661    11.96   0.000     .0596439    .0830546
       exper |   .0446856   .0037509    11.91   0.000     .0373263    .0520448
       black |  -.1821237   .0384769    -4.73   0.000    -.2576146   -.1066328
        smsa |   .1383397    .027615     5.01   0.000     .0841595    .1925199
          id |  -.0000107   .0000113    -0.95   0.343    -.0000328    .0000114
       _cons |   4.920269   .1086707    45.28   0.000     4.707059    5.133479
------------------------------------------------------------------------------

. estat hettest;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =     0.46
         Prob > chi2  =   0.4994

. vif;

    Variable |       VIF       1/VIF
-------------+----------------------
        educ |      1.55    0.646846
       exper |      1.52    0.656554
       black |      1.08    0.926777
          id |      1.07    0.933358
        smsa |      1.06    0.941766
-------------+----------------------
    Mean VIF |      1.26

. /*
> d. Now, you are worried that your measure of education is endogenous because the error
> term includes ability. Use IQ as a proxy for ability and use OLS regression to
> estimate the following model. Use robust standard errors if you found the presence of
> heteroskedasticity in part a).
```

```
> i. Dependent variable: lnwage; explanatory variables are educ exper
> black smsa IQ
> */
>
> regress lnwage educ exper black smsa IQ;

      Source |       SS       df       MS              Number of obs =    1184
-------------+------------------------------           F(  5,  1178) =   50.66
       Model |  35.9128978     5  7.18257955           Prob > F      =  0.0000
    Residual |  167.020472  1178  .141783083           R-squared     =  0.1770
-------------+------------------------------           Adj R-squared =  0.1735
       Total |   202.93337  1183   .17154131           Root MSE      =  .37654

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0644704   .0065435     9.85   0.000     .0516322    .0773087
       exper |   .0454306   .0037533    12.10   0.000     .0380667    .0527945
       black |  -.1567381   .0397053    -3.95   0.000     -.234639   -.0788372
        smsa |   .1406109   .0271062     5.19   0.000      .087429    .1937927
          IQ |   .0023745    .000928     2.56   0.011     .0005538    .0041952
       _cons |    4.74046   .1225423    38.68   0.000     4.500034    4.980885
------------------------------------------------------------------------------

. estat hettest;

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnwage

         chi2(1)      =     0.67
         Prob > chi2  =   0.4139

. vif;

    Variable |       VIF       1/VIF
-------------+----------------------
        educ |      1.87    0.535159
          IQ |      1.56    0.639176
       exper |      1.53    0.652604
       black |      1.15    0.866173
        smsa |      1.03    0.972796
-------------+----------------------
    Mean VIF |      1.43

. /*
> e. Now, return to the model in part a). Suppose that you are mostly interested in the
> effect of education on wages. Instead of trying to control for ability with a proxy, you
> decide to use IV estimation. You believe that a good instrument is whether or not a
```

```
> person lived near a 4-year college in 1966 (nearc4). Use instrumental variable
> regression to estimate the model in part a). Use ivreg2 and make sure you use the
> first and endog options in order to get the first-stage estimates and the test
> for endogeneity of educ. After you estimate that model, use ivhettest to test for
> heteroskedasticity. If you find that the errors are heteroskedastic, re-estimate the
> model using IV regression and including robust to correct for heteroskedasticity.
> */
>
> ivreg2 lnwage (educ=nearc4) exper black smsa, first endog(educ);

First-stage regressions
-----------------------

First-stage regression of educ:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =     1184
                                                      F(  4,  1179) =   164.79
                                                      Prob > F      =   0.0000
Total (centered) SS     =  6187.553209                Centered R2   =   0.3586
Total (uncentered) SS   =       243857                Uncentered R2 =   0.9837
Residual SS             =  3968.705434                Root MSE      =    1.835

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |  -.3654329   .0148109   -24.67   0.000    -.3944916   -.3363742
       black |  -.6555666    .182057    -3.60   0.000    -1.012759   -.2983747
        smsa |   .2862191   .1355166     2.11   0.035     .0203384    .5520997
      nearc4 |   .4199662    .131125     3.20   0.001     .1627017    .6772307
       _cons |   16.59676    .183997    90.20   0.000     16.23577    16.95776
------------------------------------------------------------------------------
Included instruments: exper black smsa nearc4
------------------------------------------------------------------------------
F test of excluded instruments:
  F(  1,  1179) =    10.26
  Prob > F      =   0.0014
Angrist-Pischke multivariate F test of excluded instruments:
  F(  1,  1179) =    10.26
  Prob > F      =   0.0014



Summary results for first-stage regressions
```

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

                                        (Underid)            (Weak id)
Variable     | F(  1,  1179)  P-val | AP Chi-sq(  1) P-val | AP F(  1,  1179)
educ         |      10.26    0.0014 |       10.30   0.0013 |       10.26

Stock-Yogo weak ID test critical values for single endogenous regressor:
                                  10% maximal IV size             16.38
                                  15% maximal IV size              8.96
                                  20% maximal IV size              6.66
                                  25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic       Chi-sq(1)=10.21    P-val=0.0014

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic                                    10.26

Stock-Yogo weak ID test critical values for K1=1 and L1=1:
                                  10% maximal IV size             16.38
                                  15% maximal IV size              8.96
                                  20% maximal IV size              6.66
                                  25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test           F(1,1179)=      0.30     P-val=0.5839
Anderson-Rubin Wald test           Chi-sq(1)=      0.30     P-val=0.5830
Stock-Wright LM S statistic        Chi-sq(1)=      0.30     P-val=0.5830

Number of observations               N  =       1184
Number of regressors                 K  =          5
Number of endogenous regressors      K1 =          1
Number of instruments                L  =          5
Number of excluded instruments       L1 =          1

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

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                     Number of obs =     1184
```

```
                                                      F(  4,  1179) =    24.94
                                                      Prob > F      =   0.0000
Total (centered) SS     =  202.9333701                Centered R2   =   0.1494
Total (uncentered) SS   =  48108.50864                Uncentered R2 =   0.9964
Residual SS             =   172.622443                Root MSE      =    .3818

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .037263   .0649795     0.57   0.566    -.0900946    .1646205
       exper |   .0321384   .0240597     1.34   0.182    -.0150178    .0792945
       black |  -.2120682    .057075    -3.72   0.000    -.3239332   -.1002032
        smsa |   .1564774   .0373467     4.19   0.000     .0832792    .2296756
       _cons |    5.47404   1.095629     5.00   0.000     3.326646    7.621434
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):          10.213
                                                    Chi-sq(1) P-val =    0.0014
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):               10.258
Stock-Yogo weak ID test critical values: 10% maximal IV size             16.38
                                         15% maximal IV size              8.96
                                         20% maximal IV size              6.66
                                         25% maximal IV size              5.53
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):           0.000
                                                  (equation exactly identified)
-endog- option:
Endogeneity test of endogenous regressors:                               0.287
                                                    Chi-sq(1) P-val =    0.5924
Regressors tested:    educ
------------------------------------------------------------------------------
Instrumented:         educ
Included instruments: exper black smsa
Excluded instruments: nearc4
------------------------------------------------------------------------------

. ivhettest;
IV heteroskedasticity test(s) using levels of IVs only
Ho: Disturbance is homoskedastic
    Pagan-Hall general test statistic   :   0.683  Chi-sq(4) P-value = 0.9534

. /*
> f. You decide to add another instrument to your model in part e), now you use both
> nearc4 and fatheduc (the education level of persons father) as instruments for
> education. Otherwise the model should be similar to part e). Use ivreg2 and make
> sure you use the first and endog options in order to get the first-stage
> estimates and the test for endogeneity of educ. If in part e) you found that the errors
> are heteroskedastic, include robust to correct for heteroskedasticity.
```

```
> */
>
> ivreg2 lnwage (educ=nearc4 fatheduc) exper black smsa, first endog(educ);

First-stage regressions
-----------------------

First-stage regression of educ:

OLS estimation
--------------

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =     1184
                                                      F(  5,  1178) =   162.95
                                                      Prob > F      =   0.0000
Total (centered) SS     =  6187.553209                Centered R2   =   0.4089
Total (uncentered) SS   =       243857                Uncentered R2 =   0.9850
Residual SS             =  3657.734504                Root MSE      =    1.762

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |  -.3189218   .0149648   -21.31   0.000    -.3482825    -.289561
       black |  -.3006298   .1784139    -1.69   0.092    -.6506743    .0494147
        smsa |   .1057065   .1313982     0.80   0.421    -.1520942    .3635071
      nearc4 |   .3596716   .1260805     2.85   0.004     .1123041     .607039
    fatheduc |   .1703282     .01702    10.01   0.000     .1369353    .2037212
       _cons |   14.55211   .2701329    53.87   0.000     14.02212    15.08211
------------------------------------------------------------------------------
Included instruments: exper black smsa nearc4 fatheduc
------------------------------------------------------------------------------
F test of excluded instruments:
  F(  2,  1178) =    55.64
  Prob > F      =   0.0000
Angrist-Pischke multivariate F test of excluded instruments:
  F(  2,  1178) =    55.64
  Prob > F      =   0.0000



Summary results for first-stage regressions
-------------------------------------------

                                       (Underid)             (Weak id)
Variable     | F(  2,  1178)  P-val | AP Chi-sq(  2) P-val | AP F(  2,  1178)
educ         |      55.64    0.0000 |      111.84   0.0000 |       55.64
```

```
Stock-Yogo weak ID test critical values for single endogenous regressor:
                                 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.

Underidentification test
Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)
Anderson canon. corr. LM statistic       Chi-sq(2)=102.19   P-val=0.0000

Weak identification test
Ho: equation is weakly identified
Cragg-Donald Wald F statistic                                    55.64

Stock-Yogo weak ID test critical values for K1=1 and L1=2:
                                 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.

Weak-instrument-robust inference
Tests of joint significance of endogenous regressors B1 in main equation
Ho: B1=0 and orthogonality conditions are valid
Anderson-Rubin Wald test           F(2,1178)=     10.99     P-val=0.0000
Anderson-Rubin Wald test           Chi-sq(2)=     22.10     P-val=0.0000
Stock-Wright LM S statistic        Chi-sq(2)=     21.70     P-val=0.0000

Number of observations               N  =       1184
Number of regressors                 K  =          5
Number of endogenous regressors      K1 =          1
Number of instruments                L  =          6
Number of excluded instruments       L1 =          2

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

Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

                                                      Number of obs =     1184
                                                      F(  4,  1179) =    30.85
                                                      Prob > F      =   0.0000
Total (centered) SS     =  202.9333701                Centered R2   =   0.1586
Total (uncentered) SS   =  48108.50864                Uncentered R2 =   0.9965
Residual SS             =  170.7512787                Root MSE      =    .3798
```

```
------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0978899   .0204306     4.79   0.000     .0578467    .1379332
       exper |   .0544018   .0081039     6.71   0.000     .0385185     .070285
       black |  -.1722425   .0400015    -4.31   0.000    -.2506441    -.093841
        smsa |   .1327672   .0283674     4.68   0.000     .0771681    .1883663
       _cons |   4.452306   .3460176    12.87   0.000     3.774124    5.130488
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):         102.186
                                                   Chi-sq(2) P-val =    0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                55.635
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):           0.978
                                                   Chi-sq(1) P-val =    0.3227
-endog- option:
Endogeneity test of endogenous regressors:                               1.866
                                                   Chi-sq(1) P-val =    0.1719
Regressors tested:    educ
------------------------------------------------------------------------------
Instrumented:         educ
Included instruments: exper black smsa
Excluded instruments: nearc4 fatheduc
------------------------------------------------------------------------------

. /*
> g. Use outreg2 to append your results from part f) to your table you created in part b).
> Make sure you title the table Wage Regression and that you title this column IV.
> Also include Adjusted R-squared e(r2_a), the F-test for weak instruments
> e(widstat), the p-value test of overidentifying restrictions e(sarganp) and
> the Instruments e(exexog) in your results.
> */
>
> outreg2 using wageapril, append word title("Wage Regression") ctitle("IV") e(r2_a, widstat, sarganp, exex
> ;
wageapril.rtf
dir : seeout

. outreg2 using wageapril, append excel title("Wage Regression") ctitle("IV") e(r2_a, widstat, sarganp, exe
> );
wageapril.xml
dir : seeout
```

```
. outreg2 using wageapril, append tex title("Wage Regression") ctitle("IV") e(r2_a, widstat, sarganp, exexo
wageapril.tex
dir : seeout

.
end of do-file
```

# 2 Questions

Q1 Use the results from L1 to answer this question.

a. Assuming you dont have any violations of OLS, interpret the coefficient on educ from the regression you ran in L1, part a).

b. What was the result of your test for heteroskedasticity in L1, part a)?

c. What was the result of your test for multicollinearity in L1, part a)?

d. Compare the Adjusted R-squared from your results in L1, part a) and part c). Explain the difference and why there is a problem with the model in part c).

e. Assuming you dont have any violations of OLS, interpret the coefficient on IQ from the regression you ran in L1, part d).

f. Explain why you do or do not think IQ is a good proxy for ability.