BOSTON COLLEGE
Department of Economics
EC 771: Econometrics
Spring 2009
Prof. Baum, Ms. Skira

Problem Set 4: Solutions

Point Distribution:
1): 15 points
2), 3): 10 points each
4), 5): 15 points each

1)

```
. use http://fmwww.bc.edu/ec-p/data/dmackinnon/earnings

. regress earnings group1 group2 group3, nocons

      Source |       SS       df       MS              Number of obs =    4266
-------------+------------------------------           F(  3,  4263) = 3842.49
       Model |  2.7934e+12     3  9.3114e+11           Prob > F      =  0.0000
    Residual |  1.0330e+12  4263   242327533           R-squared     =  0.7300
-------------+------------------------------           Adj R-squared =  0.7298
       Total |  3.8265e+12  4266   896967282           Root MSE      =   15567

------------------------------------------------------------------------------
    earnings |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      group1 |   22880.48   467.4505    48.95   0.000     21964.03    23796.92
      group2 |   25080.17   380.0191    66.00   0.000     24335.14    25825.21
      group3 |   27973.63   404.7784    69.11   0.000     27180.06    28767.21
------------------------------------------------------------------------------

. predict uhat, residuals

. gen uhatsq = uhat * uhat

. regress uhatsq group1 group2 group3, nocons

      Source |       SS       df       MS              Number of obs =    4266
-------------+------------------------------           F(  3,  4263) =  434.94
       Model |  2.5243e+20     3  8.4144e+19           Prob > F      =  0.0000
    Residual |  8.2472e+20  4263  1.9346e+17           R-squared     =  0.2344
```

```
-------------+------------------------------           Adj R-squared =  0.2338
       Total |  1.0772e+21  4266  2.5250e+17           Root MSE      =  4.4e+08

------------------------------------------------------------------------------
      uhatsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      group1 |   2.15e+08   1.32e+07    16.27   0.000     1.89e+08    2.41e+08
      group2 |   2.34e+08   1.07e+07    21.77   0.000     2.13e+08    2.55e+08
      group3 |   2.72e+08   1.14e+07    23.80   0.000     2.50e+08    2.95e+08
------------------------------------------------------------------------------

. test group1 = group2 = group3

 ( 1)  group1 - group2 = 0
 ( 2)  group1 - group3 = 0

       F(  2,  4263) =    5.87
            Prob > F =    0.0028

. gen group = group1 * 1 + group2 * 2 + group3 * 3

. robvar uhatsq, by(group)

            |          Summary of uhatsq
      group |        Mean   Std. Dev.       Freq.
------------+------------------------------------
          1 |   2.149e+08   4.247e+08        1109
          2 |   2.337e+08   4.243e+08        1678
          3 |   2.722e+08   4.675e+08        1479
------------+------------------------------------
      Total |   2.422e+08   4.403e+08        4266

W0  =  7.1509260   df(2, 4263)     Pr > F = 0.00079358

W50 =  4.6029258   df(2, 4263)     Pr > F = 0.01007233

W10 =  5.3990598   df(2, 4263)     Pr > F = 0.0045518
```

The squared residuals of the regression of earnings on the group dummies are regressed on these group dummies, and an F-test is performed to test for equality of the coefficients of this second regression across groups i.e. the null hypothesis of such a test is that the group does not affect the variance. This test leads us to reject this null hypothesis at all conventional levels.

Next, the robvar command carries out three tests of the hypothesis that the variance of the residuals across groups is the same. All three tests reject the null hypothesis of constant variance across groups at the 5% level.

2) First, notice that $u_{t+j} = \rho^j u_t + \sum_{i=0}^{j-1} \rho^i \epsilon_{t+j-i}$. Then,

$$\begin{aligned}
\operatorname{cov}(u_t, u_{t+j}) &= \operatorname{cov}\left(u_t, \rho^j u_t + \sum_{i=0}^{j-1} \rho^i \epsilon_{t+j-i}\right) \\
&= \operatorname{cov}\left(u_t, \rho^j u_t\right) + \operatorname{cov}\left(u_t, \sum_{i=0}^{j-1} \rho^i \epsilon_{t+j-i}\right) \\
&= \rho^j \operatorname{var}(u_t)
\end{aligned}$$

where we have used the independence of the innovation ϵ_k with respect to u_j for all $k > j$. Similarly, $\operatorname{cov}(u_t, u_{t-j}) = \rho^j \operatorname{var}(u_{t-j})$. Stationarity implies $\operatorname{var}(u_t) = \operatorname{var}(u_{t-j})$ for all integer values of j. Thus, we have that $\operatorname{cov}(u_t, u_{t-j}) = \operatorname{cov}(u_t, u_{t+j})$.

Now, $u_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}$, if we assume that $\lim_{i\to\infty} \rho^i u_{t-i} = 0$. Then,

$$\begin{aligned}
\operatorname{var}(u_t) &= \operatorname{var}\left(\sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}\right) \\
&= \sum_{i=0}^{\infty} \left(\rho^2\right)^i \operatorname{var}(\epsilon_{t-i}) \\
&= \sum_{i=0}^{\infty} \left(\rho^2\right)^i \sigma_\epsilon^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}
\end{aligned}$$

where we have used the independence assumption of the innovations to assert $\operatorname{cov}(u_{t-i}, u_{t-k}) = 0$ iff $i \neq k$.

Thus, we have that $\operatorname{cov}(u_t, u_{t-j}) = \rho^j \frac{\sigma_\epsilon^2}{1-\rho^2}$, and it follows immediately that $\operatorname{corr}(u_t, u_{t-j}) = \frac{\operatorname{cov}(u_t, u_{t-j})}{\sqrt{\operatorname{var}(u_t)\operatorname{var}(u_{t-j})}} = \rho^j$.

3) Given $c_t = \delta_0 + \delta_1 y_t + \rho u_{t-1} + \epsilon_t$ and $c_{t-1} = \delta_0 + \delta_1 y_{t-1} + u_{t-1}$, we can multiply the second equation by ρ and subtract it from the first yield

$$c_t - \rho c_{t-1} = \delta_0(1-\rho) + \delta_1 y_t - \delta_1 \rho y_{t-1} + \epsilon_t$$

Comparing coefficients we notice that this is a special case of the general model with parameters mapped as follows:

$$\alpha = \delta_0(1-\rho), \quad \beta = \rho, \quad \gamma_0 = \delta_1, \quad \gamma_1 = -\rho\delta_1$$

Since the original model has four parameters, but the special case has only three, we must lose a degree of freedom. The parameter mapping above proves this; the first three parameters $(\alpha, \beta, \gamma_0)$ are free, but the fourth parameter (γ_1) is determined by β and γ_0. Thus we conclude that there is one restriction.

4)

```
. use http://fmwww.bc.edu/ec-p/data/dmackinnon/consumption

. gen c = log(CE)
```

```
. gen y = log(YD)

. regress c L.c y L.y if year > 1952

      Source |       SS       df       MS              Number of obs =     176
-------------+------------------------------           F(  3,   172) =       .
       Model |  45.1518963     3  15.0506321           Prob > F      =  0.0000
    Residual |  .015852826   172  .000092168           R-squared     =  0.9996
-------------+------------------------------           Adj R-squared =  0.9996
       Total |  45.1677491   175  .258101424           Root MSE      =   .0096

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           c |
         L1. |   .9692255     .02231    43.44   0.000     .9251888    1.013262
           y |
         --. |   .2909892   .0551121     5.28   0.000      .182206    .3997724
         L1. |  -.2651525    .056445    -4.70   0.000    -.3765666   -.1537384
       _cons |   .0639353   .0216598     2.95   0.004      .021182    .1066886
------------------------------------------------------------------------------

. nlcom _b[_cons] / (1 - _b[L1.c])

       _nl_1:  _b[_cons] / (1 - _b[L1.c])

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |    2.07754   1.198914     1.73   0.085    -.2889387    4.444018
------------------------------------------------------------------------------

. nlcom -_b[L.y]/_b[L.c]

       _nl_1:  -_b[L.y]/_b[L.c]

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   .2735716   .0569143     4.81   0.000     .1612312    .3859119
------------------------------------------------------------------------------
```

From this OLS regression, we obtain an estimate of $\delta_0 \equiv \frac{\alpha}{1-\beta} = 2.07754$ and of $\delta_1 \equiv \gamma_0 = 0.2909892$. An alternative estimate of $\delta_1 = -\gamma_1/\beta = 0.2735716$. Also, $\rho \equiv \beta = 0.9692255$. These estimates are consistent but inefficient.

```
. gen Lc = L.c
(1 missing value generated)

. gen Ly = L.y
(1 missing value generated)

. nl (c = {delta0}*(1 - {rho}) + {delta1}*y - {delta1}*{rho}*Ly + {rho}*Lc) if year > 1952, ///
>         initial(delta0 2.0775 delta1 0.290988 rho 0.969225)  variables(Ly Lc)
(obs = 176)

Iteration 0:  residual SS =  .2033905
Iteration 1:  residual SS =  .0159923
Iteration 2:  residual SS =   .015977
Iteration 3:  residual SS =   .015977

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =       176
       Model |  45.1517721     2  22.5758861         R-squared     =    0.9996
    Residual |  .015977007   173  .000092353         Adj R-squared =    0.9996
-------------+------------------------------         Root MSE      =    .00961
       Total |  45.1677491   175  .258101424         Res. dev.     = -1138.581

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     /delta0 |   10.03034   .9619419    10.43   0.000     8.131689      11.929
        /rho |   .9950142   .0020338   489.24   0.000         .991    .9990284
     /delta1 |   .2836779   .0548059     5.18   0.000     .1755035    .3918523
------------------------------------------------------------------------------
* (SEs, P values, CIs, and correlations are asymptotic approximations)
  Parameter delta0 taken as constant term in model & ANOVA table
```

Since the restricted SSR from the nonlinear regression is 0.015977007, and the unrestricted SSR from the OLS regression is 0.015852826, the F statistic is

$$F(1,172) = \frac{0.015977007 - 0.015852826}{0.015852826/(176-4)} = 1.34733908$$

The p-value for this test statistic, based on the $F(1,172)$ distribution, is 0.247, so the null hypothesis that the restriction is valid is not rejected.

5) a)

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

. count if !NJ
   79
```

b)

```
. summ wagest

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      wagest |       390    4.615641    .3470151       4.25       5.75
```

c)

```
. ttest income, by(state)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |     331    48052.58    745.0526    13555.04    46586.93    49518.23
       2 |      78    42815.32    1189.756    10507.64    40446.21    45184.43
---------+--------------------------------------------------------------------
combined |     409    47053.78    651.6738    13179.29    45772.73    48334.84
---------+--------------------------------------------------------------------
    diff |            5237.257    1640.415                2012.512    8462.001
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =   3.1926
Ho: diff = 0                                     degrees of freedom =      407

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9992         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.0008
```

The null hypothesis that the difference in the means of income between the two states is zero is rejected at the 1% level.

d)

```
. ttest pentree == 1.39

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 pentree |     398    1.322186    .0322349    .6430845    1.258813    1.385558
------------------------------------------------------------------------------
    mean = mean(pentree)                                          t =  -2.1037
Ho: mean = 1.39                                  degrees of freedom =      397
```

```
   Ha: mean < 1.39              Ha: mean != 1.39              Ha: mean > 1.39
 Pr(T < t) = 0.0180         Pr(|T| > |t|) = 0.0360          Pr(T > t) = 0.9820
```

The null hypothesis that the price of an entree is \$1.39 is rejected at the 5% level.

e)

```
. program drop _all

. program define ols
  1.         args lnf mu sigma
  2.         qui replace `lnf' = ln(normalden(($ML_y-`mu')/`sigma'))-ln(`sigma')
  3. end

. program define olsalt
  1.         args lnf mu sigma2
  2.         qui replace `lnf' = ln(normalden(($ML_y-`mu')/sqrt(`sigma2'))) - 0.5*ln(`sigma2')
  3. end

.
. ml model lf ols (pfries = income prpblck) ()

. * ml check
. ml maximize //, trace grad

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -236.94341
rescale:       log likelihood = -236.94341
rescale eq:    log likelihood =    238.066
Iteration 0:   log likelihood =    238.066
Iteration 1:   log likelihood =  314.20716
Iteration 2:   log likelihood =  339.06104
Iteration 3:   log likelihood =  339.27361
Iteration 4:   log likelihood =   339.2744
Iteration 5:   log likelihood =   339.2744

                                                  Number of obs   =        392
                                                  Wald chi2(2)    =      31.75
Log likelihood =   339.2744                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
      income |   2.37e-06   4.34e-07     5.47   0.000     1.52e-06    3.22e-06
```

```
     prpblck |   .1114181   .0308421     3.61   0.000     .0509687    .1718675
       _cons |   .7979156   .0227235    35.11   0.000     .7533783    .8424529
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .1018316   .0036368    28.00   0.000     .0947036    .1089597
------------------------------------------------------------------------------

. ml model lf olsalt (pfries = income prpblck) ()

. * ml check
. ml maximize //, trace grad

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -298.58366
rescale:       log likelihood = -298.58366
rescale eq:    log likelihood =    238.066
Iteration 0:   log likelihood =    238.066
Iteration 1:   log likelihood =  305.94977
Iteration 2:   log likelihood =  338.19066
Iteration 3:   log likelihood =  339.27129
Iteration 4:   log likelihood =   339.2744
Iteration 5:   log likelihood =   339.2744

                                                  Number of obs   =        392
                                                  Wald chi2(2)    =      31.75
Log likelihood =   339.2744                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
      income |   2.37e-06   4.34e-07     5.47   0.000     1.52e-06    3.22e-06
     prpblck |   .1114181   .0308421     3.61   0.000     .0509687    .1718675
       _cons |   .7979156   .0227235    35.11   0.000     .7533783    .8424529
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .0103697   .0007407    14.00   0.000      .008918    .0118214
------------------------------------------------------------------------------
```

f)

```
. ml model lf ols (pfries = income prpblck) () if state == 1

. * ml check
. ml maximize //, trace grad
```

```
initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -200.93227
rescale:       log likelihood = -200.93227
rescale eq:    log likelihood =  229.72377
Iteration 0:   log likelihood =  229.72377
Iteration 1:   log likelihood =   275.9703
Iteration 2:   log likelihood =   281.7878
Iteration 3:   log likelihood =  281.89288
Iteration 4:   log likelihood =  281.89308
Iteration 5:   log likelihood =  281.89308

                                                  Number of obs   =        316
                                                  Wald chi2(2)    =       9.35
Log likelihood =  281.89308                       Prob > chi2     =     0.0093

------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
      income |   1.46e-06   4.78e-07     3.05   0.002     5.23e-07    2.40e-06
     prpblck |   .0456923   .0323636     1.41   0.158    -.0177392    .1091238
       _cons |   .8655968   .0259617    33.34   0.000     .8147128    .9164809
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .0991615   .0039444    25.14   0.000     .0914306    .1068925
------------------------------------------------------------------------------

. ml model lf olsalt (pfries = income prpblck) () if state == 1

. * ml check
. ml maximize //, trace grad

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -245.65842
rescale:       log likelihood = -245.65842
rescale eq:    log likelihood =  229.72377
Iteration 0:   log likelihood =  229.72377
Iteration 1:   log likelihood =  264.41684
Iteration 2:   log likelihood =  281.42678
Iteration 3:   log likelihood =  281.89128
Iteration 4:   log likelihood =  281.89308
Iteration 5:   log likelihood =  281.89308

                                                  Number of obs   =        316
                                                  Wald chi2(2)    =       9.35
Log likelihood =  281.89308                       Prob > chi2     =     0.0093
```

```
------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
      income |   1.46e-06   4.78e-07     3.05   0.002     5.23e-07    2.40e-06
     prpblck |   .0456923   .0323636     1.41   0.158    -.0177392    .1091238
       _cons |   .8655968   .0259617    33.34   0.000     .8147128    .9164809
-------------+----------------------------------------------------------------
eq2          |
       _cons |    .009833   .0007823    12.57   0.000     .0082998    .0113662
------------------------------------------------------------------------------
```

g)

```
. reg pfries income prpblck

      Source |       SS       df       MS              Number of obs =     392
-------------+------------------------------           F(  2,   389) =   15.75
       Model |  .329251146     2  .164625573           Prob > F      =  0.0000
    Residual |  4.06491547   389  .010449654           R-squared     =  0.0749
-------------+------------------------------           Adj R-squared =  0.0702
       Total |  4.39416662   391  .011238278           Root MSE      =  .10222

------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   2.37e-06   4.36e-07     5.45   0.000     1.52e-06    3.23e-06
     prpblck |   .1114181   .0309608     3.60   0.000     .0505467    .1722895
       _cons |   .7979156    .022811    34.98   0.000     .7530674    .8427638
------------------------------------------------------------------------------

. bootstrap: reg pfries income prpblck
(running regress on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Linear regression                               Number of obs      =       392
                                                Replications       =        50
                                                Wald chi2(2)       =     38.04
                                                Prob > chi2        =    0.0000
                                                R-squared          =    0.0749
                                                Adj R-squared      =    0.0702
                                                Root MSE           =    0.1022
```

```
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   2.37e-06   3.86e-07     6.14   0.000     1.62e-06    3.13e-06
     prpblck |   .1114181    .033164     3.36   0.001     .0464178    .1764184
       _cons |   .7979156   .0206537    38.63   0.000     .7574352     .838396
------------------------------------------------------------------------------

. reg pfries income prpblck, vce(bootstrap, reps(10000) nodots)

Linear regression                               Number of obs      =       392
                                                Replications       =     10000
                                                Wald chi2(2)       =     34.30
                                                Prob > chi2        =    0.0000
                                                R-squared          =    0.0749
                                                Adj R-squared      =    0.0702
                                                Root MSE           =    0.1022

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   2.37e-06   4.07e-07     5.82   0.000     1.57e-06    3.17e-06
     prpblck |   .1114181   .0348967     3.19   0.001     .0430218    .1798144
       _cons |   .7979156   .0219084    36.42   0.000     .7549759    .8408553
------------------------------------------------------------------------------

. reg pfries income prpblck if state == 1

      Source |       SS       df       MS              Number of obs =     316
-------------+------------------------------           F(  2,   313) =    4.63
       Model |  .091970048     2  .045985024           Prob > F      =  0.0104
    Residual |  3.10723024   313  .009927253           R-squared     =  0.0287
-------------+------------------------------           Adj R-squared =  0.0225
       Total |  3.19920029   315  .010156191           Root MSE      =  .09964

------------------------------------------------------------------------------
      pfries |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   1.46e-06   4.80e-07     3.04   0.003     5.14e-07    2.40e-06
     prpblck |   .0456923   .0325183     1.41   0.161    -.0182898    .1096745
       _cons |   .8655968   .0260859    33.18   0.000      .814271    .9169226
------------------------------------------------------------------------------
```

```
. bootstrap: reg pfries income prpblck if state == 1
(running regress on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Linear regression                               Number of obs     =        316
                                                Replications      =         50
                                                Wald chi2(2)      =      16.03
                                                Prob > chi2       =     0.0003
                                                R-squared         =     0.0287
                                                Adj R-squared     =     0.0225
                                                Root MSE          =     0.0996

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   1.46e-06   3.74e-07     3.90   0.000     7.26e-07    2.19e-06
     prpblck |   .0456923   .0354046     1.29   0.197    -.0236994    .1150841
       _cons |   .8655968   .0218178    39.67   0.000     .8228348    .9083589
------------------------------------------------------------------------------

. regress pfries income prpblck if state == 1, vce(bootstrap, reps(10000) nodots)

Linear regression                               Number of obs     =        316
                                                Replications      =      10000
                                                Wald chi2(2)      =      12.84
                                                Prob > chi2       =     0.0016
                                                R-squared         =     0.0287
                                                Adj R-squared     =     0.0225
                                                Root MSE          =     0.0996

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   1.46e-06   4.14e-07     3.53   0.000     6.48e-07    2.27e-06
     prpblck |   .0456923   .0357392     1.28   0.201    -.0243553    .1157399
       _cons |   .8655968   .0237188    36.49   0.000     .8191089    .9120848
------------------------------------------------------------------------------

. reg pfries income prpblck if state == 1, vce(bootstrap, reps(100000) nodots)

Linear regression                               Number of obs     =        316
```

```
                                                Replications      =    100000
                                                Wald chi2(2)      =     12.86
                                                Prob > chi2       =    0.0016
                                                R-squared         =    0.0287
                                                Adj R-squared     =    0.0225
                                                Root MSE          =    0.0996

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      pfries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   1.46e-06   4.14e-07     3.53   0.000     6.49e-07    2.27e-06
     prpblck |   .0456923   .0358945     1.27   0.203    -.0246596    .1160443
       _cons |   .8655968   .0237355    36.47   0.000      .819076    .9121176
------------------------------------------------------------------------------
```

First, note that the regressions yield the same parameter estimates as the ML programs above, which we expect since the ML estimator and the OLS estimator are the same when the errors are assumed to be distributed normally. The bootstrap standard errors are roughly the same as those obtained by the regression. The bootstrap standard errors are larger for the prpblck but smaller for the constant term and for income. This is true for complete sample as well as the New Jersey subset, when the number of repetitions is high (10,000 and 100,000 in the case above). If too few repetitions are used this pattern does not have to hold.

By the very nature of bootstrapping (it involves random sampling of the data with replacement), rerunning the bootstrap will yield different results every time, unless the same seed is used for the random number generator (in Stata, this can be done with the seed(#) option. Sometimes the estimates are higher than the OLS regression estimate, sometimes they are lower. An alternative way to obtain bootstrapped standard errors is to use the option vce(bootstrap) when running regress.

Increasing the number of repetitions will lower the variance of the bootstrap estimated standard errors obtained by repeated runs of the bootstrap procedure. See above for the results of 10,000 and 100,000 repetitions (the 100,000 repetitions takes a few minutes). However, if the size of the dataset is rather small, it is possible that the distribution of the data does not reflect the true distribution; bootstrapping will not solve this problem.