University of California, Santa Barbara **Olivier Deschenes**
Department of Economics **Winter 2018**

ESM 296
Individual Assignment 2

Answer Key

The questions below are from Stock and Watson and from Wooldridge

Question 1 (8 points):

Earnings functions attempt to find the determinants of earnings, using both continuous and binary variables. One of the central questions analyzed in this relationship is the returns to education.

(a) Collecting data from 253 individuals, you estimate the following relationship

$$\ln(\hat{Earn}) = \underset{(0.14)}{0.54} + \underset{(0.011)}{0.083} \times Educ\ ,\ R^2 = 0.2, SER = 0.445$$

Where *Earn* is average hourly earnings and *Educ* is years of education.

What is the effect of an additional year of schooling? If you had a strong belief that years of high school education were different from college education, how would you modify the equation? That if your theory suggested that there was a "diploma effect"?

> **ANSWER:** *One additional year of education carries an 8.3% increase, or return, on earnings. You would need additional data to see if this coefficient was different for high school versus college education. Including both variables in the regression would then allow you to test for equality of the coefficients. A "diploma effect" could be studied by creating a binary variable for a high school diploma, a junior college diploma, a B.A. or B.Sc. diploma, etc.*

(b) In Labor Economics, we teach a model of human capital investments where there are returns to on-the-job training. To approximate on-the-job training, researchers often use a potential experience variable, which is defined as *Exper* = *Age* – *Educ* – 6.

You incorporate the potential experience variable into your original regression

$$\ln(\hat{Earn}) = -0.01 + 0.101 \times Educ + 0.033 \times Exper - 0.0005 \times Exper^2 ,$$
$$\quad (0.16) \quad (0.012) \qquad (0.006) \qquad (0.0001)$$

$$R^2 = 0.34$$

Test for the statistical significance of each of the coefficients. Why has the coefficient on education changed little compared to (a)?

ANSWER: *The t-values for the coefficients on Educ (t-value: 8.42), Exper (t-value: 5.5), and Exper^2 (t-value: -5) are all significant (they are greater than 1.96, the appropriate value for testing significance at the 95% level).*

In part (a), the coefficient was 0.083, while here it is 0.101. These are relatively close to each other. This can be explained by the nature of the experience variables. Generally, workers with many years of education do not work more years (i.e. have more experience) than those with fewer years of education, so education and experience may be roughly orthogonal (uncorrelated). So in the regression in part (a), perhaps experience was not a source of omitted variable bias.

There may be a little mechanical bias induced by the rule assigning experience. If anything, 1 additional year of education generally results from not participating in the workforce (gaining 1 year of experience). After including this, the coefficient on education increases marginally, so perhaps education and experience are slightly negatively correlated. However, this effect seems to be minor.

(c) You want to find the effect of introducing two variables, gender and martial status. Accordingly, you specify a binary variable that takes on the value of one for females and is zero otherwise (*Female*), and another binary variable that is one if the worker is married but is zero otherwise (*Married*). Adding these variables to the regression results in:

$$\ln(\hat{Earn}) = 0.21 + 0.093 \times Educ + 0.032 \times Exper - 0.0005 Exper^2$$
$$\quad (0.16) \quad (0.012) \qquad (0.006) \qquad (0.0001)$$

$$- 0.289 \times Female + 0.062 \times Married$$
$$(0.049) \qquad (0.056)$$

$$R^2 = 0.43, SER = 0.378$$

Are the coefficients of the two added binary variables individually statistically significant? Are they economically important? In percentage terms, how much less do females earn per hour, controlling for education and experience? How much more do married people make? What is the percentage difference in earnings between a single male and a married female? What is the marriage differential between males and females?

ANSWER: *The coefficient for the female binary variable is statistically significant even at the 1% level. The coefficient for the married binary variable only has a t-statistic of 1.11 and is not statistically significant at the 10% level. Both coefficients indicate economic importance, since females make approximately 28.9% less than males and married people earn roughly 6.2 percent more (though this is insignificant in statistical terms). A married female earns 22.7% less (-0.289 + 0.062) than a single male. Married females earn 28.9% less than married males, the same percentage that single females earn less than single males.*

(d) In your final specification, you allow for the binary variables to interact. The results are as follows:

$$\ln(\hat{E}arn) = \underset{(0.16)}{0.14} + \underset{(0.011)}{0.093} \times Educ + \underset{(0.006)}{0.032} \times Exper - \underset{(0.0001)}{0.0005} Exper^2$$

$$- \underset{(0.075)}{0.158} \times Female + \underset{(0.080)}{0.173} \times Married - \underset{(0.097)}{0.218} \times (Female \times Married)$$

$$R^2 = 0.44, SER = 0.375$$

Repeat the exercise in (c) of calculating the various percentage differences between gender and marital status.

ANSWER: *The default is the single male. Single females earn 15.8% less. Married males earn 17.3% more. Married females earn 20.3% less (-0.158 + 0.173 – 0.218) than single males. Comparing married females with married males now results in a percentage differential of 37.6% in favor of the males (-0.158 – 0.218).*

Question 2 (8 points):

The question below requires the STATA data file "GPA2.dta" is available on the class website. The same file also available in spreadsheet format "GPA2.csv". Now consider the following regression model:

$$COLLGPA = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u$$

Where *COLLGPA* is cumulative college grade point average, *hsize* is size of high school graduating class (in hundreds), *hsperc* is academic percentile in graduating class, *sat* is combined SAT score, *female* is a binary gender variable, and *athlete* is a binary variable equal to one for student-athletes.

(a) Estimate the parameters of the regression model above by OLS. What is the estimated GPA differential between athletes and non-athletes?

ANSWER:

```
. regress colgpa hsize hsizesq hsperc sat female athlete, robust

Regression with robust standard errors                 Number of obs =    4137
                                                       F(  6,  4130) =  301.41
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2925
                                                       Root MSE      =   .5544

------------------------------------------------------------------------------
             |               Robust
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0568543   .0169123    -3.36   0.001    -.0900115   -.0236971
     hsizesq |   .0046754   .0023388     2.00   0.046       .00009    .0092608
      hsperc |  -.0132126   .0005639   -23.43   0.000    -.0143182   -.012107
         sat |   .0016464   .0000666    24.72   0.000     .0015158    .001777
      female |   .1548814    .017923     8.64   0.000     .1197427    .1900201
     athlete |   .1693064   .0369629     4.58   0.000     .0968391    .2417736
       _cons |   1.241365   .0799464    15.53   0.000     1.084627    1.398103
------------------------------------------------------------------------------
```

The estimated GPA differential between athletes and non-athletes is 0.169 GPA points in favor of athletes conditional on SAT score.

(b) Drop *sat* from the model and re-estimate the parameters of the regression model. What is the estimated GPA differential between athletes and non-athletes? Explain why the estimate is different than the one in (a).

ANSWER:

```
. regress colgpa hsize hsizesq hsperc female athlete, robust

Regression with robust standard errors                 Number of obs =    4137
                                                       F(  5,  4131) =  187.97
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1885
                                                       Root MSE      =  .59368

------------------------------------------------------------------------------
             |               Robust
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0534038   .0180048    -2.97   0.003    -.0887028   -.0181048
     hsizesq |   .0053228   .0024964     2.13   0.033     .0004285    .0102171
      hsperc |  -.0171365   .0005962   -28.74   0.000    -.0183054   -.0159675
      female |   .0581231   .0187994     3.09   0.002     .0212662    .0949801
     athlete |   .0054487   .0392878     0.14   0.890    -.0715765    .0824739
       _cons |   3.047698    .033987    89.67   0.000     2.981065    3.114331
------------------------------------------------------------------------------
```

When not controlling for SAT score, the estimated GPA differential between athletes and non-athletes is 0.005 points in favor of athletes although it is not significant. This suggests that the *variables sat and athletes are negatively correlated. (Recall the formula for omitted variable bias in Lecture 3).*

As an alterantive way to test for this, you could regress SAT on athlete and all the other variables:
`. regress sat hsize hsizesq hsperc female athlete, robust`
This gives a coefficient on athlete of roughly -100 SAT points and is very significant. This is evidence of strong correlation.

(c) Including the *sat* variable, re-estimate the model while allowing the effect of being an athlete to differ for males and females and test the null hypothesis that there is no difference in the GPA of female athletes and non-athletes. What about male athletes and non-athletes?

ANSWER:

```
. gen femaleXathlete=female*athlete

. regress colgpa hsize hsizesq hsperc sat female athlete femaleXathlete, robust

Regression with robust standard errors                 Number of obs =    4137
                                                       F(  7,  4129) =  258.25
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2925
                                                       Root MSE      =  .55446

------------------------------------------------------------------------------
             |               Robust
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0568006    .016936    -3.35   0.001    -.0900043   -.0235969
     hsizesq |   .0046699   .0023415     1.99   0.046     .0000793    .0092605
      hsperc |  -.0132114   .0005636   -23.44   0.000    -.0143164   -.0121065
         sat |   .0016462   .0000667    24.70   0.000     .0015155    .0017769
      female |   .1546151    .018304     8.45   0.000     .1187294    .1905007
     athlete |   .1674185   .0411887     4.06   0.000     .0866665    .2481705
femaleXath~e |   .0076921   .0862602     0.09   0.929    -.1614243    .1768086
       _cons |   1.241575   .0800111    15.52   0.000      1.08471    1.398439
------------------------------------------------------------------------------

. test female + athlete + femaleXathlete = female

 ( 1)  athlete + femaleXathlete = 0

       F(  1,  4129) =    5.14
            Prob > F =    0.0234
```

Holding the other variables in the regression constant, the GPA of female athletes is significantly higher than female non-athletes by 0.175 points (0.167+0.007). The coefficient on athlete is the estimated difference in the GPA of male athletes and non-athletes. The GPA of male athlete is 0.167 points higher than that of male non-athletes. The estimate is significant at the 1% level (t-stat = 4.06 > 2.58).

(d) Does the effect of *sat* on *COLLGPA* differ by gender? Justify your answer.

ANSWER:

```
. gen femaleXsat=female*sat

. regress colgpa hsize hsizesq hsperc sat female athlete femaleXathlete
femaleXsat, robust

Regression with robust standard errors                 Number of obs =    4137
                                                       F(  8,  4128) =  228.03
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2925
                                                       Root MSE      =  .55452

------------------------------------------------------------------------------
             |               Robust
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize | -.0568198   .0169421    -3.35   0.001    -.0900355   -.0236041
     hsizesq |  .0046773   .0023424     2.00   0.046     .0000849    .0092696
      hsperc | -.0132236   .0005634   -23.47   0.000    -.0143281   -.0121192
         sat |   .001624   .0000871    18.64   0.000     .0014532    .0017949
      female |  .0990198   .1328792     0.75   0.456     -.161495    .3595346
     athlete |  .1643156   .0420874     3.90   0.000     .0818016    .2468296
femaleXath~e |  .0136833     .08806     0.16   0.877    -.1589618    .1863284
  femaleXsat |  .0000539   .0001271     0.42   0.671    -.0001952    .0003031
       _cons |  1.265315   .0994457    12.72   0.000     1.070347    1.460282
------------------------------------------------------------------------------
```

The estimated coefficient on the interaction between female and sat is not significant at conventional levels, implying that the effect of sat on COLLGPA does not differ by gender.