ECO 341K **Introduction to Econometrics**
University of Texas

Homework Assignment #3 (due February 13, 2:00 p.m.)

Written problems:

1. Recall the *reverse regression* that you ran in Homework Assignment #2. Rather than the SLR of *y* on *x*, you ran a SLR of *x* on *y*.

   a. How does the R-squared value of the reverse regression (*x* on *y*) compare to the R-squared value of the original regression (*y* on *x*)? Explain.
   b. Do the slope estimates from the two regressions have the same sign? Explain.
   c. How are the magnitudes of the two slope estimates related? Explain.

2. Wooldridge: Chapter 2, Problem 6

3. Although we focused upon logarithms in class, one could use other non-linear transformations within a regression model. For the wage and education example that has been considered in lecture, suppose that we wanted to take the square root of education and relate wages to that. So the SLR model would be

$$wage = \beta_0 + \beta_1 \sqrt{educ} + u$$

   a. Figure out the formula for the effect of *educ* on *wage* by taking the derivative *dE(wage|educ)/d educ*. Note that unlike the basic SLR model, this derivative depends upon the *x* variable (here, *educ*).

   Here is the Stata for the regression of this model (using *wage1.dta*):

```
. gen sqrteduc = sqrt(educ)

. regr wage sqrteduc

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =   78.27
       Model |  930.606128     1  930.606128           Prob > F      =  0.0000
    Residual |  6229.80816   524  11.8889469           R-squared     =  0.1300
-------------+------------------------------           Adj R-squared =  0.1283
       Total |  7160.41429   525  13.6388844           Root MSE      =   3.448

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sqrteduc |   2.947042   .3331004     8.85   0.000     2.292666    3.601419
       _cons |  -4.464346   1.180639    -3.78   0.000    -6.783714   -2.144978
------------------------------------------------------------------------------
```

b. Estimate the effect of *educ* on *wage* at both *educ=12* and *educ=16* (using your formula from the previous part). How do these effects compare to the estimated effects from the original SLR model? For your reference, the original regression (*wage* on *educ*) results were:

```
. regr wage educ

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =  103.36
       Model |  1179.73204     1  1179.73204           Prob > F      =  0.0000
    Residual |  5980.68225   524  11.4135158           R-squared     =  0.1648
-------------+------------------------------           Adj R-squared =  0.1632
       Total |  7160.41429   525  13.6388844           Root MSE      =  3.3784

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5413593    .053248    10.17   0.000     .4367534    .6459651
       _cons |  -.9048516   .6849678    -1.32   0.187    -2.250472    .4407687
------------------------------------------------------------------------------
```

c. Which regression (*wage* on *educ* or *wage* on *sqrteduc*) appears to give a better overall fit? Should overall fit (as measured by R-squared) be the only reason to choose one model specification over another? Explain.

Computer problems (show any relevant Stata output):

Wooldridge: Chapter 2, Computer Exercise C6
(Note that *log(expend)* is included in the data set as the variable named *lexpend*.)

For the same dataset (MEAP93.DTA), also answer the following questions:

(vi) Plot *math10* versus *lexpend* with the fitted regression line shown.

(vii) To visualize the non-linearity described by this model, create the fitted values from the regression and then do a scatter plot of the fitted values versus *expend* (not *lexpend*). Explain how the effect of expenditures changes at higher values of expenditures.

(viii) Suppose that we wanted to measure *math10* as a fraction (a number between 0 and 1) rather than on a 0-to-100 scale. Specifically, we could do the following in Stata to re-scale the math10 variable:

```
. replace math10 = math10 / 100
```

(This replaces the original *math10* values with the values divided by 100.) If you re-ran the regression (now regressing the re-scaled *math10* upon *lexpend*), how would the following quantities change (compared to the original results)? Be specific, and try these on your own before checking your answers in Stata.

(a) R-squared
(b) SST
(c) the slope estimate
(d) the intercept estimate