Economics 312 **Spring 2012**
Daily Problem #6 **February 3**

The regressions below estimate the bivariate relationship between hourly wages (in dollars) and years of education, using a sub-sample of 1000 observations from the Current Population Survey (date of survey unknown). The education variable is in years, created by transforming the CPS education attainment variable by the following formula:

```
00 .Less than 1st grade
03 .1st,2nd,3rd,or 4th grade
03 .5th or 6th grade
08 .7th and 8th grade
09 .9th grade
10 .10th grade
11 .11th grade
12 .12th grade no diploma
12 .High school graduate - high school diploma or equivalent
13 .Some college but no degree
14 .Associate degree in college - occupation/vocation program
14 .Associate degree in college - academic program
16 .Bachelor's degree (for example: BA,AB,BS)
18 .Master's degree (for example:MA,MS,MENG,MED,MSW, MBA)
21 .Professional school degree (for example: MD,DDS,DVM,LLB,JD)
21 .Doctorate degree (for example: PHD,EDD)
```

The sample statistics for the two variables, and for the log of the wage, are below:

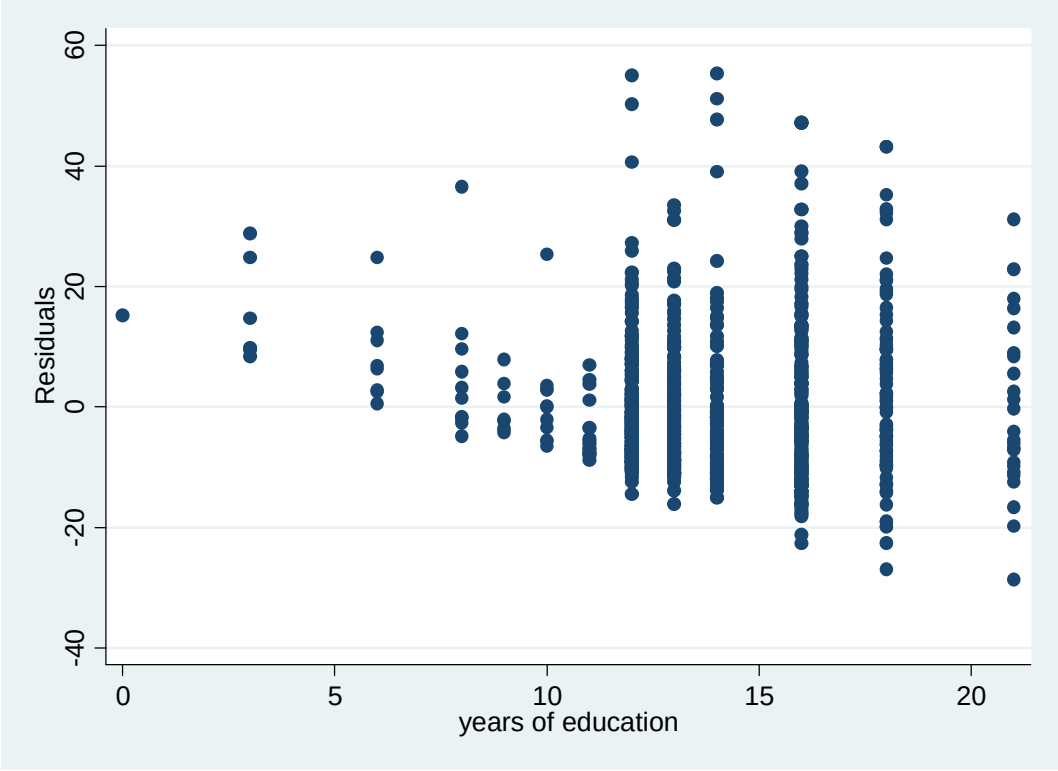
Variable	Obs	Mean	Std. Dev.	Min	Max
wage	1000	20.61566	12.83472	1.97	76.39
lnwage	1000	2.856988	.5806185	.6780335	4.335852
educ	1000	13.799	2.711079	0	21

1. Linear model. The following Stata regression table is the outcome of regressing wage on years of education. Below it is a scatter-plot of the residuals against the regressor (education).

```
. reg wage educ

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  211.66
       Model |  28794.2878     1  28794.2878           Prob > F      =  0.0000
    Residual |   135771.14   998  136.043226           R-squared     =  0.1750
-------------+------------------------------           Adj R-squared =  0.1741
       Total |  164565.428   999  164.730158           Root MSE      =  11.664

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    1.980288   .1361174    14.55   0.000     1.713178    2.247397
       _cons |   -6.710328   1.914156    -3.51   0.000    -10.46656   -2.954096
------------------------------------------------------------------------------
```



a) What is the estimated effect of 4 years of college on wage? What is the standard error of the 4-year effect. What is a 95% confidence interval for the effect?

b) Suppose that if a prospective student invested the cost of a college education, the interest she would earn would be equivalent to a $6 higher wage for her working career. Can we reject the null hypothesis that college is not worth the cost at a 5% two-tailed level of significance?

c) How good is the fit? What is the standard deviation of the residual and in what units is it measured? Does this seem large or small?

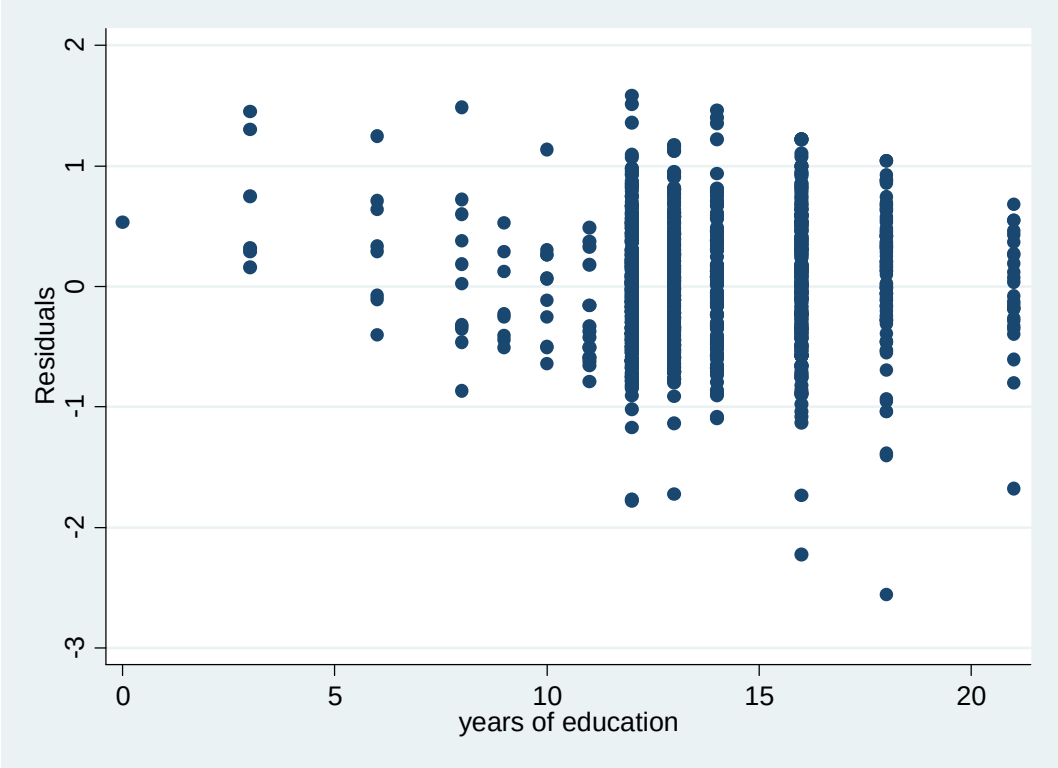
d) Does the variance of the error term seem to be constant across observations; in particular, does it seem to vary with years of education? What does this mean?

2. Log-linear model. The Stata output and residual plot below are from a regression of log(wage) on education, using the same data set.

```
. reg lnwage educ

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  216.41
       Model |   60.015841     1   60.015841           Prob > F      =  0.0000
    Residual |   276.76489   998   .27731953           R-squared     =  0.1782
-------------+------------------------------           Adj R-squared =  0.1774
       Total |  336.780731   999  .337117849           Root MSE      =  .52661

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0904082   .0061456    14.71   0.000     .0783484    .1024681
       _cons |   1.609444   .0864229    18.62   0.000     1.439853    1.779036
------------------------------------------------------------------------------
```



a) What is the estimated rate of return to education, i.e., what percentage increase in wage results from one more year of education?

b) What is the standard deviation of the residual and in what units is it measured? Does this seem large or small? How does it compare to the linear model?

c) What is the effect of four years of college on ln wage? On wage?

d) How does the validity of the homoskedasticity assumption compare to the linear model?

e) What criteria might you use to choose between the linear and the log-linear model?