Econometrics
Exercise: Multiple Regression

NOTE: You need to turn in all code and results to answer the following copied into a word file just as you have done previously.

Load the data set 'wages.csv'

1. Estimate the following model: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$ (not that you would ever want to really run this functional form most likely, but just for some practice).

a. Is $exper^2$ significant at the 1% level?

b. Use the function: $\Delta wage = (\beta_2 + 2\beta_3 \cdot exper)\Delta exper$ to find the return to the sixth year of experience (at 5 going to 6), what about the 11th (at 10 going to 11)?

c. Use the 'predict' function to see how the predicted values compare for 5 vs. 6 and 10 vs. 11 years of experience to see how good this approximation is (note it does not matter what value you put in for the level of education, so just say it = 12). If you forgot how to do this look a the notes from the end of the R exercise 'Basics of R Part 2'.

2. Consider a model where the return to education depends upon the amount of work experience (and vice versa): $log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3(educ \cdot exper) + u$

a. What is the return to another year of education ($\Delta log(wage)/\Delta educ$) holding experience fixed at 3 years?

b. State a null hypothesis that the returns to education do not depend on experience.

c. Test the null hypothesis above against the one-sided hypothesis, that the return to education increases with working experience.

d. Rerun the above equation but include a dummy variable for *Black*. Interpret the coefficient on *Black correctly.*

3. Use the housing price data 'house_price.csv'

a. Estimate the model: $log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_2 bdrms + u$
report and interpret the results.

b. Find the predicted value for log(price), when lotsize = 20,000, sqrft = 2,500, and bdrms = 4.

Predicting y

Now you have a value for your predicted $log(\hat{price})$. What if I asked for a predicted value of price? You may say well I just raise 'e' to the power $log(\hat{price})$ right (recall if $y = log(x)$ then $x = e^y$)? Well no. Lets see why you cannot do this (or shouldn't).

So we said the correct model is $ln(y) = \beta X + u$ and so $E(ln(y)|X) = \beta X + E(u|X) = \beta X$
Make sure you see why each equality holds (assuming our OLS assumptions hold)

So we can easily predict our values of $ln(y)$.

So what about y? (below instead of e^x I will write $exp(x)$)

$ln(y) = \beta X + u \rightarrow y = exp(\beta X + u) = exp(\beta X) \cdot exp(u) \rightarrow E(y|X) = exp(\beta X) \cdot E(exp(u)|X)$

And while the expectation of u conditional on X is (assumed) zero, the expectation of $exp(u)$ conditional on X is not zero. In fact $exp(u)$ is always positive and its expectation is greater than one (if the mean of u is zero than the mean of $exp(u)$ must be greater than one - to see this note $exp(x)$ is less than one if x is less than 0 and greater than 1 if x is greater than 0, but they average to greater than 1 because it is exponential growth).

So we see that our true estimate of $E(y|X)$ is $exp(\beta X) \cdot \alpha$, where α is an estimate of that expectation - so we would have underestimated our true prediction of y.

So to get a prediction of y we need $exp(\beta X)$ and an estimate of α. One simple one is to use a sample average in place of the expectation, and use our residuals in place of our unknown error terms.

$\hat{\alpha} = \frac{1}{n} \sum exp(\hat{u})$

So now:

c. Using the above method find the predicted price at the values given in b.
(Hint. the residuals are saved in your regression output as 'myregression$residuals', and you can just use the 'mean' function and the 'exp' function also)

d. How much of a difference is there if you do not make this correction?