Economics 475: Econometrics

Homework #2

This homework is due on Monday, January 30th.

1. In class we demonstrated that the OLS estimates of $\hat{B}_1$ is an unbiased estimate of β_1. Show that $\hat{B}_0$ is an unbiased estimate of β_0. (Hint: Remember $\hat{B}_0 = \overline{Y} - \hat{B}_1 \overline{X}$). What assumptions are necessary for $\hat{B}_0$ to be an unbiased estimate of β_0?

2. Open the data set, "Whatcom County Homesales" posted on my website. This data consists of observations from all home sales in the year 2000 in Whatcom County.

The data are defined as:
Area: A code for the home's location within Whatcom County
Number: The numerical portion of the home's address
Address: The street portion of the home's address
New: Binary equal to 1 if home is new
Month: The month of home sale (1 = January, 2 = February)
Price: The home's sale price
Sqft: Square footage of house
Style: A categorical variable indicating style
Yr__Built: Year the house was built
Bedrooms: # of home's bedrooms
Age: 2000 – Yr__Built
lnprice: Natural log of Price

Consider the regression:

$$lnprice_i = \beta_0 + \beta_1 sqft_i + \beta_2 sqft_i^2 + \beta_3 bedrooms_i + \beta_4 age_i$$

a. Estimate the regression above and interpret the coefficients. Carefully describe the relationship between the home price and square footage.

b. The coefficient on bedrooms turns out to be not statistically different than zero. However, it seems that people like homes with more bedrooms. What explains this odd result?

c. Use the residuals from the regression in part a and create a plot of the residuals and an independent variable (your choice) to search for heteroskedasticity. What do you find? (Question: is it appropriate to search for heterskedasticity by plotting residuals against one of the two sqft variables?)

d. Perform a Park Test on Age. Does this test indicate a heteroskedasticity problem?

e. Perform a White test on the regression in part a. Do you find heteroskedasticity? Describe the pros and cons of the White test versus the Park test.

f. Regardless of your answers to parts c through e, imagine that heteroskedasticity existed in the regression of part a. Specifically, assume that the $Var(\varepsilon) = Age_i \times \sigma^2$. Use the weighted least squares technique to correct for this type of heteroskedasticity and make comparisons to your original regression in part a.

g. Using the weighted least squares technique based upon Age in part f, has the heteroskedasticity problem been eliminated?

h. Rather than knowing the form of the heteroskedastiticy as given in part f, it is unlikely (often impossible) to know the true form of the heteroskedasticity. Using the original regression in part a, re-estimate this model using Feasible GLS. Compare this estimator to that presented in part a.

3. Using your final project data, answer the following questions.

a. Describe each variable to me. What is your dependent variable? Independent variable(s)? What do they measure? Where do they come from?

b. Estimate a regression using your variables. Show me your results. Describe what you are looking for in this regression.

c. Does your regression have heteroskedasticity?