hw04: Diagnostics and Alternative Methods of Regression

*Your name**

19 07, 2017

Question 1

Consider the following model to explain monthly beer consumption (beer):

$$\frac{beer}{inc} = \beta_0 \frac{1}{inc} + \beta_1 \frac{inc}{inc} + \beta_2 \frac{price}{inc} + \beta_3 \frac{educ}{inc} + \beta_4 \frac{female}{inc} + u'$$

$$\mathrm{E}(u|inc, price, educ, female) = 0$$

$$Var(u|inc, price, educ, female) = \sigma^2 inc^2$$

Where *inc* represents income, *price* represents price of beer, *educ* represented education in years, and *female* is a dummy variable for females.

Write the transformed equation that has a homoscedastic error term.

Question 2

Use the **hprice1.csv** data on Dropbox (use **ghfile** function we set in classroom to retrieve the dataset) to estimate the following equation:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

Variable definitions are in the **hprice1_description.txt** file.

Report your coefficient estimates, standard errors, coefficient of determination and note which estimates are significant.

Plot the residuals from the equation, and comment on whether they look homoskedastic.

Test for homoskedasticity using the BP test. Manually run the full version of White's test using the chi-squared statistic. What do you conclude?

Estimate White's heteroskedasticity consistent standard errors. Do these make a difference to your conclusions about which coefficient estimates are significant?

Estimate the equation using log values of price, lotsize and sqrft. Does this specification appear to have problems with heteroscedasticity?

*affiliation

Question 3

The file on Dropbox called **housing.csv** contains data relevant to the US housing market from 1973 to 2011. Develop a suitable linear regression model to explain annual housing starts (the number of new residential construction projects that have begun during particular time period). Housing starts are a key indicator of US macroeconomic activity.

a)

Select and appropriate functional form for your model of housing starts that is consistent with economic theory, and estimate your model. Present the results.

b)

What does theory suggest regarding the sign of your estimated coefficients? Are your estimates the correct sign?

c)

Investigate whether your model may have a problem with serial correlation, using graphical methods as well as diagnostic tests.

d)

If you reject the null hypothesis of no serial correlation, investigate whether using a lagged dependent variable in your model helps.

Question 4

If the presence of non-spherical errors (heteroskedasticity and/or serial correlation) causes our variance estimates of the OLS coefficients to be underestimated, as is usually the case, then the probability of making a type I error increases. True, false or uncertain? Explain, and refer to one of the empirical examples used in class to support your explanation.

Notes on the data files

housing.csv Description:

Variables:

1. un: Seasonally adjusted civilian unemployment rate (%)
2. hstart: New housing starts from monthly data at seasonally annual rate (thousands)
3. m2: Seasonally adjusted M2 money supply (billions of dollars)
4. mgrate: New home mortgage interest rate (%)
5. primerate: Prime interest rate charged by banks (%)
6. rgdp: Real GDP, billions of chained 2005 dollars, from quarterly data at seasonally adjusted annual rates

hprice1.csv Description:

Obs: 88

1. price: house price, $1000s
2. assess: assessed value, $1000s
3. bdrms: number of bedrooms
4. lotsize: size of lot in square feet
5. sqrft: size of house in square feet
6. colonial = 1 if home is colonial style