Exercise - Week 4

Part 1: How Does Heteroskedasticity Look?

1 Download a do-file, called *hetero.do*, from the course webpage.

2 Open the file, using the STATA Do-file Editor.

3 Discuss what this simulation is about before running the program.

4 Let's run the program and check whether the results are produced as you expect.

Part 2: Heteroskedasticity Tests

1 Load the HPRICE2.DTA data set.

2 Run ordinarily the regression:

$$\ln Hprice_i = \beta_0 + \beta_1 \ln Nox_i + \beta_2 \ln proptax_i + \beta_3 \ln dist_i + \beta_4 rooms_i + u_i$$

3 Generate the residuals from the regression:

predict e, residuals

4 Inspect the residuals with each of log Nox, log Proptax, log Dist, and Rooms:

scatter e lnox

scatter e lproptax

. . .

What do you find about heteroskedasticity?

5 Run the same regression using the option **robust**. What do you notice?

reg lprice lnox lproptax ldist rooms, robust

6 Compute the Breusch-Pagan test:

- generate the square of the residuals: **gen e2 = e^2**

- run the regression of e2 on all the independent variables:

  reg e2 lnox lproptax ldist rooms

- construct the test statistic:

$$LM = N \times R^2_{\hat{u}^2}$$

- Given the 5% significance level, what is your conclusion?

Part 3: Autocorrelation Tests

1 Load the global.dta data set.

2 For the remaining of the exercise, Stata should know that the global warming data has a time series structure. To this end, type **tsset year**. This will indicate that the variable coding for time is the variable year.

3 Estimate the following model:

$$\ln Temp_t = \beta_0 + \beta_1 \ln Co2_t + \beta_2 Year_t + u_t.$$

Get the residual of the regression: **predict e, residuals**.

- Plot the residual against time. Is there any evidence of autocorrelation?
- Test for serial correlation using the Durbin-Watson test: type **dwstat**, to get the Durbin-Watson test. What do you conclude? (given 5% signficance level, the critical values of the DW test are $dl = 1.75$ and $du = 1.79$).

4 Estimate the following model:

$$\ln Temp_t = \beta_0 + \beta_1 \ln Co2_t + \beta_2 Year_t + \beta_3 \ln Temp_{t-1} + u_t.$$

Run the regression of this model: **reg ltemp lco2 year l1.ltemp**

- Is the lagged log temperature significantly different from 0? If this variable is left out of the regression, what effect does this have on the residuals?
- Why can't use the Durbin-Watson test in this case?
- Test for autocorrelation of order 2 in the residuals using the Breusch-Godfrey test ($u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2}$). What do you conclude? (given 5% significance level, the critical value of the BG test based on the chi-square distribution is 5.99).