Computer Homework 8

This homework examines the problem of heteroskedasticity (non-common variance) in the estimation of linear regression models. Estimate the model:

$$SP = \alpha + \beta\, SQFT + \varepsilon$$

using the *Stata* data file mls.dta.

1.   Obtain OLS estimates and standard errors for β.

2.   Use the regress post-estimation command line **hettest** to obtain the Breusch-Pagan test for heteroskedasticity. Is the null hypothesis of common variance rejected?

3.   Next, rerun the model, but add the *Stata* option **robust** to the regress command. This option will provide White's heteroskedasticity consistent standard errors. Compare the coefficient estimates and standard errors with those from part (1).

4.   The *Stata* command **graph twoway scatter y x** provides a 2D plot with y on the vertical axis and x on the horizontal axis. Use this command to plot the square of floor space against the OLS residuals. Does there appear to be any evidence of heteroskedasticity?

5.   Construct the transformed variables and use OLS to obtain GLS estimates under the assumption that $Var(\varepsilon_i) = \sigma^2(SQFT_i)^2$. With this structure, the variance of the individual observations is proportional to the square of floor space. Compare the GLS estimates and standard errors with those obtained in parts (1) and (2). What are the costs and benefits of this procedure relative to using White's standard errors?

6.   Conduct a Breusch-Pagan test for heteroskedasticity on the transformed model. Is the null hypothesis of common variance rejected?

7.   Plot the square of floor space against the GLS residual (OLS residual from the transformed model). Any appearance of heteroskedasticity?

8.   Are the values of R^2 comparable across the original and transformed models? Why or why not? What limitations exist if one tries to construct R^2 using the GLS estimates and the untransformed data?

Repeat the steps above for the subsample of houses less than 2,500 square feet in size.

9.   Compare the OLS and GLS estimates of β from the full sample with those from the sample of houses less than 2,500 square feet in size. What do the results suggest about subsample analysis as a potential method of insuring common structure?

10.   How does the use of the "less than 2500 square feet" subsample affect interpretation of the coefficients?