Heteroskedasticity Exercise: Spring 2010

Exercise (Verbeek, 2008) The dataset AIRQ contains observations for 30 standard metropolitan statistical areas for California in 1972. The following variables are included:

- Airq = indicator of air quality—lower is better
- Vala = value added of companies ($1000)
- Rain = amount of rain in inches
- Coas = 1 if on the coast, 0 otherwise
- Dens = population density (per square mile)
- Medi = average income per head ($1)

a) Estimate a linear regression that explains *airq* from the other variables.
b) Test the null hypothesis that average income does not affect air quality. Test the joint hypothesis that none of the variables has an effect on air quality.
c) Test whether the variance of the error term is different for coastal and noncoastal areas. Based on this, comment on the validity of the tests in the first two parts.
d) Perform a Breusch-Pagan test for heteroskedasticity related to all five variables. Does this make you doubt the presence of heteroskedasticity?
e) Perform White's test. Comment on the appropriateness of the White test in light of the number of observations and the degrees of freedom of the test.
f) Assume you have multiplicative heteroskedasticity related to *coas* and *medi*. Estimate the model using NLLS. Test the null hypothesis that the coefficients on these two variables are zero.
g) Using the estimates of the multiplicative heteroskedastic model, estimate the linear regression using FGLS.
h) Comment on the appropriateness of R-square in the FGLS regression.