Exercises Week 3

Econometrics

1. **Exercise 7.1 in ETM**: Using the fact that $E(UU'|X) = \Omega$ in the regression $Y = X\beta + U$, show directly, without appeal to standard OLS results, that the covariance matrix of the GLS estimator, $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$, is given by

$$Var(\hat{\beta}_{GLS}) = (X'\Omega^{-1}X)^{-1}.$$

2. **(Adapted) Exercise 7.2 in ETM**. In the lecture we argued that in case of heteroskedasticity/autocorrelation, GLS is more efficient than OLS by looking at their covariance matrices. You are going to show this. That is, show that

$$Var^{-1}(\tilde{\beta}_{GLS}) - Var^{-1}(\hat{\beta}_{OLS}) = X'\Omega^{-1}X - X'X(X'\Omega X)^{-1}X'X,$$

is a positive semidefinite matrix.

Hint: Write $\Omega^{-1} = PP'$ and show that the difference can be written as $Z'M_Q Z$, where M_Q is the matrix that projects off the space generated by Q, for appropriate Z and Q matrices.

3. Generate a sample of size 50 from the model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

with $\beta_0 = 1$ and $\beta_1 = 1$. For simplicity, assume that x_t are $NID(2, 2)$. Moreover, make $u_t = N(0, \exp(0.6x_t))$; that is, the errors are heteroskedastic.

Then, do the following:

i. Estimate the model by OLS; that is, without correcting for heteroskedasticity. Compute the t-statistic associated to the test $H_0 : \beta_1 = 0$ and determine if you reject the null using a critical value of 2.009. Why use this value?

ii. Estimate the model by WLS using the true weights and compute the t-statistic associated to the test $H_0 : \beta_1 = 0$. Determine if you reject the null using the same critical value as before.

iii. Estimate the model by FGLS using the estimated weights and compute the t-statistic associated to the test $H_0 : \beta_1 = 0$. Determine if you reject the null using the same critical value as before.

iv. Repeat the exercise at least 1000 times and compute the average number of rejections of the null for all methods consider. What do you observe? Explain.

4. **Exercise 4.1 in AGME**: The dataset *Airq* from the *Ecdat* package contains observations for 30 standard metropolitan statistical areas (SMSAs) in California for 1972 on the following variables:

Variable	Description
airq	indicator for air quality (the lower the better)
vala	value added of companies (in 1000 US$)
rain	amount of rain (in inches)
coas	dummy variable, 1 for SMSAs at the coast; 0 for others
dens	population density (per square mile)
medi	average income per head (in US$)

The data are cross-sectional. The objective is to explain air quality by the other variables.

Load the data and do the following:

i. Estimate a linear regression model that explains *airq* from the other variables using ordinary least squares. Interpret the coefficient estimates.

ii. Test the null hypothesis that average income does not effect the air quality. Test the joint hypothesis that none of the variables has an effect upon air quality.

iii. Perform a Breusch–Pagan test for heteroskedasticity related to all five explanatory variables.

iv. Perform a White test for heteroskedasticity.

v. Assuming that we have heteroskedasticity related to *coas* and *medi*, estimate the coefficients by running a regression of $\log u_t^2$ upon these two variables. Test the null hypothesis of homoskedasticity on the basis of this auxiliary regression.

vi. Using the results from v., compute a FGLS estimator for the linear model. Compare your results with those obtained under i. Redo the tests from ii.