Worksheet week # 11

1. Consider the following model that explains major league baseball players' salaries:

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr+$$

$$+\beta_6 srunsyr + \beta_7 fldperc + \beta_8 allstar + \beta_9 frstbase + \beta_{10} scndbase+$$

$$+\beta_{11} thrdbase + \beta_{12} shrtstop + \beta_{13} catcher + \varepsilon \ ,$$

The variables used in the model are the following:

salary ... 1993 total salary

years ... years in the league

gamesyr ... average games played a year

bavg ... career batting average

hrunsyr ... home runs per year

rbisyr ... runs batted in per year

srunsyr ... runs scored per year

fldperc ... career fielding perc

allstar ... percentage of years as all-star

frstbase ... = 1 if playing first base

scndbase ... = 1 if playing second base

thrdbase ... = 1 if playing third base

shrtstop ... = 1 if playing shortstop

catcher ... = 1 if playing catcher.

You are given the data *baseball.xls* with the following variables:

salary ... 1993 total salary

years ... years in the league

games ... career games played

bavg ... career batting average

hruns ... career home runs

rbis ... career runs batted in

sruns ... career runs scored

fldperc ... career fielding perc

yrsallst ... years as all-star

position ... = 0 if outfield,
  = 1 if first base
  = 2 if second base
  = 3 if third base
  = 4 if shortstop
  = 5 if catcher

(a) Use the file *baseball.xls* for this exercise:
  i. Open the file in Excel, save it as *baseball.csv* (comma separated values) files.
  ii. Load the file in Gretl.

(b) Define the new variables you need for the regression.

(c) Estimate the model.

(d) Test for heteroskedasticity using the White test and Breusch-Pagan test.

(e) Reestimate the model to remedy for heteroskedasticity if it is present.

(f) Explain why dummy for outfield players is not included.

(g) Is the average salary of outfield players different from the salary of the first base players?

(h) Suppose you decided to include dummy for outfield players instead of the dummy for first base players. What regression results would you obtain in this case?

(i) Test the null hypothesis that there is no difference in average salary across positions, once other factors have been controlled for.

2. Estimate the impact of GDP on the housing prices level in the UK using quarterly time series data from Q1 1975 to Q2 2011. Consider the following model:

$$h\_price_t = \beta_0 + \beta_1 GDP_t + e_t$$

(a) Load the data *house.gdt* into Gretl and estimate the model.

(b) Test for the presence of AR(1) autocorrelation (positive or negative) in the error term. Define the hypothesis, test statistic, and interpret the results.

(c) Include lagged housing prices into the model and estimate by OLS:

$$h\_price_t = \alpha_0 + \alpha_1 GDP_t + \alpha_2 h\_price_{t-1} + u_t$$

(d) Test for autocorrelation of higher order using the analysis of residuals from the model with lagged housing prices.

(e) Reestimate the model with four lags of housing prices, and test for the autocorrelation of the error term in this model:

$$h\_price_t = \gamma_0 + \gamma_1 GDP_t + \gamma_2 h\_price_{t-1} + \gamma_3 h\_price_{t-2} + \gamma_4 h\_price_{t-3} + \gamma_5 h\_price_{t-4} + v_t$$