B203 Exercise Sheet 6

Costas Meghir

1. Explain why omitting a variable from a regression that is correlated with the included regressor is equivalent to a violation of the assumption that the error term and the included regressor are correlated.

2. Show that omitting a variable that belongs in the regression will lead to a bias in the coefficient of the included regressor *if* the included regressor is correlated with the excluded one

3. [Difficult] Using the logic of stepwise regression can you suggest a way of obtaining an expression for the OLS estimator of b_1 in

$$Y_i = a + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + u_i$$

Interpret the result you get with reference to the efficiency loss you get by including unnecessary regressors.

4. You are provided with a data set (see web site) that includes wages for women in their early thirties (all the same age) and a number of other variables. The data set is called **edqb203f.dta**.
    a. Run a regression of the female log wage on the student teacher ratio. Explain what the coefficient means.
    b. Repeat this by successively adding an ability indicator and then in addition the type of school. Comment on the way the estimated impact of the student teacher ratio changes from one regression to the next.
    c. Compute the correlation coefficients between all the regressors you have included. What do you conclude from this?

5. Use the same data set to carry out the following experiment:
    a. Regress the log wage on the student teacher ratio and on the ability indicator.
    b. Regress the student teacher ratio on the ability indicator. Compute the residuals from this regression.
    c. Regress the log wage on the residual you created in step 2. Check that the resulting coefficient on the residual is identical to the coefficient on the student teacher ratio you obtained from the first regression.