

## Problem set 11: week 15

This problem set is to be completed prior to your Stata class. The data set used, wage5.dta, is available on the EC203 website, please copy it on to a memory stick or your H: drive. I will highlight the main commands, use either: **help**, **findit** or the **search** function for guidance on each. **Note: There is no do file with this problem set as that would defeat the whole purpose.**

1. Open a do file in Stata. All the commands we use in this problem set will be copied into here. This is so you can recall what we have done, and the analysis can be repeated. It will also be useful for you to annotate the do file as you go along.
2. Load the wage5.dta dataset into Stata.
3. Open a log file to record the output from Stata. This should be written into either your memory stick or H: drive.
4. In this question your primary objective is to maximise the  $R^2$  in the regression of log wage on any of the independent variables in the data set. To do this you can also take polynomial transformations, logs, combine variables together (add them, divide them, ...), interact any independent variables together, ... etc. You can include as many variables as you like; the greater the number of variables the greater the likely value of the  $R^2$ .

Once you think you have achieved as high an  $R^2$  as possible, explain:

- How OLS calculates the estimated parameters in your model. You only need to outline the procedure.
  - Pick 2 or 3 of the estimates in your model and interpret them. Do you think your model is useful?
  - How do the CLRM assumptions relate to the model above?
5. Our primary goal in this question is to estimate whether education has an affect on wage. Run the following regressions and explain why you have entered the control variables you have and provide interpretations:

- $\ln(wage_i) = \alpha + \beta_1 educ_i + \epsilon_i$ .
- $\ln(wage_i) = \alpha + \beta_1 educ_i + \beta_2 exp_i + \beta_3 exp_i^2 + \epsilon_i$ .
- $\ln(wage_i) = \alpha + \beta_1 educ_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 black_i + \epsilon_i$ .

- $\ln(wage_i) = \alpha + \beta_1 educ_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 black_i + \beta_5 IQ + \epsilon_i$ .
  - Out of the remaining variables in the data set, pick two to include in the model, providing reasons for why you do. Interpret the results.
  - You hypothesize that the returns to education ( $\beta_1$ ) are likely to differ across different sections of the population. Pick a characteristic, and provide an argument as to why you think it will cause the return to education to change. Test your hypothesis.
  - Do you think you have estimated the causal effect of education on wage? Discuss, relating your discussion to the CIA.
6. Annotate and close your log file.