PO 7005: Assignment 4

Thomas Chadefaux

NOTE: Always justify your answer. Show R code when relevant. Late submissions will be penalized (5 points per day). Each of the subquestions is worth 10 points.

Question A

The dataset 'ps.csv', included in the same folder as this homework, provides data on per capita expenditure on public schools and per capita income by US state in 1979. Import that dataset into R and estimate the following model by OLS:

$$Expenditure = \beta_0 + \beta_1 Income + \beta_2 Income^2 + \varepsilon \quad (1)$$

1. At first sight, do you think $Income^2$ belongs in the model? In particular:
   (a) Comment on β_2's p-value.
   (b) Compare the BIC and R^2 of model (1) with a model that would exclude $Income^2$. What do you conclude?
2. Should you be concerned about heteroskedasticity? Explain what this would mean, how you would test for it (run and report the test(s)), and what you conclude.
3. Should you be concerned about serial correlation?
   (a) Explain what this would mean, how you would test for it (run and report the test(s)), and what you conclude.
   (b) Assume that you *are* in fact concerned about serial correlation. What would you do about it? Report the results of your approach.

4. Should you be concerned about outliers? Explain what this would mean, how you would test for it (run and report the test(s)), and what you conclude.

5. Based on what you have now learned, do you think $Income^2$ belongs in the model?

Question B

File 'mroz.dta', included in the same folder as this homework, provides data on the (logged) wages ('lwage') of 428 working, married women. Of interest is the effect of their education ('educ') and experience ('exper') on their (logged) wages ('lwage'). We assume that experience is exogenous, but we suspect that education might be correlated with the error term.

1. Briefly explain why we might think $cov(education, \varepsilon) \neq 0$.

2. Suppose you want to use motheduc, fatheduc, and huseduc as instruments for education. Are these good instruments? How would you know?

3. What is the OLS estimate for the effect of education? Do you think it is accurate? If not, report how you would estimate the true effect of education? Discuss your results.

Question C

Data gsswide.dta (in the Dropbox folder) has information about educational level as the DV and family background as the IV (father's prestige, parental education)

hsgrad is a binary variable referencing whether a particular individual graduated from high school. Estimate the effect of papres16 (father's prestige), paeduc (father's education), maeduc (mother's education), black (being black) and male (gender).

1. What is an individual's expected probability of graduating high school if her mother's education (maeduc) is 12 years? What if it is 16 years instead? Show how you calculated it (8 points).

2. Plot the expected probability of graduation (hsgrad) as a function of the father's prestige, for values of papres16 between 12 and 82, holding the other variables at their mean, and male and black at 0. Add another line displaying the same effect, but this time with black = 1 (8 points)