Part IIA, Paper 3, Supervision 6

A. Panel data

1. Recall the first problem of the previous supervision sheet on determinants of crime. Download the dataset CRIME4.dta. Remind yourselves what the variables mean by typing "des"

2. Observe now that we actually have a panel dataset. So we can use a first-difference or fixed effects method to eliminate omitted variable bias.

3. Why might there be an omitted variable bias (as opposed to simulteneity bias) in a regression of crime on police?

4. Regress lcrime on lpolice ldensity urban west central–what is the interpretation of the coefficient on lpolice?

5. Now run a first differenced regression by regressing clcrime on clpolice cldensity

6. How does your fd estimate of the effect of police compare with the OLS estimate above?

7. Is lpolpc strictly exogenous? If not, how do you think this will affect the estimates?

B. **Maximum Likelihood Estimation**

Suppose we want to test if a coin is fair. We toss the coin 10 times and get the following outcomes:

HHTTHTHHTH

a. Write down the likelihood function as a function of the probability of H in a single toss.

b. What is the probability of obtaining the above sample if the probability of H in a single draw was 0.5?

c What is the log likelihood? What is the value of the maximum likelihood estimator?

d. Can you estimate the variance of this estimator?

e. What is the probability of obtaining the above sample if the true probability of H in a single draw was equal to the maximum likelihood estimate? Compare with your result in part b.

f. What is your conclusion regarding fairness of the coin?

C. **Limited dependent variables**

We are interested in the determinants of childhood obesity.

1. Download the dataset fatkids.dta

2. See what the variables mean by typing "des"

3. Summarize the data using the command "summ"

4. The variable "obesec" is a dummy for whether the child is obese

5. Run a probit regression with the command "probit obesec ageyrs female white black hisp tvyest povrat"

6. What do you infer from the row corresponding to "hisp"?

7. Calculate the predicted probability of being obese for a 16 year old female white child who watched 3 hours of TV yesterday and whose family's income-to-needs ratio is 4.56.

8. How much would this probability fall if we took an identical person as above but who watched no TV yesterday? Is the fall statistically significant? How can you tell?

9. How would you test that race has no effect on the probability of being obese? Now test it using the corresponding stata command. What do you conclude?

10. What does your overall analysis tell us about whether we should discourage young children from watching TV in order to prevent them from being obese?