Problem set 4: Empirical Methods for Applied Microeconomics

General instructions. Please work in a group no larger than 3. When you write up your results, please let me know who is in your group. (Only turn in 1 completed homework.). Present your answers in a concise way (typed is highly preferred). Please include relevant Stata output and well-commented do files and ado files for all the exercises (or equivalent in the package of your choice.) Please do NOT include lots of undigested log files.
Put the do files in an appendix and make clear reference to the regression output and/or figures.

Problem 1
Multiple Testing We will start with data from the National Health Interview Survey that I have compiled for another purpose. In this data, sample adults are asked about a host of health outcomes. The NHIS also asks if anyone in the household got SNAP (food stamp) benefits. Individuals getting SNAP are less health along a host of dimensions than are not recipients. (This is not surprising as SNAP is a means-tested program, which additional has automatic eligibilty for some disability (SSI) recipients.)

Use the data and look at the relationship between use of SNAP and the following health and health care use outcomes for sample adults (samadult==1). (Each line has the variable name of a health outcome and its definition.)

- heightr: Recode of height (Height in inches without shoes)
- nobedays: No bed days in the last year
- addevr: Recode of addev (Ever told had ADHD/ADD)
- asthmaevr: Recode of asthmaev (Ever told had asthma)
- blindr: Recode of blind (Blind or unable to see at all)

- diabeticevr: Recode of diabeticev (Ever told had diabetes)
- heartattevr: Recode of heartattev (Ever told had heart attack)
- strokevr: Recode of strokev (Ever told had a stroke)
- ulcerevr: Recode of ulcerev (Ever told had ulcer)
- drinknow: Drink alcohol
- drankev:Ever drank alcohol
- smokev: Ever smoked 100 cigarettes in life
- smokenow: Smoke now

Control for gender, family size, the number of persons and kids in the HH, race/ethnicity, education, mothers' education, citizenship and years in the US, month dummies, region dummies, and the unemployment rate. You will have to create month dummies and region dummies, and you should weight by the variable sampweight.

(i) Are persons getting SNAP healthier or less healthy than the rest of the population? Do you think this is causal? (One-fourth page at most.)

(ii) Now we will adjust for multiple testing. Does the conclusion from (i) hold if you adjust for the FWER using Bonferroni? What if you use the free stepdown approach used by Anderson (you probably only want L to be 999 at most as you will be rerunning a bunch of regressions for each resample)? (Note Anderson has some example code for controlling the FWER rate at his website.)

Problem 2

SES and the wild bootstrap

Now we will use the same data to look at the effect of the unemployment rate on participation in SNAP. Control for the same Xs as in problem 1 (except of course SNAP participation). Here the measure of the unemployment rate is the regional average.

(i) First, assume the data are iid. Do more people participate when the unemployment rate is higher? What else drives participation?

(ii) Our SEs in (i) are likely wrong as the unemployment rate varies only at the region level. Cluster the SEs at the region level instead (there are 4 regions). Do you think these SEs are more appropriate? Why or why not?

(iii) Now we will try the wild bootstrap of Cameron, Gelbach, and Miller. Use Mammen weights. What do you now conclude about the effect of the unemployment rate on participation? (Note that Doug Miller has some example code at his website.)