ECON 626: Empirical Microeconomics

Problem Set 5

Department of Economics
University of Maryland
Fall 2019

Problem Set 5 is due at 5pm on Thursday, November 14.

1. **Bruhn and McKenzie.** One of the Bruhn and McKenzie (2009) recommendations is to balance on three things: baseline values of the outcome of interest (for power); geography (to ensure that any post-baseline geographic shocks are orthogonal to treatment); and any variable along which heterogeneity analysis is intended (to ensure power in that heterogeneity analysis). To answer this question, use Stata to demonstrate that last reason by using the `leaps_test_300.dta` dataset, which is student test data from Pakistan. Roughly half the students in that dataset are female. For some reasonable number of iterations, randomize a treatment assignment so that half the sample is considered "treated," and half "comparison." Do this two ways. First, without regard to anything; second, stratifying on gender. Keep track, in each iteration, of the standard error one would get if one regressed the endline math test (`math_fu`) on randomized treatment only in the subsample of girls. (Since no treatment effect is being added, this is estimation under the null. But we are only interested in standard errors.) Plot the distribution of standard errors you get from stratifying and from not stratifying. Report the means, and some percentiles, perhaps even the minimum and maximum values. Do your results validate the Bruhn and McKenzie reasoning?

2. **Regression Discontinuity Assumptions and Estimand** The file `ps5-rd-question.do` creates a dataset including potential outcomes (which you ordinarily wouldn't see all of) yielding treatment effects of several different treatments. Extend the file by doing the following steps, relating to each of the four treatments (A,B,C,D) that each are switched on when the running variable (x) crosses a threshold (zero).

   (a) Generate variables `betaA`, `betaB`, `betaC`, and `betaD` that are the differences `y1A-y0`, `y1B-y0`, `y1C-y0`, and `y1D-y0`, respectively. These are the treatment effects for each individual, since we are creating them using both of the potential outcomes relevant for each individual and each treatment. Summarize these variables.

   (b) Graph these variables as a function of `x`.

   (c) Run regression discontinuity estimation (by hand, using `rd`, or using `rdrobust`) for each of `yA`, `yB`, `yC`, and `yD` with `x` as the running variable. What is the treatment effect estimate for each?

   (d) How does the answer to part 2c relate to the answer to part 2a, especially in light of section 2.2 of Imbens and Lemieux 2008, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics* 142: 615-635 (especially their assumptions 2.1 and 2.2)?

3. **Randomization Inference (Permutation Test).** Using Stata, open the data provided in `CohenDupasSchanerT3-for626.dta`. This is a condensed version of the data used in the Cohen, Dupas, and Schaner (2015) AER paper on malaria in their Table 3. The data can be used to test whether a particular price of antimalarial (`act60`) increases the likelihood that those who get tested for malaria are actually positive (`rdt_posA`). Regressing `rdt_posA` on `act60` should yield a coefficient of about 0.187, the coefficient in the first row and column of Cohen, Dupas, and Schaner's Table 3. That is, patients facing a higher price of medication are more likely to present themselves only when they actually have malaria. The outcome and the treatment assignment are both binary, and the sample size (135) is not that large. Rather than use the p-value from the usual asymptotics (.022 or .023 depending on whether you use robust standard errors), we could instead do a permutation test, along the lines of the tea-tasting exercise in class. For some large number of iterations, randomly generate alternative treatment assignments that, like `act60`, are 1 for 64 observations and 0 for the other 71. Tallying the absolute values of coefficients on these alternative treatments, and counting those that (approximately) equal or exceed the actual coefficient, you should be able to arrive at a randomization-inference-based p-value. What value do you get? Is their coefficient still statistically significant at traditional levels ($p < 0.05$)?

4. **Wild Cluster Bootstrap p-value discreteness.** Consider the unfortunate design in which there are four clusters, two of which are assigned to treatment and two of which are assigned to comparison.

   (a) If the "Rademacher weights" are used in the Wild Cluster Bootstrap, how many possible arrangements of residuals are there?

   (b) Since every residual arrangement has a symmetric counterpart that produces exactly the same magnitude t-statistic, so the effective number of arrangements is half the answer above, what is the smallest conceivable Wild Cluster Bootstrap p-value in this situation?

   (c) If the "Webb weights" are used in the Wild Cluster Bootstrap, how many possible arrangements of residuals are there?

   (d) Thus, following the same reasoning as above, what is the smallest conceivable Wild Cluster Bootstrap p-value in this situation?

   (e) What if we used "Rademacher weights" but had ten clusters instead of four - how many possible arrangements of residuals are there, and what is the smallest conceivable Wild Cluster Bootstrap p-value in this situation?

5. **Few Clusters.** The file `ps5-wcb-question.do` simulates an environment in which four clusters out of twenty are assigned to treatment. If you run the code as-is, it demonstrates how un-clustered asymptotic standard errors are very wrongly sized, and how even clustered asymptotic standard errors are (more modestly) incorrectly sized.

   (a) Approximately what does the actual size of the two incorrectly-sized tests appear to be, in this setup? Make histograms of these two types of p-values.

   (b) Fill in the wild cluster bootstrap code everywhere it indicates YOUR CODE HERE. (Feel free to adjust the number of overall iterations or bootstrap iterations for practicality.)

(c) Does the wild cluster bootstrap p-value appear correctly sized? Make a histogram of its p-values.

(d) Install and use the user-written program `boottest` to confirm that your wild cluster bootstrap p-values are similar to the `boottest` ones; make a scatter plot of your wild cluster bootstrap p-values against those that `boottest` generates.

(e) **Optional:** change the treatment effect (line 18 of the file) from zero to something larger. See how the power of the test varies with treatment effect size.