ECON 626: Empirical Microeconomics

Problem Set 5

Department of Economics
University of Maryland
Fall 2016

Problem Set 5 is due at 5pm on Wednesday, November 16.

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), as we did in class; 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 from the Lecture 8 activity. 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. 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. **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. 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 ($p < 0.05$)?

3. **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) Thus, 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, 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?

4. **Few Clusters.** Extend the in-class Stata activity involving the Wild Cluster Bootstrap code.

(a) For even modest numbers of clusters (10, 12), the "Rademacher weights" generate reasonably fine-grained p-values, but very quickly, the exhaustive listing of possible residual arrangements can be potentially cumbersome. Make a third version of the supporting data grid in section 3 of the code that randomly draws some fixed number (the value of the local `'permutations'` for example) of Rademacher weights, analogously to how Webb weights are now generated. This should be a grid of 1's and 0's, but not a necessarily exhaustive one.

(b) Use this new grid, in combination with the `wildbsOrig` program, to plot the CDF of p-values under the null, comparing simple clustered standard errors to the Wild Cluster Bootstrap standard errors for a ten-cluster scenario (but otherwise similar to the scenario in class).

(c) Show how the clustered-versus-bootstrap-CDF comparison looks for at least two data structures: the one above (everything the same as in class but with ten clusters) and at least one other, where you change the intra-cluster correlation, the unevenness of the group sizes, the number of clusters, or any other aspect of the DGP that is of interest to you. Discuss the pattern you find.