Week 3, HW 5: Regression Discontinuity

Matt Goldman
mattgold@microsoft.com

June 30, 2017

1. **Picking Control Variables:** Imagine you have one data point from every county in the US that gives the unemployment rate in that county in 2010 and the amount of money spent by that country on job training programs in 2008. You want to know the treatment effect of money spent on job training programs onto the outcome of unemployment (as measured two years later). List as many examples as you can think of for each category of control variable:

   (a) Confounding variables:

   (b) Variance-reducing variables:

   (c) Unrelated variables:

   (d) Bad Controls:

2. **Regression Discontinuity on Incumbency Effect (Lee, 2007)**

   (a) Download the dataset (Lee2007_RDReplication.csv) from the coursework repo. There are many more columns here than you need. DWinNxt (whether or not a dem wins the next election) is the outcome variable we want to model. DemWin (whether or not dems win the current election) lets us know if democrats have the incumbency and it is our treatment of interest. DiffDPct tells us how much Dems won/lost by in the previous election, this is our running variable that gives us a discontinuity. ForgnPct, GovWkPct, BlackPct, UrbanPct, DWinPrv, YearElec, and CDNum are potential control variables about each congressional district that can help us predict winners.

   (b) Run a simple regression to try and predict DWinNxt with DemWin. Observe the coefficient on DemWin. Do you think it gives an unbiased estiamte of the incumbency effect. Why or why not?

(c) Now consider the various potential controls I mentioned above. Classify each one as a confounding, variance reducing, unrelated, or bad control. Introduce some/all of these variables to your regression and see what happens to the coefficient on DemWin.

(d) Give an additional example of a bad control variable in this setting. It does not have to be one that it is present in the dataset.

(e) Now use DifDPct as a forcing variable to run a regression discontinuity to estimate the incumbency advantage. Generate a cubic polynomial and pick a window of 10% (this just means to drop data that is more than 15% from the discontinuity) on either side of the threshold (50%). Run a regression discontinuity. How does the coefficient you got compare to in (b) and (c).

(f) Finally, install and use the package *rdd* to implement this same regression discontinuity design (this saves you from having to decide on bandwidths and code polynomials). Use the functions RDestimate and plot to do a faster regression discontinuity anlaysis. Your answer in (e) should look somewhat similar to your answer here.