Week 3, HW 4: Causal Inference with Difference in Differences (DiD) & Regression Discontinuity

Matt Goldman
mattgold@microsoft.com

June 29, 2017

Note: I've included two problems on here, but I'm not sure how long they will take. Please take your time with the first problem and if you feel you want to spend the whole day on it, that is fine and we may save the second problem for tomorrow.

1. **DiD with Card and Krueger (1993)**. Download the dataset and definitions file from the coursework repo. Note: (1) that the first interview occurred before the onset of the minimum wage and that the second occurred after and (2) this dataset is paired (each restaurant has a before and after observation) so the steps you take will be a little bit different than those discussed in class.

   (a) Before looking at the data, look just at the list outcome metrics that were tracked in each interview (EMPFT EMPPT NMGRS WAGEST INCTIME FIRSTINC BONUS PCTAFF MEALS OPEN HRSOPEN PSODA PFRY PENTREE NREGS). List the ones that you think could be impacted by the minimum wage change and give a brief explanation of why.

   (b) Compute the 'diffs'. Check to see if each outcome metric changed in a statistically significant way between the two interviews. Do this analysis separately for PA and NJ.

   (c) Now compute the "diff-in-diff" via the regression method. As I mentioned previously, this step will be a little different because the data is paired. Think carefully about the best way to do this and I'll come and chat with you guys about it after about an hour.

(d) Do you believe this analysis provides definitive evidence on the impact of the minimum wage on employment? Can you imagine a scenario that would lead to this type of estimation giving biased results?

(e) Card and Krueger create an Independent Variable called $GAP = (5.05 - W_{pre})/W_{pre}$, this is the percentage wage increase that New Jersey restaurants needed in order to meet the minimum wage. Use the variable $WAGEST$ (from before the interview) to create this variable. How might restaurants with very large or small values of this variable differ in their response to the minimum wage. Why do you think this variable is interesting? Run any other interesting analysis you can think of using this variable.

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 forcing variable that gives us a discontinuity. ForgnPct, GovWkPct, BlackPct, UrbanPct also give some basic demographics about each congressional district that can help us predict winners.

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

   (c) Now use ForgnPct, GovWkPct, BlackPct, UrbanPct as control variables? Were these significant predictors of DWinNxt. What happened to the coefficient on DemWin when you introduced these variables. Why did you think this happened?

   (d) Now use DifDPct as a forcing variable to run a regression discontinuity to estimate the incumbency advantage. Generate a cubic polynomial and pick a value of the bandwidth=15% on either side of the threshold (50%).

   (e) 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.