Regression Discontinuity Design (RDD)
Econ 672

1) The Takeaway
    a. The basis of RDD is that units around an arbitrary cutoff rule are similar to one another along a continuous running variable and we can estimate a local average treatment effect (LATE) at the cutoff
2) Pros
    a. Most credible identification strategies for observational data to identify causal effects, since the most important assumptions are director or indirectly testable
    b. It is a very visual approach to identifying the causal effect
3) Cons
    a. You need deep institutional knowledge to know about the treatment, threshold, running variables, and outcomes of interest
    b. The LATE estimate cannot be extrapolated beyond the cutoff and we are calculating the LATE for compilers around the cutoff
    c. There is a large data/observation requirement around the cutoff
4) Assumptions
    a. Continuity Assumption (Sharp and Fuzzy)
        i. This assumption says that expected potential outcomes need to be smooth continuous around the cutoff , such that $E[Y_i^0|X = c_0)$ and $E[Y_i^1|X = c_0]$ are smooth and continuous around the cutoff
    b. Exclusion Restriction Assumption (Fuzzy)
    c. Monotonicity Assumption (Fuzzy)
    d. Stable Unit Value Treatment Assumption (SUTVA) (Fuzzy)
    e. Non-zero 1st stage Assumption (Fuzzy)
        i. The running variable X must be associated with probability of assignment to treatment
5) Testable Assumptions
    a. McCrary Test:
        i. This test check to see if there is bunching in density around the cutoff. The null hypothesis is that there is no bunching around the cutoff or observations sorting around the cutoff
    b. Covariate Balance
        i. This assumption test checks to see if covariates are similar around the cutoff, such that there should be no jump in covariates around the cutoff.
    c. Placebo Tests:
        i. This tests to see if there are any discontinuous jumps at a "fake" cutoff. We can use a placebo cutoff along the running variable to test for any discontinuous jumps where there shouldn't be.
6) Sharp RDD

a. Sharp RDD means that assignment to treatment is deterministic and a discontinuous function of the running variable X
b. $\delta = \lim_{x_i \to x_c} E[Y_i|X_i = X_c] - \lim_{x_c < -x_i} E[Y_i|X_i = X_c]$

7) Fuzzy RDD
   a. Fuzzy RDD means that assignment to treatment is probabilistic upon the running variable X and there is a discontinuous jump in probability to assignment to treatment at the cutoff
   b. $\delta_{FuzzyRDD} = \frac{\lim_{x \to x_c} E[Y|X=X_c] - \lim_{x_c < -x} E[Y|X=X_c]}{\lim_{x \to x_c} E[D|X=X_c] - \lim_{x_c < -x} E[D|X=X_c]}$
8) Challenges to Identification: Violations of the continuity assumption
   a. Assignment rules are known in advance
   b. Agents are interested in adjusting
   c. Agents have time to adjust
   d. The cutoff is associated/endogenous with factors that independently cause potential shifts in outcomes
   e. There is nonrandom heaping along the running variable