6 Analysis of a Regression Discontinuity Design

1. Identification in the RDD with constant treatment effect.

   Consider the model, as in e.g. Ludwig and Miller (2007),

$$Y_i = \beta_0 + \beta_{1i} D_i + \delta_1 X + U_i \tag{1}$$
$$D_i = \mathbb{1}[X \geq c] \tag{2}$$

   where

   - Y is the outcome, e.g. child mortality
   - D is the binary treatment status, e.g. grant-writing assistance
   - β is the treatment effect
   - X is the assignment variable, e.g. poverty rate in 1960
   - $D = \mathbb{1}[X \geq c]$ is the treatment assignment, e.g. grant-writing assistance if high poverty rate
   - c: is the cutoff, i.e. the OECD cutoff
   - U are unobserable variables, other factors that affect Y

   Assume that we have constant treatment effects, i.e.:

$$\beta_i = \beta \ \forall \ i \tag{3}$$

   And assume that

$$\mathbb{E}(U|X = x) \text{ is continuous in } x \tag{4}$$

   (a) Show that these two assumptions are sufficient to show that the RD estimator

$$\frac{E(Y|X = c) - lim_{x \uparrow c} E(Y|X = x)}{E(D|X = c) - lim_{x \uparrow c} E(D|X = x)} \tag{5}$$

   identifies the common treatment effect β.

   (b) What changes if we have a sharp design?

2. Fuzzy RDD is IV.

   Start with the following simple RD setup: we have a running variable X_i and an outcome Y_i. Assume that the probability of treatment has a discontinuity at $X_i = c$. Thus we have:

$$Pr(D_i = 1|X_i) \quad = \quad \begin{cases} g_1(X_i) & \text{if} \quad X_i \geq c \\ g_0(X_i) & \text{if} \quad X_i < c \end{cases} \tag{6}$$

   and assume that $g_1(c) > g_0(c)$ (such that $X_i > c$ makes treatment more likely). For simplicity, assume that $g_0(X_i)$ and $g_1(X_i)$ can be well approximated by a linear function each. Explain how you could estimate the RD using 2SLS.

   (a) Write down the structural model.

   (b) Write down the first-stage regression and plug this into the structural model.

3. Replicate Ludwig and Miller (2007): Does Head Improve Children's Life Chances?

   The U.S. "Head Start" program has been providing pre-school, health and other social services to poor children and their families since 1965. Evaluating whether the program improves children's life chances is challenging because participation is likely correlated with outcomes. Ludwig and Miller (2007) exploit the fact that in its initial stage, only the poorest 300 counties in the U.S. received federal support in applying for the program. In what follows, I invite you to replicate the parametric (OLS) estimates they get on mortality reduction, one of the main results of their study. What follows is a suggested roadmap:

   (a) (*Not Mandatory!*) To familiarize yourself with the setting, I strongly suggest you read the relevant parts of the paper; you can find it in the PS6 OLAT folder. You can safely ignore sections VIII and IX for our purposes.

   (b) Replicate their Figures II and III on the relationship between poverty rate, Head Start funding, and other social spending. In particular, you should show the raw cell means, confidence intervals, and flexible quadratic fits (instead of their nonparametric estimates of the regression function you can show flexible cubic fits). (Hint: use the floor() and egen commands in Stata to construct your figure).

   (c) Replicate their parametric results on Head Start spending in 1968 and 1972, and on other social spending in 1972 in Table II (last two columns for bandwidths 8 and 16).

   (d) Replicate their Figure IV. Again, you should show the raw cell means, confidence intervals, and flexible quadratic fits (instead of their nonparametric estimates of the regression function you can show flexible cubic fits).

   (e) Replicate the parametric results from Table III that correspond to the mortality outcomes in Figure IV (last two columns for bandwidths 8 and 16).

   (f) Estimate impacts on the four mortality outcomes for different bandwidths and polynomials in the 1960 poverty rate. The choice of the specifications is yours; choose 4 of them. Report your replication and new estimates in a table of the same format as Table III.