Regression Discontinuity Methods
Class Notes

Manuel Arellano

January 2008

1. Introduction and examples

- In the matching context we make the conditional exogeneity assumption
$$(Y_1, Y_0) \perp D \mid X$$
whereas in the IV context we assume
$$(Y_1, Y_0) \perp Z \mid X \quad \text{(independence)}$$
$$D \not\perp Z \mid X \quad \text{(relevance).}$$
The relevance condition can also be expressed as saying that for some $z \neq z'$
$$\Pr(D = 1 \mid Z = z) \neq \Pr(D = 1 \mid Z = z').$$

- In regression discontinuity we consider a situation where there is a continuous variable Z that is not necessarily a valid instrument (it does not satisfy the exogeneity assumption), but such that treatment assignment is a discontinuous function of Z.

- The basic asymmetry on which identification rests is discontinuity in the dependence of D on Z but continuity in the dependence of (Y_1, Y_0) on Z.

- RD methods have much potential in economic applications because geographic boundaries or program rules often create usable discontinuities.

Examples

- Effect of class size on test scores ("Maimonides' rule" in Israel, Angrist & Lavy, 1999):

$$
\begin{aligned}
Y_{is} &: \quad \text{average score at class } i \text{ in school } s \\
D_{is} &: \quad \text{size of class } i \text{ (not binary)} \\
Z_s &: \quad \text{beginning of year enrollment in school } s
\end{aligned}
$$

  Maimonides' rule allows enrollment cohorts of 1–40 to be grouped in a single class, but enrollment groups of 41–80 are split into two classes of average size 20.5–40, enrollment groups of 81–120 are split into three classes of average size 27–40, etc. In practice, the rule was not exact: class size predicted by the rule differed from actual size.

Examples (continued)

- Effect of financial aid offers on students' enrollment decisions (van der Klaauw, 2002)

  Y_i : decision of student i to enroll in college "X" (binary)

  D_i : amount of financial aid offer to student i

  Z_i : index that aggregates SAT score and high school GPA

  Applicants for aid were divided into four groups on the basis of the interval the index Z fell into. Average aid offers as a function of Z contained jumps at the cutoff points for the different ranks, with those scoring just below a cutoff point receiving much less on average than those who scored just above the cutoff.

- Do parties matter for economic outcomes? (Petterson-Lidbom, 2006; Arellano & Bentolila, in progress):

  Y_i : economic outcome in area i

  D_i : party control indicator in local government i

  Z_i : vote share

2. The fundamental RD assumption

- We can now state the basic RD assumption more formally. Namely, discontinuity in treatment assignment but continuity in potential outcomes: There is at least a known value $z = z_0$ such that

$$\lim_{z \to z_0^+} \Pr(D = 1 \mid Z = z) \neq \lim_{z \to z_0^-} \Pr(D = 1 \mid Z = z) \tag{1}$$

$$\lim_{z \to z_0^+} \Pr(Y_j \leq r \mid Z = z) = \lim_{z \to z_0^-} \Pr(Y_j \leq r \mid Z = z) \quad (j = 0, 1) \tag{2}$$

  Implicit regularity conditions are: (i) the existence of the limits, and (ii) that Z has positive density in a neighborhood of z_0.

- We abstract from conditioning covariates for the time being for simplicity.

Sharp and fuzzy designs

- The early RD literature in psychology (Cook & Campbell 1979) distinguished between "sharp" and "fuzzy" designs. In the former, D is a deterministic function of Z:

$$D = 1(Z \geq z_0)$$

  whereas in the latter is not.

- The sharp design can be regarded as a special case of the fuzzy design, but one that has different implications for identification of treatment effects. In the sharp design

$$\lim_{z \to z_0^+} E(D \mid Z = z) = 1, \quad \lim_{z \to z_0^-} E(D \mid Z = z) = 0.$$

3. Homogeneous treatment effects

- Like in the IV setting, the case of homogeneous treatment effects is useful to present the basic RD estimand. Suppose that $\alpha = Y_1 - Y_0$ is constant, so that
$$Y_i = \alpha D_i + Y_{0i}$$
- Taking conditional expectations given $Z = z$ and left- and right-side limits:
$$\begin{aligned}
\lim_{z \to z_0^+} E\left(Y \mid Z = z\right) &= \alpha \lim_{z \to z_0^+} E\left(D \mid Z = z\right) + \lim_{z \to z_0^+} E\left(Y_0 \mid Z = z\right) \\
\lim_{z \to z_0^-} E\left(Y \mid Z = z\right) &= \alpha \lim_{z \to z_0^-} E\left(D \mid Z = z\right) + \lim_{z \to z_0^-} E\left(Y_0 \mid Z = z\right).
\end{aligned}$$
- The RD assumption then leads to consideration of the following RD parameter
$$\gamma = \frac{\lim_{z \to z_0^+} E\left(Y \mid Z = z\right) - \lim_{z \to z_0^-} E\left(Y \mid Z = z\right)}{\lim_{z \to z_0^+} E\left(D \mid Z = z\right) - \lim_{z \to z_0^-} E\left(D \mid Z = z\right)}$$
which is determined provided the "relevance part" (1) of the RD assumption is satisfied, and equals α provided the "independence part" (2) of the RD assumption holds.

- In the case of a sharp design, the denominator is unity so that
$$\gamma = \lim_{z \to z_0^+} E\left(Y \mid Z = z\right) - \lim_{z \to z_0^-} E\left(Y \mid Z = z\right), \tag{3}$$
which can be regarded as a matching-type situation, in the same way that the general case can be regarded as an IV-type situation.

- So the basic idea is to obtain a treatment effect by comparing the average outcome left of the discontinuity with the average outcome to the right of discontinuity, relative to the difference between the left and right propensity scores.

- Intuitively, considering units within a small interval around the cutoff point is similar to a randomized experiment at the cutoff point.

4. Heterogeneous treatment effects

- Now suppose that

$$Y_i = \alpha_i D_i + Y_{0i}$$

- In the *sharp design* since $D_i = 1(Z \geq z_0)$ we have

$$E(Y \mid Z = z) = E(\alpha \mid Z = z) 1(z \geq z_0) + E(Y_0 \mid Z = z).$$

- Therefore, the situation is one of selection on observables. That is, letting

$$k(z) = E(Y_0 \mid Z = z) + [E(\alpha \mid Z = z) - E(\alpha \mid Z = z_0)] 1(z \geq z_0)$$

we have

$$E(Y \mid Z = z) = E(\alpha \mid Z = z_0) 1(z \geq z_0) + k(z)$$

where $k(z)$ is continuous at $z = z_0$.

- Therefore, the OLS population coefficient on D in the equation

$$Y = \gamma D + k(z) + w \tag{4}$$

coincides with γ, which in turn equals $E(\alpha \mid Z = z_0)$.

- The control function $k(z)$ is nonparametrically identified. To see this, first note that γ is identified from (3). Then $k(z)$ is identifiable as the nonparametric regression $E(Y - \gamma D \mid Z = z)$. Note that if the treatment effect is homogeneous $k(z)$ coincides with $E(Y_0 \mid Z = z)$, but not in general.

- If $\mu(z) \equiv E(Y_0 \mid Z = z)$ was known (e.g. using data from a setting in which no program was present) then we could consider a regression of Y on D and $\mu(z)$. It turns out that the coefficient on D in such a regression is $E(\alpha \mid z \geq z_0)$.
- In the *fuzzy design*, D not only depends on $1(Z \geq z_0)$ but also on other unobserved variables. Thus, D is an endogenous variable in equation (4). However, we can still use $1(Z \geq z_0)$ as an instrument for D in such equation to identify γ, at least in the homogeneous case.
- The connection between the fuzzy design and the instrumental variables perspective was first made explicit in van der Klaauw (2002).
- Next, we discuss the interpretation of γ in the fuzzy design with heterogeneous treatment effects, under two different assumptions.

Conditional independence near z_0

- Let us first consider the weak conditional independence assumption
$$D \perp (Y_0, Y_1) \mid Z = z$$
for z near z_0, i.e. for $z = z_0 \pm e$ where $e > 0$ denotes an arbitrarily small number, or
$$\Pr(Y_j \leq r \mid D = 1, Z = z_0 \pm e) = \Pr(Y_j \leq r \mid Z = z_0 \pm e) \quad (j = 0, 1).$$

- That is, we are assuming that treatment assignment is exogenous in a neighborhood of z_0. An implication is
$$E(\alpha D \mid Z = z_0 \pm e) = E(\alpha \mid Z = z_0 \pm e) E(D \mid Z = z_0 \pm e).$$

- Proceeding as before, we have
$$\begin{aligned}
\lim_{z \to z_0^+} E(Y \mid Z = z) &= \lim_{z \to z_0^+} E(\alpha \mid D = 1, Z = z) \lim_{z \to z_0^+} \Pr(D = 1 \mid Z = z) \\
&\quad + \lim_{z \to z_0^+} E(Y_0 \mid Z = z) \\
\lim_{z \to z_0^-} E(Y \mid Z = z) &= \lim_{z \to z_0^-} E(\alpha \mid D = 1, Z = z) \lim_{z \to z_0^-} \Pr(D = 1 \mid Z = z) \\
&\quad + \lim_{z \to z_0^-} E(Y_0 \mid Z = z)
\end{aligned}$$
and
$$\begin{aligned}
\lim_{z \to z_0^+} E(Y \mid Z = z) &= E(\alpha \mid Z = z_0) \lim_{z \to z_0^+} \Pr(D = 1 \mid Z = z) + \lim_{z \to z_0^+} E(Y_0 \mid Z = z) \\
\lim_{z \to z_0^-} E(Y \mid Z = z) &= E(\alpha \mid Z = z_0) \lim_{z \to z_0^-} \Pr(D = 1 \mid Z = z) + \lim_{z \to z_0^-} E(Y_0 \mid Z = z).
\end{aligned}$$

- Subtracting

$$\lim_{z \to z_0^+} E(Y \mid Z = z) - \lim_{z \to z_0^-} E(Y \mid Z = z)$$
$$= \left[\lim_{z \to z_0^+} \Pr(D = 1 \mid Z = z) - \lim_{z \to z_0^-} \Pr(D = 1 \mid Z = z)\right] E(\alpha \mid Z = z_0).$$

- Thus, it emerges that

$$\gamma = E(Y_1 - Y_0 \mid Z = z_0).$$

  That is, the RD parameter can be interpreted as the average treatment effect at z_0.

Monotonicity near z_0

- Hahn, Todd, and van der Klaauw (2001) also consider an alternative LATE-type of assumption. Let D_z be the potential assignment indicator associated with $Z = z$, and for some $\overline{\varepsilon} > 0$ and any pair $(z_0 - \varepsilon, z_0 + \varepsilon)$ with $0 < \varepsilon < \overline{\varepsilon}$ suppose the local monotonicity assumption

$$D_{z_0+\varepsilon} \geq D_{z_0-\varepsilon} \text{ for all units in the population.}$$

- An example is a population of cities where Z denotes voting share and D_z is an indicator of party control when $Z = z$. In this case the local conditional independence assumption could be problematic but the monotonicity assumption is not.

- In such case, it can be shown that γ identifies the local average treatment effect at $z = z_0$:

$$\gamma = \lim_{\varepsilon \to 0^+} E\left(Y_1 - Y_0 \mid D_{z_0+\varepsilon} - D_{z_0-\varepsilon} = 1\right)$$

i.e. the ATE for the units for whom treatment changes discontinuously at z_0.

- If the policy is a small change in the threshold for program entry, the LATE parameter delivers the treatment effect for the subpopulation affected by the change, so that in that case it would be the parameter of policy interest.

5. Estimation strategies

- There are parametric and semiparametric strategies.

A nonparametric Wald estimator

- Hahn-Todd-van der Klaauw suggested the following local Wald estimator. Let $S_i \equiv 1(z_0 - h < Z_i < z_0 + h)$ where $h > 0$ denotes the bandwidth, and consider the subsample such that $S_i = 1$.
- The proposed estimator is the IV regression of Y_i on D_i using $W_i \equiv 1(z_0 < Z_i < z_0 + h)$ as an instrument, applied to the subsample with $S_i = 1$:
$$\widehat{\gamma} = \frac{\widehat{E}(Y_i \mid W_i = 1, S_i = 1) - \widehat{E}(Y_i \mid W_i = 0, S_i = 1)}{\widehat{E}(D_i \mid W_i = 1, S_i = 1) - \widehat{E}(D_i \mid W_i = 0, S_i = 1)}.$$
- This estimator has nevertheless a poor boundary performance. An alternative suggested by HTV is a local linear regression method.

Parametric and semiparametric alternatives

- Suppose
$$E(D \mid Z) = g(Z) + \delta 1(Z \geq z_0)$$
and
$$E(Y_0 \mid Z) = k(Z).$$

- A control function regression-based approach is based in the control function augmented equation that replaces D by the propensity score $E(D \mid Z)$:
$$Y = \gamma E(D \mid Z) + k(Z) + w$$

- In a parametric approach, we assume functional forms for $g(Z)$ and $k(Z)$. van der Klaauw (2002) considered a semiparametric approach using a power series approximation for $k(Z)$.

- If $g(Z) = k(Z)$, then we can do 2SLS using as instrumental variables
$$\{1(Z \geq z_0), g(Z)\},$$
where $g(Z)$ is the "included" instrument and $1(Z \geq z_0)$ is the "excluded" instrument.

- These methods of estimation, which are not local to data points near the threshold, are implicitly predicated on the assumption of homogeneous treatment effects.

6. Distributional effects

- For some function $h(.)$, consider the outcome

$$W = h(Y)\,D = \begin{cases} W_1 = h(Y_1) & \text{if } D = 1 \\ W_0 = 0 & \text{if } D = 0 \end{cases}$$

- Using $h(Y) = 1(Y \leq r)$, the RD parameter for the outcome $W(r) = 1(Y \leq r)\,D$ delivers

$$\Pr(Y_1 \leq r \mid Z = z_0) = \frac{\lim_{z \to z_0^+} E(W(r) \mid Z = z) - \lim_{z \to z_0^-} E(W(r) \mid Z = z)}{\lim_{z \to z_0^+} E(D \mid Z = z) - \lim_{z \to z_0^-} E(D \mid Z = z)}$$

under the local conditional independence assumption.

- A similar strategy can be followed to obtain $\Pr(Y_0 \leq r \mid Z = z_0)$. In that case we consider

$$V = h(Y)(1 - D) = \begin{cases} V_1 = h(Y_0) & \text{if } 1 - D = 1 \\ V_0 = 0 & \text{if } 1 - D = 0 \end{cases}.$$

- The RD parameter for the outcome $V(r) = 1(Y \leq r)(1 - D)$ delivers

$$\Pr(Y_0 \leq r \mid Z = z_0) = \frac{\lim_{z \to z_0^+} E(V(r) \mid Z = z) - \lim_{z \to z_0^-} E(V(r) \mid Z = z)}{\lim_{z \to z_0^+} E(D \mid Z = z) - \lim_{z \to z_0^-} E(D \mid Z = z)}.$$

7. Conditioning on covariates

- Even if the RD assumption is satisfied unconditionally, conditioning on covariates may mitigate the heterogeneity in treatment effects, hence contributing to the relevance of RD estimated parameters.
- Covariates may also make the local conditional exogeneity assumption more credible.
- This would also be true of within-group estimation in a panel data context (see Hoxby, QJE, 2000, 1239–1285, for an application).