Chapter 5. Regression Discontinuity

Joan Llull

Quantitative Statistical Methods II
Barcelona GSE

The Fundamental RD Assumption

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:

$$\lim_{z \to z_0^+} P(D_i = 1|Z_i = z) \neq \lim_{z \to z_0^-} P(D_i = 1|Z_i = z)$$

$$\lim_{z \to z_0^+} P(Y_{ji} \leq r|Z_i = z) = \lim_{z \to z_0^-} P(Y_{ji} \leq r|Z_i = z) \quad (j = 0, 1)$$

which are **relevance** and **orthogonality** conditions respectively.

Implicit regularity conditions are:

- existence of the limits,
- Z_i has positive density in a neighborhood of z_0.

For now we abstract from **conditioning covariates** for simplicity.

Sharp and Fuzzy Designs

Early RD literature in Psychology (Cook and Campbell, 1979) distinguishes between:

- **Sharp design**: $D_i = \mathbb{1}\{Z_i \geq z_0\}$, with:

$$\lim_{z \to z_0^+} \mathbb{E}[D_i | Z_i = z] = 1$$
$$\lim_{z \to z_0^-} \mathbb{E}[D_i | Z_i = z] = 0.$$

- **Fuzzy design**: $0 < P(D_i = 1 | Z_i \geq z_0) < 1$, with:

$$P(D_i = 1 | Z_i = z_0 - \varepsilon) \neq P(D_i = 1 | Z_i = z_0 + \varepsilon)$$

Homogeneous Treatment Effects

Suppose that $\alpha_i = Y_{1i} - Y_{0i}$ is **constant**, so that $Y_i = \alpha D_i + Y_{0i}$.

Conditional expectations given $Z_i = z$ and left- and right-side limits:

$$\lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] = \alpha \lim_{z \to z_0^+} \mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^+} \mathbb{E}[Y_{0i}|Z_i = z]$$
$$\lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z] = \alpha \lim_{z \to z_0^-} \mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^-} \mathbb{E}[Y_{0i}|Z_i = z],$$

which leads to the consideration of the following **RD parameter**:

$$\alpha = \frac{\lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z]}{\lim_{z \to z_0^+} \mathbb{E}[D_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[D_i|Z_i = z]}.$$

determined by **relevance** and **orthogonality** conditions above.

In the case of a sharp design, the denominator is unity so that:

$$\alpha = \lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z],$$

Sharp corresponds to **matching** and fuzzy corresponds to **IV**.

Intuitively, **randomized experiment** at the cut-off point (with or without perfect compliance).

Heterogeneous Treatment Effects: Sharp

Now suppose that: $Y_i = \alpha_i D_i + Y_{0i}$.

In the **sharp** design since $D_i = \mathbb{1}\{Z_i \geq z_0\}$ we have:

$$\mathbb{E}[Y_i|Z_i = z] = \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{1}\{z \geq z_0\} + \mathbb{E}[Y_{0i}|Z = z].$$

Average treatment effect for individuals **at the threshold** value z_0:

$$\alpha_{RD} \equiv \mathbb{E}[\alpha_i|Z_i = z_0].$$

Thus, we can rewrite the above expression as:

$$\begin{aligned}
\mathbb{E}[Y_i|Z_i = z] \quad &= \alpha_{RD}\,\mathbb{1}\{z \geq z_0\} + \mathbb{E}[Y_{0i}|Z_i = z] \\
&\qquad\qquad\qquad + (\mathbb{E}[\alpha_i|Z_i = z] - \mathbb{E}[\alpha_i|Z_i = z_0])\,\mathbb{1}\{z \geq z_0\} \\
&\equiv \alpha_{RD} D_i + k_{z_0}(z).
\end{aligned}$$

$\Rightarrow$ the situation is one of **selection on observables**.

Control function approach: the OLS population coefficient on D_i in the equation:

$$Y = \alpha_{RD} D + k(z) + w$$

equals $\mathbb{E}[\alpha_i|Z_i = z_0]$.

Heterogeneous Treatment Effects: Fuzzy

In the **fuzzy design**, D_i not only depends on $\mathbb{1}\{Z_i \geq z_0\}$, but also on other unobserved variables. Thus, D_i is an endogenous variable in the above regression.

We can use $\mathbb{1}\{Z_i \geq z_0\}$ as an **instrument** for D_i in such equation to identify α_{RD}, at least in the homogeneous case (connection with IV was first made explicit by van der Klaauw (2002)).

Below we discuss two alternative assumptions we can make for identification fuzzy designs: **conditional independence** near z_0, and **monotonicity**.

Conditional independence near z_0:

Weak conditional independence: $Y_{1i}, Y_{0i} \perp\!\!\!\perp D_i | Z_i = z$ for z near z_0, i.e. for $z = z_0 \pm e$, where e is arbitrarily small positive number, or:

$$F(Y_{ji}|D_i = 1, Z_i = z_0 \pm e) = F(Y_{ji}|Z_i = z_0 \pm e) \quad (j = 0, 1).$$

An **implication** is:

$$\mathbb{E}[\alpha_i D_i|Z_i = z_0 \pm e] = \mathbb{E}[\alpha_i|Z_i = z_0 \pm e]\,\mathbb{E}[D_i|Z_i = z_0 \pm e].$$

Proceeding as before, we have:

$$\lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] = \lim_{z \to z_0^+} \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^+} \mathbb{E}[Y_{0i}|Z_i = z]$$
$$\lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z] = \lim_{z \to z_0^-} \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^-} \mathbb{E}[Y_{0i}|Z_i = z].$$

Noting that $\lim_{z \to z_0^+} \mathbb{E}[\alpha_i|Z_i = z] = \lim_{z \to z_0^-} \mathbb{E}[\alpha_i|Z_i = z] = \alpha_{RD}$:

$$\alpha_{RD} \equiv \mathbb{E}[Y_{1i} - Y_{0i}|Z_i = z_0] = \frac{\lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z]}{\lim_{z \to z_0^+} \mathbb{E}[D_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[D_i|Z_i = z]}.$$

That is, the RD parameter can be interpreted as the **average TE** at z_0.

Monotonicity near z_0:

Alternative assumption: **local monotonicity** (Hahn et al., 2001):

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

for some $\bar{\varepsilon} > 0$ and any pair $(z_0 - \varepsilon, z_0 + \varepsilon)$ with $0 < \varepsilon < \bar{\varepsilon}$, where D_{zi} is the potential assignment indicator associated with $Z_i = z$.

Sometimes, **conditional independence** is problematic and **local monotonicity** not.

In such cases, α_{RD} identifies the **local average treatment effect** at $z = z_0$:

$$\alpha_{RD} = \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_{1i} - Y_{0i} | D_{z_0+\varepsilon,i} - D_{z_0-\varepsilon,i} = 1]$$

that is, the ATE for the units for whom treatment changes discontinuously at z_0.

Estimation Strategies

Hahn et al. (2001): Let $S_i \equiv \mathbb{1}\{z_0 - h < Z_i < z_0 + h\}$ where $h > 0$ denotes the bandwidth, and consider the subsample such that $S_i = 1$, and define $W_i \equiv \mathbb{1}\{z_0 < Z_i < z_0 + h\}$ as an instrument, applied to the subsample with $S_i = 1$:

$$\widehat{\alpha}_{RD} = \frac{\widehat{\mathbb{E}}[Y_i|W_i = 1, S_i = 1] - \widehat{\mathbb{E}}[Y_i|W_i = 0, S_i = 1]}{\widehat{\mathbb{E}}[D_i|W_i = 1, S_i = 1] - \widehat{\mathbb{E}}[D_i|W_i = 0, S_i = 1]}.$$

Alternative by the same authors, **control function**:

- **Sharp design**: OLS on $Y_i = \alpha_{RD} D_i + k(Z_i) + w_i$
- **Fuzzy design**: IV on $Y_i = \alpha_{RD} D_i + k(Z_i) + w_i$ using $\mathbb{1}\{Z_i \geq z_0\}$ as the excluded instrument.

Semiparametric approach (van der Klaaw, 2002): power series approximation for $k(Z)$.

The latter methods of estimation, not local to data points near the threshold, are implicitly predicated on the assumption of **homogeneous TE**.

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, which otherwise are "very local".

Covariates may also make the **local conditional exogeneity** assumption more credible.