Chapter 5. Regression Discontinuity

Joan Llull

Quantitative & Statistical Methods II
Master in Economics of Public Policy
Barcelona GSE

I. The fundamental RD assumption

In what we have seen so far, the main assumption in the matching context is conditional independence, $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i | X_i$, whereas in the IV context we assume orthogonality and relevance of the instrument, $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i | X_i$ and $D_i \not\perp\!\!\!\perp Z_i | X_i$ respectively. The relevance condition can also be expressed as $P(D_i = 1|Z_i = z) \neq P(D_i = 1|Z_i = z')$ for some $z \neq z'$. In ***regression discontinuity*** (RD) we consider a situation where there is a continuous variable Z_i that is not necessarily a valid instrument (it does not satisfy the exogeneity assumption), but that it is such that treatment assignment is a discontinuous function of Z_i. The basic asymmetry on which identification rests is discontinuity in the dependence of D_i on Z_i but continuity in the dependence of (Y_{1i}, Y_{0i}) on Z_i. RD methods have much potential in economic applications because geographic boundaries or program rules (e.g. eligibility thresholds) often create usable discontinuities.

More formally, discontinuity in treatment assignment but continuity in potential outcomes means that there is at least a known value $z = z_0$ such that:

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

$$\lim_{z \to z_0^+} P(Y_{ji} \leq r|Z_i = z) = \lim_{z \to z_0^-} P(Y_{ji} \leq r|Z_i = z) \quad (j = 0, 1) \tag{2}$$

Implicit regularity conditions are: (i) the existence of the limits, and (ii) that Z_i has positive density in a neighborhood of z_0. We abstract from conditioning covariates for the time being for simplicity.

Early RD literature in Psychology (e.g. Cook and Campbell, 1979) distinguishes between ***sharp*** and ***fuzzy*** designs. In the former, D_i is a deterministic function of Z_i:

$$D_i = \mathbb{1}\{Z_i \geq z_0\}, \tag{3}$$

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:

$$\begin{aligned}\lim_{z\to z_0^+}\mathbb{E}[D_i|Z_i=z]&=1\\ \lim_{z\to z_0^-}\mathbb{E}[D_i|Z_i=z]&=0.\end{aligned}\tag{4}$$

II. Homogeneous Treatment Effects

Like in the IV setting, the case of homogeneous treatment effects is useful to present the basic RD estimator. Suppose that $\alpha = Y_{1i} - Y_{0i}$ is constant, so that:

$$Y_i = \alpha D_i + Y_{0i}\tag{5}$$

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

$$\begin{aligned}\lim_{z\to z_0^+}\mathbb{E}[Y_i|Z_i=z]&=\alpha\lim_{z\to z_0^+}\mathbb{E}[D_i|Z_i=z]+\lim_{z\to z_0^+}\mathbb{E}[Y_{0i}|Z_i=z]\\ \lim_{z\to z_0^-}\mathbb{E}[Y_i|Z_i=z]&=\alpha\lim_{z\to z_0^-}\mathbb{E}[D_i|Z_i=z]+\lim_{z\to z_0^-}\mathbb{E}[Y_{0i}|Z_i=z],\end{aligned}\tag{6}$$

which leads to the consideration of the following RD parameter:

$$\alpha=\frac{\lim_{z\to z_0^+}\mathbb{E}[Y_i|Z_i=z]-\lim_{z\to z_0^-}\mathbb{E}[Y_i|Z_i=z]}{\lim_{z\to z_0^+}\mathbb{E}[D_i|Z_i=z]-\lim_{z\to z_0^-}\mathbb{E}[D_i|Z_i=z]},\tag{7}$$

which is determined provided the relevance condition in Equation (1) is satisfied, and equals α provided the independence condition in Equation (2) holds.

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

$$\alpha=\lim_{z\to z_0^+}\mathbb{E}[Y_i|Z_i=z]-\lim_{z\to z_0^-}\mathbb{E}[Y_i|Z_i=z],\tag{8}$$

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.

III. Heterogeneous Treatment Effects

Now suppose that:

$$Y_i = \alpha_i D_i + Y_{0i}.\tag{9}$$

It is useful again to distinguish sharp and fuzzy designs.

A. Sharp design

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

$$\mathbb{E}[Y_i|Z_i = z] = \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{1}\{z \geq z_0\} + \mathbb{E}[Y_{0i}|Z_i = z]. \tag{10}$$

In other words, conditioning on a value z for Z_i, individuals are treated if $z \geq z_0$, and thus we observe $Y_i = Y_{1i} = \alpha_i + Y_{0i}$, and untreated if $z \leq z_0$, in which case we observe $Y_i = Y_{0i}$. Thus, to obtain an average treatment effect for individuals at the threshold value z_0, that is, α_{RD} defined as:

$$\alpha_{RD} \equiv \mathbb{E}[\alpha_i|Z_i = z_0], \tag{11}$$

we rewrite (10) as:

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = z] &= \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{1}\{z \geq z_0\} + \mathbb{E}[Y_{0i}|Z_i = z] \pm \mathbb{E}[\alpha_i|Z_i = z_0]\,\mathbb{1}\{z \geq z_0\} \\ &= \alpha_{RD}\,\mathbb{1}\{z \geq z_0\} + \mathbb{E}[Y_{0i}|Z_i = z] \\ &\qquad\qquad + (\mathbb{E}[\alpha_i|Z_i = z] - \mathbb{E}[\alpha_i|Z_i = z_0])\,\mathbb{1}\{z \geq z_0\} \\ &\equiv \alpha_{RD} D_i + k_{z_0}(z). \end{aligned} \tag{12}$$

This equation corresponds to a situation of selection on observables, and the term $k_{z_0}(z)$ "controls" for the selection bias (this type of functions are indeed known as a control functions, and including them in the regression is know as a ***control function approach***). Therefore, the OLS population coefficient on D_i in the equation:

$$Y_i = \alpha_{RD} D_i + k_{z_0}(Z_i) + w_i \tag{13}$$

equals $\mathbb{E}[\alpha_i|Z_i = z_0]$, which is the causal effect of interest (an average treatment effect for individuals with Z_i right below or above the discontinuity).

The control function $k_{z_0}(z)$ is nonparametrically identified (e.g. including a high-order polynomial in Z_i —or $Z_i - z_0$— in the OLS regression interacted with a dummy $\mathbb{1}\{Z_i \geq z_0\}$). Note that if the treatment effect is homogeneous, $k(z)$ coincides with $\mathbb{E}[Y_{0i}|Z_i = z]$, but not in general.

B. Fuzzy design

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 Equation (13). However, we can still 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. The connection between the fuzzy design and the instrumental variables perspective was first made explicit in van der Klaaw (2002).

Next, we discuss the interpretation of α_{RD} in the fuzzy design with heterogeneous treatment effects, under two different assumptions. Consider first the weak conditional independence assumption:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i | Z_i = z \quad \text{for } z \text{ near } z_0, \tag{14}$$

that is, for $z = z_0 \pm e$, where e is an arbitrarily small positive number, or simply:

$$F(Y_{ji}|D_i = 1, Z_i = z_0 \pm e) = F(Y_{ji}|Z_i = z_0 \pm e) \quad (j = 0, 1). \tag{15}$$

Thus, we are assuming that treatment assignment is exogenous in the neighborhood of z_0. An implication is:

$$\mathbb{E}[\alpha_i D_i|Z_i = z_0 \pm e] = \mathbb{E}[\alpha_i|Z_i = z_0 \pm e]\,\mathbb{E}[D_i|Z_i = z_0 \pm e]. \tag{16}$$

Proceeding as before, we have:

$$\begin{aligned} \lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] &= \lim_{z \to z_0^+} \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^+} \mathbb{E}[Y_{0i}|Z_i = z] \\ \lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z] &= \lim_{z \to z_0^-} \mathbb{E}[\alpha_i|Z_i = z]\,\mathbb{E}[D_i|Z_i = z] + \lim_{z \to z_0^-} \mathbb{E}[Y_{0i}|Z_i = z]. \end{aligned} \tag{17}$$

Noting that $\lim_{z \to z_0^+} \mathbb{E}[\alpha_i|Z_i = z] = \lim_{z \to z_0^-} \mathbb{E}[\alpha_i|Z_i = z] = \alpha_{RD}$, subtracting one equation from the other, and rearranging the terms we obtain:

$$\begin{aligned} \alpha_{RD} &\equiv \mathbb{E}[Y_{1i} - Y_{0i}|Z_i = z_0] \\ &= \frac{\lim_{z \to z_0^+} \mathbb{E}[Y_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[Y_i|Z_i = z]}{\lim_{z \to z_0^+} \mathbb{E}[D_i|Z_i = z] - \lim_{z \to z_0^-} \mathbb{E}[D_i|Z_i = z]}. \end{aligned} \tag{18}$$

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

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

$$D_{z_0+\varepsilon,i} \geq D_{z_0-\varepsilon,i} \text{ for all units } i \text{ in the population.} \tag{19}$$

Sometimes, the local conditional independence assumption could be problematic, especially in fuzzy designs, but the monotonicity assumption is not. In such case, it can be shown that α_{RD} identifies the local average treatment effect at $z = z_0$:

$$\alpha_{RD} = \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_1 - Y_0 | D_{z_0+\varepsilon} - D_{z_0-\varepsilon} = 1] \tag{20}$$

that is, 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.

IV. Estimation Strategies

There are parametric and semiparametric estimation strategies. Hahn *et al.* (2001) suggested the following local estimator. 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 $S_i = 1$. The proposed estimator is the IV regression of Y_i on D_i using $W_i \equiv \mathbb{1}\{z_0 < Z_i < z_0 + h\}$ as an instrument, applied to the subsample with $S_i = 1$:

$$\widehat{\alpha}_{RD} = \frac{\widehat{\mathbb{E}}[Y_i | W_i = 1, S_i = 1] - \widehat{\mathbb{E}}[Y_i | W_i = 0, S_i = 1]}{\widehat{\mathbb{E}}[D_i | W_i = 1, S_i = 1] - \widehat{\mathbb{E}}[D_i | W_i = 0, S_i = 1]}. \tag{21}$$

In sharp designs, the denominator is equal to 1. This estimator has nevertheless a poor boundary performance. An alternative is based on Equation (13). In the case of a sharp design, OLS provides consistent estimates of α_{RD}, but in the fuzzy design D_i is endogenous. In that context, we would typically use $\mathbb{1}\{Z_i \geq z_0\}$ as an instrument for D_i. These regression methods, not local to data points near the threshold, are implicitly predicated on the assumption of homogeneous treatment effects.

V. 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.