

Regression Discontinuity

Erick F Molina

Summer, 2024

ITAM

Introduction to RDD

- The Regression Discontinuity Design (RDD) is a quasi-experimental design.
- It is used to estimate the causal effect of interventions by assigning a cutoff point above or below which treatment is assigned.

Examples:

- Do politicians seeking re-election have an advantage in elections?
- What is the effect of attending a school when we know students are admitted based on an exam score cutoff?

Introduction to RDD

- Thistlethwaite and Campbell (1960) are credited with the first use of this design.
- Other authors and reviews include:
 - Black (1999).
 - Ludwig and Miller (2007).
 - Imbens and Lemieux (2008).
 - Lee and Lemieux (2010).
 - Cattaneo, Idrobo, and Titiunik (2020, 2021).

Sharp Regression Discontinuity

Sharp Regression Discontinuity

Let us use the potential outcomes framework.

- An individual is untreated if $T_i = 0$.
- An individual is treated if $T_i = 1$.
- The individual has outcome Y_0 if untreated and Y_1 if treated.
- The treatment effect for an individual is $\delta = Y_1 - Y_0$, which is random.
- The Conditional Average Treatment Effect (ATE) for the subpopulation with $X = x$ is $\delta(x) = E[\delta \mid X = x]$.

Sharp Regression Discontinuity

- A **Sharp Regression Discontinuity Design** occurs when treatment is determined by a threshold rule in X , for example, $T = 1\{X \geq c\}$.
- In most applications, the threshold c is determined by policy or by rule.
- The covariate X that determines treatment is typically called the **running variable**. The threshold c is often called the **cutoff point**.

Sharp Regression Discontinuity

Example:

- Ludwig and Miller (2007) used this design to evaluate a U.S. federal anti-poverty program called Head Start.
- The program began in 1965 to provide preschool, health, and other social services to poor children ages three to five and their families.
- 300 counties were selected to start the program.
- The 300 counties were selected based on the poverty rate measured by the 1960 U.S. Census.
- They wanted to see whether this threshold had a measurable effect on infant mortality eight to eighteen years later in the same counties, compared to counties that did not receive assistance.

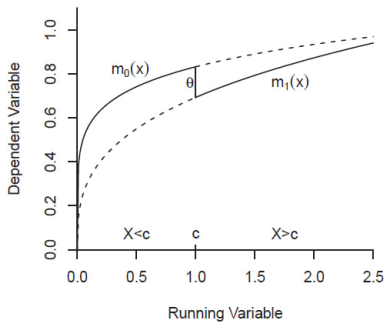
Identification

- Recall that $\delta(x) = E[\delta \mid X = x]$ is the conditional ATE.
- Let $\bar{\delta} = \delta(c)$ denote the conditional ATE for the subpopulation at the cutoff (the subpopulation marginally affected by setting c).
- We say that $\bar{\delta}$ is identified by the regression discontinuity design under certain assumptions.
- Let $m(x) = E[Y \mid X = x]$, $m_0(x) = E[Y_0 \mid X = x]$, and $m_1(x) = E[Y_1 \mid X = x]$.
- Then $\delta(x) = m_1(x) - m_0(x)$.
- Let $m(x+) = \lim_{z \downarrow x} m(z)$ and $m(x-) = \lim_{z \uparrow x} m(z)$.

Theorem: Suppose treatment is assigned as $T = 1\{X \geq c\}$. Suppose $m_0(x)$ and $m_1(x)$ are continuous at $x = c$. Then $\delta = m(c+) - m(c-)$.

- Continuity of $m_0(x)$ and $m_1(x)$ means that the conditional expectation of the untreated and treated outcomes varies smoothly with the running variable.
- In the Head Start example, $m_0(x)$ is the average mortality rate given the poverty rate for counties that did not receive assistance.
- $m_1(x)$ is the average mortality rate for counties that did receive assistance.
- There is no reason to expect a discontinuity in either function.

Figure 1: Sharp Regression Discontinuity



(a) Sharp Regression Discontinuity



(b) Effect of Head Start on Childhood Mortality

To estimate δ , it is common to center the running variable around the cutoff point c .

$$Y_i = \alpha + \beta(X_i - c) + \gamma(X_i - c) \times T_i + \delta T_i + \varepsilon_i \quad (1)$$

- In general, we can consider designs with nonlinear relationships.
- In those cases, we can use local regressions or kernel estimators.

- In general, under local regressions, the estimator $\hat{\delta}_{RDD}$ has an asymptotically normal distribution.
- Since nonparametric methods are used, there is a bias that we cannot directly measure.
- Two recommendations to reduce bias are:
 - Use a common bandwidth for both local regressions.
 - Use a smaller bandwidth than the optimal bandwidth.

- Bandwidth choice is critical in nonparametric estimation.
- Data-driven bandwidth rules should be computed before estimation.
- One suggestion is the Rule-of-Thumb (ROT) bandwidth from Fan and Gijbels (1996).
- The reference model is a polynomial of order q plus a level shift.

- The ROT bandwidth formula is:

$$h_{rot} = 0.58 \left(\frac{\hat{\sigma}^2(\xi_2 - \xi_1)}{\hat{B}} \right)^{1/5} n^{-1/5}$$

- Another useful method is cross-validation (CV).
- CV minimizes leave-one-out residuals and selects the bandwidth that minimizes this criterion.
- CV can be noisy, especially near the cutoff.

Fuzzy RDD

Fuzzy regression discontinuity (Fuzzy RD) exploits discontinuities in the probability (or expected value) of treatment conditional on a covariate.

Treatment Probability

Let d_i denote treatment, and suppose there is a jump in the probability of treatment at x_0 , such that:

$$P[d_i = 1 \mid x_i] = \begin{cases} g_0(x_i) & \text{if } x_i \leq x_0 \\ g_1(x_i) & \text{if } x_i > x_0 \end{cases}$$

We can write the relationship between treatment probability and x_i as:

$$E[d_i \mid x_i] = P[d_i = 1 \mid x_i] = g_0(x_i) + [g_1(x_i) - g_0(x_i)]t_i$$

where $t_i = 1(x_i \geq x_0)$.

Fuzzy RD naturally leads to a two-stage least squares (2SLS) estimation strategy:

$$E[d_i | x_i] = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \cdots + \gamma_p x_i^p) t_i$$

First stage:

$$d_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \cdots + \alpha_p x_i^p + \delta t_i + \nu_i$$

where t_i is the excluded instrument.

Reduced form: This is obtained by substituting into the outcome equation:

$$y_i = \pi_0 + \pi_1 x_i + \pi_2 x_i^2 + \cdots + \pi_p x_i^p + \tau t_i + \epsilon_i$$

where τ is the treatment effect.

Treatment effect estimates that vary with x_i can be constructed via 2SLS:

$$d_i = \alpha_0 + \alpha_1 \tilde{x}_i + \alpha_2 \tilde{x}_i^2 + \cdots + \alpha_p \tilde{x}_i^p + \gamma_0 t_i + \gamma_1 \tilde{x}_i t_i + \gamma_2 \tilde{x}_i^2 t_i + \cdots + \gamma_p \tilde{x}_i^p t_i + \nu_i$$

The nonparametric version of Fuzzy RD is IV estimation in a neighborhood close to the discontinuity:

$$\frac{E[y_i \mid x_0 < x_i < x_0 + h] - E[y_i \mid x_0 - h < x_i < x_0]}{E[d_i \mid x_0 < x_i < x_0 + h] - E[d_i \mid x_0 - h < x_i < x_0]}$$