Regression Discontinuity

Charlie Gibbons
Political Science 236

October 31, 2008

Outline

Overview

The regression discontinuity strategy assumes that the probability of receiving treatment changes discontinuously as a function of underlying variables.

In practice, it is difficult to find a good application of RD. But, if you can find an appropriate design, then this method can be very persuasive.

The model

We can think of the following model:

$$\begin{aligned} Y_i &= Y_0 + D_i(z_i)\beta \\ z_i &= X_i\gamma + \nu_i \end{aligned}$$

Y_i is the outcome variable for individual i,
Y_0 is the no-treatment outcome (could be individual-specific),
z_i is the covariate that induces a discontinuous change in treatment probability at z_0,
$D_i(z_i)$ is an indicator of individual i's treatment status as a function of his value of z_i,
X_i is a row vector of (potentially unobserved) covariates,
ν_i is an error term.

We want to estimate the treatment effect $\beta \equiv Y_1 - Y_0$.

The model

A *sharp* design has:

$$D_i(z_i) = D(z_i) = \begin{cases} 1 & z_i \geq z_0 \\ 0 & z_i < z_0 \end{cases}$$

The model

A *sharp* design has:

$$D_i(z_i) = D(z_i) = \begin{cases} 1 & z_i \geq z_0 \\ 0 & z_i < z_0 \end{cases}$$

A *fuzzy* design has:

$$\begin{aligned} \mathbb{E}[D_i(z_i)|z_i \geq z_0] &= \Pr[D_i(z_i) = 1|z_i \geq z_0] > \\ \Pr[D_i(z_i) &= 1|z_i < z_0] = \mathbb{E}[D_i(z_i)|z_i < z_0] \end{aligned}$$

Estimator

Before we get to the assumptions necessary for identification in this model, let's see what the estimate of β will be:

$$
\begin{aligned}
\hat{\beta} &= \frac{\hat{Y}^{+} - \hat{Y}^{-}}{\hat{D}_i(z_i)^{+} - \hat{D}_i(z_i)^{-}} \\
&= \frac{\frac{1}{\#\{i : z_i \geq z_0\}} \sum_{\{i : z_i \geq z_0\}} Y_i - \frac{1}{\#\{i : z_i < z_0\}} \sum_{\{i : z_i < z_0\}} Y_i}{\frac{1}{\#\{i : z_i \geq z_0\}} \sum_{\{i : z_i \geq z_0\}} D_i(z_i) - \frac{1}{\#\{i : z_i < z_0\}} \sum_{\{i : z_i < z_0\}} D_i(z_i)}
\end{aligned}
$$

Estimator

Before we get to the assumptions necessary for identification in this model, let's see what the estimate of β will be:

$$\begin{aligned}
\hat{\beta} &= \frac{\hat{Y}^{+} - \hat{Y}^{-}}{\hat{D}_i(z_i)^{+} - \hat{D}_i(z_i)^{-}} \\
&= \frac{\frac{1}{\#\{i:\, z_i \geq z_0\}} \sum_{\{i:\, z_i \geq z_0\}} Y_i - \frac{1}{\#\{i:\, z_i < z_0\}} \sum_{\{i:\, z_i < z_0\}} Y_i}{\frac{1}{\#\{i:\, z_i \geq z_0\}} \sum_{\{i:\, z_i \geq z_0\}} D_i(z_i) - \frac{1}{\#\{i:\, z_i < z_0\}} \sum_{\{i:\, z_i < z_0\}} D_i(z_i)}
\end{aligned}$$

What does this remind you of?

Estimator

Before we get to the assumptions necessary for identification in this model, let's see what the estimate of β will be:

$$\begin{aligned}\hat{\beta} &= \frac{\hat{Y}^{+} - \hat{Y}^{-}}{\hat{D}_i(z_i)^{+} - \hat{D}_i(z_i)^{-}} \\ &= \frac{\frac{1}{\#\{i:\, z_i \geq z_0\}} \sum_{\{i:\, z_i \geq z_0\}} Y_i - \frac{1}{\#\{i:\, z_i < z_0\}} \sum_{\{i:\, z_i < z_0\}} Y_i}{\frac{1}{\#\{i:\, z_i \geq z_0\}} \sum_{\{i:\, z_i \geq z_0\}} D_i(z_i) - \frac{1}{\#\{i:\, z_i < z_0\}} \sum_{\{i:\, z_i < z_0\}} D_i(z_i)}\end{aligned}$$

What does this remind you of?

This estimator is *numerically equivalent* to an IV estimator with $\mathbb{I}\{z_i \geq z_0\}$ as an instrument.

Estimator

Before we get to the assumptions necessary for identification in this model, let's see what the estimate of β will be:

$$\begin{aligned}\hat{\beta} &= \frac{\hat{Y}^{+} - \hat{Y}^{-}}{\hat{D}_i(z_i)^{+} - \hat{D}_i(z_i)^{-}} \\ &= \frac{\frac{1}{\#\{i: z_i \geq z_0\}} \sum_{\{i: z_i \geq z_0\}} Y_i - \frac{1}{\#\{i: z_i < z_0\}} \sum_{\{i: z_i < z_0\}} Y_i}{\frac{1}{\#\{i: z_i \geq z_0\}} \sum_{\{i: z_i \geq z_0\}} D_i(z_i) - \frac{1}{\#\{i: z_i < z_0\}} \sum_{\{i: z_i < z_0\}} D_i(z_i)}\end{aligned}$$

What does this remind you of?

This estimator is *numerically equivalent* to an IV estimator with $\mathbb{I}\{z_i \geq z_0\}$ as an instrument.

Some features of the *design* give different identifying assumptions, however.

Assumptions of IV

Recall the assumptions of IV:

- SUTVA holds
- The instrument is ignorably assigned
- The instrument satisfies the exclusion restriction
- The instrument's average effect on D is non-zero (*i.e.*, it satisfies the inclusion restriction)

$$\mathbb{E}[D_i(1) - D_i(0)|X] \neq 0$$

- It satisfies the monotonicity assumption

$$D_i(1) \geq D_i(0) \text{ or } D_i(1) \leq D_i(0) \quad \forall\, i$$

SUTVA

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA)

$$(Y_i(0), Y_i(1)) \perp T_j \ \forall \ j \neq i$$

Existence of discontinuity

Assumption 2: Existence of a discontinuity in treatment probabilities

$$\lim_{z \to z_0^+} \mathbb{E}[D_i(z_i)|z_i = z] \neq \lim_{z \to z_0^-} \mathbb{E}[D_i(z_i)|z_i = z]$$

assuming that both of these limits exist.

Existence of discontinuity

Assumption 2: Existence of a discontinuity in treatment probabilities

$$\lim_{z \to z_0^+} \mathbb{E}[D_i(z_i)|z_i = z] \neq \lim_{z \to z_0^-} \mathbb{E}[D_i(z_i)|z_i = z]$$

assuming that both of these limits exist.

The probability of receiving treatment must be discontinuous at z_0.

Existence of discontinuity

Assumption 2: Existence of a discontinuity in treatment probabilities

$$\lim_{z \to z_0^+} \mathbb{E}[D_i(z_i)|z_i = z] \neq \lim_{z \to z_0^-} \mathbb{E}[D_i(z_i)|z_i = z]$$

assuming that both of these limits exist.

The probability of receiving treatment must be discontinuous at z_0.

This assumption combines the inclusion restriction and monotonicity assumptions from IV; *i.e.*, having a z_i above the threshold must change your probability of receiving treatment and everyone must move from less likely to more likely to receive treatment.

Continuity of Y_0

Assumption 3a: Y_0 is continuous

$\mathbb{E}[Y_0|z_i = z]$ is continuous at $z = z_0$

Y_1 is continuous

Assumption 3b: Y_1 is continuous

$\mathbb{E}[Y_1|z_i = z]$ is continuous at $z = z_0$

Y_1 is continuous

Assumption 3b: Y_1 is continuous

$\mathbb{E}[Y_1|z_i = z]$ is continuous at $z = z_0$

This assumption can arise in two ways:

- Homogeneous treatment effect: $Y_{1i} - Y_{0i} = \beta \quad \forall\ i$
  Then, Assumption 3b is implied by Assumption 3a.
- Treatment effect continuous at z_0
  $\mathbb{E}[\beta_i|z_i = z]$ is continuous at z_0

Selection

Assumption 4: Selection

There are three possible assumptions to overcome selection bias. The first two assume selection problems away and estimate the ATE:

- Homogeneous treatment effect: $Y_{1i} - Y_{0i} = \beta \quad \forall\, i$
  Selection not a concern
- $\beta_i \perp D_i(z_i)|z_i$
  $i.e.$, $(Y_i, Y_0) \perp D_i(z_i)|X_i, z_i$
  Selection on observables

Selection

Assumption 4: Selection

There are three possible assumptions to overcome selection bias. The first two assume selection problems away and estimate the ATE:

- Homogeneous treatment effect: $Y_{1i} - Y_{0i} = \beta \quad \forall\, i$
  Selection not a concern
- $\beta_i \perp D_i(z_i) | z_i$
  i.e., $(Y_i, Y_0) \perp D_i(z_i) | X_i, z_i$
  Selection on observables

Selection

Assumption 4: Selection

There are three possible assumptions to overcome selection bias. The first two assume selection problems away and estimate the ATE:

- Homogeneous treatment effect: $Y_{1i} - Y_{0i} = \beta \quad \forall\, i$
  Selection not a concern
- $\beta_i \perp D_i(z_i)|z_i$
  $i.e.$, $(Y_i, Y_0) \perp D_i(z_i)|X_i, z_i$
  Selection on observables

Selection

If you believe that selection is a problem, then you can look at the LATE near z_0 rather than the ATE. There are two ways to think of the assumption that is necessary here:

- z_i has a random component and the distribution of z_i is continuous near z_0 for all i (Lee)
- $(\beta_i, T_i) \perp z_i$ for z_i near z_0 (Hahn, Todd, and van der Klaauw)

Selection

If you believe that selection is a problem, then you can look at the LATE near z_0 rather than the ATE. There are two ways to think of the assumption that is necessary here:

- z_i has a random component and the distribution of z_i is continuous near z_0 for all i (Lee)
- $(\beta_i, T_i) \perp z_i$ for z_i near z_0 (Hahn, Todd, and van der Klaauw)

Selection

If you believe that selection is a problem, then you can look at the LATE near z_0 rather than the ATE. There are two ways to think of the assumption that is necessary here:

- z_i has a random component and the distribution of z_i is continuous near z_0 for all i (Lee)
- $(\beta_i, T_i) \perp z_i$ for z_i near z_0 (Hahn, Todd, and van der Klaauw)

Selection

If you believe that selection is a problem, then you can look at the LATE near z_0 rather than the ATE. There are two ways to think of the assumption that is necessary here:

- z_i has a random component and the distribution of z_i is continuous near z_0 for all i (Lee)
- $(\beta_i, T_i) \perp z_i$ for z_i near z_0 (Hahn, Todd, and van der Klaauw)

These are analogous to the IV assumption of ignorability of treatment assignment.

Selection

The first assumption states that, if you look at z's near z_i, then the probability of that z arising doesn't change that much from the probability of z_i.

Recall that z_i is a function of (potentially unobserved) covariates. If the probability of getting some z changes a little as you change z_i, then as you change the covariates a little, the probability of that new z arising doesn't change much. This suggests that the covariates will be balanced around z_0.

Selection

The second assumption says that the value of treatment to the individual (β_i) is independent of z_i—*i.e.*, *there is no selection around* z_0.

The second part of this assumption states that receipt of treatment is independent of z_i—*i.e.*, getting treatment for individuals with z_i around z_0 is based upon random variation.

Selection

RD finds exogeneity in spite of selection by assuming the following: Because there is random error, the distribution of treatment around a given point (*e.g.*, z_0) will be random.

Selection

RD finds exogeneity in spite of selection by assuming the following: Because there is random error, the distribution of treatment around a given point (*e.g.*, z_0) will be random.

Essentially, we are saying that the group that just receives treatment is just like the group that just doesn't receive treatment—we assume that there is no selection around z_0.

Selection

RD finds exogeneity in spite of selection by assuming the following: Because there is random error, the distribution of treatment around a given point (*e.g.*, z_0) will be random.

Essentially, we are saying that the group that just receives treatment is just like the group that just doesn't receive treatment—we assume that there is no selection around z_0.

This assumption gets worse the further that we get from z_0.

Selection

RD finds exogeneity in spite of selection by assuming the following: Because there is random error, the distribution of treatment around a given point (*e.g.*, z_0) will be random.

Essentially, we are saying that the group that just receives treatment is just like the group that just doesn't receive treatment—we assume that there is no selection around z_0.

This assumption gets worse the further that we get from z_0.

Note that there is an "auxiliary prediction" that the distributions of the covariates will be equal in the treatment and control groups.

Selection

RD finds exogeneity in spite of selection by assuming the following: Because there is random error, the distribution of treatment around a given point (*e.g.*, z_0) will be random.

Essentially, we are saying that the group that just receives treatment is just like the group that just doesn't receive treatment—we assume that there is no selection around z_0.

This assumption gets worse the further that we get from z_0.

Note that there is an "auxiliary prediction" that the distributions of the covariates will be equal in the treatment and control groups.

In summary these assumptions assume that the average outcome for individuals just below z_0 must be a reasonable counterfactual for those just above the threshold.

Non-strategic response

Assumption 5: Non-strategic response

$z_i \perp Z_0$, where Z_0 is the set of possible z_0

An individual does not change his value of the measure as the threshold changes.

This is like the exclusion restriction of IV.

Advantages over IV

We've seen that RD has very similar assumptions to IV; indeed, we can place RD easily into the IV framework.

Advantages over IV

We've seen that RD has very similar assumptions to IV; indeed, we can place RD easily into the IV framework.

Then why is RD a superior strategy?

Advantages over IV

We've seen that RD has very similar assumptions to IV; indeed, we can place RD easily into the IV framework.

Then why is RD a superior strategy?

- RD can be non-parametric
- The subpopulation for which the LATE is estimated is identifiable in RD, but not in IV
- Ignorability of treatment assignment can be examined in RD, but not in IV

Advantages over IV

We've seen that RD has very similar assumptions to IV; indeed, we can place RD easily into the IV framework.

Then why is RD a superior strategy?

- RD can be non-parametric
- The subpopulation for which the LATE is estimated is identifiable in RD, but not in IV
- Ignorability of treatment assignment can be examined in RD, but not in IV

Advantages over IV

We've seen that RD has very similar assumptions to IV; indeed, we can place RD easily into the IV framework.

Then why is RD a superior strategy?

- RD can be non-parametric
- The subpopulation for which the LATE is estimated is identifiable in RD, but not in IV
- Ignorability of treatment assignment can be examined in RD, but not in IV

References

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica.* 69(1): 201-209.

Lee, David S. 2007. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics.*