An Alternative Assumption to Identify LATE in Regression Discontinuity Designs

Yingying Dong

UC Irvine

May 2015 @ SETA 2015

Regression Discontinuity (RD) Designs

x_i is a **binary treatment indicator**: 1 if treated, 0 if untreated.
e.g., welfare program participation, winning a scholarship, pollution regulation

y_i is the **observed outcome**
e.g. employment, college choice, housing price

z_i is a continuous "**running variable**."
At a **known threshold** of z_i, say z_0, the treatment prob. changes discontinuously.
e.g., income for welfare eligibility, test score to determine scholarship, pollution level

Regression Discontinuity (RD) Design

Under certain conditions, RD models identify a local average treatment effect (LATE):

$$\text{RD LATE} = \frac{\lim_{z \downarrow z_0} E(y_i \mid z_i = z) - \lim_{z \uparrow z_0} E(y_i \mid z_i = z)}{\lim_{z \downarrow z_0} E(x_i \mid z_i = z) - \lim_{z \uparrow z_0} E(x_i \mid z_i = z)}.$$

Sharp design: Treatment prob. jumps from 0 to 1 at z_0.
Fuzzy design: Treatment prob. changes less than 1 at z_0.

Intuition: If everything is smooth without treatment, the mean outcome change should be caused by the treatment change at z_0.

Hahn, Todd and van der Klaauw, (hereafter HTV, 2001):

The first paper to provide RD identifying assumptions, show that RD designs identify a LATE, and discuss local linear estimation.

y_{1i} (y_{0i}) is potential outcome when i is treated (not treated).

The observed outcome $y_i = \alpha_i + \beta_i x_i$, with $\alpha_i \equiv y_{0i}$ and $\beta_i \equiv y_{1i} - y_{0i}$.

A key assumption (the HTV independence assumption):

- $(\beta_i, x_i(z)) \perp z_i$ in a neighborhood of z_0 (Theorem 2 of HTV 2001): practically requires that **the treatment effect is independent of the running variable** near the cutoff.

HTV(2001)

Alternative assumptions to the independence assumption :

Constant treatment effects: β_i is a constant.
Treatment effects independent of the treatment: $\beta_i \perp x_i$ conditional on z_i near z_0 (ruling out self-selection based on idiosyncratic gains).

Identification additionally requires 1) continuity of $E(\alpha_i \mid z_i = z)$ at $z = z_0$, 2) a discontinuity at z_0 (compliers), and 3) monotonicity (no defiers).

A Simple Illustration

Assume the cutoff $z_0 = 0$.

Assume $y_i = a + bz_i + \tau x_i + \tau_1 z_i x_i + e_i$.

- The treatment effect is $\beta_i = \tau + \tau_1 z_i$.
- The HTV independence assumption requires that $\beta_i \perp z_i$, so $\tau_1 = 0$.
- In sharp design, $\tau_1 = 0$ means no slope change at the cutoff.

Why care about relaxing the assumption?

The independence assumption

- places a restriction on treatment effect heterogeneity.
  - In practice, treatment effect β_i may directly depend on the running variable or is correlated with it.
- may not hold in many empirical applications of RD models:
  - E.g., Lee (2008), Goodman (2008), Yörük and Yörük (2011) etc.
- restrict the "treatment effect derivative" (TED) to be zero; TED can in fact be nonparametrically identified and useful in several ways (Dong and Lewbel 2014).

Example 1

Lee (2008): SRD estimating the electoral advantage of incumbency in the US House election

x_i = Democrats being the incumbent party.
z_i = Democratic Party's winning margin in election t with $z_0 = 0$.
y_i = Democrats winning the next election $(t+1)$.

- The HTV independence assumption requires that the incumbency advantage does not depend on winning margin.
- When the incumbent party won by a larger margin, their prob. of winning the next election may be greater.

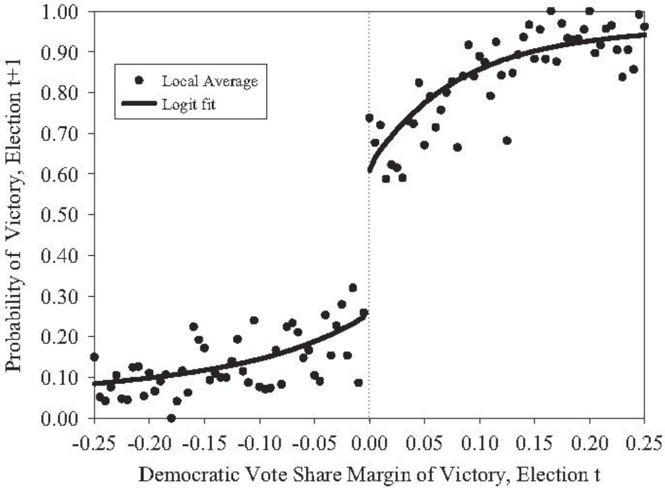

Figure: Probability of the Democratic Party winning election t+1 against its winning margin in election t

Example 2

Goodman (2008): estimating impacts of the Adams Scholarship on college choices.

$x_i =$ being eligible for Adams Scholarship, and hence tuition waiver at MA's public colleges.
$z_i =$ distance (number of grade points) to the eligibility threshold.
$y_i =$ whether one chooses a 4-year public college vs. private college.

- The HTV independence assumption requires that students' responses to Adams Scholarship does not depend on their test scores, and hence outside opportunities.

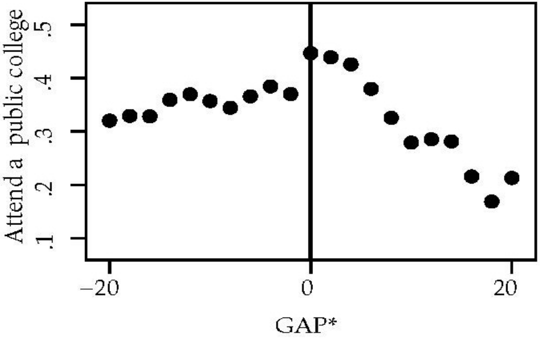

Figure: Prob. of choosing a public college against distance to eligibility threshold

Dramatic downward slope change at the cutoff:

- Marginal winners respond much more strongly to the scholarship.
- Quality - Price tradeoff: highly skilled students can gain admission to private colleges of much higher quality, and hence face a much larger quality drop in taking the Scholarship.

Hypothetical example 3

x_i = remedial education
z_i = test score; z_0=threshold failing grade
y_i = students' later outcomes (see, e.g., Jacob and Lefgren 2004)

- The HTV independence assumption requires the effectiveness of remedial education not to depend on one's pre-treatment test score near the threshold.
- The smoothness assumption in this paper only requires that no students have precise manipulation of their test scores and no other discrete changes at the cutoff.

This paper

- Shows that both SRD and FRD can be identified by smoothness assumptions instead of the HTV independence assumption.
- Provides a weak behavioral assumption (in spirit of Lee 2008) that leads to the smoothness required for identification of SRD and FRD.
  - empirically plausible and partially testable.
- Results provide formal support for the popular practice of performing McCrary's (2008) density test to assess fuzzy design RD.
- Provides a test for the HTV independence assumption.
- An empirical application showing that smoothness plausibly holds, but the HTV independence assumption does not.

Identification under Alternative Assumptions

Focus on fuzzy design, treating sharp design as a special case.

Define potential treatment below or above the cutoff as $x_{0i}(z)$ and $x_{1i}(z)$, respectively.
One of them is not observed, or the counterfactual treatment.

Define four types of individuals (Angrist, Imbens and Rubin, 1996):
Always taker: $A = \{i : x_{0i}(z) = x_{1i}(z) = 1\}$
Never taker: $N = \{i : x_{0i}(z) = x_{1i}(z) = 0\}$
Complier: $C = \{i : x_{0i}(z) = 0, x_{1i}(z) = 1\}$
Defier: $D = \{i : x_{0i}(z) = 1, x_{1i}(z) = 0\}$

Standard RD models identify a LATE for compliers at $z = z_0$, i.e., individuals with $x_{0i}(z_0) = 0$ and $x_{1i}(z_0) = 1$.

Identification under Alternative Assumptions

Suppress the argument to simply use x_{0i} and x_{1i}.
Define the random vector $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{0i}, x_{1i})$.

- An individual i can be seen as defined by the vector $\mathbf{w}_i$.
- Given her draw of the running variable, $\mathbf{w}_i$ completely determines her treatment status and outcome y_i.

Identification under Alternative Assumptions

ASSUMPTION A1a (Smoothness): $f_{z|\mathbf{w}}(z \mid \mathbf{w})$ is continuous in a neighborhood of $z = z_0$ for all $\mathbf{w} \in supp\,(\mathbf{w}_i)$, and $f_z(z)$ is continuous and strictly positive in a neighborhood of $z = z_0$.

- For each individual defined by **w** the density of the running variable is continuous.
- Implies that individuals cannot precisely manipulate the running variable to be just above or below the cutoff.
- A1a leads to the smoothness required for identifying both sharp and fuzzy design RD.

Identification under Alternative Assumptions

ASSUMPTION A1b (Smoothness): $E(y_{ti} \mid x_{0i} = t_0, x_{1i} = t_1, z_i = z)$ and $Pr(x_{0i} = t_0, x_{1i} = t_1 \mid z_i = z)$, for $t = 0, 1$, $t_0 = 0, 1$ and $t_1 = 0, 1$ are continuous in z at $z = z_0$.

- For fuzzy design, A1b requires
  - conditional means of y_{0i} and y_{1i} are continuous for each type of individuals, and
  - the probabilities of types are continuous.
- For sharp design, everyone is a complier; A1b reduces to that $E(y_{ti} \mid z_i = z)$ for $t = 0, 1$ is continuous.

- **LEMMA**: If A1a holds, then A1b holds.

Identification under Alternative Assumptions

ASSUMPTION A2 (RD, existence of compliers): $x^+ \neq x^-$.

ASSUMPTION A3 (Monotonicity, no defiers): $x_{1i}(z) \geq x_{0i}(z)$ for all i and $z \in (z_0 - \varepsilon, z_0 + \varepsilon)$.

THEOREM: Given assumptions A1b, A2 and A3, the local average treatment effect for compliers at $z_i = z_0$ is identified and is given by $E[y_{1i} - y_{0i} \mid z_i = z_0, C] = (y^+ - y^-) / (x^+ - x^-)$.

Discussion

- A1b suffices for identification; A1a is stronger than necessary, but A1a is more interpretable, and has testable implications.
- $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{0i}, x_{1i})$ puts no restrictions on treatment effect heterogeneity or selection into treatment based on idiosyncratic gains and hence selection into different types.
- Individuals can self-select to be compliers, as long as the probability of being a complier is smooth at the RD cutoff.
- Independence is a strong assumption; in empirical applications, the smoothness assumption is more likely to hold.

Testing the HTV Independence Assumption

Extend the previous A1a or A1b to require continuous differentiability instead of continuity.
One can then estimate the treatment effect derivative w.r.t. the running variable.
The independence assumption by HTV implies the treatment effect derivative (TED) to be zero:

$$\frac{\partial E(\beta_i \mid z_i = z, C)}{\partial z} \mid_{z=z_0} = 0.$$

- Can test HTV Independence by testing whether TED is significant or not.

Empirical Application: Lee (2008) Incumbency Advantages

Table 1 RD estimates of the treatment effect and treatment effect derivative (TED)

Bandwidth	Treatment effect		Treatment effect derivative		Optimal order of polynomial
	Dependent var.: Winning in election t+1				
1.00	0.385	(0.039)***	1.293	(0.514)**	4
0.50	0.370	(0.043)***	1.542	(0.678)**	3
0.45	0.363	(0.046)***	1.574	(0.801)**	3
0.40	0.407	(0.036)***	1.186	(0.368)***	2
0.35	0.393	(0.038)***	1.150	(0.451)**	2
0.30	0.381	(0.042)***	1.446	(0.577)**	2
0.25	0.375	(0.047)***	1.664	(0.791)**	2
0.20	0.423	(0.033)***	0.988	(0.263)***	1
0.15	0.409	(0.040)***	0.998	(0.424)**	1
[0.172]	0.417	(0.037)***	1.134	(0.339)***	1
	Dependent var.: Density of winning margin in election t				
0.50	0.125	(0.131)	-1.290	(1.064)	2
	Dependent var.: Vote share in election t-1				
0.50	-0.010	(0.014)	0.177	(0.252)	3

Note: The optimal bandwidth for the local linear regression (the last row of the top

Empirical Application: Goodman (2008) Impacts of Adams Scholarship on College Choices

Table 2 RD estimates of the effect of Adams Scholarship on college choices

	4 year Pubic		4 year Private		Any college	
			A: $\|X\| <= 10$			
$\tau(c)$	0.081	0.082	-0.080	-0.071	0.012	0.019
	(0.015)***	(0.015)***	(0.015)***	(0.015)***	(0.009)	(0.008)**
TED	-0.019	-0.019	0.018	0.017	-0.004	-0.004
	(0.003)***	(0.002)***	(0.002)***	(0.002)***	(0.001)***	(0.001)***
			B: $\|X\| <= 20$			
$\tau(c)$	0.075	0.076	-0.061	-0.056	0.023	0.027
	(0.011)***	(0.011)***	(0.012)***	(0.011)***	(0.008)***	(0.006)***
TED	-0.017	-0.017	0.013	0.012	-0.003	-0.003
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Covariates	N	Y	N	Y	N	Y

Note: The sample size for the top panel is 18,456, and for the bottom panel is 27,885; Tre
ment effect - new refers to the RD treatment effect if the eligibility threshold were margin
lowered by 2 grade points. Robust standard errors are in the parentheses; * significant at
10% level, ** significant at the 5% level, *** significant at the 1% level.

Related Work

Assume smoothness, in particular $f_{z|\mathbf{w}}(z \mid \mathbf{w})$ is continuously differentiable,

- Dong and Lewbel (2014) identifies how RD treatment effect would change if the RD threshold changes.
- Dong (2014) generalizes the standard RD design to identify a causal effect when there is no jump, but instead a kink (a slope change) in the treatment prob.
  - *not* a RKD; RKD requires a continuous treatment. Here treatment is binary.
- Dong (2014) also discusses a general model with possibly either a jump, or a kink, or both and provide robust estimation.

Conclusions

"But when you start exercising those rules, all sorts of processes start to happen and you start to find out all sorts of stuff about people...**It's just a way of thinking about a problem**, which lets the shape of the problem begin to emerge. The more rules, the tinier the rules, **the more arbitrary they are, the better**."
– Douglas Adams, *Mostly Harmless*