Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models

Yingying Dong and Arthur Lewbel
University California Irvine and Boston College

UCI&BC

Regression Discontinuity (RD) Models

T is a treatment indicator: $T = 1$ if treated, $T = 0$ if untreated.
e.g., winning a scholarship, being eligible for Superfund cleanups

Y is an observed outcome.
e.g. college choices, housing price

X is a continuous running variable determining treatment assignment.
e.g., test score, hazard level

c is a known threshold constant
e.g., qualifying threshold score for a scholardship; ceiling poullution level for pollution abatement mandate.

Regression Discontinuity (RD) Models

Given a policy rule such the treatment probability changes discontinuously when X exceeds c, RD models identify a local average treatment effect (LATE) under weak conditons.

$$\text{RD LATE} = \frac{\text{Mean outcome discontinuity at } X = c}{\text{Treatment prob. discontinuity at } X = c}$$

Example: Thistlethwaite and Campbell (1960) among many others

Sharp design: Everybody is a *complier*; Treatment prob. jumps from 0 to 1 at c. Fuzzy design: Treatment prob. changes less than 1 at c.

RD Intuition: individuals just above or just below the cutoff are comparable, and so mean outcome change is due to treatment prob change at c.

This paper

1) The derivative of the RD treatment effect with respect to the running variable at the cutoff (TED) is nonparametrically identified and easily estimated, under smoothness conditions already assumed in empirical applications of RD models.

2) The change in the RD treatment effect resulting from a marginal change in the RD threshold (MTTE) can also be nonparametrically identified and easily estimated (MTTE = TED), given local policy invariance.

3) Applies these results to Goodman (2008) and Greenstone and Gallagher (2008).

Why Care about Identifying TED and MTTE?

RD models:

1) Strong internal validity

- How about external validity?
- Useful to know what the treatment effect would be at points other than the cutoff c.

2) Identifying a LATE

- Nothing is known about how the treatment effect would change if the threshold is changed.
- Many policy debates center on threshold changes. Examples:
  - Income levels for means-tested welfare program.
  - Test core threshold for remedial education, diploma or scholarships
  - Pollution ceiling for pollution abatement mandates

Quotes from RD Literature

Hahn, Todd, and Van der Klaauw (2001): "A limitation of the approach is that it only identifies treatment effects locally at the point ... It would be of interest, for example, if the policy change being considered is a small change in the program rules, such as lowering or raising the threshold for program entry..."

Imbens and Lemieux (2008): "One important aspect of both the SRD and FRD designs is that they, at best, provide estimates of the average effect for a subpopulation..."

Greenstone and Gallagher (2008) : "it is important to highlight that the RD approach only provides estimates of the treatment effect at the regulatory discontinuity... To extend the external validity of the RD estimates to the full 1982 HRS Sample, it is necessary to assume a homogeneous treatment effect in that sample."

Why Issues TED and MTTE can Address?

TED

- Investigate treatment effect heterogeneity with respect to the running variable
- Evaluate external validity or generality of the estimated LATE
- Test for locally constant treatment effects; $TED = 0$ is a necessary condition.

MTTE

- Approximate the impact on treatment effects of a small discrete change in the threshold.
- Sign of MTTE tells whether the estimated LATE would increase or decrease if the threshold were marginally increased or lowered.

Literature Review

General RD identification, estimation and inference: Thistlethwaite and Campbell (1960), Hahn, Todd, and Van der Klaauw (2001), Porter (2003), Battistin and Rettore (2008), Lee (2008), McCrary (2008), Ludwig and Miller (2008), Imbens and Kalyanaraman (2011), Frandsen et al. (2012), Cattaneo et al. (2014), DiNardo and Lee (2011), Angrist and Pischke (2008), Van der Klaauw (2008), Imbens and Lemieux (2008), Lee and Lemieux (2010), Dong (2014).

RD derivatives: Card, Lee, Pei, and Weber (2012), Dong (2013).

Identification away from the cutoff: Angrist and Rokkanen (2012), Ai et al (2011), Maynard (2013), Jackson (2010).

Discrete or mismeasured running variable: Lee and Card (2008), Barreca et al (2010), and Yu (2012), Dong (2013).
Many others...

Sharp Design TED

Suppose sharp design; Suppose holding the RD threshold c fixed.

Let $Y(t)$ for $t = 1, 0$ be potential outcomes when treated or not treated (Rubin 1974), so $Y = Y(1)T + Y(0)(1 - T)$.

Define $D(X) = E[Y(1) - Y(0) \mid X]$ and $D'(X) = \partial D(X) / \partial X$.

- $D(x)$ is the ATE (for compliers) when the running variable $X = x$.
- $D'(x)$ is change the ATE from a tiny change in x.

The ATE identified by sharp RD is $D(c)$.
Treatment Effect Derivative (TED) is defined as $D'(c)$.

- Can be taken as the coefficient of $(X - c)T$ in a (local) linear regression of Y on a constant, T, $X - c$ and $(X - c)T$.

Sharp Deisgn MTTE

Now consider possibly changing the RD threshold:

Standard RD models identify a LATE, denoted at $\tau(c)$.

- Define the Marginal Threshold Treatment Effect (MTTE) to be $\tau'(c) = \frac{\partial \tau(c)}{\partial c}$.
- MTTE is the change in the LATE resulting from a marginal change in the threshold.

TED and MTTE in Empirical RD Applications

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

$X =$ the Democratic Party's winning margin in election t
$T = T^* =$ Democrats being the incumbent party, determined by $X > 0$.
$Y =$Democrats winning the next election $(t+1)$

- TED $D'(c)$ tells how the incumbency advantage depends on winning margin.
- Likely the incumbent party who won by a larger margin has a greater probability of winning the next election (Figure 1)

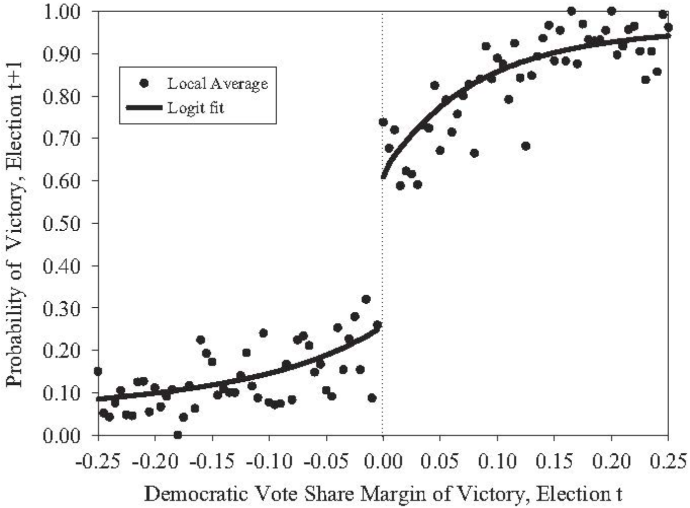

Figure: Probability of the Democratic Party winning election t+1 against its winning margin in election t

TED and MTTE in Empirical RD Applications

Greenstone and Gallagher (2008): Impacts of Superfund cleanups on nearby housing prices

X = a county's HRS score (measuring the hazard level of a waste site)
T = being eligible for Superfund cleanups, largely determined by the 1982 HRS score exceeding a threshold c.
Y = median housing prices in the surrounding area.

- TED $D'(c)$ shows how the effect of Superfund cleanups would change with the HRS score, and thereby provides info. on external validity of the estimate at the current regulatory threshold.
- MTTE $\tau'(c)$ tells how the effect of Superfund cleanups would change if the threshold changed slightly.

Why is identification of TED and MTTE questionable?

- We have no info about the treatment effect anywhere but at c.
- We never observe a threshold change.

Parametric Illustration

Suppose true model were $Y = a + Xb + X^2 d + T\beta + XT\gamma + X^2 T\delta + e$, where $E(e \mid X) = 0$.
Suppose sharp design, so $T = T^*$.

- $D(x) = E(Y \mid T = 1, X = x) - E(Y \mid T = 0, X = x) = \beta + x\gamma + x^2\delta$.
- so TED is $D'(c) = \gamma + 2c\delta$.

- LATE at $x = c$ is $\tau(c) = \beta + c\gamma + c^2\delta$, so MTTE is $\tau'(c) = \gamma + 2c\delta$.

- Here functional form identifies the treatment effect function $D(x)$ both at $x = c$ and for values $x \neq c$, permitting identification of $D'(c)$.
- But can TED be nonparametrically identified? Yes.

Sharp Design TED Identification

ASSUMPTION A1: For each individual we observe Y, T, X.

ASSUMPTION A2 (sharp design): $T = I(X \geq c)$ for some known constant threshold c. The support of X includes a neighborhood of c. For $t = 0, 1$, $E(Y(t) \mid X = x)$ is continuously differentiable in x in a neighborhood of $x = c$.

RD only needs continuity, not differentiability of $E(Y(t) \mid X = x)$.

However, nonparametric estimators of $g(x)$ (e.g., kernel or local linear regression) assume differentiability, and parametric specifications are also differentiable. We will use this differentiability.

Sharp Design TED Identification

- Define $g(x) = E(Y \mid X = x)$, then standard sharp design RD result:

$$\tau(c) = g_{+}(c) - g_{-}(c), \tag{1}$$

where $g_{+}(c) = \lim_{\varepsilon \longrightarrow 0} g(c+\varepsilon)$, $g_{-}(c) = \lim_{\varepsilon \longrightarrow 0} g(c-\varepsilon)$..

THEOREM 1: If Assumptions A1 and A2 hold then the treatment effect $D(c)$ holding the threshold fixed at c is identified by equation (1) and TED defined by $D'(c)$ is identified by

$$D'(c) = g'_{+}(c) - g'_{-}(c), \tag{2}$$

where $g'_{+}(x) = \lim_{\varepsilon \downarrow 0} \frac{g(x+\varepsilon) - g(x)}{\varepsilon}$ and $g'_{-}(x) = \lim_{\varepsilon \downarrow 0} \frac{g(x) - g(x-\varepsilon)}{\varepsilon}$.

Proof Sketch of Theorem 1

Let $h_t(x) = E(Y(t) \mid X = x)$ for $t = 0, 1$.

For $x > c$ have $h_1(x) = g(x)$ so $h'_{1+}(x) = g'_+(x)$ for any $x > c$.

By differentiability, $h'_{1+}(x) = h'_1(x)$.

By continuity of $h'_1(x)$, having above hold for all $x > c$ in a neighborhood of c implies they hold at $x = c$, so $g'_+(c) = h'_1(c)$.

The same argument based on $x < c$ leads to $g'_-(c) = h'_0(c)$.

Sharp Design MTTE

- In general $Y(1)$ and $Y(0)$ can depend on both the running variable X and a given cutoff C, so define
$$S(X, C) = E(Y(1) - Y(0) \mid X, C).$$
- Given the true threshold c, then $\tau(c) = S(c, c)$ while $D(x) = S(x, c)$.
- Define derivatives of $S(X, C)$ to be $S_X(X, C) = \frac{\partial S(X,C)}{\partial X}$, $S_C(X, C) = \frac{\partial S(X,C)}{\partial C}$.
- Then TED $D'(c)$ and MTTE $\tau'(c)$ are given by
$$D'(c) = S_X(c, c) \quad \text{and} \quad \tau'(c) = D'(c) + S_C(c, c). \qquad (3)$$

COROLLARY 1: Let Assumptions A1, A2, and the local policy invariance condition $S_C(c, c) = 0$ hold. Then MTTE is nonparametrically identified by $\tau'(c) = D'(c)$.

Local Policy Invariance

Local policy invariance assumes that $D(x)$, the treatment effect as a function of the running variable, does not change when the policy threshold changes infinitesimally for x near c.

- A limit version of *policy invariance (*Abbring and Heckman, 2007):
  - "...assumption that an agent's outcome only depends on the treatment assigned to the agent and not separately on the mechanism used to assign treatments."
- Commobly used in reduced form program evaluation
  - Almost always required in any counterfactual policy impact evaluation or extrapolation.
  - E.g., Heckman (2010) and Carneiro, Heckman, and Vytlacil (2010)...
- It does not place any restriction on how the treatment effect depends on the running variable.
- Plausibility depends on context.

Fuzzy Design TED

Analogous to $D(x)$ in sharp designs, define

$$D_f(x) = E\left[Y(1) - Y(0) \mid X = x, T(0) < T(1)\right] \tag{4}$$

Standard fuzzy design RD identification result:

$$D_f(c) = \frac{g_+(c) - g_-(c)}{f_+(c) - f_-(c)}, \tag{5}$$

where $f_+(c) = \lim_{\varepsilon \longrightarrow 0} f(c+\varepsilon)$, $f_-(c) = \lim_{\varepsilon \longrightarrow 0} f(c-\varepsilon)$

ASSUMPTION B2 (Fuzzy design): The support of X includes a neighborhood of a known constant threshold c. $T(0) \leq T(1)$. The conditional means $E(Y(t) \mid T(0) < T(1), X = x)$ and $E(Y(t) \mid T(0) = T(1) = t, X = x)$ as well as the probabilities $\Pr(T(0) < T(1) \mid X = x)$ and $\Pr(T(0) = T(1) = t \mid X = x)$ for $t = 0, 1$ are continuously differentiable in x in a neighborhood of $x = c$. $\Pr(T(0) < T(1) \mid X = x)$ is strictly positive at $x = c$.

Fuzzy Design TED Identification

THEOREM 2: If Assumptions A1 and B2 hold then the treatment effect $D_f(c)$ is identified by equation (5) and the Fuzzy Design RD TED $D'_f(c)$ is identified by

$$D'_f(c) = \frac{g'_+(c) - g'_-(c) - [f'_+(c) - f'_-(c)]\, D_f(c)}{f_+(c) - f_-(c)}.$$

Proof Sketch of Theorem 2

First show that given smoothness in Assumption 2, $D_f(x) = [g_+(x) - g_-(x)] / p(x)$, where $p(x) = \Pr(T(0) < T(1) \mid X = x)$.

$$
\begin{aligned}
& g_+(x) - g_-(x) \\
= \; & \lim_{\varepsilon \longrightarrow 0} E(Y|X = x + \varepsilon) - \lim_{\varepsilon \longrightarrow 0} E(Y|X = x - \varepsilon) \\
= \; & [E(Y(1) - E(Y(0))|X = x, T(0) < T(1)) \Pr(T(0) < T(1)|X = \\
= \; & D_f(x) p(x)
\end{aligned}
$$

With continuous differentiability, can take the derivative of the above expression and evaluate at $x = c$:

$$
D_f'(c) = \frac{g_+'(c) - g_-'(c)}{p(x)} - D_f(x) \frac{p'(x)}{p(x)}, \tag{6}
$$

Applying Theorem 1 replacing Y with T then gives $p(c) = f_+(c) - f_-(c)$ and $p'(c) = f_+'(c) - f_-'(c)$. Plug these in to get Theorem 2.

Fuzzy Design MTTE

Analogous to the way S and its derivatives are defined in sharp design, define S_f and its derivatives as

$$S_f(X, C) = E(Y(1) - Y(0) \mid X, C, T(0) < T(1)), \text{ and}$$

$$S_{fX}(X, C) = \frac{\partial S_f(X, C)}{\partial X}, S_{fC}(X, C) = \frac{\partial S_f(X, C)}{\partial C}.$$

The fuzzy design TED and MTTE are given by

$$D'_f(c) = S_{fC}(c, c), \quad \text{and} \quad \tau'_f(c) = D'_f(c) + S_{fC}(c, c) \qquad (7)$$

Fuzzy design local policy invariance: $S_{fC}(c, c) = 0$.

COROLLARY 2: Let Assumptions A1, B2, and the local policy invariance condition $S_{fC}(c, c) = 0$ hold. Then the fuzzy design MTTE $\tau'_f(c)$ is nonparametrically identified by $\tau'_f(c) = D'_f(c)$.

Estimation of Sharp Design TED and MTTE

Can estimate TED or MTTE by local linear or quadratic regressions.

For example, when applying a local linear regression to data satisfying $c - \varepsilon \leq X_i \leq c + \varepsilon$ for a small $\varepsilon > 0$ with a uniform kernel,

$$Y_i = \alpha + (X_i - c)\,\beta + T_i^* \gamma_0 + (X_i - c)\,T_i^* \gamma_1 + e_i.$$

By Theorem 1, the sharp design treatment effect and TED (MTTE) are

$$\begin{aligned} \widehat{D}(c) &= plim\left(\widehat{\gamma}_0\right) = g_+(c) - g_-(c) & (8) \\ \widehat{D}'(c) &= plim\left(\widehat{\gamma}_1\right) = g'_+(c) - g'_-(c). & (9) \end{aligned}$$

Given MTTE, use a Taylor approximation

$$\tau(c_{new}) \approx \tau(c) + (c_{new} - c)\,\tau'(c).$$

Estimation of Fuzzy Design TED and MTTE

For fuzzy design, need to similarly estimate the treatment eq.

$$T_i = \alpha^T + (X_i - c)\,\beta^T + T_i^* \gamma_0^T + (X_i - c)\, T_i^* \gamma_1^T + e_i^T \qquad (10)$$

Then

$$plim\left(\widehat{\gamma}_0^T\right) = f_+(c) - f_-(c) \text{ and } plim\left(\widehat{\gamma}_1^T\right) = f'_+(c) - f'_-(c) \qquad (11)$$

By Theorem 2, the fuzzy design treatment effect and TED (MTTE) is

$$\widehat{D}'_f(c)\,\widehat{D}_f(c) = \widehat{\gamma}_0 / \widehat{\gamma}_0^T \text{ and } \widehat{D}'_f(c) = \left(\widehat{\gamma}_1 - \widehat{\gamma}_1^T \widehat{D}_f(c)\right) / \widehat{\gamma}_0^T$$

Empirical Application

Goodman (2008): Impacts of Adams Scholarship in MA on college choices.

X is the minimum distance (number of grade points) to the eligibility threshold, so $c = 0$.
Y is whether one chooses a 4-year public college vs. private college.
$T = T^* = I(X \geq 0)$, being eligible for Adams Scholarship, which provides tuition waiver at the MA's public colleges.

- TED $D'(c)$ tells how students' responses to Adams Scholarship depends on their test scores, and hence outside opportunities.
- MTTE $\tau'(c)$ tells how the average effects of Adams Scholarship program would change if the qualifying threshold were marginally raised or lowered.

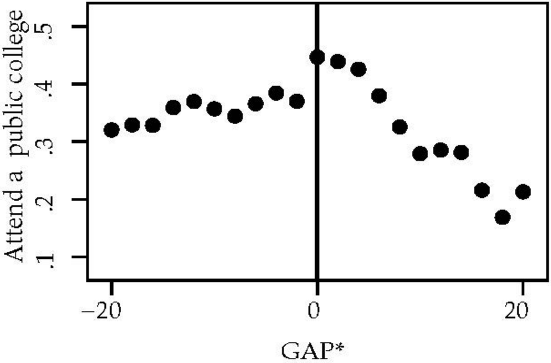

Figure: Prob of choosing a public college against distance to the eligiblity threshold

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 up the Scholarship and choosing MA's public colleges.

Local Policy Invariance

Local policy invariance

- requires that a student's choice of college with or without an Adams scholarship would not change if the eligibility threshold itself were marginally changed.
- still allows these choices to depend on the student's skill level (test scores).

Would be violated

- if a marginal change in threshold changed the perceived value of attending a public college vs alternatives,
- or through peer effects, e.g., if seeing more of one's friends qualify for an Adams Scholarship changed one's own college choice.

Table 1 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)**
MTTE(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)**
$\tau(c_{new})$	0.120	0.120	-0.117	-0.105	0.019	0.027
	(0.015)***	(0.015)***	(0.015)***	(0.015)***	(0.009)**	(0.000)**
			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)**
MTTE(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)**
$\tau(c_{new})$	0.110	0.110	-0.086	-0.079	0.028	0.033
	(0.011)***	(0.011)***	(0.012)***	(0.011)***	(0.008)***	(0.006)**
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 referes 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.

Conclusions

- In RD models the treatment effect derivative (TED) is nonparametrically identified and easily estimated, under smoothness already assumed.
- TED estimate can be useful
  - in investigating treatment effect heterogeneity with respect to the running variable
  - in investigating external validity near the RD threshold.
  - in assessing locally constant treatment effects.
- Given local policy invariance (commonly used in counterfactual policy forecasting or extraplolation), TED $=$MTTE, measuring changes in treatment effects if the policy threshold were marginally changed.
- Empirical applications to evaluation of the Adams Scholarship program: TED is neg. and significant.

Conclusions

Another fuzzy design empirical application estimates the effect of Suplerfund cleanups on nearby housing prices (Greenstone and Gallagher, 2008):

- Estimated TED is numerically small, implying estimated effects of Superfund cleanups in Greenstone and Gallagher (2008) holds more generally at hazard levels other than the chosen regulatory threshold
- Policy Invariance plausibly holds, so estimated TED is also MTTE.
- Raising or lowering the regulatory threshold would not result in a significant change in the estimated effects of Superfund cleanups on nearby housing prices.