# Discontinuity Designs

## ECON 31720 Applied Microeconometrics

Francesco Ruggieri

The University of Chicago

November 18, 2020

# Framework for Sharp Regression Discontinuity Designs

- $Y \in \mathbb{R}$ is a scalar **outcome** of interest, $D \in \{0, 1\}$ is a **binary treatment**
- $D$ and $Y$ are linked by **potential outcomes** $Y(0), Y(1)$
- $R \in \mathbb{R}$ is a **running variable**, not necessarily continuously distributed everywhere
- The **dependency** of $D$ and $R$ with $Y(0), Y(1)$ is **not restricted**
- There exists a **cutoff** $c \in \mathbb{R}$ such that $D = \mathbb{I}[R \geq c]$
  - The treatment is a **deterministic function** of the running variable
- $\mathbb{E}[Y(d)|R = r]$ is **continuous** at $r = c$ for $d = 0, 1$

# Framework for Sharp Regression Discontinuity Designs

- Since $D = 1$ if and only if $R \geq c$ and $D = 0$ if and only if $R < c$:

$$\mathbb{E}[Y|R = r] = \mathbb{E}[Y|R = r, D = 1] = \mathbb{E}[Y(1)|R = r] \quad \text{for every } r \geq c$$
$$\mathbb{E}[Y|R = r] = \mathbb{E}[Y|R = r, D = 0] = \mathbb{E}[Y(0)|R = r] \quad \text{for every } r < c$$

- Taking **limits** for $r \downarrow c$ and $r \uparrow c$:

$$\lim_{r \downarrow c} \mathbb{E}[Y|R = r] = \lim_{r \downarrow c} \mathbb{E}[Y(1)|R = r] = \mathbb{E}[Y(1)|R = c]$$
$$\lim_{r \uparrow c} \mathbb{E}[Y|R = r] = \lim_{r \uparrow c} \mathbb{E}[Y(0)|R = r] = \mathbb{E}[Y(0)|R = c]$$

where the last equality follows in both cases from **continuity** of $\mathbb{E}[Y(d)|R = r]$ at $r = c$

- These limits can be **differenced out** to point identify

$$\text{ATE}(c) \equiv \mathbb{E}[Y(1) - Y(0)|R = c] = \lim_{r \downarrow c} \mathbb{E}[Y|R = r] - \lim_{r \uparrow c} \mathbb{E}[Y|R = r]$$

# Sharp Regression Discontinuity Designs: Extensions

1. **Multiple Cutoffs**: the cutoff is a discrete random variable, $C$, rather than a constant, $c$
   - Example: plurality voting in elections with more than two competing candidates
   - Example: state or local governments setting eligibility cutoffs for a federal program
2. **Multiple Running Variables**: $R \in \mathbb{R}^{d_r}$, with $d_r > 1$, as opposed to $R \in \mathbb{R}$
   - Example: a scholarship awarded to students who score above two subject-specific thresholds
   - Example: counties that require voting by mail vs. counties that allow in-person voting

1. Framework for Sharp Regression Discontinuity Designs

2. Sharp Regression Discontinuity Designs: Extensions
   - Multiple Cutoffs
   - Multiple Running Variables

3. Framework for Regression Kink Designs

4. Regression Probability Kink Designs

5. Summary

# Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016)

- $Y \in \mathbb{R}$ is a scalar **outcome** of interest, $D \in \{0, 1\}$ is a **binary treatment**
- $C$ is a **cutoff random variable** with support $\mathcal{C} = \left\{c_1, \ldots, c_{\bar{j}}\right\}$
  - The probability of each cutoff realization is $p_c \equiv \mathbb{P}(C = c) \in [0, 1]$
- $R \in \mathbb{R}$ is a continuously distributed **running variable** with density $f_R(r)$
  - $f_{R|C}(r|c_j)$ denotes the density of $R$ conditional on $C = c_j$ for $j = 1, \ldots, \bar{j}$
- In this setting, different agents may **face different cutoffs**

# Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016)

- For simplicity, focus on the **sharp design**, so $D = \mathbb{I}[R \geq C]$

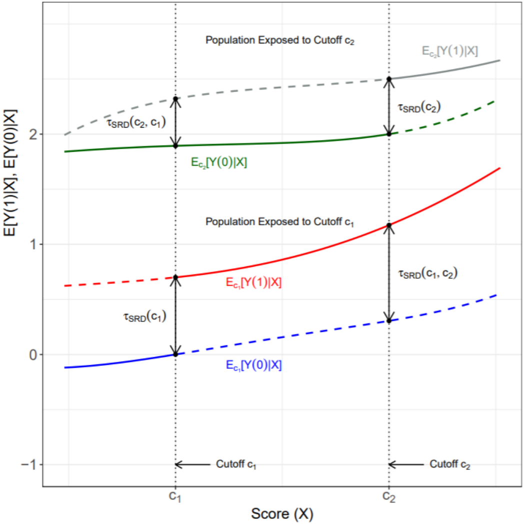
- $D$ and $Y$ are linked by **potential outcomes** $Y(0), Y(1)$, which are also functions of $C$:

$$Y(0,C), Y(1,C) \quad \text{s.t.} \quad Y = \sum_{j=1}^{\bar{j}} \mathbb{I}[C = c_j] \times [DY(1,c_j) + (1-D)Y(0,c_j)]$$

  $Y(d, c_j)$ highlights that potential outcomes may be **affected** by the realization of $C$

- Further assumptions:
  1. $\mathbb{E}[Y(d,c)|R=r, C=c]$ is **continuous in $r$** at $r = c$ for every $c \in \mathcal{C}$ and $d = 0, 1$
  2. The densities $f_{R|C}(r|c)$ are **positive and continuous in $r$** at $r = c$ for every $c \in \mathcal{C}$

# Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016)



# Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016)

Rather than estimating **cutoff-specific effects**, one may choose to **normalize and pool**:

1. **Normalize**: define the **normalized running variable** $\widetilde{R} \equiv R - C$
   - In words, **center** each agent's running variable around their **cutoff realization**

2. **Pool**: identify a target parameter using **standard** regression discontinuity arguments
   - This approach ignores the **heterogeneity** in the distributions of $Y(0)$ and $Y(1)$ in terms of $C$
   - The point-identified target parameter is a **pooled estimand**:

$$\tau^P \equiv \sum_{j=1}^{\bar{j}} \mathbb{E}\left[Y(1, c_j) - Y(0, c_j) | R = c_j, C = c_j\right] \times \omega(c_j)$$

where

$$\omega(c) \equiv \frac{f_{R|C}\left(c|c\right) \times \mathbb{P}\left(C = c\right)}{\sum_{c' \in \mathcal{C}} f_{R|C}\left(c'|c'\right) \times \mathbb{P}\left(C = c'\right)}$$

# Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016)

1. **Constant treatment effects**: if $Y(1, c_j) - Y(0, c_j) = \tau(c_j)$ for $j = 1, \dots, \bar{j}$, $\tau^P \equiv \sum_{j=1}^{\bar{j}} \tau(c_j) \times \omega(c_j)$ is a weighted average of **cutoff-specific constants**

2. **Ignorable $R$**: if $\mathbb{E}[Y(1, c_j) - Y(0, c_j)|R = c_j, C = c_j] = \mathbb{E}[Y(1, c_j) - Y(0, c_j)|C = c_j]$,

$$\tau^P \equiv \sum_{j=1}^{\bar{j}} \mathbb{E}[Y(1, c_j) - Y(0, c_j)|C = c_j] \times \omega(c_j)$$

   so $\tau^P$ may be estimated with **global polynomial** techniques

3. **Ignorable $C$**: if $\mathbb{E}[Y(1, c_j) - Y(0, c_j)|R = c_j, C = c_j] = \mathbb{E}[Y(1, c_j) - Y(0, c_j)|R = c_j]$,

$$\tau^P \equiv \sum_{j=1}^{\bar{j}} \mathbb{E}[Y(1, c_j) - Y(0, c_j)|R = c_j] \times \mathbb{P}(C = c_j)$$

   so $\tau^P$ is a weighted average of "local" (in terms of $R$) **average treatment effects**

# Keele and Titiunik (2015)

- $Y \in \mathbb{R}$ is a scalar **outcome of interest**, $D \in \{0, 1\}$ is a **binary treatment**
- $D$ and $Y$ are linked by **potential outcomes** $Y(0), Y(1)$
- Treatment assignment **changes discontinuously** at a **border** $\mathcal{B}$
  - $\mathcal{B}$ is geographic boundary that separates a **treated area** $(\mathcal{B}_t)$ from a **control area** $(\mathcal{B}_c)$
- $R \in \mathbb{R}^2$ is a pair of **running variables** usually denoting **latitude and longitude**
- For simplicity, focus on the **sharp design**, so

$$D = \mathbb{I}[R_1 \geq b_1] \times \mathbb{I}[R_2 \geq b_2]$$

where $(b_1, b_2) \in \mathcal{B}$ is a **boundary point**

# Keele and Titiunik (2015)

- As in the scalar case, **average potential outcomes** are **continuous** at the **border**:

$$\mathbb{E}\left[Y(d)|\left(R_1, R_2\right) = \left(r_1, r_2\right)\right] \text{ is continuous in } r_1, r_2 \text{ at } r_1 = b_1,\ r_2 = b_2$$

for every $(b_1, b_2) \in \mathcal{B}$ and $d = 0, 1$

- Let $\mathbf{R} = (R_1, R_2)$, $\mathbf{r} = (r_1, r_2)$, and $\mathbf{b} = (b_1, b_2)$. Then

$$\lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_t} \mathbb{E}\left[Y|\mathbf{R} = \mathbf{r}\right] = \lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_t} \mathbb{E}\left[Y(1)|\mathbf{R} = \mathbf{r}\right] = \mathbb{E}\left[Y(1)|\mathbf{R} = \mathbf{b}\right]$$
$$\lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_c} \mathbb{E}\left[Y|\mathbf{R} = \mathbf{r}\right] = \lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_c} \mathbb{E}\left[Y(0)|\mathbf{R} = \mathbf{r}\right] = \mathbb{E}\left[Y(0)|\mathbf{R} = \mathbf{b}\right]$$

where the last equality follows in both cases from **continuity** of $\mathbb{E}\left[Y(d)|\mathbf{R} = \mathbf{r}\right]$ at $\mathbf{r} = \mathbf{b}$

# Keele and Titiunik (2015)

- As in the scalar case, these limits can be **differenced out** to identify

$$\mathbb{E}\left[Y(1)-Y(0)|\mathbf{R}=\mathbf{b}\right]=\lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_t}\mathbb{E}\left[Y|\mathbf{R}=\mathbf{r}\right]-\lim_{\mathbf{r}\to\mathbf{b};\mathbf{r}\in\mathcal{B}_c}\mathbb{E}\left[Y|\mathbf{R}=\mathbf{r}\right]$$

  where $\mathbb{E}\left[Y(1)-Y(0)|\mathbf{R}=\mathbf{b}\right]$ is the $\mathbf{ATE}$ of $D$ on $Y$ at the **border point** $\mathbf{b}\in\mathcal{B}$

- In practice, one may construct a **scalar running variable** as the **Euclidean distance**

$$D(b_1,b_2)=\sqrt{(R_1-b_1)^2+(R_2-b_2)^2}$$

  which reduces the design to a standard **unidimensional** regression discontinuity:

$$\mathbb{E}\left[Y(1)-Y(0)|D(\mathbf{b})=0\right]=\lim_{d\downarrow 0}\mathbb{E}\left[Y|D(\mathbf{b})=d\right]-\lim_{d\uparrow 0}\mathbb{E}\left[Y|D(\mathbf{b})=d\right],\quad \mathbf{b}\in\mathcal{B}$$

- Rather than estimating **b**-specific effects, one may again choose to **normalize and pool**

# Framework for Regression Kink Designs

- $Y \in \mathbb{R}$ is a scalar **outcome** of interest
- $D \in \mathbb{R}$ is a **continuously distributed treatment**
- $R \in \mathbb{R}$ is a continuously distributed **running variable**
- $U \in \mathbb{R}$ is a continuous **latent variable** denoting the **unobserved determinants** of $Y$
- Consider an **all-causes model** of the outcome variable: $Y \equiv g(D, U)$
  - $g(\cdot)$ is an **unknown function** of the observed and unobserved determinants of $Y$
- The **dependency** between $R$ and $U$ is **not restricted**
  - The **distribution** function $f_{U|R}(u|r)$ is **continuously differentiable** in $r$ at $r = c$

# Framework for Regression Kink Designs

- Analogously to regression discontinuity designs, **two** scenarios are possible:
  1. **Sharp**: the treatment is $D = h(R)$, where $h(\cdot)$ is a **known function**
  2. **Fuzzy**: the treatment is $D = h(R, U)$, where $h(\cdot)$ is an **unknown function** ($U$ is latent)
- In **both** cases, $g(\cdot)$ is **continuously differentiable** at the threshold
  1. **Sharp**: $g(d, u)$ is continuously differentiable in $d$ at $d = h(c)$
  2. **Fuzzy**: $g(d, u)$ is continuously differentiable in $d$ at $d = h(c, u)$ **for every** $\boldsymbol{u}$
- In **both** cases, $h(\cdot)$ is **continuous**, but its **derivative** is **discontinuous** at the threshold
  1. **Sharp**: $h(r)$ is continuous, but $h'(r)$ is discontinuous at $r = c$
  2. **Fuzzy**: $h(r, u)$ is continuous **for every** $\boldsymbol{u}$, but $h'(r, u)$ is discontinuous at $r = c$ **for every** $\boldsymbol{u}$

# Framework for Regression Kink Designs

To identify a target parameter in a **sharp regression kink design**, consider any $r \neq c$ and

$$\begin{aligned}
\frac{\partial \mathbb{E}\left[Y|R=r\right]}{\partial r} &= \frac{\partial}{\partial r}\mathbb{E}\left[g\left(h(r),U\right)|R=r\right] \qquad (Y \equiv g(D,U)) \\
&= \frac{\partial}{\partial r}\int g\left(h(r),u\right) f_{U|R}\left(u|r\right) du \qquad (\text{definition of } \mathbb{E}\left[\cdot\right]) \\
&= \int \frac{\partial}{\partial r}\left[g\left(h(r),u\right)\right] f_{U|R}\left(u|r\right) du \qquad (\text{Fubini's Theorem}) \\
&= h'(r)\int \left(\frac{\partial}{\partial d} g\left(h(r),u\right)\right) f_{U|R}\left(u|r\right) du + \int g\left(h(r),u\right)\left(\frac{\partial}{\partial r} f_{U|R}\left(u|r\right)\right) du
\end{aligned}$$

where the last equality follows from an application of the chain rule

# Framework for Regression Kink Designs

- By assumption, $h'(r)$ is **discontinuous** at $r = c$
- Take the **limits** of $\frac{\partial \mathbb{E}[Y|R=r]}{\partial r}$ as $r \downarrow c$ and $r \uparrow c$:

$$\lim_{r \downarrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} = \lim_{r \downarrow c} h'(r) \times \int \left( \frac{\partial}{\partial d} g\left(h(c), u\right) \right) f_{U|R}\left(u|c\right) du + \int g\left(h(c), u\right) \left( \frac{\partial}{\partial r} f_{U|R}\left(u|c\right) \right) du$$

$$\lim_{r \uparrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} = \lim_{r \uparrow c} h'(r) \times \int \left( \frac{\partial}{\partial d} g\left(h(c), u\right) \right) f_{U|R}\left(u|c\right) du + \int g\left(h(c), u\right) \left( \frac{\partial}{\partial r} f_{U|R}\left(u|c\right) \right) du$$

# Framework for Regression Kink Designs

These limits can be **differenced out**:

$$\lim_{r \downarrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} - \lim_{r \uparrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} = \left( \lim_{r \downarrow c} h'(r) - \lim_{r \uparrow c} h'(r) \right) \times \mathbb{E}\left[ \frac{\partial}{\partial d} g\left(h(c), U\right) \middle| R = c \right]$$

Rearranging terms, the **Local Average Response** (LAR) of $Y$ to $D$ is

$$\mathbb{E}\left[ \frac{\partial}{\partial d} g\left(h(c), U\right) \middle| R = c \right] = \frac{\lim_{r \downarrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} - \lim_{r \uparrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r}}{\lim_{r \downarrow c} h'(r) - \lim_{r \uparrow c} h'(r)}$$

This is the **average effect** of a **marginal increase in $D$** on $Y$ at $R = c$

- The LAR **averages** marginal effects over the distribution of $U$ among agents with $R = c$

# Framework for Regression Kink Designs

- A similar derivation in the case of a **fuzzy regression kink design** leads to

$$\mathbb{E}\left[\frac{\partial}{\partial d} g\left(h(c,U),U\right) \times \omega\left(c,U\right) \Big| R=c\right] = \frac{\lim_{r\downarrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} - \lim_{r\uparrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r}}{\lim_{r\downarrow c} \frac{\partial \mathbb{E}[D|R=r]}{\partial r} - \lim_{r\uparrow c} \frac{\partial \mathbb{E}[D|R=r]}{\partial r}}$$

  where $\omega(c,U)$ is **proportional** to the **size of the kink** (analogously to an IV first stage)

- Recall that $D = h(R,U)$. If a **monotonicity** assumption holds, i.e.,

$$\lim_{r\downarrow c} \frac{\partial h(r,U)}{\partial r} \geq \lim_{r\uparrow c} \frac{\partial h(r,U)}{\partial r} \quad \text{with} \quad \mathbb{P}\left(\lim_{r\downarrow c} \frac{\partial h(r,U)}{\partial r} > \lim_{r\uparrow c} \frac{\partial h(r,U)}{\partial r}\right) > 0$$

  then the target parameter has a similar interpretation to the **LATE**

  - Weights $\omega(c,U)$ are **non-zero** for agents whom the kink **induces** to choose more $D$

# Dong (2018)

- This class of discontinuity designs is based on a working paper by Dong (2018)
- $Y \in \mathbb{R}$ is a scalar **outcome** of interest, $D \in \{0, 1\}$ is a **binary treatment**
- $D$ and $Y$ are linked by **potential outcomes** $Y(0), Y(1)$
- $R \in \mathbb{R}$ is a continuously distributed **running variable**
- Suppose that **compliance** with the treatment is **one-sided**, so that

$$D \times (\mathbb{I}[R \geq c] - \mathbb{I}[R < c]) \geq 0 \qquad \text{with probability one}$$

where $c \in \mathbb{R}$. $D = 1$ is **not available** to agents for which $R < c$

# Dong (2018)

- For clearer intuition, construct a **binary instrumental variable** $Z \equiv \mathbb{I}[R \geq c]$
- Temporarily define the **propensity score** as $p(Z) \equiv \mathbb{P}(D = 1|Z)$
- **One-sided noncompliance** implies that

$$\begin{aligned} p(1) \equiv \mathbb{P}(D = 1|Z = 1) &= \mathbb{P}(D = 1|R \geq c) \\ &\geq \mathbb{P}(D = 1|R < c) = \mathbb{P}(D = 1|Z = 0) \equiv p(0) = 0 \end{aligned}$$

# Dong (2018)

- As usual, denote **potential treatments** with $D(z)$, $z \in \{0, 1\}$
- Because $p(0) = 0$, **always-takers** and **defiers** can be safely assumed away
- **Agent types** can be characterized as follows:

$$T \equiv \begin{cases} n, & \text{if } D(0) = D(1) = 0 \\ cp, & \text{if } D(0) = 0 \text{ and } D(1) = 1 \end{cases}$$

- The definition of $T$ **completely partitions** the set of realizations of $(D(0), D(1))$
- $D(1) \geq D(0)$ almost surely, so the Imbens and Angrist **monotonicity** condition holds

# Dong (2018)

- The propensity score was previously defined as $p(Z) \equiv \mathbb{P}(D = 1|Z)$
- This definition is **unnecessarily restrictive** because $R$ may predict the treatment state
- Define the **propensity score** as $p(Z, R) \equiv \mathbb{P}(D = 1|Z, R)$
  - $Z = \mathbb{I}[R \geq c]$ is a **deterministic function** of $R$, so $p(Z, R) = p(R) \equiv \mathbb{P}(D = 1|R)$
- $Z$ being a **function** of $R$ additionally implies the **conditional exogeneity** assumption

$$(Y(0), Y(1), D(0), D(1)) \perp\!\!\!\perp Z|R = r \quad \forall r$$

- Vytlacil (2002)'s **equivalence result** can be used to derive a **nonparametric Roy** model
- This model meets all of the **Imbens and Angrist** assumptions

# Dong (2018)

- Let $I$ denote any open or closed interval and define a continuous random variable $V_I$
- $V_I \perp\!\!\!\perp (Y(0), Y(1), D(0), D(1), Z)$ is uniformly distributed over $I$
- Define a **random variable $U$** conditional on each element in the support of $R$:

$$(U|R = r) \equiv \mathbb{I}[T = cp, R = r] \, V_{[0,p(r)]} + \mathbb{I}[T = n, R = r] \, V_{(p(r),1]}$$

- For every $r$, $(U|R = r)$ is a **continuously distributed** random variable with support $[0, 1]$
- $U$ can be used to construct the **selection model**

$$D = D(r) = \mathbb{I}[U \leq p(r)] \quad \forall \, r$$

where $D(R) = D(R, Z) = D(Z)$ indicates the **potential treatment** associated with $R$

# Dong (2018)

- For each element in the support of $R$, the **propensity score** can be expressed as

$$
\begin{aligned}
p(r) &\equiv \mathbb{P}(D=1|R=r) \\
&= \mathbb{P}(U \leq p(R)|R=r) \qquad (D(R) = \mathbb{I}[U \leq p(R)]) \\
&= \mathbb{P}(U \leq p(r)) \\
&= \mathbb{P}(F_U(U) \leq F_U(p(r))) \qquad (U \text{ is continuous}) \\
&= \mathbb{P}\left(\tilde{U} \leq F_U(p(r))\right) \qquad \text{with } \tilde{U} \sim \mathcal{U}[0,1] \\
&= F_U(p(r))
\end{aligned}
$$

- Thus, the **selection model** can be written as

$$
D(r) = \mathbb{I}[U \leq p(r)] = \mathbb{I}[F_U(U) \leq F_U(p(r))] = \mathbb{I}\left[\tilde{U} \leq p(r)\right] \quad \forall\, r
$$

where $\tilde{U} \sim \mathcal{U}[0,1]$. For ease of notation, $\tilde{U}$ is denoted with as $U$

# Dong (2018)

This nonparametric Roy model is **nested** in the Imbens and Angrist model:

1. $(Y(0), Y(1), D(0), D(1)) \perp\!\!\!\perp Z|R = r \; \forall \, r$, because $Z$ is a deterministic function of $R$

2. $D(1) \geq D(0)$ **almost surely**, because compliance with the treatment is one-sided

3. $U \perp\!\!\!\perp Z|R = r \; \forall \, r$, because $U$ is a function of $D(Z)$ and the completely idiosyncratic $V_I$

4. **Potential treatments** conditional on any $R = r$ are equal in the two models:

$$\begin{aligned}
r \geq c \text{ and } T = n &\implies (U|R = r) = V_{(p(r),1]} > p(r) \implies D(r) = 0 \\
r \geq c \text{ and } T = cp &\implies (U|R = r) = V_{[0,p(r)]} \leq p(r) \implies D(r) = 1 \\
r < c \text{ and } T = n &\implies (U|R = r) = V_{(p(r),1]} = V_{(0,1]} > p(r) = 0 \implies D(r) = 0 \\
r < c \text{ and } T = cp &\implies (U|R = r) = V_{[0,p(r)]} = V_{[0,0]} = 0 = p(r) \implies D(r) = 0
\end{aligned}$$

where the last scenario is a knife-edge case (see footnote 3 in the MTE supplement)

# Dong (2018)

- Let us make a few additional assumptions:
  - $\mathbb{E}\left[Y(d)|R=r, U=u\right]$ is a **continuously differentiable** function of $(r, u)$ for $d = 0, 1$
  - The propensity score, $p(r)$, is **continuous** and **differentiable** at $r = c$
  - The **derivative** of $p(r)$ is **discontinuous** at $r = c$
- This setting is similar to a **fuzzy regression kink design**
- But the treatment is **binary** as opposed to **continuously distributed**
- A target parameter may be identified with a standard argument from the **MTE framework**...

# Dong (2018)

As usual, let us express the **mean** of $\boldsymbol{Y}$ conditional on $D = 1$ and $R = r$:

$$\begin{aligned}
\mathbb{E}[Y|D=1,R=r] &= \mathbb{E}[DY(1)+(1-D)Y(0)|D=1,R=r] \\
&= \mathbb{E}[Y(1)|D=1,R=r] \\
&= \mathbb{E}[Y(1)|U \leq p(R), R=r] \qquad (D = \mathbb{I}[U \leq p(R)]) \\
&= \mathbb{E}[Y(1)|U \leq p(r), R=r] \\
&= \frac{1}{p(r)} \int_0^{p(r)} \mathbb{E}[Y(1)|U=u,R=r]\, du \qquad (U \sim \mathcal{U}[0,1])
\end{aligned}$$

Analogously, the **mean** of $\boldsymbol{Y}$ conditional on $D = 0$ and $R = r$ is

$$\mathbb{E}[Y|D=0,R=r] = \frac{1}{1-p(r)} \int_{p(r)}^{1} \mathbb{E}[Y(0)|U=u,R=r]\, du$$

# Dong (2018)

The **Law of Iterated Expectations** further implies that

$$\begin{aligned}
\mathbb{E}[Y|R=r] &= \mathbb{E}[Y|D=1,R=r]\times\mathbb{P}(D=1|R=r)\\
&+\mathbb{E}[Y|D=0,R=r]\times\mathbb{P}(D=0|R=r)\\
&=\mathbb{E}[Y|D=1,R=r]\times p(r)\\
&+\mathbb{E}[Y|D=0,R=r]\times(1-p(r))\\
&=\int_0^{p(r)}\mathbb{E}[Y(1)|U=u,R=r]\,du+\int_{p(r)}^1\mathbb{E}[Y(0)|U=u,R=r]\,du
\end{aligned}$$

# Dong (2018)

**Leibniz's rule** implies that the **derivative** of $\mathbb{E}[Y|R=r]$ with respect to $R$, at $r \neq c$, is

$$\begin{aligned}
\frac{\partial \mathbb{E}[Y|R=r]}{\partial r} &= \frac{\partial p(r)}{\partial r} \times \mathbb{E}[Y(1)|U=p(r), R=r] + \int_0^{p(r)} \frac{\partial \mathbb{E}[Y(1)|U=p(r), R=r]}{\partial r} du \\
&- \frac{\partial p(r)}{\partial r} \times \mathbb{E}[Y(0)|U=p(r), R=r] + \int_{p(r)}^{1} \frac{\partial \mathbb{E}[Y(0)|U=p(r), R=r]}{\partial r} du \\
&= \frac{\partial p(r)}{\partial r} \times \mathbb{E}[Y(1)-Y(0)|U=p(r), R=r] \\
&+ \int_0^{p(r)} \frac{\partial \mathbb{E}[Y(1)|U=p(r), R=r]}{\partial r} du + \int_{p(r)}^{1} \frac{\partial \mathbb{E}[Y(0)|U=p(r), R=r]}{\partial r} du
\end{aligned}$$

where $\mathbb{E}[Y(1)-Y(0)|U=p(r), R=r]$ is the $\mathbf{MTE}$ of $D$ on $Y$ at $U=p(r)$ and $R=r$

# Dong (2018)

- By assumption, the **derivative** of the **propensity score** is **discontinuous** at $r = c$

- Take the **limits** of $\frac{\partial \mathbb{E}[Y|R=r]}{\partial r}$ as $r \downarrow c$ and $r \uparrow c$:

$$\lim_{r \downarrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} = \lim_{r \downarrow c} \frac{\partial p(r)}{\partial r} \times \mathbb{E}[Y(1) - Y(0)|R = c, U = p(c)] + \int_0^{p(c)} \frac{\partial \mathbb{E}[Y(1)|R = c, U = u]}{\partial r} du + \int_{p(c)}^1 \frac{\partial \mathbb{E}[Y(0)|R = c, U = u]}{\partial r} du$$

$$\lim_{r \uparrow c} \frac{\partial \mathbb{E}[Y|R=r]}{\partial r} = \lim_{r \uparrow c} \frac{\partial p(r)}{\partial r} \times \mathbb{E}[Y(1) - Y(0)|R = c, U = p(c)] + \int_0^{p(c)} \frac{\partial \mathbb{E}[Y(1)|R = c, U = u]}{\partial r} du + \int_{p(c)}^1 \frac{\partial \mathbb{E}[Y(0)|R = c, U = u]}{\partial r} du$$

# Dong (2018)

These limits can be **differenced out**:

$$\lim_{r\downarrow c}\frac{\partial\mathbb{E}[Y|R=r]}{\partial r}-\lim_{r\uparrow c}\frac{\partial\mathbb{E}[Y|R=r]}{\partial r}=\left(\lim_{r\downarrow c}\frac{\partial p(r)}{\partial r}-\lim_{r\uparrow c}\frac{\partial p(r)}{\partial r}\right)\times \mathrm{MTE}\left(U=p(c),R=c\right)$$

Rearranging terms, the **Marginal Treatment Effect** of $D$ on $Y$ at $U=p(c)$ and $R=c$ is

$$\mathbb{E}\left[Y(1)-Y(0)|U=p(c),R=c\right]=\frac{\lim_{r\downarrow c}\frac{\partial\mathbb{E}[Y|R=r]}{\partial r}-\lim_{r\uparrow c}\frac{\partial\mathbb{E}[Y|R=r]}{\partial r}}{\lim_{r\downarrow c}\frac{\partial p(r)}{\partial r}-\lim_{r\uparrow c}\frac{\partial p(r)}{\partial r}}$$

This point-identified parameter is the $\mathrm{MTE}$ for agents

- Whose realization of the **running variable** is $R=c$, and
- Who are **at the margin** of choosing $D=1$ if $R=c$

# Summary

- RD designs with **multiple cutoffs** or **multiple running variables** typically require an empiricist to choose whether to estimate **cutoff-specific effects** or **normalize and pool**

- **Regression probability kink designs** allow a researcher to derive a nonparametric Roy model and point identify a "local" (in terms of $R$) **marginal treatment effect**