Noise-Induced Randomization in Regression Discontinuity Designs

Dean Eckles, Nikolaos Ignatiadis, Stefan Wager, Han Wu

Presented by: Sai Zhang

November 18, 2022

Outline

Introduction

RD Identification

$$\underbrace{Z_i}_{\text{running variable}}$$

RD Identification

$\underbrace{Z_i}_{\text{running variable}}$ $\underbrace{W_i}_{\text{treatment}}$

RD Identification

$$\underbrace{Z_i}_{\text{running variable}} \xrightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}}$$

RD Identification

$$\underbrace{Z_i}_{\text{running variable}} \xrightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}}$$

RD Identification

$$\underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

RD Identification

$$\underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

---

test scores admission outcomes

RD Identification

$$\underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

test scores	admission	outcomes
test results	medication	outcomes

RD Identification: Continuity Argument

For potential outcomes $\{Y_i(0), Y_i(1)\}$: $Y_i = Y_i(W_i)$, a weighted **causal effect** can be identified as

$$\tau_c = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid Z_i = c\right]$$

RD Identification: Continuity Argument

For potential outcomes $\{Y_i(0), Y_i(1)\}$: $Y_i = Y_i(W_i)$, a weighted **causal effect** can be identified as

$$
\begin{aligned}
\tau_c &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid Z_i = c\right] \\
&= \lim_{z \downarrow c} \mathbb{E}\left[Y \mid Z = z\right] - \lim_{z \uparrow c} \mathbb{E}\left[Y \mid Z = z\right]
\end{aligned}
$$

RD Identification: Continuity Argument

For potential outcomes $\{Y_i(0), Y_i(1)\}$: $Y_i = Y_i(W_i)$, a weighted **causal effect** can be identified as

$$\begin{aligned}\tau_c &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid Z_i = c\right] \\ &= \lim_{z \downarrow c} \mathbb{E}\left[Y \mid Z = z\right] - \lim_{z \uparrow c} \mathbb{E}\left[Y \mid Z = z\right]\end{aligned}$$

assuming

- the conditional response functions $\mu_w(z) = \mathbb{E}\left[Y(w) \mid Z = z\right]$ are **continuous**

RD Identification: Continuity Argument

For potential outcomes $\{Y_i(0), Y_i(1)\}$: $Y_i = Y_i(W_i)$, a weighted **causal effect** can be identified as

$$\begin{aligned}\tau_c &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid Z_i = c\right] \\ &= \lim_{z \downarrow c} \mathbb{E}\left[Y \mid Z = z\right] - \lim_{z \uparrow c} \mathbb{E}\left[Y \mid Z = z\right]\end{aligned}$$

assuming

- the conditional response functions $\mu_w(z) = \mathbb{E}\left[Y(w) \mid Z = z\right]$ are **continuous**
- $\mu_w(z)$ to have a uniformly **bounded 2nd derivative** for CIs (Armstrong and Kolesár, 2018, 2020)

RD Identification: Problems of Continuity Argument

Assumption: **continuous** $\mu_w(z) = \mathbb{E}\left[Y(w) \mid Z = z\right]$

$$\tau_c = \lim_{z \downarrow c} \mathbb{E}\left[Y \mid Z = z\right] - \lim_{z \uparrow c} \mathbb{E}\left[Y \mid Z = z\right]$$

RD Identification: Problems of Continuity Argument

Assumption: **continuous** $\mu_w(z) = \mathbb{E}[Y(w) \mid Z = z]$

$$\tau_c = \lim_{z \downarrow c} \mathbb{E}[Y \mid Z = z] - \lim_{z \uparrow c} \mathbb{E}[Y \mid Z = z]$$

Where does this continuity come from?

RD Identification: Problems of Continuity Argument

Assumption: **continuous** $\mu_w(z) = \mathbb{E}[Y(w) \mid Z = z]$

$$\tau_c = \lim_{z \downarrow c} \mathbb{E}[Y \mid Z = z] - \lim_{z \uparrow c} \mathbb{E}[Y \mid Z = z]$$

Where does this continuity come from?

Lee (2008): **continuous measurement error** in the running variable by units

RD Identification: Measurement Error

$$\underbrace{U_i}_{\text{latent variable}} \qquad \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

---

	test scores	admission	outcomes
	test results	medication	outcomes

RD Identification: Measurement Error

$$\underbrace{U_i}_{\text{latent variable}} \qquad \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

ability	test scores	admission	outcomes
condition	test results	medication	outcomes

RD Identification: Measurement Error

$$\underbrace{U_i}_{\text{latent variable}} \xRightarrow{Z_i = U_i + e_i} \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

ability	test scores	admission	outcomes
condition	test results	medication	outcomes

RD Identification: Measurement Error

$$\underbrace{U_i}_{\text{latent variable}} \xRightarrow{Z_i = U_i + e_i} \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

---

ability	test scores	admission	outcomes
condition	test results	medication	outcomes

Why don't we take advantage of the **measurement error** itself for inference?

This Paper

$$\underbrace{U_i}_{\text{latent variable}} \xRightarrow{Z_i = U_i + e_i} \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

Weighted treatment effects can be estimated if the measurement error in Z_i

This Paper

$$\underbrace{U_i}_{\text{latent variable}} \xRightarrow{Z_i = U_i + e_i} \underbrace{Z_i}_{\text{running variable}} \xRightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

Weighted treatment effects can be estimated if the measurement error in Z_i

- has a **known distribution**

This Paper

$$\underbrace{U_i}_{\text{latent variable}} \xrightarrow{Z_i = U_i + e_i} \underbrace{Z_i}_{\text{running variable}} \xrightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}} \Rightarrow \underbrace{Y_i}_{\text{outcome}}$$

Weighted treatment effects can be estimated if the measurement error in Z_i

- has a **known distribution**
- is **conditionally (on U_i) independent** of potential outcomes

References I

Armstrong, T. B., & Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica, 86*(2), 655–683.

Armstrong, T. B., & Kolesár, M. (2020). Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics, 11*(1), 1–39.

Eckles, D., Ignatiadis, N., Wager, S., & Wu, H. (2020). Noise-induced randomization in regression discontinuity designs. *arXiv preprint arXiv:2004.09458.*

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics, 142*(2), 675–697.

Thank you!