

# Noise-Induced Randomization in Regression Discontinuity Designs

Dean Eckles, Nikolaos Ignatiadis, Stefan Wager, Han Wu

Presented by: Sai Zhang

November 18, 2022

# Outline

- 1 Introduction
- 2 Key Argument
- 3 Estimation
- 4 Confidence Intervals
- 5 Applications
- 6 Discussion

# Introduction

# RD Identification

$Z_i$   
running variable

# RD Identification

$Z_i$   
running variable

$W_i$   
treatment

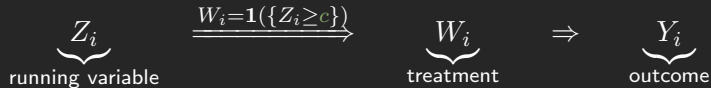
# RD Identification

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

# RD Identification

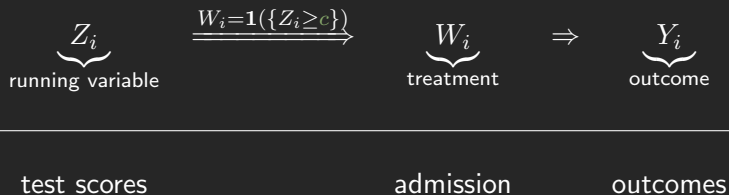
$$\underbrace{Z_i}_{\text{running variable}} \xrightarrow{W_i = \mathbf{1}(\{Z_i \geq c\})} \underbrace{W_i}_{\text{treatment}}$$

# RD Identification

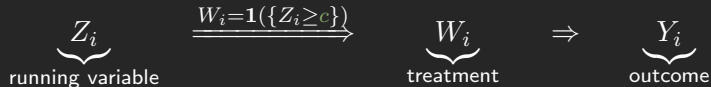




# RD Identification



# RD Identification




---

test scores  
test results

admission  
medication

outcomes  
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

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

# 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

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

# 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

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

assuming

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

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

assuming

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

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

Where does this continuity come from?

# RD Identification: Problems of Continuity Argument

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

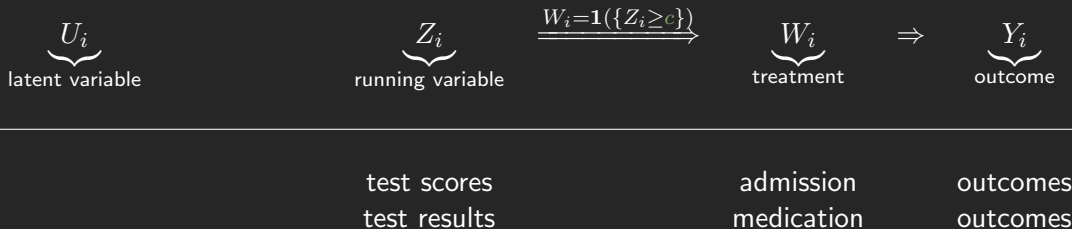
$$\tau_c = \lim_{z \downarrow c} \mathbb{E}[Y \mid Z = z] - \lim_{z \uparrow c} \mathbb{E}[Y \mid Z = z]$$

Where does this continuity come from?

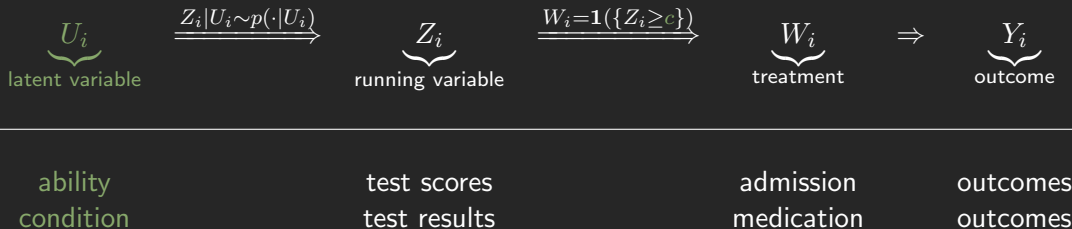
Lee (2008): continuous measurement error in the running variable by units



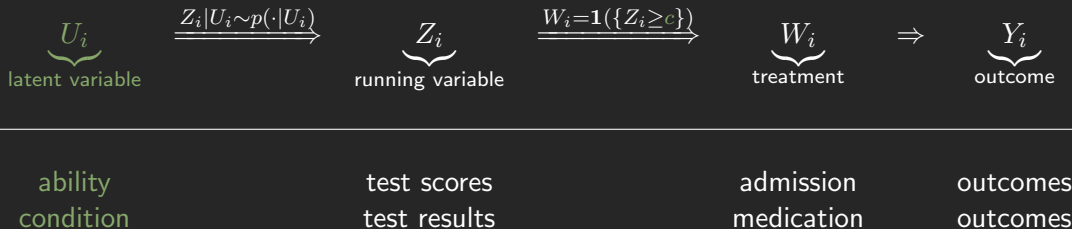
# RD Identification: Measurement Error



# RD Identification: Measurement Error

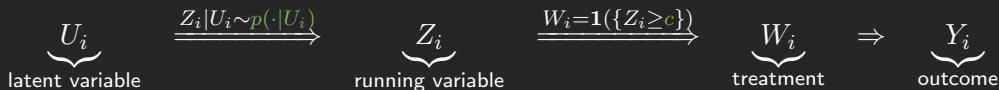


# RD Identification: Measurement Error



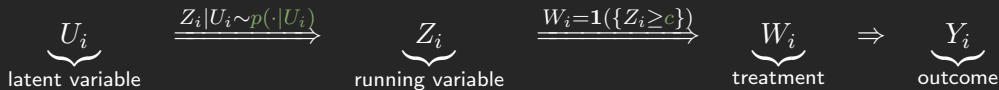
Why don't we take advantage of the measurement error itself for inference?

# This Paper



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

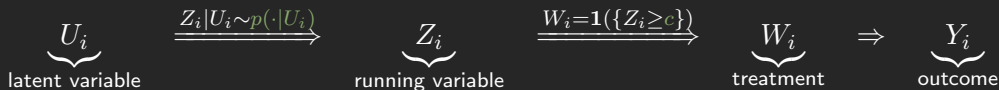
# This Paper



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

- has a known distribution

# This Paper

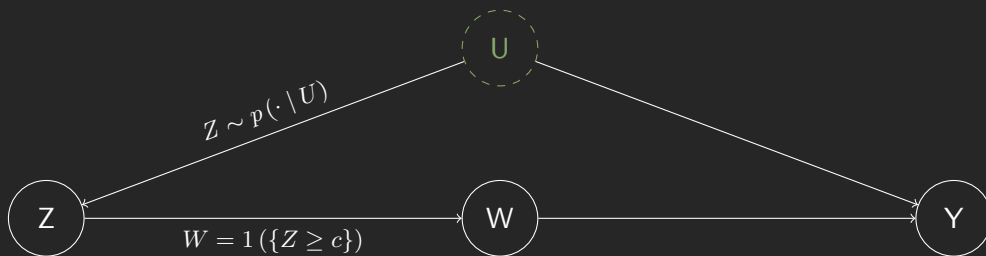


Weighted treatment effects can be estimated if the measurement error in  $Z_i$

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

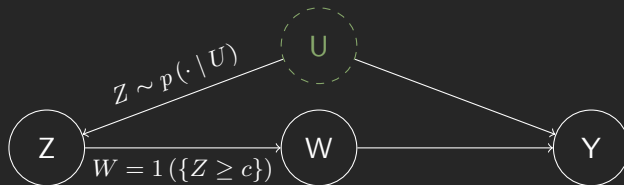
# Key Argument

# Sharp RD Design with A Noisy Running Variable





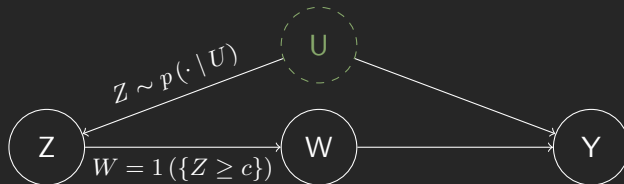
# Sharp RD Design with A Noisy Running Variable



## Assumption 1: Sharp RD design

- I.I.D. samples  $\{Y_i(0), Y_i(1), Z_i\} \in \mathbb{R}^3, i = 1, \dots, n$
- treatment assignment:  $W_i = 1(\{Z_i \geq c\})$ , where  $c \in \mathbb{R}$  is the cutoff
- observation:  $\{Y_i, Z_i\}$  where  $Y_i = Y_i(W_i)$

# Sharp RD Design with A Noisy Running Variable

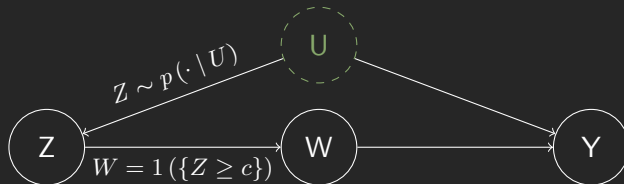


## Assumption 2: Noisy running variable

$$Z_i \mid U_i \sim p(\cdot \mid U_i)$$

where  $p(\cdot \mid \cdot)$  is a **known** conditional density w.r.t. to a measure  $\lambda$ , the latent variable  $U_i$  has an **unknown** distribution  $G$

# Sharp RD Design with A Noisy Running Variable

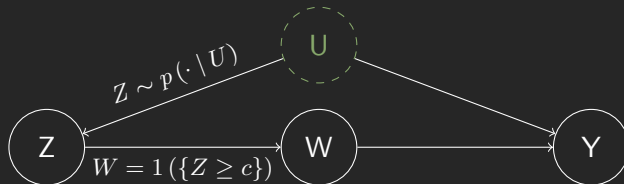


## Assumption 2: Noisy running variable

$$Z_i | U_i \sim \mathcal{N}(U_i, \nu^2), \nu > 0$$

where  $p(\cdot | \cdot)$  is a **known** conditional density w.r.t. to a measure  $\lambda$ , the latent variable  $U_i$  has an **unknown** distribution  $G$

# Sharp RD Design with A Noisy Running Variable

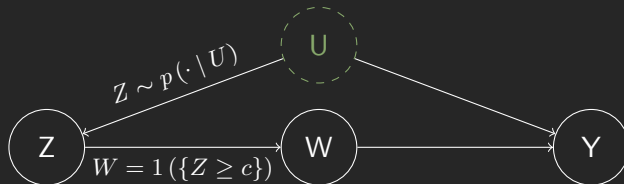


## Assumption 2: Noisy running variable

$$Z_i | U_i \sim \text{Binomial}(K, U_i), K \in \mathbb{N}$$

where  $p(\cdot | \cdot)$  is a **known** conditional density w.r.t. to a measure  $\lambda$ , the latent variable  $U_i$  has an **unknown** distribution  $G$

# Sharp RD Design with A Noisy Running Variable

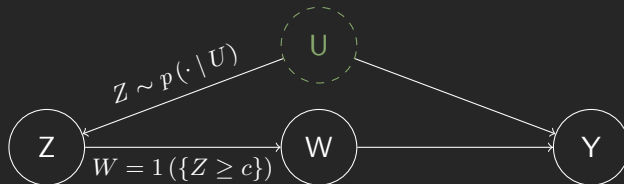


## Assumption 3: Exogeneity

$$[\{Y_i(0), Y_i(1)\} \perp Z_i] \mid U_i$$

which implies  $\mathbb{E}[Y_i \mid U_i, Z_i] = \alpha_{(W_i)}(u)$

# Sharp RD Design with A Noisy Running Variable

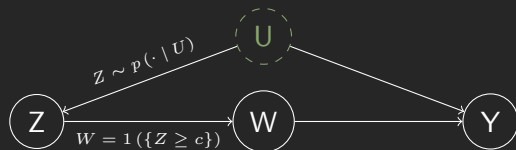


## Assumption 3: Exogeneity

$$[\{Y_i(0), Y_i(1)\} \perp Z_i] \mid U_i$$

which implies  $\mathbb{E}[Y_i \mid U_i, Z_i] = \alpha_{(W_i)}(u)$ , where  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) \mid U_i = u]$  is the **response functions** for the potential outcomes conditional on the latent variable  $u$

# Sharp RD Design with A Noisy Running Variable

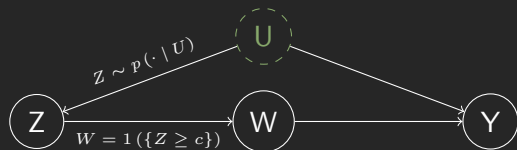


A1 Sharp RD

A2 Noisy  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 Exogeneity:  
 $[\{Y_i(0), Y_i(1)\} \perp Z_i | U_i]$

# Sharp RD Design with A Noisy Running Variable



A1 **Sharp** RD

A2 **Noisy**  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 **Exogeneity**:

$$[\{Y_i(0), Y_i(1)\} \perp Z_i | U_i]$$

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z | u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$



# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

$$\blacksquare \mathbb{E}[Y^2], \mathbb{E}[\gamma_-(Z)^2], \mathbb{E}[\gamma_+(Z)^2] < \infty$$

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

- $\mathbb{E}[Y^2], \mathbb{E}[\gamma_-(Z)^2], \mathbb{E}[\gamma_+(Z)^2] < \infty$
- $\gamma_+(\cdot), \gamma_-(\cdot)$  are weighting functions s.t.

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

- $\mathbb{E}[Y^2], \mathbb{E}[\gamma_-(Z)^2], \mathbb{E}[\gamma_+(Z)^2] < \infty$
- $\gamma_+(\cdot), \gamma_-(\cdot)$  are weighting functions s.t.
  - $\gamma_+(z) = 0$  for  $z < c$ : assign non-zero weights only to **treated** units

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

- $\mathbb{E}[Y^2], \mathbb{E}[\gamma_-(Z)^2], \mathbb{E}[\gamma_+(Z)^2] < \infty$
- $\gamma_+(\cdot), \gamma_-(\cdot)$  are weighting functions s.t.
  - $\gamma_+(z) = 0$  for  $z < c$ : assign non-zero weights only to **treated** units
  - $\gamma_-(z) = 0$  for  $z \geq c$ : assign non-zero weights only to **control** units

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

Proof:

$$\mathbb{E}[\gamma_+(Z)Y | U]$$

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

Proof:

$$\mathbb{E}[\gamma_+(Z)Y | U] = \mathbb{E}[\gamma_+(Z)Y \cdot \mathbf{1}(\{Z \geq c\}) | U]$$

# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

Proof:

$$\begin{aligned} \mathbb{E}[\gamma_+(Z)Y | U] &= \mathbb{E}[\gamma_+(Z)Y \cdot \mathbf{1}(\{Z \geq c\}) | U] \\ &= \mathbb{E}[\gamma_+(Z)Y(1) \cdot \mathbf{1}(\{Z \geq c\}) | U] \end{aligned}$$



# Sharp RD Design with A Noisy Running Variable

## Proposition 1

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

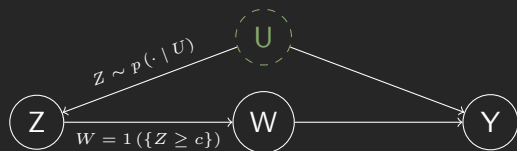
$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

Proof:

$$\begin{aligned} \mathbb{E}[\gamma_+(Z)Y | U] &= \mathbb{E}[\gamma_+(Z)Y \cdot \mathbf{1}(\{Z \geq c\}) | U] \\ &= \mathbb{E}[\gamma_+(Z)Y(1) \cdot \mathbf{1}(\{Z \geq c\}) | U] \\ &= \underbrace{\mathbb{E}[Y(1) | U]}_{=\alpha_{(1)}(U)} \cdot \underbrace{\mathbb{E}[\gamma_+(Z)\mathbf{1}(\{Z \geq c\}) | U]}_{=\mathbb{E}[\gamma_+(Z)|U]=\int \gamma_+(z)p(z|U)d\lambda(z)=h(U, \gamma_+)} \end{aligned}$$

# Sharp RD Design with A Noisy Running Variable



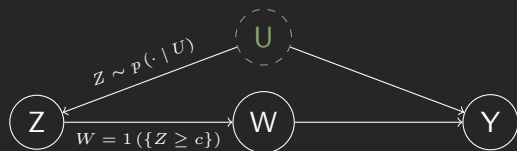
A1 Sharp RD

A2 Noisy  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 Exogeneity:

$$[\{Y_i(0), Y_i(1)\} \perp Z_i] | U_i$$

# Sharp RD Design with A Noisy Running Variable



A1 Sharp RD

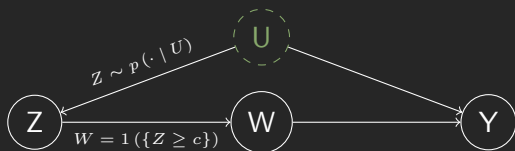
A2 Noisy  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 Exogeneity:

$$[\{Y_i(0), Y_i(1)\} \perp Z_i] | U_i$$

- No need to know  $G$  (distribution of  $U$ )
- Need to know  $p(z | u)$  (conditional distribution of the noise)

# Sharp RD Design with A Noisy Running Variable



A1 Sharp RD

A2 Noisy  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 Exogeneity:

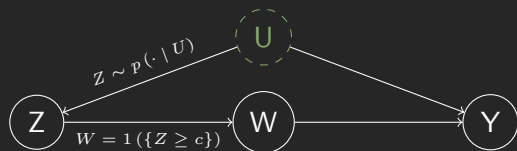
$$[\{Y_i(0), Y_i(1)\} \perp Z_i] | U_i$$

■ No need to know  $G$  (distribution of  $U$ )

■ Need to know  $p(z | u)$  (conditional distribution of the noise)

- test-retest data, prior modelling of responses to tests, physical model of the measurement device, biomedical knowledge, etc.

# Sharp RD Design with A Noisy Running Variable



A1 Sharp RD

A2 Noisy  $Z_i$ :  $Z_i | U_i \sim p(\cdot | U_i)$

A3 Exogeneity:

$$[\{Y_i(0), Y_i(1)\} \perp Z_i] | U_i$$

■ No need to know  $G$  (distribution of  $U$ )

■ Need to know  $p(z | u)$  (conditional distribution of the noise)

- test-retest data, prior modelling of responses to tests, physical model of the measurement device, biomedical knowledge, etc.
- still valid when **underestimating** the true noise level

# Estimation

# Estimation of Weighted Treatment Effects

## Proposition: The Key Argument

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E} [\gamma_+ (Z) Y] = \mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_+)] , \quad \mathbb{E} [\gamma_- (Z) Y] = \mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_-)]$$

where  $h (u, \gamma) := \int \gamma (z) p (z | u) d\lambda (z)$ ,  $\alpha_{(w)} (u) = \mathbb{E} [Y_i (w) | U_i = u]$

ratio-form estimators:

$$\begin{aligned} \hat{\tau}_\gamma &= \hat{\mu}_{\gamma,+} \\ &= \frac{\sum_i \gamma_+ (Z_i) Y_i}{\sum_i \underbrace{\gamma_+ (Z_i)}} \end{aligned}$$

$$\begin{aligned} & - \hat{\mu}_{\gamma,-} \\ &= \frac{\sum_i \gamma_- (Z_i) Y_i}{\sum_i \underbrace{\gamma_- (Z_i)}} \end{aligned}$$

# Estimation of Weighted Treatment Effects

## Proposition: The Key Argument

Let  $\gamma_+(\cdot), \gamma_-(\cdot)$  be measurable functions of  $Z$ , then under A1-A3:

$$\mathbb{E}[\gamma_+(Z)Y] = \mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)], \quad \mathbb{E}[\gamma_-(Z)Y] = \mathbb{E}[\alpha_{(0)}(U)h(U, \gamma_-)]$$

where  $h(u, \gamma) := \int \gamma(z)p(z|u)d\lambda(z)$ ,  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

ratio-form estimators:

$$\begin{aligned} \hat{\tau}_\gamma &= \hat{\mu}_{\gamma,+} \\ &= \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \underbrace{\gamma_+(Z_i)}_{\gamma_+(z)=0, z < c}} \end{aligned}$$

$$\begin{aligned} & - \hat{\mu}_{\gamma,-} \\ &= \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \underbrace{\gamma_-(Z_i)}_{\gamma_-(z)=0, z \geq c}} \end{aligned}$$



# Estimation of Weighted Treatment Effects

Ratio-form estimators:

$$\hat{\tau}_{\gamma} = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

What's the weighted treatment effects to conduct inference for?

# Estimation of Weighted Treatment Effects

Ratio-form estimators:

$$\hat{\tau}_{\gamma} = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

What's the weighted treatment effects to conduct inference for?

$$\tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u), w(\cdot) \geq 0$$

# Estimation of Weighted Treatment Effects

Ratio-form estimators:

$$\hat{\tau}_{\gamma} = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

What's the weighted treatment effects to conduct inference for?

$$\tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u), w(\cdot) \geq 0$$

where  $\tau(u)$  (**Conditional Average Treatment Effects**) is

$$\tau(u) = \mathbb{E}[Y_i(1) - Y_i(0) \mid U_i = u] = \alpha_{(1)}(u) - \alpha_{(0)}(u)$$

# Weighted Treatment Effects: Example

$$\tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u), w(\cdot) \geq 0$$

where  $\tau(u)$  (**CATE**) is  $\tau(u) = \mathbb{E}[Y_i(1) - Y_i(0) \mid U_i = u] = \alpha_{(1)}(u) - \alpha_{(0)}(u)$

■ RD paramater:

$$\begin{aligned} \tau_c &= \mathbb{E}[Y_i(1) - Y_i(0) \mid Z_i = c] = \mathbb{E}[\tau(U_i) \mid Z_i = c] \\ &= \int \frac{p(c \mid u)}{\int p(c \mid u) dG(u)} \tau(u) dG(u) \end{aligned}$$

## Weighted Treatment Effects: Example

$$\tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u), w(\cdot) \geq 0$$

where  $\tau(u)$  (**CATE**) is  $\tau(u) = \mathbb{E}[Y_i(1) - Y_i(0) \mid U_i = u] = \alpha_{(1)}(u) - \alpha_{(0)}(u)$

■ RD paramater:

$$\begin{aligned} \tau_c &= \mathbb{E}[Y_i(1) - Y_i(0) \mid Z_i = c] = \mathbb{E}[\tau(U_i) \mid Z_i = c] \\ &= \int \frac{p(c \mid u)}{\underbrace{\int p(c \mid u) dG(u)}_{=f_G(c)}} \tau(u) dG(u) \end{aligned}$$

# Asymptotic Bias

## Theorem: Asymptotic Limit of $\hat{\tau}_\gamma$

$$\hat{\tau}_\gamma = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

# Asymptotic Bias

## Theorem: Asymptotic Limit of $\hat{\tau}_\gamma$

$$\hat{\tau}_\gamma = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)} \xrightarrow{p} \frac{\mathbb{E}[\gamma_+(Z)Y]}{\mathbb{E}[\gamma_+(Z)]} - \frac{\mathbb{E}[\gamma_-(Z)Y]}{\mathbb{E}[\gamma_-(Z)]}$$

# Asymptotic Bias

## Theorem: Asymptotic Limit of $\hat{\tau}_\gamma$

$$\begin{aligned}
 \hat{\tau}_\gamma = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} &= \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)} \\
 &\xrightarrow{p} \frac{\mathbb{E}[\gamma_+(Z)Y]}{\mathbb{E}[\gamma_+(Z)]} - \frac{\mathbb{E}[\gamma_-(Z)Y]}{\mathbb{E}[\gamma_-(Z)]} \\
 \text{(Prop.1)} &= \frac{\mathbb{E}[\alpha_{(1)}(U) h(U, \gamma_+)]}{\mathbb{E}[h(U, \gamma_+)]} - \frac{\mathbb{E}[\alpha_{(0)}(U) h(U, \gamma_-)]}{\mathbb{E}[h(U, \gamma_-)]} = \mu_{\gamma,+} - \mu_{\gamma,-} \equiv \theta_\gamma
 \end{aligned}$$

where

$$h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z), \quad \alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$$



# Asymptotic Bias

## Theorem: Asymptotic Limit of $\hat{\tau}_\gamma$

$$\hat{\tau}_\gamma = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)} \xrightarrow{p} \frac{\mathbb{E} [\alpha_{(1)}(U) h(U, \gamma_+)]}{\mathbb{E} [h(U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)}(U) h(U, \gamma_-)]}{\mathbb{E} [h(U, \gamma_-)]} = \mu_{\gamma,+} - \mu_{\gamma,-} \equiv \theta_\gamma$$

How biased is this asymptotic limit? Comparing to

$$\tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u), w(\cdot) \geq 0$$

# Asymptotic Bias: Decomposition

$$a\text{Bias} \left[ \gamma_{\pm}, \tau_w; \alpha_{(0)}(\cdot), \tau(\cdot), G \right] = \theta_{\gamma} - \tau_w$$

# Asymptotic Bias: Decomposition

$$a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] = \theta_{\gamma} - \tau_w$$

$$= \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_+)]}{\mathbb{E} [h (U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_-)]}{\mathbb{E} [h (U, \gamma_-)]} - \int \frac{w (u)}{\mathbb{E}_G [w (U)]} \tau (u) \, dG (u)$$

# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_+)]}{\mathbb{E} [h (U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_-)]}{\mathbb{E} [h (U, \gamma_-)]} - \int \frac{w (u)}{\mathbb{E}_G [w (U)]} \tau (u) dG (u) \\
 &= \int \left( \frac{h (u, \gamma_+)}{\mathbb{E}_G [h (U, \gamma_+)]} \right) \alpha_{(1)} (u) dG (u) - \int \left( \frac{h (u, \gamma_-)}{\mathbb{E}_G [h (U, \gamma_-)]} \right) \alpha_{(0)} (u) dG (u) \\
 &\quad - \int \frac{w (u)}{\mathbb{E}_G [w (U)]} \tau (u) dG (u)
 \end{aligned}$$

# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_+)]}{\mathbb{E} [h (U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_-)]}{\mathbb{E} [h (U, \gamma_-)]} - \int \frac{w (u)}{\mathbb{E}_G [w (U)]} \tau (u) dG (u) \\
 &= \int \left( \frac{h (u, \gamma_+)}{\mathbb{E}_G [h (U, \gamma_+)]} \right) \alpha_{(1)} (u) dG (u) - \int \left( \frac{h (u, \gamma_-)}{\mathbb{E}_G [h (U, \gamma_-)]} \right) \alpha_{(0)} (u) dG (u) \\
 &\quad - \int \frac{w (u)}{\mathbb{E}_G [w (U)]} \tau (u) dG (u)
 \end{aligned}$$

Remember?  $\tau(u)$  **(Conditional Average Treatment Effects)** is

$$\tau (u) = \mathbb{E} [Y_i (1) - Y_i (0) \mid U_i = u] = \alpha_{(1)} (u) - \alpha_{(0)} (u) \Rightarrow \boxed{\alpha_{(1)} (u) = \tau (u) + \alpha_{(0)} (u)}$$

# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)}(\cdot), \tau(\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} \right) \underbrace{\alpha_{(1)}(u)}_{=\tau(u) + \alpha_{(0)}(u)} dG(u) \\
 &\quad - \int \left( \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) - \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u)
 \end{aligned}$$

# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} \right) \underbrace{\alpha_{(1)}(u)}_{=\tau(u) + \alpha_{(0)}(u)} dG(u) \\
 &\quad - \int \left( \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) - \int \frac{w(u)}{\mathbb{E}_G [w(U)]} \tau(u) dG(u) \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) \\
 &\quad + \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G [w(U)]} \right) \tau(u) dG(u)
 \end{aligned}$$

# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} \right) \underbrace{\alpha_{(1)}(u)}_{=\tau(u) + \alpha_{(0)}(u)} dG(u) \\
 &\quad - \int \left( \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) - \int \frac{w(u)}{\mathbb{E}_G [w(U)]} \tau(u) dG(u) \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) \\
 &\quad + \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G [w(U)]} \right) \tau(u) dG(u)
 \end{aligned}$$

**Confounding bias**



# Asymptotic Bias: Decomposition

$$\begin{aligned}
 a\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_{(0)} (\cdot), \tau (\cdot), G] &= \theta_{\gamma} - \tau_w \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} \right) \underbrace{\alpha_{(1)}(u)}_{=\tau(u) + \alpha_{(0)}(u)} dG(u) \\
 &\quad - \int \left( \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) - \int \frac{w(u)}{\mathbb{E}_G [w(U)]} \tau(u) dG(u) \\
 &= \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G [h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) \\
 &\quad + \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G [h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G [w(U)]} \right) \tau(u) dG(u)
 \end{aligned}$$

**Confounding bias**

**CATE heterogeneity bias**

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**Confounding bias**

**CATE heterogeneity bias**

How to minimize them?

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**Confounding bias**

**CATE heterogeneity bias**

How to minimize them?

- **Confounding bias:**  $h(\cdot, \gamma_+) \approx h(\cdot, \gamma_-)$

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

**Confounding bias**

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**CATE heterogeneity bias**

How to minimize them?

- **Confounding bias:**  $h(\cdot, \gamma_+) \approx h(\cdot, \gamma_-)$  where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$   
How well the units are **balanced** via the latent variable  $u$

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

**Confounding bias**

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**CATE heterogeneity bias**

How to minimize them?

- **Confounding bias:**  $h(\cdot, \gamma_+) \approx h(\cdot, \gamma_-)$  where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$   
How well the units are **balanced** via the latent variable  $u$
- **CATE heterogeneity bias:**

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

**Confounding bias**

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**CATE heterogeneity bias**

How to minimize them?

- **Confounding bias:**  $h(\cdot, \gamma_+) \approx h(\cdot, \gamma_-)$  where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$   
How well the units are **balanced** via the latent variable  $u$
- **CATE heterogeneity bias:**
  - $\tau(u)$  being constant w.r.t.  $u$ , a constant conditional treatment effect

# Asymptotic Bias: Decomposition

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)$$

**Confounding bias**

$$\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)$$

**CATE heterogeneity bias**

How to minimize them?

- **Confounding bias:**  $h(\cdot, \gamma_+) \approx h(\cdot, \gamma_-)$  where  $h(u, \gamma) := \int \gamma(z) p(z | u) d\lambda(z)$

How well the units are **balanced** via the latent variable  $u$

- **CATE heterogeneity bias:**

- $\tau(u)$  being constant w.r.t.  $u$ , a constant conditional treatment effect
- $h(u, \gamma_+) = w(u), \forall u$ , an absolutely “correct” weighting function

# Confidence Intervals



# Asymptotic Normality

$$\hat{\tau} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

$$\hat{\tau}_\gamma \xrightarrow{p} \theta_\gamma = \frac{\mathbb{E} [\alpha_{(1)}(U) h(U, \gamma_+)]}{\mathbb{E} [h(U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)}(U) h(U, \gamma_-)]}{\mathbb{E} [h(U, \gamma_-)]}$$

# Asymptotic Normality

## Theorem: Asymptotic Normality of $\hat{\tau}$

Suppose the sequence of weighting kernels  $\gamma_+^{(n)}$  and  $\gamma_-^{(n)}$  is deterministic, and  $\exists \beta \in (0, \frac{1}{2})$ ,  $C, C' > 0$  s.t.  $\forall n$  large enough:

$$\sup_z \left| \gamma_\diamond^{(n)}(z) \right| < C n^\beta \mathbb{E} \left[ \gamma_\diamond^{(n)}(Z_i) \right] \quad \sup_u \left| h(u, \gamma_\diamond^{(n)}) \right| < C' \mathbb{E} \left[ \gamma_\diamond^{(n)}(Z_i) \right], \quad \diamond = \{+, -\}$$

Then

$$\frac{\sqrt{n} (\hat{\tau}_\gamma - \theta_\gamma)}{\sqrt{V_\gamma}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where

$$V_\gamma = \frac{\mathbb{E} \left[ \gamma_+^2(Z_i) (Y_i - \mu_{\gamma,+})^2 \right]}{\mathbb{E} [\gamma_+(Z_i)]^2} + \frac{\mathbb{E} \left[ \gamma_-^2(Z_i) (Y_i - \mu_{\gamma,-})^2 \right]}{\mathbb{E} [\gamma_-(Z_i)]^2}$$

# Asymptotic Normality

## Theorem: Asymptotic Normality of $\hat{\tau}$

Suppose the sequence of weighting kernels  $\gamma_+^{(n)}$  and  $\gamma_-^{(n)}$  is deterministic, and  $\exists \beta \in (0, \frac{1}{2})$ ,  $C, C' > 0$  s.t.  $\forall n$  large enough:  $\sup_z |\gamma_\diamond^{(n)}(z)| < Cn^\beta \mathbb{E}[\gamma_\diamond^{(n)}(Z_i)]$ ,  $\sup_u |h(u, \gamma_\diamond^{(n)})| < C' \mathbb{E}[\gamma_\diamond^{(n)}(Z_i)]$  where  $\diamond = \{+, -\}$  Then

$$\frac{\sqrt{n}(\hat{\tau}_\gamma - \theta_\gamma)}{\sqrt{V_\gamma}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where  $V_\gamma = \frac{\mathbb{E}[\gamma_+^2(Z_i)(Y_i - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2} + \frac{\mathbb{E}[\gamma_-^2(Z_i)(Y_i - \mu_{\gamma,-})^2]}{\mathbb{E}[\gamma_-(Z_i)]^2}$

Assumption:

- The response  $Y_i$  is bounded:  $Y_i \in [0, 1]$
- $\inf_z \text{Var}[Y_i | Z_i = z] > 0$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sqrt{n} [(\hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-}) - (\mu_{\gamma,+} - \mu_{\gamma,-})]}{\sqrt{\frac{\mathbb{E}[\gamma_+^2(Z_i)(Y_i - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2} + \frac{\mathbb{E}[\gamma_-^2(Z_i)(Y_i - \mu_{\gamma,-})^2]}{\mathbb{E}[\gamma_-(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sqrt{n} [(\hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-}) - (\mu_{\gamma,+} - \mu_{\gamma,-})]}{\sqrt{\frac{\mathbb{E}[\gamma_+^2(Z_i)(Y_i - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2} + \frac{\mathbb{E}[\gamma_-^2(Z_i)(Y_i - \mu_{\gamma,-})^2]}{\mathbb{E}[\gamma_-(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\frac{\sqrt{n} (\hat{\mu}_{\gamma,+} - \mu_{\gamma,+})}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sqrt{n} [(\hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-}) - (\mu_{\gamma,+} - \mu_{\gamma,-})]}{\sqrt{\frac{\mathbb{E}[\gamma_+^2(Z_i)(Y_i - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2} + \frac{\mathbb{E}[\gamma_-^2(Z_i)(Y_i - \mu_{\gamma,-})^2]}{\mathbb{E}[\gamma_-(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\frac{\sqrt{n} (\hat{\mu}_{\gamma,+} - \mu_{\gamma,+})}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\frac{\sqrt{n} \left( \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \mu_{\gamma,+} \right)}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sqrt{n} \left( \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \mu_{\gamma,+} \right)}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2 (Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sqrt{n} \left( \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \mu_{\gamma,+} \right)}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2 (Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\frac{\frac{\sum_i \gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})}{\sum_i \gamma_+(Z_i)}}{\sqrt{\frac{n \mathbb{E}[\gamma_+(Z_i)^2 (Y_i(1) - \mu_{\gamma,+})^2]}{n \mathbb{E}[\gamma_+(Z_i)]}}} = \underbrace{\frac{\sum_i \gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})}{\sqrt{n \mathbb{E}[\gamma_+(Z_i)^2 (Y_i(1) - \mu_{\gamma,+})^2]}}}_{\xrightarrow{d} \mathcal{N}(0,1)} \cdot \underbrace{\frac{1}{\frac{\frac{1}{n} \sum_i \gamma_+(Z_i)}{\mathbb{E}[\gamma_+(Z_i)]}}}_{=1+o_p(1)} \xrightarrow{d} \mathcal{N}(0, 1)$$



# Asymptotic Normality: A Sketch of the Proof

$$\frac{\sum_i \gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})}{\sqrt{n \mathbb{E} [\gamma_+(Z_i)^2 (Y_i(1) - \mu_{\gamma,+})^2]}} \xrightarrow{d} \mathcal{N}(0, 1)$$

- $\underline{\sigma}^2 = \inf_z \text{Var} [Y_i \mid Z_i = z] > 0$ :  $\text{Var} [\gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})] \geq \underline{\sigma}^2 \mathbb{E} [\gamma_+(Z_i)^2]$
- $\sup_u \left| h(u, \gamma_\diamond^{(n)}) \right| < C' \mathbb{E} [\gamma_\diamond^{(n)}(Z_i)]$  and  $Y_i \in [0, 1]$ :  $|\mu_{\gamma,+}| = \left| \frac{\mathbb{E}[\alpha_{(1)}(U)h(U, \gamma_+)]}{\mathbb{E}[\gamma_+(Z_i)]} \right| \leq C'$
- $\sup_z \left| \gamma_\diamond^{(n)}(z) \right| < C n^\beta \mathbb{E} [\gamma_\diamond^{(n)}(Z_i)]$  and Lyapunov CLT condition:

$$\frac{n \mathbb{E} [|\gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})|^{2+q}]}{(n \text{Var} [\gamma_+(Z_i) (Y_i(1) - \mu_{\gamma,+})])^{\frac{2+q}{2}}} \leq \left( \frac{C' + 1}{\underline{\sigma}} \right)^{2+q} \left( C n^{\beta - \frac{1}{2}} \right)^q \xrightarrow{0 \rightarrow \infty} 0$$

# Asymptotic Normality: A Sketch of the Proof

$$\frac{\frac{1}{n} \sum_i \gamma_+(Z_i)}{\mathbb{E}[\gamma_+(Z_i)]} \xrightarrow{p} 1$$

■  $\left[ \sup_z \left| \gamma_{\diamond}^{(n)}(z) \right| < C n^{\beta} \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right] \right]$  and Chebyshev's inequality:

$$\begin{aligned} \mathbf{P} \{ |\mathbb{E}_n[\gamma_+(Z_i)] - \mathbb{E}[\gamma_+(Z_i)]| \geq \epsilon \mathbb{E}[\gamma_+(Z_i)] \} &\leq \frac{\text{Var}[\gamma_+(Z_i)]}{n \epsilon^2 \mathbb{E}[\gamma_+(Z_i)]^2} \\ &\leq \left( \frac{C}{\epsilon} \cdot n^{\beta - \frac{1}{2}} \right)^2 \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

# Asymptotic Normality: A Sketch of the Proof

$$\begin{aligned}
 \frac{\frac{\sum_i \gamma_+(Z_i)(Y_i(1) - \mu_{\gamma,+})}{\sum_i \gamma_+(Z_i)}}{\sqrt{\frac{n\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}{n\mathbb{E}[\gamma_+(Z_i)]}}} &= \underbrace{\frac{\sum_i \gamma_+(Z_i)(Y_i(1) - \mu_{\gamma,+})}{\sqrt{n\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}}}_{\xrightarrow{d} \mathcal{N}(0,1)} \cdot \underbrace{\frac{1}{\frac{\frac{1}{n} \sum_i \gamma_+(Z_i)}{\mathbb{E}[\gamma_+(Z_i)]}}}_{=1+o_p(1)} \xrightarrow{d} \mathcal{N}(0,1) \\
 \Rightarrow \frac{\sqrt{n} \left( \frac{\sum_i \gamma_+(Z_i)Y_i}{\sum_i \gamma_+(Z_i)} - \mu_{\gamma,+} \right)}{\sqrt{\frac{\mathbb{E}[\gamma_+(Z_i)^2(Y_i(1) - \mu_{\gamma,+})^2]}{\mathbb{E}[\gamma_+(Z_i)]^2}}} &\xrightarrow{d} \mathcal{N}(0,1) \Rightarrow \frac{\sqrt{n}(\hat{\tau}_\gamma - \theta_\gamma)}{\sqrt{V_\gamma}} \xrightarrow{d} \mathcal{N}(0,1)
 \end{aligned}$$

## Plug-in Estimator for $V_\gamma$

$$V_\gamma = \frac{\mathbb{E} \left[ \gamma_+^2 (Z_i) (Y_i - \mu_{\gamma,+})^2 \right]}{\mathbb{E} [\gamma_+ (Z_i)]^2} + \frac{\mathbb{E} \left[ \gamma_-^2 (Z_i) (Y_i - \mu_{\gamma,-})^2 \right]}{\mathbb{E} [\gamma_- (Z_i)]^2}$$

## Plug-in Estimator for $V_\gamma$

$$V_\gamma = \frac{\mathbb{E} \left[ \gamma_+^2 (Z_i) (Y_i - \mu_{\gamma,+})^2 \right]}{\mathbb{E} [\gamma_+ (Z_i)]^2} + \frac{\mathbb{E} \left[ \gamma_-^2 (Z_i) (Y_i - \mu_{\gamma,-})^2 \right]}{\mathbb{E} [\gamma_- (Z_i)]^2}$$

### Proposition: Plug-in Estimator $\hat{V}_\gamma$

Under the same assumptions, we have  $\frac{\hat{V}_\gamma}{V_\gamma} = 1 + o_p(1)$  where

$$\hat{V}_\gamma = \frac{\frac{1}{n} \sum_{i=1}^n \gamma_+^2 (Z_i) (Y_i - \hat{\mu}_{\gamma,+})^2}{\left[ \frac{1}{n} \sum_{i=1}^n \gamma_+ (Z_i) \right]^2} + \frac{\frac{1}{n} \sum_{i=1}^n \gamma_-^2 (Z_i) (Y_i - \hat{\mu}_{\gamma,-})^2}{\left[ \frac{1}{n} \sum_{i=1}^n \gamma_- (Z_i) \right]^2}$$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

$$\begin{aligned}
 \text{aBias} = \theta_\gamma - \tau_w = & \underbrace{\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) \, dG(u)}_{\text{Confounding bias}} \\
 & + \underbrace{\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) \, dG(u)}_{\text{CATE heterogeneity bias}}
 \end{aligned}$$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

$$\begin{aligned}
 \text{aBias} = \theta_\gamma - \tau_w = & \underbrace{\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u)}_{\text{Confounding bias}} \\
 & + \underbrace{\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)}_{\text{CATE heterogeneity bias}}
 \end{aligned}$$

The distribution of  $G(\cdot)$  is **unknown**

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

Bound the **worst-case** bias:



# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

Bound the **worst-case** bias:

- Back out the class of latent variable distribution from  $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Z_i \leq t)$ , the empirical distribution of  $Z_i$  (Massart, 1990):

$$\mathcal{G}_n = \left\{ G(\cdot) : \sup_{t \in \mathbb{R}} |F_G(t) - \hat{F}_n(t)| \leq \sqrt{\frac{\log(2/\alpha_n)}{2n}} \right\}, \quad \alpha_n = \min \{0.05, n^{-1/4}\}$$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

Bound the **worst-case** bias:

- Back out the class of latent variable distribution from  $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Z_i \leq t)$ , the empirical distribution of  $Z_i$  (Massart, 1990):

$$\mathcal{G}_n = \left\{ G(\cdot) : \sup_{t \in \mathbb{R}} |F_G(t) - \hat{F}_n(t)| \leq \sqrt{\frac{\log(2/\alpha_n)}{2n}} \right\}, \quad \alpha_n = \min \{0.05, n^{-1/4}\}$$

- Take treatment effect heterogeneity into consideration:

$$\mathcal{T}_M = \{\tau(\cdot) \mid \tau(u) = \bar{\tau} + \Delta(u), \bar{\tau} \in \mathbb{R}, |\Delta(u)| \leq M\}, \quad M \in [0, 1]$$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

Bound the **worst-case** bias:

- Back out the class of latent variable distribution from  $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Z_i \leq t)$ , the empirical distribution of  $Z_i$  (Massart, 1990):

$$\mathcal{G}_n = \left\{ G(\cdot) : \sup_{t \in \mathbb{R}} |F_G(t) - \hat{F}_n(t)| \leq \sqrt{\frac{\log(2/\alpha_n)}{2n}} \right\}, \quad \alpha_n = \min \{0.05, n^{-1/4}\}$$

- Take treatment effect heterogeneity into consideration:

$$\mathcal{T}_M = \{\tau(\cdot) \mid \tau(u) = \bar{\tau} + \Delta(u), \bar{\tau} \in \mathbb{R}, |\Delta(u)| \leq M\}, \quad M \in [0, 1]$$

- $\mathcal{T}_0$  ( $M = 0$ ): constant CATE
- $\mathcal{T}_1$  ( $M = 1$ ): no assumptions
- $\mathcal{T}_{1/2}$  ( $M = 0$ ): a conservative choice for a monotonicity restriction

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

## Proposition: Upper Bound for the Potential Bias $\hat{B}_{\gamma,M}$

Under asymptotic normality and all necessary assumptions, for  $\tau(\cdot) \in \mathcal{T}_M$ , the upper bound of bias is

$$\hat{B}_{\gamma,M} = \sup \left\{ |\text{Bias}[\gamma_\pm, \tau_w; \alpha_0(\cdot), \tau(\cdot), G]| : G \in \mathcal{G}_n, \alpha_{(0)}(\cdot) \in [0, 1], \tau(\cdot) \in \mathcal{T}_M \right\}$$

then  $\mathbf{P} \left( |b_\gamma| \leq \hat{B}_{\gamma,M} \right) \xrightarrow{n \rightarrow \infty} 1$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

## Proposition: Upper Bound for the Potential Bias $\hat{B}_{\gamma,M}$

Under asymptotic normality and all necessary assumptions, for  $\tau(\cdot) \in \mathcal{T}_M$ , the upper bound of bias is

$$\hat{B}_{\gamma,M} = \sup \left\{ |\text{Bias}[\gamma_\pm, \tau_w; \alpha_0(\cdot), \tau(\cdot), G]| : G \in \mathcal{G}_n, \alpha_{(0)}(\cdot) \in [0, 1], \tau(\cdot) \in \mathcal{T}_M \right\}$$

then  $\mathbf{P}(|b_\gamma| \leq \hat{B}_{\gamma,M}) \xrightarrow{n \rightarrow \infty} 1$

$$\{G \in \mathcal{G}_n\} \subset \{|b_\gamma| \leq \hat{B}_{\gamma,M}\} \Rightarrow \mathbf{P}(|b_\gamma| \leq \hat{B}_{\gamma,M}) \geq \mathbf{P}(G \in \mathcal{G}_n)$$

# Upper Bound for the Potential Bias $|b_\gamma| = |\theta_\gamma - \tau_w|$

## Proposition: Upper Bound for the Potential Bias $\hat{B}_{\gamma,M}$

Under asymptotic normality and all necessary assumptions, for  $\tau(\cdot) \in \mathcal{T}_M$ , the upper bound of bias is

$$\hat{B}_{\gamma,M} = \sup \left\{ |\text{Bias}[\gamma_\pm, \tau_w; \alpha_0(\cdot), \tau(\cdot), G]| : G \in \mathcal{G}_n, \alpha_{(0)}(\cdot) \in [0, 1], \tau(\cdot) \in \mathcal{T}_M \right\}$$

then  $\mathbf{P}(|b_\gamma| \leq \hat{B}_{\gamma,M}) \xrightarrow{n \rightarrow \infty} 1$

$$\{G \in \mathcal{G}_n\} \subset \{|b_\gamma| \leq \hat{B}_{\gamma,M}\} \Rightarrow \mathbf{P}(|b_\gamma| \leq \hat{B}_{\gamma,M}) \geq \mathbf{P}(G \in \mathcal{G}_n)$$

and for  $G \in \mathcal{G}_n$ , Dvoretzky-Kiefer-Wolfowitz (DKW) inequality gives:

$$\mathbf{P}(G \in \mathcal{G}_n) \geq \mathbf{P} \left[ \sup_{t \in \mathbb{R}} |F_G(t) - \hat{F}_n(t)| \leq \sqrt{\frac{\log(2/\alpha_n)}{2n}} \right] \geq 1 - \alpha_n \xrightarrow{n \rightarrow \infty} 1$$

# Recap: What We Have Derived

## Recap: What We Have Derived

- Asymptotic limit:

$$\hat{\tau}_{\gamma} \xrightarrow{p} \theta_{\gamma} = \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_+)]}{\mathbb{E} [h (U, \gamma_+)]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_-)]}{\mathbb{E} [h (U, \gamma_-)]}$$



# Recap: What We Have Derived

## ■ Asymptotic limit:

$$\hat{\tau}_{\gamma} \xrightarrow{p} \theta_{\gamma} = \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_{+})]}{\mathbb{E} [h (U, \gamma_{+})]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_{-})]}{\mathbb{E} [h (U, \gamma_{-})]}$$

## ■ Asymptotic normality:

$$\frac{\sqrt{n} (\hat{\tau}_{\gamma} - \theta_{\gamma})}{\sqrt{\hat{V}_{\gamma}}} \xrightarrow{d} \mathcal{N} (0, 1)$$

## Recap: What We Have Derived

- Asymptotic limit:

$$\hat{\tau}_{\gamma} \xrightarrow{p} \theta_{\gamma} = \frac{\mathbb{E} [\alpha_{(1)} (U) h (U, \gamma_{+})]}{\mathbb{E} [h (U, \gamma_{+})]} - \frac{\mathbb{E} [\alpha_{(0)} (U) h (U, \gamma_{-})]}{\mathbb{E} [h (U, \gamma_{-})]}$$

- Asymptotic normality:

$$\frac{\sqrt{n} (\hat{\tau}_{\gamma} - \theta_{\gamma})}{\sqrt{\hat{V}_{\gamma}}} \xrightarrow{d} \mathcal{N} (0, 1)$$

- Upper bound of the asymptotic bias:

$$\hat{B}_{\gamma, M} = \sup \left\{ |\text{Bias} [\gamma_{\pm}, \tau_w; \alpha_0 (\cdot), \tau (\cdot), G]| : G \in \mathcal{G}_n, \alpha_{(0)} (\cdot) \in [0, 1], \tau (\cdot) \in \mathcal{T}_M \right\}$$

# Bias-aware Confidence Intervals

## Corollary: Valid Confidence Intervals

Under asymptotic normality and all necessary assumptions, for  $\tau(\cdot) \in \mathcal{T}_M$ , consider the CIs

$$\hat{\tau}_\gamma \pm l_\alpha, \quad l_\alpha = \min \left\{ l : \mathbf{P} \left[ \left| b + n^{-\frac{1}{2}} \hat{V}_\gamma^{\frac{1}{2}} \tilde{Z} \right| \leq l \right] \geq 1 - \alpha, \forall |b| \leq \hat{B}_{\gamma, M} \right\}$$

where

- $\tilde{Z}$  is a standard Gaussian random variable
- $\alpha \in (0, 1)$  is the significant level
- $\hat{V}_\gamma$  is an estimate of the sampling variance  $V_\gamma$

then

$$\liminf_{n \rightarrow \infty} \mathbf{P} [\tau_w \in \hat{\tau}_\gamma \pm l_\alpha] \geq 1 - \alpha$$

# Bias-aware Confidence Intervals

CLT (with bias considered) is

$$\frac{\sqrt{n}(\hat{\tau}_\gamma - \theta_\gamma)}{\hat{V}_\gamma^{1/2}} = \frac{\sqrt{n}(\hat{\tau}_\gamma - \tau_w - b_\gamma)}{\hat{V}_\gamma^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where  $b_\gamma = \theta_\gamma - \tau_w$ , then let  $\tilde{Z} \sim \mathcal{N}(0, 1)$  we have

$$\begin{aligned} \mathbf{P}[\tau_w \in \hat{\tau}_\gamma \pm l_\alpha] &= \mathbf{P}[-l_\alpha - b_\gamma \leq \hat{\tau}_\gamma - \tau_w - b_\gamma \leq l_\alpha - b_\gamma] \\ &= \mathbf{P}\left[-\sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha + b_\gamma) \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(\hat{\tau}_\gamma - \tau_w - b_\gamma) \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha - b_\gamma)\right] \\ &= \mathbb{E}\left(\mathbf{P}\left[-\sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha + b_\gamma) \leq \tilde{Z} \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha - b_\gamma)\right] \mid \hat{V}_\gamma, \hat{B}_{\gamma,M}, \hat{\tau}_\gamma\right) + o(1) \\ &= \mathbb{E}\left[\mathbf{P}\left(-l_\alpha \leq n^{-1/2}\hat{V}_\gamma^{1/2}\tilde{Z} + b_\gamma \leq l_\alpha\right) \mid \hat{V}_\gamma, \hat{B}_{\gamma,M}, \hat{\tau}_\gamma\right] + o(1) \\ &= 1 - \alpha + o(1) \end{aligned}$$

# Bias-aware Confidence Intervals

CLT (with bias considered) is

$$\frac{\sqrt{n}(\hat{\tau}_\gamma - \theta_\gamma)}{\hat{V}_\gamma^{1/2}} = \frac{\sqrt{n}(\hat{\tau}_\gamma - \tau_w - b_\gamma)}{\hat{V}_\gamma^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

where  $b_\gamma = \theta_\gamma - \tau_w$ , then let  $\tilde{Z} \sim \mathcal{N}(0, 1)$  we have

$$\begin{aligned} \mathbf{P}[\tau_w \in \hat{\tau}_\gamma \pm l_\alpha] &= \mathbf{P}[-l_\alpha - b_\gamma \leq \hat{\tau}_\gamma - \tau_w - b_\gamma \leq l_\alpha - b_\gamma] \\ &= \mathbf{P}\left[-\sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha + b_\gamma) \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(\hat{\tau}_\gamma - \tau_w - b_\gamma) \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha - b_\gamma)\right] \\ &= \mathbb{E}\left(\mathbf{P}\left[-\sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha + b_\gamma) \leq \tilde{Z} \leq \sqrt{n}\hat{V}_\gamma^{-1/2}(l_\alpha - b_\gamma)\right] \mid \hat{V}_\gamma, \hat{B}_{\gamma, M}, \hat{\tau}_\gamma\right) + o(1) \\ &= \mathbb{E}\left[\mathbf{P}\left(-l_\alpha \leq n^{-1/2}\hat{V}_\gamma^{1/2}\tilde{Z} + b_\gamma \leq l_\alpha\right) \mid \hat{V}_\gamma, \hat{B}_{\gamma, M}, \hat{\tau}_\gamma\right] + o(1) \\ &= 1 - \alpha + o(1) \end{aligned}$$

# Robustness to CATE Heterogeneity Misspecification

$$\hat{\tau}_{\gamma} \pm l_{\alpha}, l_{\alpha} = \min \left\{ l : \mathbf{P} \left[ \left| b + n^{-\frac{1}{2}} \hat{V}_{\gamma}^{\frac{1}{2}} \tilde{Z} \right| \leq l \right] \geq 1 - \alpha, \forall |b| \leq \hat{B}_{\gamma, M} \right\}$$

where

$$\mathcal{T}_M = \{ \tau(\cdot) \mid \tau(u) = \bar{\tau} + \Delta(u), \bar{\tau} \in \mathbb{R}, |\Delta(u)| \leq M \}, \quad M \in [0, 1]$$

Consider an extreme misspecification of CATE heterogeneity:  $M = 0$ , are the CIs robust?

## Robustness to CATE Heterogeneity Misspecification

$$\hat{\tau}_{\gamma} \pm l_{\alpha}, l_{\alpha} = \min \left\{ l : \mathbf{P} \left[ \left| b + n^{-\frac{1}{2}} \hat{V}_{\gamma}^{\frac{1}{2}} \tilde{Z} \right| \leq l \right] \geq 1 - \alpha, \forall |b| \leq \hat{B}_{\gamma, M} \right\}$$

where

$$\mathcal{T}_M = \{ \tau(\cdot) \mid \tau(u) = \bar{\tau} + \Delta(u), \bar{\tau} \in \mathbb{R}, |\Delta(u)| \leq M \}, \quad M \in [0, 1]$$

### Corollary: Robustness to CATE Heterogeneity Misspecification

The CIs under the misspecification of  $M = 0$  is still valid, but only for the convenience-weighted treatment effect:

$$\tau_{h,+} := \int \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} \tau(u) \, dG(u)$$

# Robustness to CATE Heterogeneity Misspecification

## Corollary: Robustness to CATE Heterogeneity Misspecification

The CIs under the misspecification of  $M = 0$  is still valid, but only for:

$$\tau_{h,+} := \int \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} \tau(u) dG(u)$$

$$\begin{aligned} a\text{Bias} = \theta_\gamma - \tau_w = & \int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{h(u, \gamma_-)}{\mathbb{E}_G[h(U, \gamma_-)]} \right) \alpha_{(0)}(u) dG(u) \\ & + \underbrace{\int \left( \frac{h(u, \gamma_+)}{\mathbb{E}_G[h(U, \gamma_+)]} - \frac{w(u)}{\mathbb{E}_G[w(U)]} \right) \tau(u) dG(u)}_{\text{CATE heterogeneity bias}} \end{aligned}$$



# Applications

# Design Estimators

The goal: Make the confidence intervals **shorter**

$$\hat{\tau}_{\gamma} \pm l_{\alpha}, \quad l_{\alpha} = \min \left\{ l : \mathbf{P} \left[ \left| b + n^{-\frac{1}{2}} \hat{V}_{\gamma}^{\frac{1}{2}} \tilde{Z} \right| \leq l \right] \geq 1 - \alpha, \forall |b| \leq \hat{B}_{\gamma, M} \right\}$$

by minimizing the worst-case MSE of

$$\hat{\tau} = \hat{\mu}_{\gamma,+} - \hat{\mu}_{\gamma,-} = \frac{\sum_i \gamma_+(Z_i) Y_i}{\sum_i \gamma_+(Z_i)} - \frac{\sum_i \gamma_-(Z_i) Y_i}{\sum_i \gamma_-(Z_i)}$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

Using the fact

$$\text{Var} [\gamma_{\diamond}(Z_i)Y_i] \leq \int \gamma_{\diamond}^2(z) dF(z), \quad \diamond \in \{+, -\}$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$|h(u, \gamma_{+}) - h(u, \gamma_{-})| \leq t_1, \quad \forall u$$

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

$$\int \gamma_{+}(z) d\bar{F}(z) = \int \gamma_{-}(z) d\bar{F}(z) = 1$$

$$\gamma_{-}(z) = 0, \quad z \geq c$$

$$\gamma_{+}(z) = 0, \quad z < c$$

$$|\gamma_{\diamond}(z)| \leq Cn^{\beta}, \quad \forall z, \diamond \in \{\pm\}$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$|h(u, \gamma_{+}) - h(u, \gamma_{-})| \leq t_1, \quad \forall u$$

confounding bias

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

CATE-heterogeneity bias

$$\int \gamma_{+}(z) d\bar{F}(z) = \int \gamma_{-}(z) d\bar{F}(z) = 1$$

$$\gamma_{-}(z) = 0, \quad z \geq c$$

$$\gamma_{+}(z) = 0, \quad z < c$$

$$|\gamma_{\diamond}(z)| \leq Cn^{\beta}, \quad \forall z, \diamond \in \{\pm\}$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_-^2(z) d\bar{F}(z) + \int \gamma_+^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$|h(u, \gamma_+) - h(u, \gamma_-)| \leq t_1, \quad \forall u$$

confounding bias

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

CATE-heterogeneity bias

$$\int \gamma_+(z) d\bar{F}(z) = \int \gamma_-(z) d\bar{F}(z) = 1$$

normalization constraint

$$\gamma_-(z) = 0, \quad z \geq c$$

Sharp RD

$$\gamma_+(z) = 0, \quad z < c$$

$$|\gamma_{\diamond}(z)| \leq Cn^{\beta}, \quad \forall z, \diamond \in \{\pm\}$$

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_-^2(z) d\bar{F}(z) + \int \gamma_+^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$|h(u, \gamma_+) - h(u, \gamma_-)| \leq t_1, \quad \forall u$$

confounding bias

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

CATE-heterogeneity bias

$$\int \gamma_+(z) d\bar{F}(z) = \int \gamma_-(z) d\bar{F}(z) = 1$$

normalization constraint

$$\gamma_-(z) = 0, \quad z \geq c$$

Sharp RD

$$\gamma_+(z) = 0, \quad z < c$$

$$|\gamma_{\diamond}(z)| \leq Cn^{\beta}, \quad \forall z, \diamond \in \{\pm\} \quad \text{no observation is given excessive influence}$$



# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

CATE-heterogeneity bias

$$\int \gamma_{+}(z) d\bar{F}(z) = \int \gamma_{-}(z) d\bar{F}(z) = 1$$

normalization constraint

# Design Estimators: Quadratic Programming

Solve

$$\min_{\gamma_{\pm}(\cdot)} \frac{1}{n} \left( \int \gamma_{-}^2(z) d\bar{F}(z) + \int \gamma_{+}^2(z) d\bar{F}(z) \right) + (t_1 + t_2)^2$$

s.t.

$$M |h(u, \gamma_{\diamond}) - \bar{w}(u)| \leq t_2, \quad \forall u, \diamond \in \{\pm\}$$

CATE-heterogeneity bias

$$\int \gamma_{+}(z) d\bar{F}(z) = \int \gamma_{-}(z) d\bar{F}(z) = 1$$

normalization constraint

$$\bar{F}(\cdot) : \quad F_G(t) = \int \mathbf{1}(\{z \leq t\}) \int p(z | u) dG(u) d\lambda(z)$$

$$\bar{w}(\cdot) : \quad \tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u)$$

# Design Estimators: Quadratic Programming

$$\bar{F}(\cdot) : \quad F_G(t) = \int \mathbf{1}(\{z \leq t\}) \int p(z | u) dG(u) d\lambda(z)$$

$$\bar{w}(\cdot) : \quad \tau_w = \int \frac{w(u)}{\mathbb{E}_G[w(U)]} \tau(u) dG(u)$$

- $\bar{F}(\cdot)$  assigns non-trivial mass to  $[c, \infty)$  and  $\bar{w}(\cdot)$  is bounded:  $\exists k > 1$  s.t.

$$\mathbb{P} \left[ \frac{1}{k} < \bar{F}([c, \infty)) < 1 - \frac{1}{k}, \sup_u |\bar{w}(u)| < k \right] \xrightarrow{n \rightarrow \infty} 1$$

- $\int \gamma_{\diamond}^{(n)}(z) dF(z)$  is asymptotically lower bounded by a strictly positive number:

$$\exists \delta > 0 \text{ s.t. } \mathbb{P} \left[ \int \gamma_{\diamond}^{(n)}(z) dF(z) > \delta \right] \xrightarrow{n \rightarrow \infty} 1$$

# Design Estimators: Quadratic Programming

$$\frac{1}{k} < \bar{F}([c, \infty)) < 1 - \frac{1}{k}, \sup_u |\bar{w}(u)| < k \quad \Rightarrow \quad \sup_z \left| \gamma_{\diamond}^{(n)}(z) \right| < C n^{\beta} \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right]$$

$$\int \gamma_{\diamond}^{(n)}(z) dF(z) > \delta \quad \Rightarrow \quad \sup_u \left| h(u, \gamma_{\diamond}^{(n)}) \right| < C' \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right]$$

## Design Estimators: Quadratic Programming

$$\frac{1}{k} < \bar{F}([c, \infty)) < 1 - \frac{1}{k}, \sup_u |\bar{w}(u)| < k \quad \Rightarrow \quad \sup_z \left| \gamma_{\diamond}^{(n)}(z) \right| < C n^{\beta} \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right] \\ \int \gamma_{\diamond}^{(n)}(z) dF(z) > \delta \quad \Rightarrow \quad \sup_u \left| h(u, \gamma_{\diamond}^{(n)}) \right| < C' \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right] \Rightarrow$$

### Theorem: Asymptotic Normality of $\hat{\tau}$

Suppose the sequence of weighting kernels  $\gamma_+^{(n)}$  and  $\gamma_-^{(n)}$  is deterministic, and  $\exists \beta \in (0, \frac{1}{2})$ ,  $C, C' > 0$  s.t.  $\forall n$  large enough:  $\sup_z \left| \gamma_{\diamond}^{(n)}(z) \right| < C n^{\beta} \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right]$ ,  $\sup_u \left| h(u, \gamma_{\diamond}^{(n)}) \right| < C' \mathbb{E} \left[ \gamma_{\diamond}^{(n)}(Z_i) \right]$  where  $\diamond = \{+, -\}$ . Then

$$\frac{\sqrt{n}(\hat{\tau}_{\gamma} - \theta_{\gamma})}{\sqrt{V_{\gamma}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

# Design Estimators: Procedure

## ■ Input:

- samples  $\{Z_i, Y_i, W_i\}$  and cutoff  $c$
- sensitivity model  $\mathcal{T}_M$ , estimand of interest  $\tau_w$
- nominal significance level  $\alpha$

# Design Estimators: Procedure

## ■ Input:

- samples  $\{Z_i, Y_i, W_i\}$  and cutoff  $c$
- sensitivity model  $\mathcal{T}_M$ , estimand of interest  $\tau_w$
- nominal significance level  $\alpha$

## ■ Procedure:

**S1** guess/estimate  $\bar{F}(\cdot)$  and  $\bar{w}(\cdot)$  via nonparametric maximum likelihood

**S2** solve the minimax program, get  $\gamma_+, \gamma_-$

**S3** form the point estimate  $\hat{\tau}_\gamma$  and its variance  $\hat{V}_\gamma$

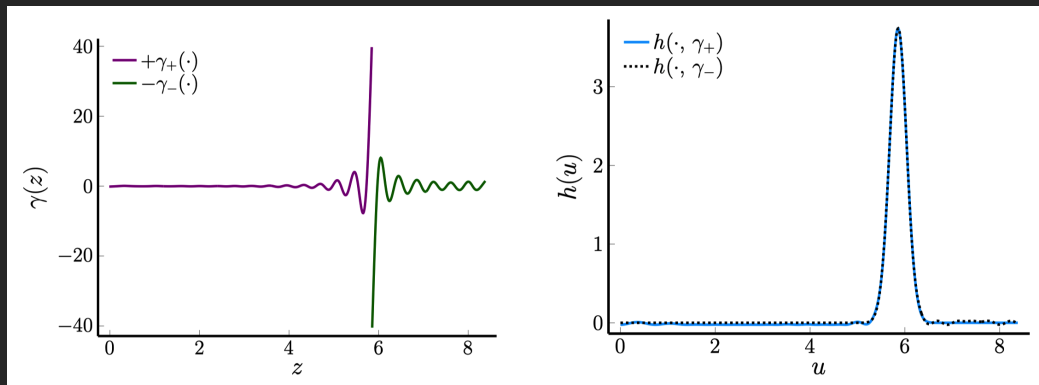
**S4** estimate the worst-case bias

$$\hat{B}_\gamma = \sup \left\{ |\text{Bias}[\gamma_\pm, \tau_w; \alpha_0(\cdot), \tau(\cdot), G]| : G \in \mathcal{G}_n, \alpha_{(0)}(\cdot) \in [0, 1], \tau(\cdot) \in \mathcal{T}_M \right\}$$

**S5** form the bias-aware CIs at level  $\alpha$  as

$$\hat{\tau}_\gamma \pm l_\alpha, l_\alpha = \min \left\{ l : \mathbf{P} \left[ \left| b + n^{-\frac{1}{2}} \hat{V}_\gamma^{\frac{1}{2}} \tilde{Z} \right| \leq l \right] \geq 1 - \alpha, \forall |b| \leq \hat{B}_{\gamma, M} \right\}$$

# Design Estimators: Example





# Discussion

## Literature: Continuity-Based RD

Most popular: local linear regression (Hahn et al., 2001; G. W. Imbens and Lemieux, 2008)

$$\hat{\tau}_c = \arg \min_{\tau} \left\{ \sum_{i=1}^n \underbrace{K}_{\text{weighting}} \left( \underbrace{\frac{|Z_i - c|}{h_n}}_{\text{bandwidth}} \right) (Y_i - a - \tau W_i - \beta_- (Z_i - c)_- - \beta_+ (Z_i - c)_+)^2 \right\}$$

# Literature: Continuity-Based RD

Most popular: local linear regression (Hahn et al., 2001; G. W. Imbens and Lemieux, 2008)

$$\hat{\tau}_c = \arg \min_{\tau} \left\{ \sum_{i=1}^n \underbrace{K}_{\text{weighting}} \left( \underbrace{\frac{|Z_i - c|}{h_n}}_{\text{bandwidth}} \right) (Y_i - a - \tau W_i - \beta_- (Z_i - c)_- - \beta_+ (Z_i - c)_+)^2 \right\}$$

## Literature: Continuity-Based RD

Most popular: local linear regression (Hahn et al., 2001; G. W. Imbens and Lemieux, 2008)

$$\hat{\tau}_c = \arg \min_{\tau} \left\{ \sum_{i=1}^n \underbrace{K}_{\text{weighting}} \left( \underbrace{\frac{|Z_i - c|}{h_n}}_{\text{bandwidth}} \right) (Y_i - a - \tau W_i - \beta_- (Z_i - c)_- - \beta_+ (Z_i - c)_+)^2 \right\}$$

- $\mu_{(w)}(z) = \mathbb{E}[Y(w) \mid Z = z]$  is **smooth**
- $h_n$  decays at an **appropriate** rate

## Literature: Continuity-Based RD

Most popular: local linear regression (Hahn et al., 2001; G. W. Imbens and Lemieux, 2008)

$$\hat{\tau}_c = \arg \min_{\tau} \left\{ \sum_{i=1}^n \underbrace{K}_{\text{weighting}} \left( \underbrace{\frac{|Z_i - c|}{h_n}}_{\text{bandwidth}} \right) (Y_i - a - \tau W_i - \beta_- (Z_i - c)_- - \beta_+ (Z_i - c)_+)^2 \right\}$$

- $\mu_{(w)}(z) = \mathbb{E}[Y(w) \mid Z = z]$  is **smooth**
- $h_n$  decays at an **appropriate** rate

Robust CIs (Armstrong and Kolesár, 2020; Calonico et al., 2014; Kolesár and Rothe, 2018);

## Literature: Continuity-Based RD

Most popular: local linear regression (Hahn et al., 2001; G. W. Imbens and Lemieux, 2008)

$$\hat{\tau}_c = \arg \min_{\tau} \left\{ \sum_{i=1}^n \underbrace{K}_{\text{weighting}} \left( \underbrace{\frac{|Z_i - c|}{h_n}}_{\text{bandwidth}} \right) (Y_i - a - \tau W_i - \beta_- (Z_i - c)_- - \beta_+ (Z_i - c)_+)^2 \right\}$$

- $\mu_{(w)}(z) = \mathbb{E}[Y(w) \mid Z = z]$  is **smooth**
- $h_n$  decays at an **appropriate** rate

Robust CIs (Armstrong and Kolesár, 2020; Calonico et al., 2014; Kolesár and Rothe, 2018); Data-adaptive bandwidths (G. Imbens and Kalyanaraman, 2012)

## Literature: Continuity-Based RD extended

$$\mu_{(w)}(z) = \mathbb{E}[Y(w) \mid Z = z]$$

If further assume **convexity** of  $\mu_{(w)}(z)$ , e.g.:

$$\left| \mu''_{(w)}(z) \right| \leq B, \forall z \in \mathbb{R}$$

# Literature: Continuity-Based RD extended

$$\mu_{(w)}(z) = \mathbb{E}[Y(w) \mid Z = z]$$

If further assume **convexity** of  $\mu_{(w)}(z)$ , e.g.:

$$\left| \mu''_{(w)}(z) \right| \leq B, \forall z \in \mathbb{R}$$

**Optimization-based RD**: the treatment effect  $\tau_c$  can be estimated (minimax linear estimation) via **numerical convex optimization** (Armstrong and Kolesár, 2018; G. Imbens and Wager, 2019)



# Link Noise-Induced RD and Continuity-Based RD

$$\begin{aligned}\mu_{(w)}(z) &= \mathbb{E}[Y_i(w) \mid Z_i = z] \\ &= \frac{\int \mathbb{E}[Y_i(w) \mid U_i = u, Z_i = z] p(z \mid u) dG(u)}{f_G(z)} = \frac{\int \alpha_{(w)}(u) p(z \mid u) dG(u)}{\int p(z \mid u) dG(u)}\end{aligned}$$

# Link Noise-Induced RD and Continuity-Based RD

$$\begin{aligned}\mu_{(w)}(z) &= \mathbb{E}[Y_i(w) \mid Z_i = z] \\ &= \frac{\int \mathbb{E}[Y_i(w) \mid U_i = u, Z_i = z] p(z \mid u) dG(u)}{f_G(z)} = \frac{\int \boxed{\alpha_{(w)}(u)} p(z \mid u) dG(u)}{\int p(z \mid u) dG(u)}\end{aligned}$$

# Link Noise-Induced RD and Continuity-Based RD

$$\begin{aligned}\mu_{(w)}(z) &= \mathbb{E}[Y_i(w) \mid Z_i = z] \\ &= \frac{\int \mathbb{E}[Y_i(w) \mid U_i = u, Z_i = z] p(z \mid u) dG(u)}{f_G(z)} = \frac{\int \boxed{\alpha_{(w)}(u)} p(z \mid u) dG(u)}{\int p(z \mid u) dG(u)}\end{aligned}$$

**Convexity** assumption on  $\mu_{(w)}(z)$ :

$$\left| \mu''_{(w)}(z) \right| \leq B, \forall z \in \mathbb{R}$$

# Link Noise-Induced RD and Continuity-Based RD

$$\begin{aligned}\mu_{(w)}(z) &= \mathbb{E}[Y_i(w) \mid Z_i = z] \\ &= \frac{\int \mathbb{E}[Y_i(w) \mid U_i = u, Z_i = z] p(z \mid u) dG(u)}{f_G(z)} = \frac{\int \boxed{\alpha_{(w)}(u)} p(z \mid u) dG(u)}{\int p(z \mid u) dG(u)}\end{aligned}$$

**Convexity** assumption on  $\mu_{(w)}(z)$ :

$$\left| \mu''_{(w)}(z) \right| \leq B, \forall z \in \mathbb{R}$$

Then the worst-case possible curvature is:

$$\text{Curv}(z, \rho, p) = \sup \left\{ \left| \frac{d^2 \mu_{(w)}(z)}{dz^2} \right| : f_G(z) = \int p(z \mid u) dG(u) \geq \rho > 0, \alpha_{(w)}(\cdot) \in [0, 1] \right\}$$

## Link Noise-Induced RD and Continuity-Based RD

$$\begin{aligned}\mu_{(w)}(z) &= \mathbb{E}[Y_i(w) \mid Z_i = z] \\ &= \frac{\int \mathbb{E}[Y_i(w) \mid U_i = u, Z_i = z] p(z \mid u) dG(u)}{f_G(z)} = \frac{\int \boxed{\alpha_{(w)}(u)} p(z \mid u) dG(u)}{\int p(z \mid u) dG(u)}\end{aligned}$$

Then the worst-case possible curvature is:

$$\text{Curv}(z, \rho, p) = \sup \left\{ \left| \frac{d^2 \mu_{(w)}(z)}{dz^2} \right| : f_G(z) = \int p(z \mid u) dG(u) \geq \rho > 0, \alpha_{(w)}(\cdot) \in [0, 1] \right\}$$

Armstrong and Kolesár (2020): fit 4th-degree polynomials to  $\mu_{(0)}(z)$  and  $\mu_{(1)}(z)$ , and take the largest estimated curvature obtained anywhere

# Literature: Randomization Inference RD

Posit a non-trivial interval  $\mathcal{I}$  with  $c \in \mathcal{I}$  s.t.

$$\{Y_i(0), Y_i(1)\} \perp Z_i \mid \{Z_i \in \mathcal{I}\}$$

then focus on this interval, perform classical **randomized** study inference

# Literature: Randomization Inference RD

Posit a non-trivial interval  $\mathcal{I}$  with  $c \in \mathcal{I}$  s.t.

$$\{Y_i(0), Y_i(1)\} \perp Z_i \mid \{Z_i \in \mathcal{I}\}$$

then focus on this interval, perform classical **randomized** study inference

- Design-based approach (Rubin, 2008)

# Literature: Randomization Inference RD

Posit a non-trivial interval  $\mathcal{I}$  with  $c \in \mathcal{I}$  s.t.

$$\{Y_i(0), Y_i(1)\} \perp Z_i \mid \{Z_i \in \mathcal{I}\}$$

then focus on this interval, perform classical **randomized** study inference

- Design-based approach (Rubin, 2008)

- Strong assumption

No **data-driven way** of choosing  $\mathcal{I}$

If the interval  $\mathcal{I}$  is known a-priori, the problem collapses to a **RCT**



# Measurement Error Induced RD

Rokkanen (2015) considers a similar approach, assuming:

# Measurement Error Induced RD

Rokkanen (2015) considers a similar approach, assuming:

- noisy running variables **(A2)** and exogeneity of the noise **(A3)**
- **NOT** assuming prior knowledge of the noise distribution  $p(\cdot | u)$

# Measurement Error Induced RD

Rokkanen (2015) considers a similar approach, assuming:

- noisy running variables **(A2)** and exogeneity of the noise **(A3)**
- **NOT** assuming prior knowledge of the noise distribution  $p(\cdot | u)$
- A stronger assumption: observing at least 3 noisy measurements of the latent variable  $U_i$ ,  $\{Z_i, Z'_i, Z''_i\}$

# Measurement Error Induced RD

Rokkanen (2015) considers a similar approach, assuming:

- noisy running variables **(A2)** and exogeneity of the noise **(A3)**
- **NOT** assuming prior knowledge of the noise distribution  $p(\cdot | u)$
- A stronger assumption: observing at least 3 noisy measurements of the latent variable  $U_i$ ,  $\{Z_i, Z'_i, Z''_i\}$

# Measurement Error Induced RD

Rokkanen (2015) considers a similar approach, assuming:

- noisy running variables **(A2)** and exogeneity of the noise **(A3)**
- **NOT** assuming prior knowledge of the noise distribution  $p(\cdot | u)$
- A stronger assumption: observing at least 3 noisy measurements of the latent variable  $U_i$ ,  $\{Z_i, Z'_i, Z''_i\}$ 
  - $(U_i, Z_i, Z'_i, Z''_i)$  is **joint normal**
  - $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$  is **linear** w.r.t.  $u$

## RD with Ordinal Running Variables

Similarly, ordinal  $Z_i$  (bond rating, custody security score, etc.) can be seen as a noisy measurement of a latent variable  $U_i$  (Li et al., 2021)

Li et al. (2021) assume

$$U_i = \mathbf{X}_i\beta + \epsilon_i$$

then use **inverse-propensity weighting** with estimated propensities  $e(u) = \mathbb{P}[Z_i \geq c \mid U_i = u]$  for inference.

## RD with Ordinal Running Variables

Similarly, ordinal  $Z_i$  (bond rating, custody security score, etc.) can be seen as a noisy measurement of a latent variable  $U_i$  (Li et al., 2021)

Li et al. (2021) assume

$$U_i = \mathbf{X}_i\beta + \epsilon_i$$

then use **inverse-propensity weighting** with estimated propensities  $e(u) = \mathbb{P}[Z_i \geq c \mid U_i = u]$  for inference.

**Assuming:**  $U_i$  can be well predicted by observable  $\mathbf{X}_i$

# Measurement Errors

- The **running variable** is unobserved, only a noisy measurement is observed  
Bartalotti et al. (2021), Davezies and Le Barbanchon (2017), Dong and Kolesár (2021), and Pei and Shen (2017)
- Measurement error in causal inference beyond RD  
Jiang and Ding (2020), Kuroki and Pearl (2014), and Pearl (2012)



# A Comparison

## RD designs      Assumptions

Noise-induced RD  
Noise-induced RD (Rokkanen, 2015)

a **known distribution** of the measurement error  $p(\cdot \mid u)$   
**multiple joint-normal** noisy measurements  $(U_i, Z_i, Z'_i, Z''_i)$   
**linear**  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) \mid U_i = u]$

Continuity-based RD  
OPTimization-based RD  
Randomization inference RD  
RD with ordinal  $Z_i$

$\mu_{(w)} = \mathbb{E}[Y(w) \mid Z = z]$  is **smooth**  
**convexity** of  $\mu_{(w)}(z)$ :  $|\mu''_{(w)}(z)| \leq B, \forall z \in \mathbb{R}$   
an **"RCT"** interval  $\mathcal{I}$ :  $\{Y_i(0), Y_i(1)\} \perp Z_i \mid \{Z_i \in \mathcal{I}\}$   
 $U_i$  can be observed, and well predicted by  $\mathbf{X}_i$

# A Comparison

## RD designs      Assumptions

Noise-induced RD  
Noise-induced RD (Rokkanen, 2015)

a **known distribution** of the measurement error  $p(\cdot | u)$   
**multiple joint-normal** noisy measurements  $(U_i, Z_i, Z'_i, Z''_i)$   
**linear**  $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) | U_i = u]$

Continuity-based RD  
OPTimization-based RD  
Randomization inference RD  
RD with ordinal  $Z_i$

$\mu_{(w)} = \mathbb{E}[Y(w) | Z = z]$  is **smooth**  
**convexity** of  $\mu_{(w)}(z)$ :  $|\mu''_{(w)}(z)| \leq B, \forall z \in \mathbb{R}$   
an "**RCT**" interval  $\mathcal{I}$ :  $\{Y_i(0), Y_i(1)\} \perp Z_i | \{Z_i \in \mathcal{I}\}$   
 $U_i$  can be observed, and well predicted by  $\mathbf{X}_i$

# 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.

Bartalotti, O., Brummet, Q., & Dieterle, S. (2021). A correction for regression discontinuity designs with group-specific mismeasurement of the running variable. *Journal of Business & Economic Statistics*, 39(3), 833–848.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.

Davezies, L., & Le Barbanchon, T. (2017). Regression discontinuity design with continuous measurement error in the running variable. *Journal of econometrics*, 200(2), 260–281.

Dong, Y., & Kolesár, M. (2021). When can we ignore measurement error in the running variable? *arXiv preprint arXiv:2111.07388*.

## References II

- Eckles, D., Ignatiadis, N., Wager, S., & Wu, H. (2020). Noise-induced randomization in regression discontinuity designs. *arXiv preprint arXiv:2004.09458*.
- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201–209.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615–635.
- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), 933–959.
- Imbens, G., & Wager, S. (2019). Optimized regression discontinuity designs. *Review of Economics and Statistics*, 101(2), 264–278.
- Jiang, Z., & Ding, P. (2020). Measurement errors in the binary instrumental variable model. *Biometrika*, 107(1), 238–245.
- Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8), 2277–2304.

## References III

- Kuroki, M., & Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2), 423–437.
- Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2), 675–697.
- Li, F., Mercatanti, A., Mäkinen, T., & Silvestrini, A. (2021). A regression discontinuity design for ordinal running variables: Evaluating central bank purchases of corporate bonds. *The Annals of Applied Statistics*, 15(1), 304–322.
- Massart, P. (1990). The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, 1269–1283.
- Pearl, J. (2012). On measurement bias in causal inference. *arXiv preprint arXiv:1203.3504*.
- Pei, Z., & Shen, Y. (2017). The devil is in the tails: Regression discontinuity design with measurement error in the assignment variable. In *Regression discontinuity designs*. Emerald Publishing Limited.
- Rokkanen, M. A. (2015). Exam schools, ability, and the effects of affirmative action: Latent factor extrapolation in the regression discontinuity design.

## References IV

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The annals of applied statistics*, 2(3), 808–840.

Thank you!