Noise-Induced Randomization in Regression Discontinuity Designs

Dean Eckles, Nikolaos Ignatiadis, Stefan Wager, Han Wu

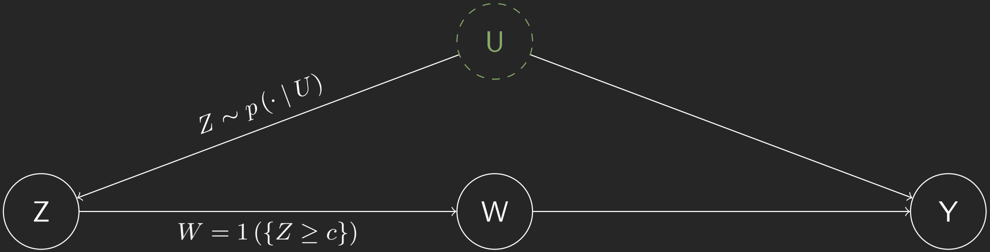
Presented by: Sai Zhang

November 18, 2022

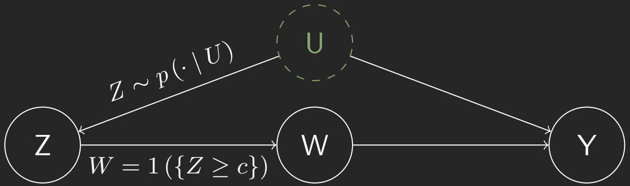
Outline

Key Argument

Sharp RD Design with A Noisy Running Variable



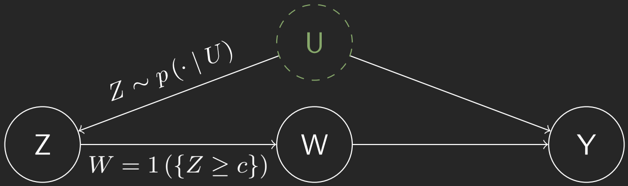
Sharp RD Design with A Noisy Running Variable



Assumption 1: Sharp RD design

- **I.I.D.** samples $\{Y_i(0), Y_i(1), Z_i\} \in \mathbb{R}^3, i = 1, \cdots, n$
- 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

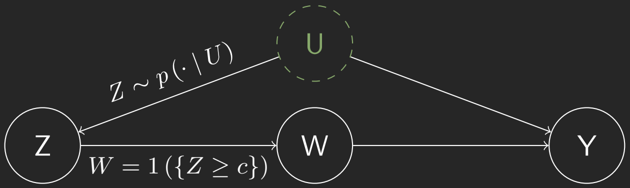


Assumption 2: Noisy running variable

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

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

Sharp RD Design with A Noisy Running Variable

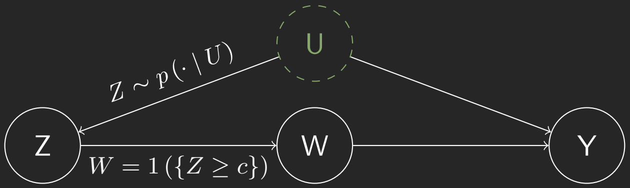


Assumption 2: Noisy running variable

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

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

Sharp RD Design with A Noisy Running Variable

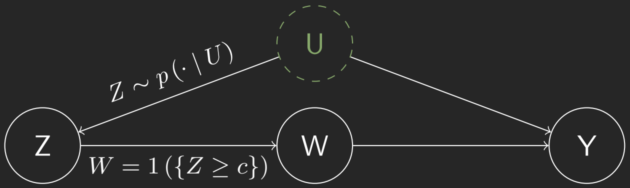


Assumption 2: Noisy running variable

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

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

Sharp RD Design with A Noisy Running Variable

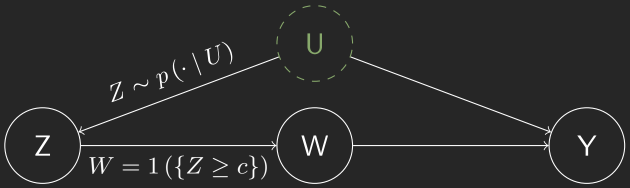


Assumption 3: Exogeneity

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

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

Sharp RD Design with A Noisy Running Variable

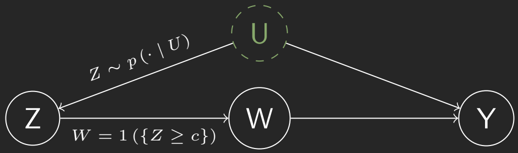


Assumption 3: Exogeneity

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

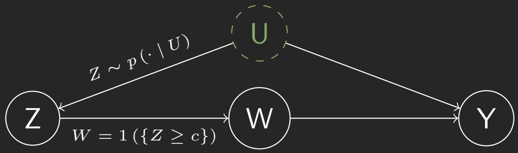
which implies $\mathbb{E}[Y_i \mid U_i, Z_i] = \alpha_{(W_i)}(u)$, where $\alpha_{(w)}(u) = \mathbb{E}[Y_i(w) \mid U_i = u]$ is the **response functions** for the potential oucomes conditional on the latent variable u

Sharp RD Design with A Noisy Running Variable



- A1 **Sharp** RD
- A2 **Noisy** Z_i: $Z_i \mid U_i \sim p(\cdot \mid U_i)$
- A3 **Exogeneity**:
  $[\{Y_i(0), Y_i(1)\} \perp Z_i] \mid U_i$

Sharp RD Design with A Noisy Running Variable



A1 **Sharp** RD

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

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

Proposition 1

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

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

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

Sharp RD Design with A Noisy Running Variable

Proposition 1

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

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

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

Sharp RD Design with A Noisy Running Variable

Proposition 1

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

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

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

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

Sharp RD Design with A Noisy Running Variable

Proposition 1

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

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

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

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

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

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

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

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

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

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

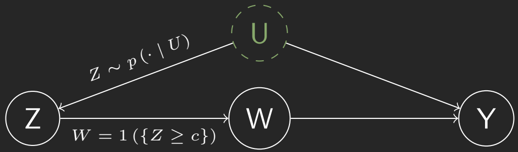
$$\mathbb{E}\left[\gamma_+(Z)\,Y\right] = \mathbb{E}\left[\alpha_{(1)}(U)\,h(U,\gamma_+)\right], \qquad \mathbb{E}\left[\gamma_-(Z)\,Y\right] = \mathbb{E}\left[\alpha_{(0)}(U)\,h(U,\gamma_-)\right]$$

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

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

To achieve balance in the latent variable: $h(\cdot,\gamma_+) \approx h(\cdot,\gamma_-)$

Sharp RD Design with A Noisy Running Variable

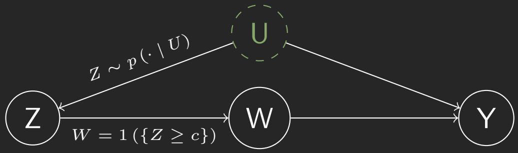


A1 **Sharp** RD

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

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

Sharp RD Design with A Noisy Running Variable



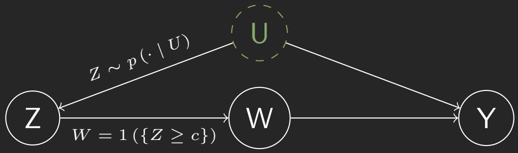
A1 **Sharp** RD

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

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

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

Sharp RD Design with A Noisy Running Variable



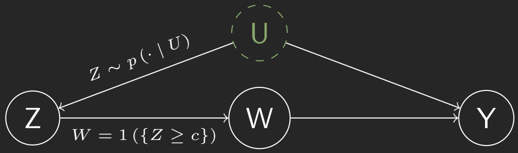
A1 **Sharp** RD

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

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

- **No need** to know G (distribution of U)
- **Need** to know $p(z \mid u)$ (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: $Z_i \mid U_i \sim p(\cdot \mid U_i)$

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

- **No need** to know G (distribution of U)
- **Need** to know $p(z \mid u)$ (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

References I

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

Thank you!