

# Nonlinear and Nonseparable Structural Functions in Fuzzy Regression Discontinuity Designs

Haitian Xie

PhD Candidate

Department of Economics  
University of California San Diego

December, 2022

# Motivation

- Regression discontinuity (RD) design - one of the most important non-experimental methods for causal inference
- Treatment changes discontinuously as a function of some underlying index which we call the running variable
- Most theoretical results study binary treatment variable
  - either 0 or 1
- In practice, the treatment may well be continuous
  - taking values inside an interval

# RD first stage: binary treatment

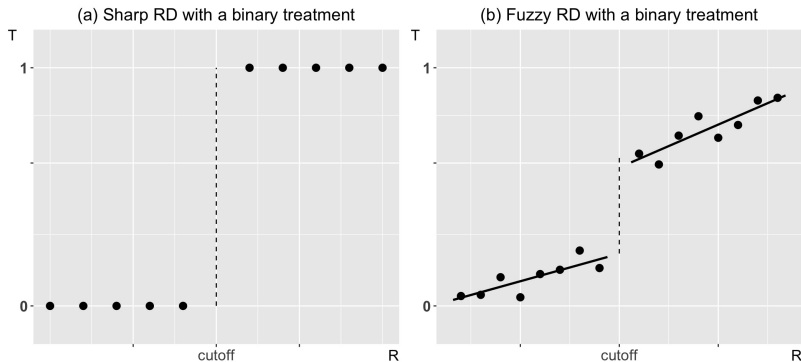


Figure: Demonstration of RD designs with a binary treatment.

# RD first stage: continuous treatment

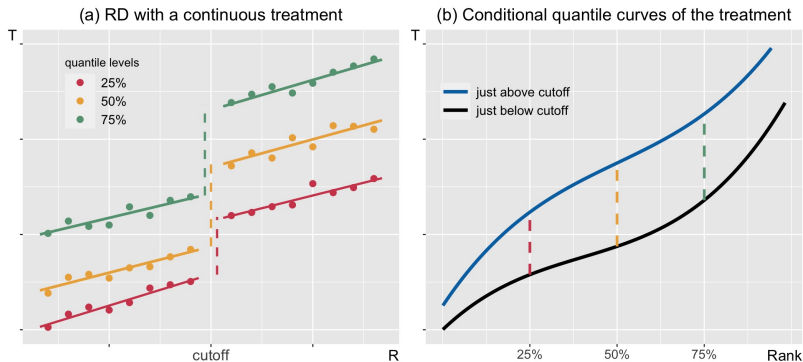


Figure: Demonstration of RD designs with a continuous treatment.

# Causal object of interest

- The structural function (SF)

$$Y = g^*(T, R, \varepsilon)$$

- We use the RD to identify the SF at the cutoff  $\bar{r}$

$$g^*(t, \bar{r}, \varepsilon), \text{ for all } t$$

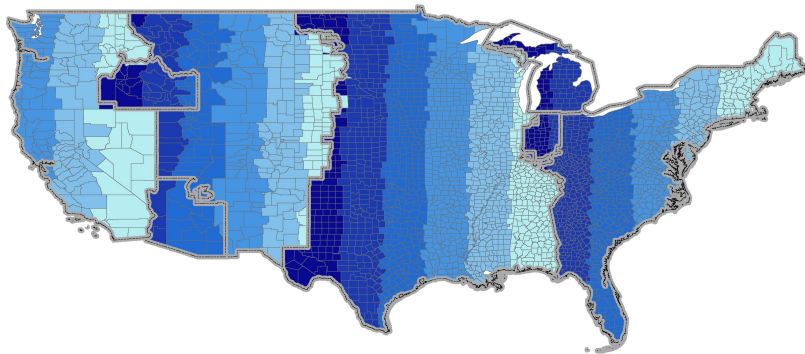
- With a binary treatment, we care about the scalar treatment effect

$$g^*(1, \bar{r}, \varepsilon) - g^*(0, \bar{r}, \varepsilon)$$

# Main contributions

- General specification of the structural function:  
nonlinear and nonseparable
- Nonparametric identification result under shape  
restrictions (monotonicity and smoothness)
- Semiparametric estimator
  - same convergence rate as with a binary treatment ( $n^{-2/5}$ )
  - asymptotic normality

# Example: time zone boundaries



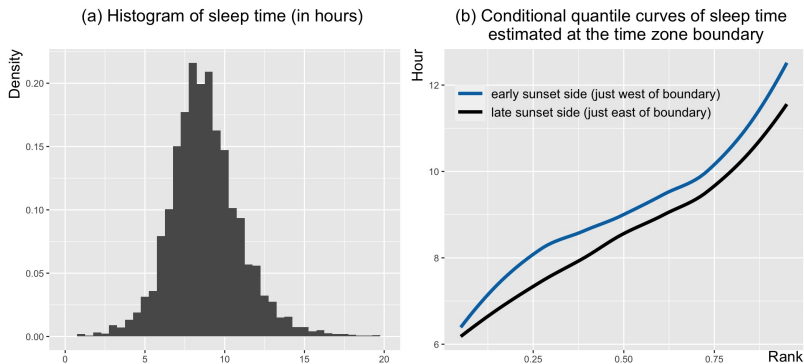
**Figure:** Time Zones and Average Sunset Time. The darker the color, the later the average sunset time.

## Example: time zone boundaries (cont'd)

- Giuntella and Mazzonna (2019)
- Sleep time  $\rightarrow$  Health
- Treatment  $T$ : sleep time (continuous variable)
- Outcome of interest  $Y$ : health (measured by BMI)
- Running variable  $R$ : distance to the time zone boundary (boundary is the cutoff/threshold  $\bar{r}$ )



# Example: time zone boundaries (cont'd)



**Figure:** Empirical illustration of RD designs with a continuous treatment.

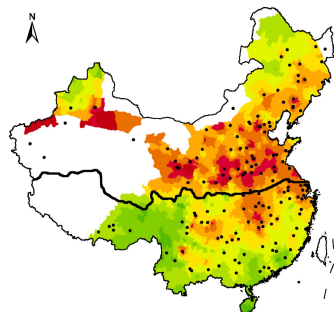
## Example: time zone boundaries (cont'd)

Why general specification of SF?

- nonlinearity: undersleeping and oversleeping are both bad
- nonseparability: effect heterogeneity due to eating habits
- determining the optimal sleep time needs a nonlinear specification

# More empirical examples

- China's Huai River winter heating policy
- Chen et al. (2013), Ebenstein et al. (2017)
- $Y$  – life expectancy
- $T$  – air pollution
- $R$  – distance to Huai River



Huai River and PM<sub>10</sub> level

# More empirical examples

- Very low birth weight threshold
- Almond et al. (2010),  
Barreca et al. (2011),  
Bharadwaj et al. (2013)
- $Y$  – infants health
- $T$  – medical spending,  
days of hospitalization
- $R$  – birth weight



# More empirical examples

- Brazil tax redistribution
- Litschig and Morrison (2010)
- $Y$  – incumbent's reelection
- $T$  – local government spending
- $R$  – district population



# More empirical examples

- Child-related tax benefits
- Barr et al. (2021), Cole (2021)
- $Y$  – academic achievements
- $T$  – family income
- $R$  – birth date



# Roadmap

- 1 Introduction
- 2 The model and comparison with the literature
- 3 Nonparametric identification
- 4 Semiparametric estimation
- 5 Numerical results

# Setup

## A triangular system

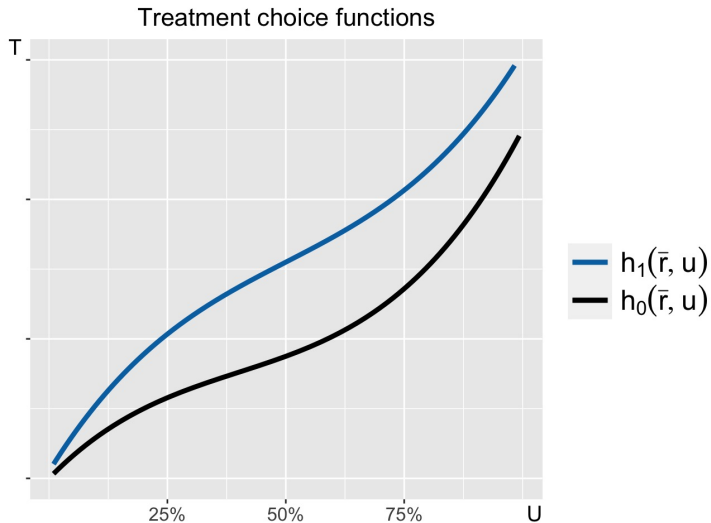
Outcome Equation:  $Y = g^*(T, R, \varepsilon)$

$$\text{Treatment Choice: } T = h(R, U) = \begin{cases} h_0(R, U), & R < \bar{r} \\ h_1(R, U), & R \geq \bar{r} \end{cases}$$

- $h$  - treatment choice function (conditional quantile)
- $U$  - conditional rank of  $T$  given  $R$
- $\bar{r}$  - cutoff/threshold
- $\varepsilon$  can correlate with  $T, R$



# Back to the graph



# Common practice: TSLS & Wald ratio

- Empirical studies use TSLS to estimate the Wald ratio

$$\text{WR} = \frac{\lim_{r \uparrow \bar{r}} \mathbb{E}[Y|R=r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[Y|R=r]}{\lim_{r \uparrow \bar{r}} \mathbb{E}[T|R=r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[T|R=r]}$$

- Binary treatment (Hahn et al., 2001):
  - $\text{WR} = (\text{local})$  average treatment effect
- Continuous treatment (Lee and Lemieux, 2010)  
if  $g^*$  is linear and separable

$$g^*(T, R, \varepsilon) = \beta T + \delta R + \varepsilon$$

then  $\text{WR} = \beta$

- General SF:  $\text{WR} \neq \text{SF}$

# Literature: RD with a continuous treatment

Dong et al. (2021)

- Quantile specific LATE

$$\frac{\lim_{r \uparrow \bar{r}} \mathbb{E}[Y|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[Y|U = u, R = r]}{\lim_{r \uparrow \bar{r}} \mathbb{E}[T|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[T|U = u, R = r]}$$

# Literature: RD with a continuous treatment

Dong et al. (2021)

- Quantile specific LATE

$$\frac{\lim_{r \uparrow \bar{r}} \mathbb{E}[Y|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[Y|U = u, R = r]}{\lim_{r \uparrow \bar{r}} \mathbb{E}[T|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[T|U = u, R = r]}$$

$$= \int_{h_0(\bar{r}, u)}^{h_1(\bar{r}, u)} \frac{\mathbb{E}[\frac{\partial}{\partial t} g^*(t, \bar{r}, \varepsilon) | U = u, R = \bar{r}]}{h_1(\bar{r}, u) - h_0(\bar{r}, u)} dt$$

- a weighted average derivative of  $g^*$
- my paper focuses on how to directly identify the structural function  $g^*(\cdot, \bar{r}, \cdot)$  at the cutoff

# Literature: IV

- Instruments with small support
  - Torgovitsky (2015, 2016),  
D'Haultfoeuille and Février (2015), Ishihara (2021)
  - Need independent instrument:  $R \perp \varepsilon$   
and exclusion restriction:  $R$  excluded from  $g^*, h_0, h_1$
- Challenges we face in RD designs
  - $R$  can violate independence and exclusion restrictions
  - estimation theory is more complicated (local to the cutoff)

# Literature: nonseparable structural functions

- Matzkin (2003, 2013, 2016),  
Hoderlein and Mammen (2007, 2009),  
Sasaki (2015), Su et al. (2015, 2019)
- Need independent treatment:  $T \perp \varepsilon | R$

# Roadmap

- 1 Introduction
- 2 The model and comparison with the literature
- 3 Nonparametric identification**
- 4 Semiparametric estimation
- 5 Numerical results

# Identification assumptions

## Dual monotonicity

- $g^*(T, R, \varepsilon)$  is strictly increasing in  $\varepsilon$  ( $\varepsilon$  is one-dimensional)
- $h(R, U)$  is strictly increasing in  $U$
- one-to-one mapping  $(Y, T) \leftrightarrow (\varepsilon, U)$  given  $R$

## Smoothness

- $g^*, h_0, h_1$  continuous
- $F_{\varepsilon|U,R}(e|u, r)$  is strictly increasing in  $e$  and

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = \lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r)$$



# Identification assumptions (cont'd)

## Fuzzy RD

- There are treatment levels that are taken both below and above the cutoff.
- Overlapping support:

$$\text{Supp}(h_0(\bar{r}, U)) \cap \text{Supp}(h_1(\bar{r}, U)) \neq \emptyset$$

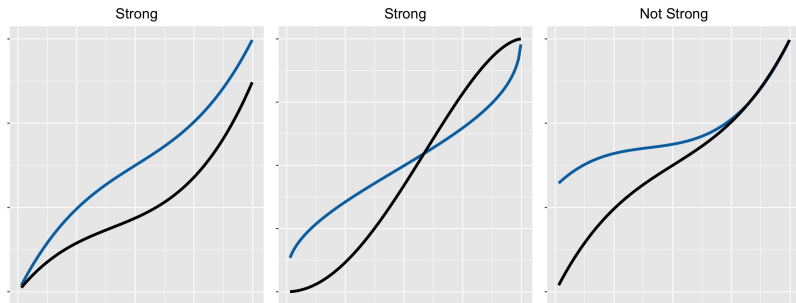
## Support invariance

- $\text{Supp}(\varepsilon|U = u, R = r)$  does not depend on  $u$  or  $r$
- still allows  $\varepsilon$  to be correlated with  $U$  or  $R$  arbitrarily

# Identification assumptions (cont'd)

## Strong discontinuity

- $h_0(\bar{r}, \cdot)$  and  $h_1(\bar{r}, \cdot)$  intersect but do not overlap



# Identification equation

## Smoothness

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r)$$

$$||$$

$$\lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r)$$

# Identification equation

Smoothness

Dual Monotonicity

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = F_{Y|T,R}^{-}(g^*(h_0(u), e)|h_0(u))$$

||

$$\lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r)$$

# Identification equation

Smoothness

Dual Monotonicity

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = F_{Y|T,R}^{-}(g^*(h_0(u), e)|h_0(u))$$

||

$$\lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = F_{Y|T,R}^{+}(g^*(h_1(u), e)|h_1(u))$$

# Identification equation

Smoothness

Dual Monotonicity

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = F_{Y|T,R}^{-}(g^*(h_0(u), e)|h_0(u))$$

||

$$\lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = F_{Y|T,R}^{+}(g^*(h_1(u), e)|h_1(u))$$

- $F_{Y|T,R}$  and  $(h_0, h_1)$  are identified  
 $\implies$  restrictions on the structural function
- These restrictions identify a unique structural function

Proof

# Nonparametric identification

## Theorem 1

*If a qualifying  $g$  satisfies the following condition*

$$F_{Y|T,R}^{-}(g(h_0(u), e)|h_0(u)) = F_{Y|T,R}^{+}(g(h_1(u), e)|h_1(u)),$$

*then there exists a continuous and strictly increasing function  $\lambda^g$  such that*

$$g^*(T, \bar{r}, \varepsilon) = g(T, \bar{r}, \lambda^g(\varepsilon)).$$

- identification at the cutoff ( $\bar{r}$  suppressed in the notation)
- identification up to a monotone transformation in  $\varepsilon$

# Roadmap

- 1 Introduction
- 2 The model and comparison with the literature
- 3 Nonparametric identification
- 4 Semiparametric estimation**
- 5 Numerical results



# Parametrize the structural function

Nonlinear:  $g_\gamma(T, \bar{r}, \varepsilon) = \gamma_1 T + \gamma_2 T^2 + \varepsilon$

Nonseparable:  $g_\gamma(T, \bar{r}, \varepsilon) = \gamma_1 T + \gamma_3 T \varepsilon + \varepsilon$

Both:  $g_\gamma(T, \bar{r}, \varepsilon) = \gamma_1 T + \gamma_2 T^2 + \gamma_3 T \varepsilon + \varepsilon$

More general: in the paper

# Semiparametric estimation

## Three-step estimation procedure

- 1 estimate conditional quantiles of  $T|R$ :  $\hat{h}_0(\bar{r}, u)$ ,  $\hat{h}_1(\bar{r}, u)$
- 2 local linear regression estimates  $\hat{F}_{Y|T,R}^{\pm}(y|t, \bar{r})$
- 3  $\hat{\gamma}$  is the minimizer of the criterion function based on the nonparametric identification equation

# Criterion function

Rearrange the identification equation

$$D_{\gamma,h}(e,u) = \int_0^u \left( F_{Y|T,R}^-(g_\gamma(h_0(v),e)|h_0(v)) - F_{Y|T,R}^+(g_\gamma(h_1(v),e)|h_1(v)) \right) dv.$$

Take  $L^2$ -norm

$$\|D_{\gamma,h}\|_w = \left( \int_0^1 \int_{\mathbb{R}} |D_{\gamma,h}(e,u)|^2 w(e,u) dedu \right)^{1/2},$$

where  $w(e,u)$  is a positive weighting function on  $\mathbb{R} \times [0,1]$ .

# Asymptotic normality

## Theorem 2

*Under regularity conditions, the following term is asymptotically standard normal*

$$n^{2/5}(\Sigma_- + \Sigma_+)^{-1/2} \Delta(\hat{\gamma} - \gamma^*) - (B_- - B_+)$$

Regularity conditions

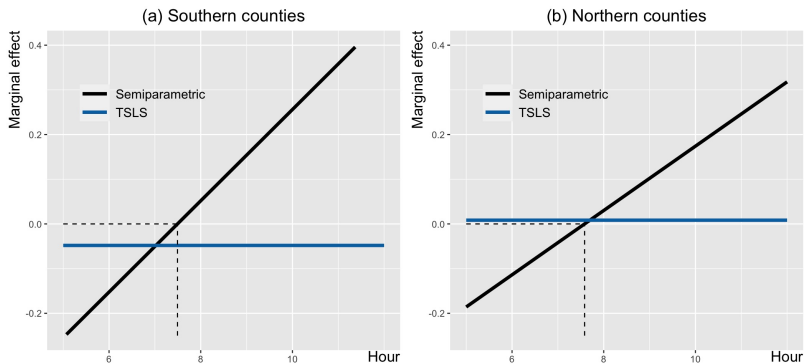
Proof

- same convergence rate as in the binary case
- second step estimators  $\hat{F}_{Y|T,R}^\pm(y|t, \bar{r})$  converges slower
- analogy: a  $\sqrt{n}$ -semiparametric estimator with a nonparametric first step being  $o(n^{-1/4})$

# Roadmap

- 1 Introduction
- 2 The model and comparison with the literature
- 3 Nonparametric identification
- 4 Semiparametric estimation
- 5 Numerical results**

# Empirical study: sleep time



**Figure:** Estimated marginal effects (structural derivative) of sleep time on BMI ( $\text{kg/m}^2$ ).

# Empirical findings

- Marginal effects are increasing, indicating a nonlinear (U-shaped) structural function.
- Optimal sleep time is between 7 and 8 hours.
- The TSLS is not capable of demonstrating such results.

# Simulations

- Semiparametric estimator vs TSLS
- marginal effect:

$$\frac{\partial}{\partial t} g_{\gamma}(t, \bar{r}, \varepsilon) = \gamma_1 + 2\gamma_2 t + \gamma_3 \varepsilon$$

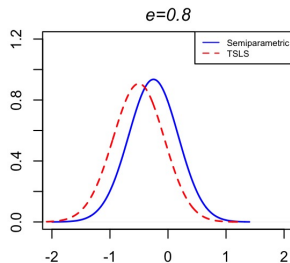
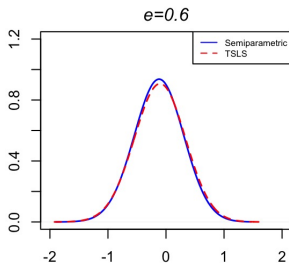
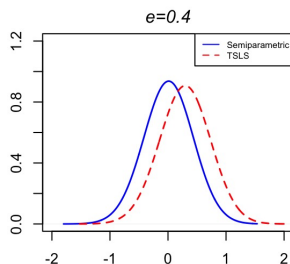
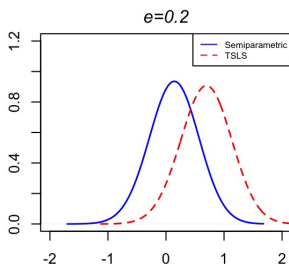
- estimated marginal effect

$$\text{SP: } \hat{\gamma}_1 + 2\hat{\gamma}_2 t + \hat{\gamma}_3 \varepsilon$$

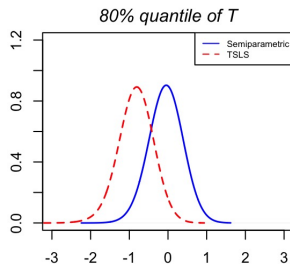
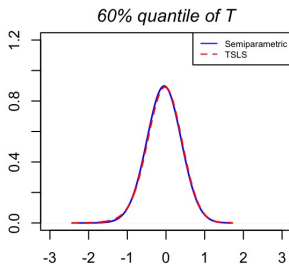
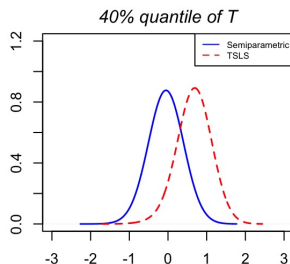
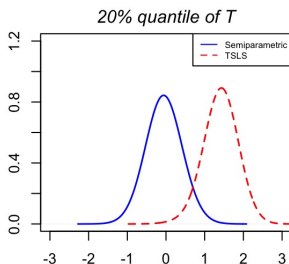
$$\text{TSLS: } \widehat{\text{WR}}$$



# Nonseparable SF

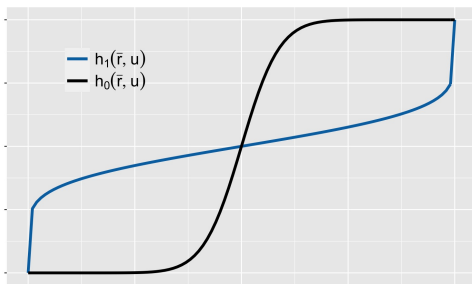


# Nonlinear SF

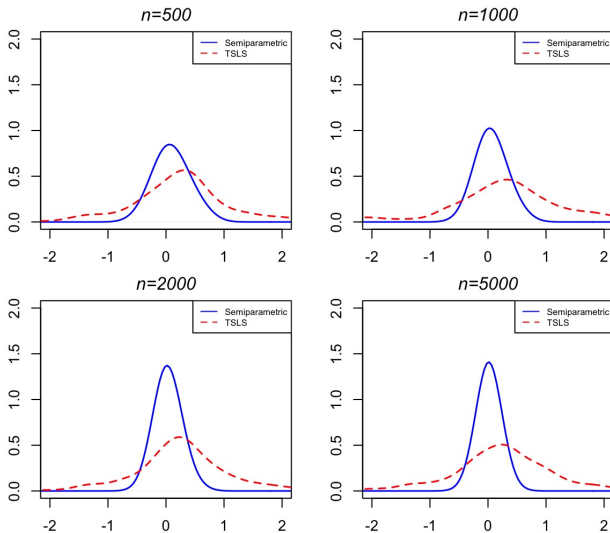


# Linear SF: weakly identified TSLS

- $\hat{\gamma}$  can outperform TSLS even with a linear  $g^*$
- weak identification in the difference in means
- $h_0(\bar{r}, \cdot)$  – quantile of  $Beta(0.1, 0.1)$ ,  
 $h_1(\bar{r}, \cdot)$  – quantile of  $Beta(10, 10)$



# Linear SF: weakly identified TSLS



# Conclusion

- RD designs with a continuous treatment
- Nonlinear and nonseparable structural function
- Nonparametric identification
- Semiparametric estimation
- Future work
  - partial identification when some assumptions fail
  - extrapolation away from the cutoff
  - Regression kink designs

*Thank You!*

# Sequencing approach

Prove that the identification equation pins down a unique structural function. [Back](#)

- If  $g$  satisfies the identification equation, then

$$g^{-1}(h_0(u), g^*(h_0(u), e)) = g^{-1}(h_1(u), g^*(h_1(u), e))$$

- Define  $\pi(t) = h_1(h_0^{-1}(t))$ , then

$$\begin{aligned} g^{-1}(t, g^*(t, e)) &= g^{-1}(\pi(t), g^*(\pi(t), e)) \\ &= g^{-1}(\pi(\pi(t)), g^*(\pi(\pi(t)), e)) \\ &= g^{-1}(\pi^m(t), g^*(\pi^m(t), e)) \end{aligned}$$

- The sequence  $t, \pi(t), \pi(\pi(t)), \dots, \pi^m(t), \dots$  is convergent.
- $g^{-1}(t, g^*(t, e))$  does not depend on  $t$ .

# Regularity conditions

- Distributions of  $Y, T$ , and  $R$  [detail](#)
- Complexity of the parametric model [detail](#)
- Kernel functions [detail](#)
- Bandwidth [detail](#)
- First-step conditional quantile estimators [detail](#)

[Back](#)



# Proof

- 1 uniform convergence results on the LLR estimators
- 2 consistency of  $\hat{\gamma}$
- 3 initial bound on the convergence rate of  $\hat{\gamma}$
- 4 stochastic equicontinuity of the criterion function
- 5 linearization of the criterion function
- 6 asymptotic normality of the minimizer of the linearized criterion function
- 7 asymptotic normality of  $\hat{\gamma}$

Back

- ① The support of  $T$  does not vary with  $R$  except when crossing the cutoff  $\bar{r}$ , i.e.,  $\text{Supp}(T|R = r) = [t'_0, t''_0]$  for  $r < \bar{r}$  and  $\text{Supp}(T|R = r) = [t'_1, t''_1]$  for  $r > \bar{r}$ . The density functions  $f_{T,R}^-$  and  $f_{T,R}^+$  are bound away from zero.
- ② The density functions  $f_{T,R}^-$  and  $f_{T,R}^+$  are twice continuously differentiable, and  $\frac{\partial^2}{\partial T^2} f_{T,R}^-(t, \bar{r})$  and  $\frac{\partial^2}{\partial T^2} f_{T,R}^+(t, \bar{r})$  are Lipschitz continuous with respect to  $t$ .
- ③ The support of  $Y$ ,  $\mathcal{Y}$ , is compact. The conditional distribution functions  $F_{Y|T,R}^-$  and  $F_{Y|T,R}^+$  are three-times continuously differentiable over  $\mathcal{Y} \times [t'_0, t''_0] \times [r_0, \bar{r}]$  and  $\mathcal{Y} \times [t'_1, t''_1] \times [\bar{r}, r_1]$ , respectively.

Back

The parametrization  $\{g_\gamma(\cdot, \bar{r}, \cdot) : \gamma \in \Gamma\}$  satisfies the following conditions.

- 1 The parameter space  $\Gamma$  is compact.
- 2 The class of functions  $\{T \mapsto g_\gamma(T + v, \bar{r}, e) : \gamma \in \Gamma, v \in (-1, 1), e \in \mathcal{E}\}$  is finite-dimensional.
- 3 The function  $g_\gamma(t, \bar{r}, e)$  is twice continuously differentiable over  $\gamma \in \Gamma$ ,  $t \in [t'_0, t''_0] \cup [t'_1, t''_1]$ , and  $e \in \mathcal{E}$ .
- 4 The gradient  $\nabla_\gamma D_{\gamma^*, h^*}(e, u)$  is a vector of linearly independent functions of  $(e, u)$ .

Back

- ① The kernel functions  $k_T$  and  $k_R$  are (1) supported on  $[-1, 1]$ , (2) strictly greater than zero in the interior of the support, (3) of bounded variation, (4) continuously differentiable on  $\mathbb{R}$ .
- ② The kernel function  $k_Y$  is (1) nonnegative and (2) integrable on  $\mathbb{R}$  with  $\int k_Y(y)dy = 1$  and satisfies (3)  $\int y k_Y(y)dy = 0$ .

Back

# Bandwidth

The bandwidth  $b_1$  and  $b_2$  satisfy the following conditions:

- ①  $b_1 \asymp b_2$ .
- ②  $(n \log n) b_1^6 = o(1)$ .
- ③  $n b_1^{\frac{13}{3} + \epsilon} \rightarrow \infty$ , for some sufficiently small  $\epsilon > 0$ .

In particular, the optimal bandwidth  $n^{-1/5}$  is allowed.

Back

# First-step conditional quantile estimators (1)

Monotonicity and smoothness: for every  $n$  sufficiently large, there exist  $C > 0$  and deterministic and finite partitions  $\mathcal{P}_0^n$  and  $\mathcal{P}_1^n$  on  $(0, 1)$  such that

$$\mathbb{P}\left(\hat{h}_0(\bar{r}, \cdot) \notin \mathcal{H}_0(\mathcal{P}_0^n)\right), \mathbb{P}\left(\hat{h}_1(\bar{r}, \cdot) \notin \mathcal{H}_1(\mathcal{P}_1^n)\right) = O(\sqrt{b_1}),$$

where  $\mathcal{H}_0(\mathcal{P}_0^n)$  includes functions that are strictly increasing with a three-times continuously differentiable inverse on each element of  $\mathcal{P}_0^n$ .  $\mathcal{H}_1(\mathcal{P}_1^n)$  is defined analogously by replacing  $\mathcal{P}_0^n$  with  $\mathcal{P}_1^n$ .

Back

## First-step conditional quantile estimators (2)

Uniform Bahadur representation:

$$\begin{aligned}\hat{h}_0(\bar{r}, u) - h_0(\bar{r}, u) &= b_1^2 \nu_0(u) + O_p(b_1^3) \\ &+ \frac{1}{nb_1} \sum_{i=1}^n q_0(T_i, R_i; u) k_{Q,0} \left( \frac{R_i - \bar{r}}{b_1} \right) \mathbf{1}\{R_i < \bar{r}\} + o_p(1/\sqrt{nb_1}),\end{aligned}$$

uniformly over  $u \in (0, 1)$ . The same for  $\hat{h}_1$ . The functions  $\nu_0$  and  $\nu_1$  are bounded. The functions  $q_0$  and  $q_1$  are (1) bounded, (2) centered, that is,

$\mathbb{E}[q_0(T, R; u)|T, R] = \mathbb{E}[q_1(T, R; u)|T, R] = 0$ , and (3) does not vary with  $n$ . The functions  $k_{Q,0}$  and  $k_{Q,1}$  are bounded.

Back

## First-step conditional quantile estimators (3)

Uniform convergence rate:

$$\begin{aligned}\|\hat{h} - h\|_{\infty} &= \sup_{u \in (0,1)} |\hat{h}_0(\bar{r}, u) - h_0(\bar{r}, u)| \vee |\hat{h}_1(\bar{r}, u) - h_1(\bar{r}, u)| \\ &= O_p \left( \sqrt{\log n / (nb_1)} + b_1^2 \right).\end{aligned}$$

Back