

Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments

Yikun Zhang

Joint work with *Professor Yen-Chi Chen*

Department of Statistics,
University of Washington

TGIF Meeting
February 7, 2025

① Introduction

② Inference Theory for $\theta(t)$ Under Positivity

③ Inference Theory for $\theta(t)$ Without Positivity

④ Simulations and Case Study

⑤ Discussion



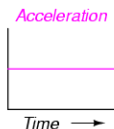
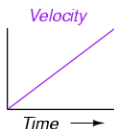
The Notion of Derivative

The derivative $f'(t) = \lim_{\Delta \rightarrow 0} \frac{f(t+\Delta) - f(t)}{\Delta}$ signals an instantaneous rate of change of a function f with respect to the input variable t .

The Notion of Derivative

The derivative $f'(t) = \lim_{\Delta \rightarrow 0} \frac{f(t+\Delta) - f(t)}{\Delta}$ signals an instantaneous rate of change of a function f with respect to the input variable t .

- Physics:



Position $f(t)$ $\xRightarrow{\text{derivative}}$

Velocity $v(t) = f'(t)$

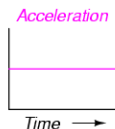
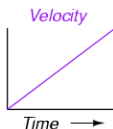
$\xRightarrow{\text{derivative}}$

Acceleration $a(t) = v'(t)$

The Notion of Derivative

The derivative $f'(t) = \lim_{\Delta \rightarrow 0} \frac{f(t+\Delta) - f(t)}{\Delta}$ signals an instantaneous rate of change of a function f with respect to the input variable t .

- Physics:



Position $f(t)$ $\xRightarrow{\text{derivative}}$ Velocity $v(t) = f'(t)$
 $\xRightarrow{\text{derivative}}$ Acceleration $a(t) = v'(t)$

- Economics:** marginal cost, marginal revenue, marginal propensity to consume ([Haavelmo, 1947](#)) are all related to derivatives.

Derivative and Causation

Derivatives measure rates of change over infinitesimal neighborhoods.

$$\begin{array}{lll} \text{Position } f(t) & \xRightarrow{\text{derivative}} & \text{Velocity } v(t) = f'(t) \\ & \xRightarrow{\text{derivative}} & \text{Acceleration } a(t) = v'(t) \end{array}$$

Derivative and Causation

Derivatives measure rates of change over infinitesimal neighborhoods.

$$\begin{array}{lcl} \text{Position } f(t) & \xRightarrow{\text{derivative}} & \text{Velocity } v(t) = f'(t) \\ & \xRightarrow{\text{derivative}} & \text{Acceleration } a(t) = v'(t) \end{array}$$

Given the values $v(t_0)$ and $f(t_0)$,

$$\begin{array}{lcl} \text{Acceleration } a(t) = v'(t) & \xRightarrow{\text{"cause"}} & \text{Velocity } v(t) \text{ over } [t_0, t_1], \\ \text{Velocity } v(t) = f'(t) & \xRightarrow{\text{"cause"}} & \text{Position } f(t) \text{ over } [t_0, t_1]. \end{array}$$

Derivative and Causation

Derivatives measure rates of change over infinitesimal neighborhoods.

$$\begin{array}{lcl} \text{Position } f(t) & \xRightarrow{\text{derivative}} & \text{Velocity } v(t) = f'(t) \\ & \xRightarrow{\text{derivative}} & \text{Acceleration } a(t) = v'(t) \end{array}$$

Given the values $v(t_0)$ and $f(t_0)$,

$$\begin{array}{lcl} \text{Acceleration } a(t) = v'(t) & \xRightarrow{\text{"cause"}} & \text{Velocity } v(t) \text{ over } [t_0, t_1], \\ \text{Velocity } v(t) = f'(t) & \xRightarrow{\text{"cause"}} & \text{Position } f(t) \text{ over } [t_0, t_1]. \end{array}$$

"The fundamental causal laws must use present properties and past neighborhood properties to determine future neighborhood properties ... the fundamental laws ... must involve some neighbourhood properties as well. And the most natural sort of neighbourhood property appears to be derivative."

Brit. J. Phil. Sci. **65** (2014), 845–862

Why Physics Uses Second Derivatives

Kenny Easwaran

Quoted from pp.857 of [Easwaran \(2014\)](#). This view is also defended in Chapter 1 of [Lange \(2002\)](#).

The Role of Derivatives in Causal Inference

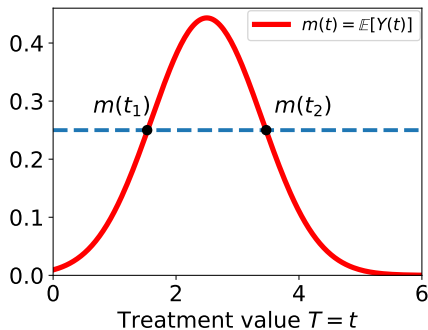
Goal: We want to study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.

- $\mathbb{E}[Y(t)]$ = mean potential outcome under a static intervention $T = t$.

The Role of Derivatives in Causal Inference

Goal: We want to study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.

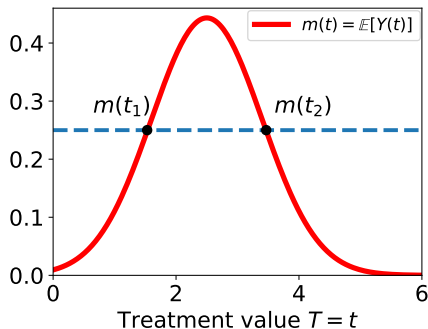
- $\mathbb{E}[Y(t)]$ = mean potential outcome under a static intervention $T = t$.
- When t varies in a continuous space, $t \mapsto \mathbb{E}[Y(t)] := m(t)$ is a curve!



The Role of Derivatives in Causal Inference

Goal: We want to study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.

- $\mathbb{E}[Y(t)]$ = mean potential outcome under a static intervention $T = t$.
- When t varies in a continuous space, $t \mapsto \mathbb{E}[Y(t)] := m(t)$ is a curve!



- While $m(t_1) = m(t_2)$, the derivative effects $m'(t_1), m'(t_2)$ are distinct!
- The derivative effect curve $\theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ is a continuous generalization to the average treatment effect $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$.

Our causal estimand of interest is the **derivative effect curve**

$$t \mapsto \theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] \quad \text{for } t \in \mathcal{T}.$$

Our causal estimand of interest is the **derivative effect curve**

$$t \mapsto \theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] \quad \text{for } t \in \mathcal{T}.$$

Problem: $\theta(t)$ is non-regular and cannot be estimated in a \sqrt{n} -rate.

Our causal estimand of interest is the **derivative effect curve**

$$t \mapsto \theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] \quad \text{for } t \in \mathcal{T}.$$

Problem: $\theta(t)$ is non-regular and cannot be estimated in a \sqrt{n} -rate.

There are some closely related but distinct estimands:

- *Incremental Causal/Treatment Effect* (Kennedy, 2019; Rothenhäusler and Yu, 2019):

$$\mathbb{E}[Y(T + \delta)] - \mathbb{E}[Y(T)]$$

for some small deterministic $\delta > 0$.

Our causal estimand of interest is the **derivative effect curve**

$$t \mapsto \theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] \quad \text{for } t \in \mathcal{T}.$$

Problem: $\theta(t)$ is non-regular and cannot be estimated in a \sqrt{n} -rate.

There are some closely related but distinct estimands:

- *Incremental Causal/Treatment Effect* (Kennedy, 2019; Rothenhäusler and Yu, 2019):

$$\mathbb{E}[Y(T + \delta)] - \mathbb{E}[Y(T)]$$

for some small deterministic $\delta > 0$.

- *Average Derivative/Partial Effect* (Powell et al., 1989; Newey and Stoker, 1993):

$$\mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}(Y|T, \mathbf{S}) \right] = \mathbb{E}[\theta(T)],$$

where $\mathbf{S} \in \mathcal{S} \subset \mathbb{R}^d$ is a covariate vector.

Basic Framework and Assumptions

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$, the following assumptions are generally imposed.

Assumption (Identification Conditions)

- 1 (Consistency) $Y_i = Y_i(t)$ whenever $T_i = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y_i(t) \perp\!\!\!\perp T_i \mid \mathbf{S}_i$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

¹Some mild assumptions are needed; see Theorem 1.1 in (Shao, 2003).

Basic Framework and Assumptions

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, the following assumptions are generally imposed.

Assumption (Identification Conditions)

- 1 (Consistency) $Y_i = Y_i(t)$ whenever $T_i = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y_i(t) \perp\!\!\!\perp T_i \mid S_i$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|S}(t|s) \geq p_{\min} > 0$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

$$\theta(t) = \frac{d}{dt} \mathbb{E} [\mu(t, S)] \stackrel{(*)^1}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, S) \right] \quad \text{with} \quad \mu(t, s) = \mathbb{E}(Y|T = t, S = s).$$

¹Some mild assumptions are needed; see Theorem 1.1 in (Shao, 2003).

Basic Framework and Assumptions

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$, the following assumptions are generally imposed.

Assumption (Identification Conditions)

- ① (*Consistency*) $Y_i = Y_i(t)$ whenever $T_i = t \in \mathcal{T}$.
- ② (*Ignorability or Unconfoundedness*) $Y_i(t) \perp\!\!\!\perp T_i \mid \mathbf{S}_i$ for all $t \in \mathcal{T}$.
- ③ (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

$$\theta(t) = \frac{d}{dt} \mathbb{E} [\mu(t, \mathbf{S})] \stackrel{(*)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right] \quad \text{with} \quad \mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s}).$$

- ① Estimating (partial) derivatives is a challenging problem ([Dai et al., 2016](#)).
 - Data generally come from $Y_i = \mu(T_i, \mathbf{S}_i) + \epsilon_i$ but not $Y'_i = \frac{\partial}{\partial t} \mu(T_i, \mathbf{S}_i) + \epsilon'_i$.

¹Some mild assumptions are needed; see Theorem 1.1 in ([Shao, 2003](#)).

Basic Framework and Assumptions

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, the following assumptions are generally imposed.

Assumption (Identification Conditions)

- 1 (Consistency) $Y_i = Y_i(t)$ whenever $T_i = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y_i(t) \perp\!\!\!\perp T_i \mid S_i$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|S}(t|s) \geq p_{\min} > 0$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

$$\theta(t) = \frac{d}{dt} \mathbb{E} [\mu(t, S)] \stackrel{(*)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, S) \right] \quad \text{with} \quad \mu(t, s) = \mathbb{E}(Y|T = t, S = s).$$

- 1 Estimating (partial) derivatives is a challenging problem ([Dai et al., 2016](#)).
 - Data generally come from $Y_i = \mu(T_i, S_i) + \epsilon_i$ but not $Y'_i = \frac{\partial}{\partial t} \mu(T_i, S_i) + \epsilon'_i$.
- 2 Positivity is a strong assumption with continuous treatments!

¹Some mild assumptions are needed; see Theorem 1.1 in ([Shao, 2003](#)).

An Example of the Positivity Violation

Assumption (Positivity Condition)

There exists a constant $p_{\min} > 0$ such that $p_{T|S}(t|s) \geq p_{\min}$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

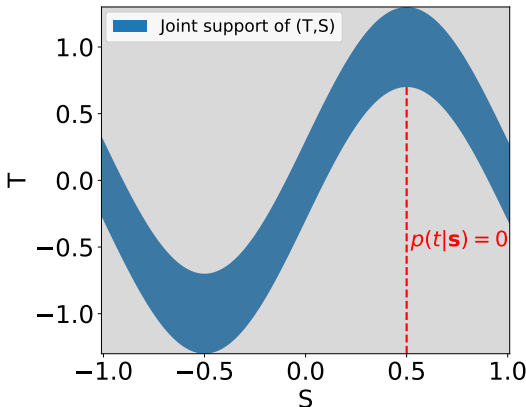
$$T = \sin(\pi S) + E, \quad E \sim \text{Unif}[-0.3, 0.3], \quad S \sim \text{Unif}[-1, 1], \quad \text{and} \quad E \perp\!\!\!\perp S.$$

An Example of the Positivity Violation

Assumption (Positivity Condition)

There exists a constant $p_{\min} > 0$ such that $p_{T|S}(t|s) \geq p_{\min}$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

$T = \sin(\pi S) + E$, $E \sim \text{Unif}[-0.3, 0.3]$, $S \sim \text{Unif}[-1, 1]$, and $E \perp\!\!\!\perp S$.



► **Note:** $p(t|s) = 0$ in the gray regions, and the positivity condition fails.

Highlights of Today's Talk

Under the positivity condition:

- ① We propose doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.
 - Deriving a DR estimator for $\theta(t)$ is more intricate than for $m(t)$.

Highlights of Today's Talk

Under the positivity condition:

- ① We propose doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.
 - Deriving a DR estimator for $\theta(t)$ is more intricate than for $m(t)$.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{ \begin{array}{l} \Rightarrow \\ \nRightarrow \end{array} \right.$ DR

Highlights of Today's Talk

Under the positivity condition:

- ① We propose doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.
 - Deriving a DR estimator for $\theta(t)$ is more intricate than for $m(t)$.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{ \begin{array}{l} \Rightarrow \\ \nRightarrow \end{array} \right.$ DR

Without the positivity condition:

- ② $m(t)$ and $\theta(t)$ are identifiable with an additive structural assumption:

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon. \quad (1)$$

Highlights of Today's Talk

Under the positivity condition:

- 1 We propose doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.
 - Deriving a DR estimator for $\theta(t)$ is more intricate than for $m(t)$.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{ \begin{array}{l} \Rightarrow \\ \nRightarrow \end{array} \right.$ DR

Without the positivity condition:

- 2 $m(t)$ and $\theta(t)$ are identifiable with an additive structural assumption:

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon. \quad (1)$$

- 3 However, the usual IPW estimators of $m(t)$ and $\theta(t)$ are still biased even under model (1).
 - These biases are due to the support discrepancy.

Highlights of Today's Talk

Under the positivity condition:

- 1 We propose doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.
 - Deriving a DR estimator for $\theta(t)$ is more intricate than for $m(t)$.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{ \begin{array}{l} \Rightarrow \\ \nRightarrow \end{array} \right.$ DR

Without the positivity condition:

- 2 $m(t)$ and $\theta(t)$ are identifiable with an additive structural assumption:

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon. \quad (1)$$

- 3 However, the usual IPW estimators of $m(t)$ and $\theta(t)$ are still biased even under model (1).
 - These biases are due to the support discrepancy.
- 4 We propose our bias-corrected IPW and DR estimators of $\theta(t)$.
 - Our approach establishes an interesting connection to nonparametric support and level set estimation problems.

- ① Introduction
- ② Inference Theory for $\theta(t)$ Under Positivity
- ③ Inference Theory for $\theta(t)$ Without Positivity
- ④ Simulations and Case Study
- ⑤ Discussion



Assumption (Identification Conditions)

- ① (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- ② (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.
- ③ (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

Given that $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$, we have

RA or G-computation:
$$\begin{cases} m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})], \\ \theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \frac{d}{dt} \mathbb{E}[\mu(t, \mathbf{S})] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, \mathbf{S})\right]. \end{cases}$$

Assumption (Identification Conditions)

- 1 (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

Given that $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$, we have

RA or G-computation:
$$\begin{cases} m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})], \\ \theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \frac{d}{dt} \mathbb{E}[\mu(t, \mathbf{S})] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, \mathbf{S})\right]. \end{cases}$$

IPW:
$$\begin{cases} m(t) = \mathbb{E}[Y(t)] = \lim_{h \rightarrow 0} \mathbb{E}\left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})}\right], \\ \theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = ??? \end{cases}$$

Here, $K: \mathbb{R} \rightarrow [0, \infty)$ is a kernel function and $h > 0$ is a smoothing bandwidth parameter.

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right]$$

from the data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$.

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right]$$

from the data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$.

① **RA Estimator** (Robins, 1986; Gill and Robins, 2001):

$$\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).$$

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right]$$

from the data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$.

① **RA Estimator** (Robins, 1986; Gill and Robins, 2001):

$$\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).$$

② **IPW Estimator** (Hirano and Imbens, 2004; Imai and van Dyk, 2004):

$$\hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot Y_i.$$

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right]$$

from the data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$.

① **RA Estimator** (Robins, 1986; Gill and Robins, 2001):

$$\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).$$

② **IPW Estimator** (Hirano and Imbens, 2004; Imai and van Dyk, 2004):

$$\hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot Y_i.$$

③ **DR Estimator** (Kallus and Zhou, 2018; Colangelo and Lee, 2020):

$$\hat{m}_{\text{DR}}(t) = \frac{1}{nh} \sum_{i=1}^n \left\{ \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot [Y_i - \hat{\mu}(t, \mathbf{S}_i)] + h \cdot \hat{\mu}(t, \mathbf{S}_i) \right\}.$$

RA and IPW Estimators of $\theta(t)$ Under Positivity

To estimate $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, \mathbf{S})\right]$ from $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$, we also have three strategies:

1 RA Estimator:

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{with} \quad \beta(t, \mathbf{s}) = \frac{\partial}{\partial t} \mu(t, \mathbf{s}).$$

RA and IPW Estimators of $\theta(t)$ Under Positivity

To estimate $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, \mathbf{S})\right]$ from $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$, we also have three strategies:

① RA Estimator:

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{with} \quad \beta(t, \mathbf{s}) = \frac{\partial}{\partial t} \mu(t, \mathbf{s}).$$

Question: How to generalize the IPW form $m(t) = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right]$ to identifying $\theta(t)$?

RA and IPW Estimators of $\theta(t)$ Under Positivity

To estimate $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, \mathbf{S})\right]$ from $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$, we also have three strategies:

1 RA Estimator:

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{with} \quad \beta(t, \mathbf{s}) = \frac{\partial}{\partial t} \mu(t, \mathbf{s}).$$

Question: How to generalize the IPW form $m(t) = \lim_{h \rightarrow 0} \mathbb{E}\left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})}\right]$ to identifying $\theta(t)$?

2 IPW Estimator: Inspired by the derivative estimator in [Mack and Müller \(1989\)](#), we propose

$$\hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \quad \text{with} \quad \kappa_2 = \int u^2 K(u) du.$$

Challenges of Deriving a DR Estimator of $\theta(t)$

The usual approach to construct a DR (or AIPW) estimator is as follows:

$$\begin{aligned}\hat{m}_{\text{RA}}(t) &= \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i) & \text{"+"} & \hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot Y_i \\ \implies \hat{m}_{\text{DR}}(t) &= \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot [Y_i - \hat{\mu}(t, \mathbf{S}_i)] + \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).\end{aligned}$$

Challenges of Deriving a DR Estimator of $\theta(t)$

The usual approach to construct a DR (or AIPW) estimator is as follows:

$$\begin{aligned}\hat{m}_{\text{RA}}(t) &= \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i) & \text{"+"} & \hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i \\ \implies \hat{m}_{\text{DR}}(t) &= \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot [Y_i - \hat{\mu}(t, \mathbf{S}_i)] + \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).\end{aligned}$$

This “naive” combining approach does not work for constructing a DR estimator of $\theta(t)$:

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{"+"} \quad \hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i$$

Challenges of Deriving a DR Estimator of $\theta(t)$

The usual approach to construct a DR (or AIPW) estimator is as follows:

$$\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i) \quad \text{"+"} \quad \hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i$$

$$\Rightarrow \hat{m}_{\text{DR}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot [Y_i - \hat{\mu}(t, \mathbf{S}_i)] + \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i).$$

This “naive” combining approach does not work for constructing a DR estimator of $\theta(t)$:

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{"+"} \quad \hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i$$

$$\Rightarrow$$

- $\hat{\theta}_{\text{AIPW},1}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \left[Y_i - \hat{\beta}(t, \mathbf{S}_i) \right] + \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i);$
- $\hat{\theta}_{\text{AIPW},2}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i-t}{h}\right)}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)} \left[\frac{Y_i}{h \cdot \kappa_2} \left(\frac{T_i-t}{h}\right) - \hat{\beta}(t, \mathbf{S}_i) \right] + \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i); \text{ etc.}$

Remark: All these AIPW estimators are, at best, singly robust!!

Doubly Robust Estimator of $\theta(t)$ Under Positivity

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{"+"} \quad \hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i | \mathbf{S}_i)} \cdot Y_i$$

\Rightarrow

Doubly Robust Estimator of $\theta(t)$ Under Positivity

$$\begin{aligned}\hat{\theta}_{\text{RA}}(t) &= \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) & \text{“+”} & \quad \hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i \\ \Rightarrow & & & \\ \hat{\theta}_{\text{DR}}(t) &= \underbrace{\frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \left[Y_i - \hat{\mu}(t, \mathbf{S}_i) - (T_i - t) \cdot \hat{\beta}(t, \mathbf{S}_i) \right]}_{\text{IPW component}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i)}_{\text{RA component}}.\end{aligned}$$

- 1 The “IPW component” leverages a local polynomial approximation to push the residual to (roughly) second order.
 - Neyman orthogonality (Neyman, 1959; Chernozhukov et al., 2018) holds as $h \rightarrow 0$.

Doubly Robust Estimator of $\theta(t)$ Under Positivity

$$\hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i) \quad \text{“+”} \quad \hat{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \cdot Y_i$$

\Rightarrow

$$\hat{\theta}_{\text{DR}}(t) = \underbrace{\frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right)}{\kappa_2 \cdot \hat{p}_{T|S}(T_i|\mathbf{S}_i)} \left[Y_i - \hat{\mu}(t, \mathbf{S}_i) - (T_i - t) \cdot \hat{\beta}(t, \mathbf{S}_i) \right]}_{\text{IPW component}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, \mathbf{S}_i)}_{\text{RA component}}.$$

- 1 The “IPW component” leverages a local polynomial approximation to push the residual to (roughly) second order.
 - Neyman orthogonality (Neyman, 1959; Chernozhukov et al., 2018) holds as $h \rightarrow 0$.
- 2 Different from $\hat{m}_{\text{IPW}}(t)$ and $\hat{m}_{\text{DR}}(t)$, we must compute the inverse probability weights as $\frac{1}{\hat{p}_{T|S}(T_i|\mathbf{S}_i)}$ but not $\frac{1}{\hat{p}_{T|S}(t|\mathbf{S}_i)}$ for $i = 1, \dots, n$.

Theorem (Theorem 1 in Zhang and Chen 2025)

Under some regularity assumptions and

① $\hat{\mu}, \hat{\beta}, \hat{p}_{T|S}$ are estimated on a dataset independent of $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$;

② at least one of the model specification conditions hold:

- $\hat{p}_{T|S}(t|\mathbf{s}) \xrightarrow{P} \bar{p}_{T|S}(t|\mathbf{s}) = p_{T|S}(t|\mathbf{s})$ (conditional density model),
- $\hat{\mu}(t, \mathbf{s}) \xrightarrow{P} \bar{\mu}(t, \mathbf{s}) = \mu(t, \mathbf{s})$ and $\hat{\beta}(t, \mathbf{s}) \xrightarrow{P} \bar{\beta}(t, \mathbf{s}) = \beta(t, \mathbf{s})$ (outcome model);

③
$$\sup_{|u-t| \leq h} \left\| \hat{p}_{T|S}(u|\mathbf{S}) - p_{T|S}(u|\mathbf{S}) \right\|_{L_2} \left[\left\| \hat{\mu}(t, \mathbf{S}) - \mu(t, \mathbf{S}) \right\|_{L_2} + h \left\| \hat{\beta}(t, \mathbf{S}) - \beta(t, \mathbf{S}) \right\|_{L_2} \right] = o_P \left(\frac{1}{\sqrt{nh}} \right),$$

we prove that

Theorem (Theorem 1 in Zhang and Chen 2025)

Under some regularity assumptions and

- ① $\hat{\mu}, \hat{\beta}, \hat{p}_{T|S}$ are estimated on a dataset independent of $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$;
- ② at least one of the model specification conditions hold:
 - $\hat{p}_{T|S}(t|s) \xrightarrow{P} \bar{p}_{T|S}(t|s) = p_{T|S}(t|s)$ (conditional density model),
 - $\hat{\mu}(t, s) \xrightarrow{P} \bar{\mu}(t, s) = \mu(t, s)$ and $\hat{\beta}(t, s) \xrightarrow{P} \bar{\beta}(t, s) = \beta(t, s)$ (outcome model);
- ③ $\sup_{|u-t| \leq h} \|\hat{p}_{T|S}(u|S) - p_{T|S}(u|S)\|_{L_2} \left[\|\hat{\mu}(t, S) - \mu(t, S)\|_{L_2} + h \|\hat{\beta}(t, S) - \beta(t, S)\|_{L_2} \right] = o_P\left(\frac{1}{\sqrt{nh}}\right),$

we prove that

- $\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{h,t}(Y_i, T_i, \mathbf{S}_i; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) + o_P(1).$
- $\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) - h^2 B_{\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{\theta}(t)).$

We can conduct asymptotically valid inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ using

$$\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) - h^2 B_{\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{\theta}(t)).$$

Statistical Inference on $\theta(t)$

We can conduct asymptotically valid inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ using

$$\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) - h^2 B_{\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{\theta}(t)).$$

① We estimate $V_{\theta}(t) = \mathbb{E} \left[\phi_{h,t}^2(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) \right]$ with

$$\phi_{h,t}(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) = \frac{\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{\sqrt{h} \cdot \kappa_2 \cdot \bar{p}_{T|S}(T|\mathbf{S})} \cdot [Y - \bar{\mu}(t, \mathbf{S}) - (T-t) \cdot \bar{\beta}(t, \mathbf{S})]$$

$$\text{by } \hat{V}_{\theta}(t) = \frac{1}{n} \sum_{i=1}^n \phi_{h,t}^2(Y, T, \mathbf{S}; \hat{\mu}, \hat{\beta}, \hat{p}_{T|S}).$$

Statistical Inference on $\theta(t)$

We can conduct asymptotically valid inference on $\theta(t) = \frac{d}{dt} \mathbb{E} [Y(t)]$ using

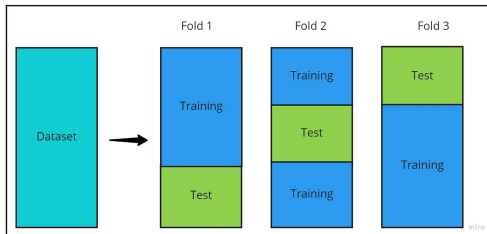
$$\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) - h^2 B_{\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{\theta}(t)).$$

- ① We estimate $V_{\theta}(t) = \mathbb{E} \left[\phi_{h,t}^2(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) \right]$ with

$$\phi_{h,t}(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) = \frac{\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{\sqrt{h} \cdot \kappa_2 \cdot \bar{p}_{T|S}(T|S)} \cdot [Y - \bar{\mu}(t, \mathbf{S}) - (T - t) \cdot \bar{\beta}(t, \mathbf{S})]$$

$$\text{by } \hat{V}_{\theta}(t) = \frac{1}{n} \sum_{i=1}^n \phi_{h,t}^2(Y, T, \mathbf{S}; \hat{\mu}, \hat{\beta}, \hat{p}_{T|S}).$$

- ② $\hat{\mu}, \hat{\beta}, \hat{p}_{T|S}$ can be estimated via sample-splitting or cross-fitting.



Statistical Inference on $\theta(t)$

We can conduct asymptotically valid inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ using

$$\sqrt{nh^3} \left[\hat{\theta}_{\text{DR}}(t) - \theta(t) - h^2 B_{\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{\theta}(t)).$$

- ① We estimate $V_{\theta}(t) = \mathbb{E} \left[\phi_{h,t}^2(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) \right]$ with

$$\phi_{h,t}(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) = \frac{\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{\sqrt{h} \cdot \kappa_2 \cdot \bar{p}_{T|S}(T|\mathbf{S})} \cdot [Y - \bar{\mu}(t, \mathbf{S}) - (T-t) \cdot \bar{\beta}(t, \mathbf{S})]$$

$$\text{by } \hat{V}_{\theta}(t) = \frac{1}{n} \sum_{i=1}^n \phi_{h,t}^2(Y, T, \mathbf{S}; \hat{\mu}, \hat{\beta}, \hat{p}_{T|S}).$$

- ② $\hat{\mu}, \hat{\beta}, \hat{p}_{T|S}$ can be estimated via sample-splitting or cross-fitting.
- ③ The explicit form of $B_{\theta}(t)$ is complicated, but $h^2 B_{\theta}(t)$ is asymptotically negligible when $h = O\left(n^{-\frac{1}{5}}\right)$.
- This order aligns with the outputs from usual bandwidth selection methods (Wand and Jones, 1994; Wasserman, 2006).

- ① Introduction
- ② Inference Theory for $\theta(t)$ Under Positivity
- ③ Inference Theory for $\theta(t)$ Without Positivity
- ④ Simulations and Case Study
- ⑤ Discussion



Why Do We Need Positivity?

Assumption (Identification Conditions)

- 1 (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

The RA (or G-computation) formulae are given by

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] \quad \text{and} \quad \theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right].$$

The IPW approaches also rely on the following identities:

$$\lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right] = \mathbb{E}[\mu(t, \mathbf{S})] \quad \text{and} \quad \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{\kappa_2 h^2 p_{T|\mathbf{S}}(T|\mathbf{S})} \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right].$$

Why Do We Need Positivity?

Assumption (Identification Conditions)

- 1 (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.
- 3 (**Positivity**) $p_{T|\mathbf{S}}(t|\mathbf{s}) \geq p_{\min} > 0$ for all $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$.

The RA (or G-computation) formulae are given by

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})] \quad \text{and} \quad \theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right].$$

The IPW approaches also rely on the following identities:

$$\lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right] = \mathbb{E}[\mu(t, \mathbf{S})] \quad \text{and} \quad \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{\kappa_2 h^2 p_{T|\mathbf{S}}(T|\mathbf{S})} \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right].$$

Identification Issue: Without positivity, $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$ is *not well-defined* outside the support $\mathcal{J} \subset \mathcal{T} \times \mathcal{S}$ of the joint density $p(t, \mathbf{s})$.

Assumption (Identification Conditions)

- 1 (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.

Identification Strategy Without Positivity

Assumption (Identification Conditions)

- 1 (Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- 2 (Ignorability or Unconfoundedness) $Y(t) \perp\!\!\!\perp T \mid \mathbf{S}$ for all $t \in \mathcal{T}$.

Assumption (Extrapolation condition; Zhang et al. 2024)

Suppose that at least one of the following conditions are valid.

- 1 The function $\mathbb{E}[Y(t)|\mathbf{S} = \mathbf{s}]$ is continuously differentiable with respect to t for any $(t, \mathbf{s}) \in \mathcal{T} \times \mathcal{S}$ with $p(\mathbf{s}|t) > 0$ and

$$\theta(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \mid T = t \right].$$

- 2 The potential outcome $Y(t)$ is continuously differentiable with respect to t and

$$\theta(t) = \mathbb{E} \left[\mathbb{E} \left[\frac{\partial}{\partial t} Y(t) \mid \mathbf{S} \right] \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{\partial}{\partial t} Y(t) \mid \mathbf{S} \right] \mid T = t \right].$$

Additionally, it holds true that $\mathbb{E}(Y) = \mathbb{E}[m(T)]$.

Identification Strategy Without Positivity

If $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \mid T = t \right]$ holds true, then

Identification Strategy Without Positivity

If $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T = t \right]$ holds true, then

$$\theta(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T = t \right]$$

$$\stackrel{(*)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|T = t, S] \middle| T = t \right] \quad (*) \text{ Ignorability}$$

$$\stackrel{(**)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}(Y|T = t, S) \middle| T = t \right] \quad (**) \text{ Consistency}$$

$$= \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, S) \middle| T = t \right] := \theta_C(t).$$

Identification Strategy Without Positivity

If $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S]\right] = \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \Big| T = t\right]$ holds true, then

$$\begin{aligned}\theta(t) &= \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \Big| T = t\right] \\ &\stackrel{(*)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|T = t, S] \Big| T = t\right] && (*) \text{ Ignorability} \\ &\stackrel{(**)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}(Y|T = t, S) \Big| T = t\right] && (**) \text{ Consistency} \\ &= \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, S) \Big| T = t\right] := \theta_C(t).\end{aligned}$$

- For any $t \in \mathcal{T}$, the fundamental theorem of calculus reveals that

$$m(t) = m(T) + \int_{\tilde{T}=T}^{\tilde{T}=t} m'(\tilde{t}) d\tilde{t} = m(T) + \int_{\tilde{T}=T}^{\tilde{T}=t} \theta(\tilde{t}) d\tilde{t}.$$

Identification Strategy Without Positivity

If $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T = t \right]$ holds true, then

$$\begin{aligned}\theta(t) &= \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T = t \right] \\ &\stackrel{(*)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|T = t, S] \middle| T = t \right] && (*) \text{ Ignorability} \\ &\stackrel{(**)}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}(Y|T = t, S) \middle| T = t \right] && (**) \text{ Consistency} \\ &= \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, S) \middle| T = t \right] := \theta_C(t).\end{aligned}$$

- For any $t \in \mathcal{T}$, the fundamental theorem of calculus reveals that

$$m(t) = m(T) + \int_{\tilde{T}=T}^{\tilde{T}=t} m'(\tilde{t}) d\tilde{t} = m(T) + \int_{\tilde{T}=T}^{\tilde{T}=t} \theta(\tilde{t}) d\tilde{t}.$$

- Taking the expectation on both sides of the above equality yields that

$$m(t) = \mathbb{E}(Y) + \mathbb{E} \left\{ \int_{\tilde{T}=T}^{\tilde{T}=t} \mathbb{E} \left[\frac{\partial}{\partial t} \mu(\tilde{t}, S) \middle| T = \tilde{t} \right] d\tilde{t} \right\}.$$

Validity of Our Identification Strategies Without Positivity

We identify $\theta(t)$ through

$$\theta_C(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}(Y|T = t, \mathbf{S}) \middle| T = t \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \middle| T = t \right].$$

Validity of Our Identification Strategies Without Positivity

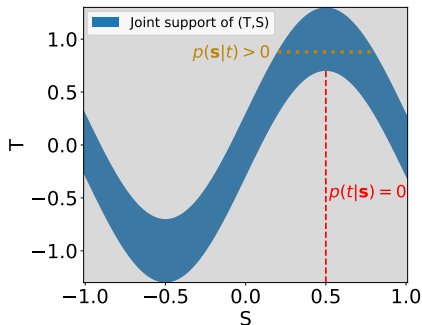
We identify $\theta(t)$ through

$$\theta_C(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}(Y|T = t, \mathbf{S}) \middle| T = t \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \middle| T = t \right].$$

- In contrast to the identification via $\mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, \mathbf{S}) \right]$ under positivity, we only need

$$\frac{\partial}{\partial t} \mu(t, \mathbf{s}) = \frac{\partial}{\partial t} \mathbb{E}(Y|T = t, \mathbf{s})$$

to be well-defined when $p(\mathbf{s}|t) > 0$.



Key Example: Additive Confounding Model

$$\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \mid T = t \right].$$

Key Example: Additive Confounding Model

$$\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \mid T = t \right].$$

Proposition 2 in [Zhang et al. \(2024\)](#) shows that the above equality holds under an additive structural assumption

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon.$$

- $\bar{m} : \mathcal{T} \rightarrow \mathbb{R}$ and $\eta : \mathcal{S} \rightarrow \mathbb{R}$ are deterministic functions.
- $\epsilon \in \mathbb{R}$ is an independent noise variable with $\mathbb{E}(\epsilon) = 0$ and $\text{Var}(\epsilon) > 0$.
- $m(t) = \mathbb{E}[Y(t)] = \bar{m}(t) + \mathbb{E}[\eta(\mathbf{S})]$ and $\theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \bar{m}'(t)$.

Key Example: Additive Confounding Model

$$\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \mid T = t \right].$$

Proposition 2 in [Zhang et al. \(2024\)](#) shows that the above equality holds under an additive structural assumption

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon.$$

- $\bar{m} : \mathcal{T} \rightarrow \mathbb{R}$ and $\eta : \mathcal{S} \rightarrow \mathbb{R}$ are deterministic functions.
- $\epsilon \in \mathbb{R}$ is an independent noise variable with $\mathbb{E}(\epsilon) = 0$ and $\text{Var}(\epsilon) > 0$.
- $m(t) = \mathbb{E}[Y(t)] = \bar{m}(t) + \mathbb{E}[\eta(\mathbf{S})]$ and $\theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \bar{m}'(t)$.
- **Identification:**

$$m(t) = \mathbb{E} \left[Y + \int_{\tilde{t}=T}^{\tilde{t}=t} \theta(\tilde{t}) d\tilde{t} \right], \quad \theta(t) = \int \frac{\partial}{\partial t} \mu(t, \mathbf{s}) dF_{\mathbf{S}|T}(\mathbf{s}|t).$$

Key Example: Additive Confounding Model

$$\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \right] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|\mathbf{S}] \mid T = t \right].$$

Proposition 2 in [Zhang et al. \(2024\)](#) shows that the above equality holds under an additive structural assumption

$$Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon.$$

- $\bar{m} : \mathcal{T} \rightarrow \mathbb{R}$ and $\eta : \mathcal{S} \rightarrow \mathbb{R}$ are deterministic functions.
- $\epsilon \in \mathbb{R}$ is an independent noise variable with $\mathbb{E}(\epsilon) = 0$ and $\text{Var}(\epsilon) > 0$.
- $m(t) = \mathbb{E}[Y(t)] = \bar{m}(t) + \mathbb{E}[\eta(\mathbf{S})]$ and $\theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[Y(t)] = \bar{m}'(t)$.
- **Identification:**

$$m(t) = \mathbb{E} \left[Y + \int_{\tilde{t}=T}^{\tilde{t}=t} \theta(\tilde{t}) d\tilde{t} \right], \quad \theta(t) = \int \frac{\partial}{\partial t} \mu(t, \mathbf{s}) dF_{\mathbf{S}|T}(\mathbf{s}|t).$$

- **RA estimator without positivity** ([Zhang et al., 2024](#)):

$$\hat{m}_{\text{C,RA}}(t) = \frac{1}{n} \sum_{i=1}^n \left[Y_i + \int_{\tilde{t}=T_i}^{\tilde{t}=t} \hat{\theta}_{\text{C,RA}}(\tilde{t}) d\tilde{t} \right], \quad \hat{\theta}_{\text{C,RA}}(t) = \int \hat{\beta}(t, \mathbf{s}) d\hat{F}_{\mathbf{S}|T}(\mathbf{s}|t).$$

Estimation Biases of IPW Estimators Without Positivity

Question: How about IPW and DR estimators of $\theta(t)$ without positivity?

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon$.

Estimation Biases of IPW Estimators Without Positivity

Question: How about IPW and DR estimators of $\theta(t)$ without positivity?

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon$.
- Consider usual (oracle) IPW estimators of $m(t)$ and $\theta(t)$ as:

$$\tilde{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{Y_i \cdot K\left(\frac{T_i - t}{h}\right)}{p_{T|\mathbf{S}}(T_i|\mathbf{S}_i)}, \quad \tilde{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot p_{T|\mathbf{S}}(T_i|\mathbf{S}_i)}.$$

Estimation Biases of IPW Estimators Without Positivity

Question: How about IPW and DR estimators of $\theta(t)$ without positivity?

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon$.
- Consider usual (oracle) IPW estimators of $m(t)$ and $\theta(t)$ as:

$$\tilde{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{Y_i \cdot K\left(\frac{T_i - t}{h}\right)}{p_{T|\mathbf{S}}(T_i|\mathbf{S}_i)}, \quad \tilde{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot p_{T|\mathbf{S}}(T_i|\mathbf{S}_i)}.$$

- We show in Proposition 2 of [Zhang and Chen \(2025\)](#) that

$$\lim_{h \rightarrow 0} \mathbb{E} [\tilde{m}_{\text{IPW}}(t)] = \bar{m}(t) \cdot \rho(t) + \omega(t) \neq m(t),$$
$$\lim_{h \rightarrow 0} \mathbb{E} [\tilde{\theta}_{\text{IPW}}(t)] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

where $\rho(t) = \mathbb{P}(\mathbf{S} \in \mathcal{S}(t))$ and $\omega(t) = \mathbb{E} [\eta(\mathbf{S}) \mathbb{1}_{\{\mathbf{S} \in \mathcal{S}(t)\}}]$.

Estimation Biases of IPW Estimators Without Positivity

Question: How about IPW and DR estimators of $\theta(t)$ without positivity?

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon$.
- Consider usual (oracle) IPW estimators of $m(t)$ and $\theta(t)$ as:

$$\tilde{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{Y_i \cdot K\left(\frac{T_i - t}{h}\right)}{p_{T|S}(T_i | \mathbf{S}_i)}, \quad \tilde{\theta}_{\text{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot p_{T|S}(T_i | \mathbf{S}_i)}.$$

- We show in Proposition 2 of [Zhang and Chen \(2025\)](#) that

$$\lim_{h \rightarrow 0} \mathbb{E} [\tilde{m}_{\text{IPW}}(t)] = \bar{m}(t) \cdot \rho(t) + \omega(t) \neq m(t),$$
$$\lim_{h \rightarrow 0} \mathbb{E} [\tilde{\theta}_{\text{IPW}}(t)] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

where $\rho(t) = \mathbb{P}(\mathbf{S} \in \mathcal{S}(t))$ and $\omega(t) = \mathbb{E} [\eta(\mathbf{S}) \mathbb{1}_{\{\mathbf{S} \in \mathcal{S}(t)\}}]$.

Key Issue: The conditional support $\mathcal{S}(t)$ of $p_{S|T}(\mathbf{s}|t)$ and the marginal support \mathcal{S} of $p_S(\mathbf{s})$ are different!!

Bias-Corrected IPW Estimator of $\theta(t)$

$$\lim_{h \rightarrow 0} \mathbb{E} \left[\tilde{\theta}_{\text{IPW}}(t) \right] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right)}{h^2 \cdot \kappa_2 \cdot p_{T|\mathbf{S}}(T|\mathbf{S})} \right] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

where $\rho(t) = \mathbb{P}(\mathbf{S} \in \mathcal{S}(t))$ and $\omega(t) = \mathbb{E} [\eta(\mathbf{S}) \mathbb{1}_{\{\mathbf{S} \in \mathcal{S}(t)\}}]$.

Bias-Corrected IPW Estimator of $\theta(t)$

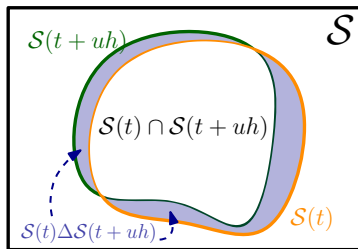
$$\lim_{h \rightarrow 0} \mathbb{E} \left[\tilde{\theta}_{\text{IPW}}(t) \right] = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S)} \right] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

where $\rho(t) = \mathbb{P}(S \in \mathcal{S}(t))$ and $\omega(t) = \mathbb{E} [\eta(S) \mathbb{1}_{\{S \in \mathcal{S}(t)\}}]$.

① We first want to disentangle $\theta(t) = \bar{m}'(t)$ from the bias term:

$$\mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right) p_{S|T}(S|T)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S) \cdot p_S(S)} \right] = \bar{m}'(t) + O(h^2)$$

$$+ \underbrace{\int_{\mathbb{R}} \mathbb{E} \left\{ [\bar{m}(t+uh) + \eta(S)] [\mathbb{1}_{\{S \in \mathcal{S}(t+uh) \setminus \mathcal{S}(t)\}} - \mathbb{1}_{\{S \in \mathcal{S}(t) \setminus \mathcal{S}(t+uh)\}}] \middle| T=t \right\} u \cdot K(u) du}_{\text{Non-vanishing Bias}}.$$



Bias-Corrected IPW Estimator of $\theta(t)$

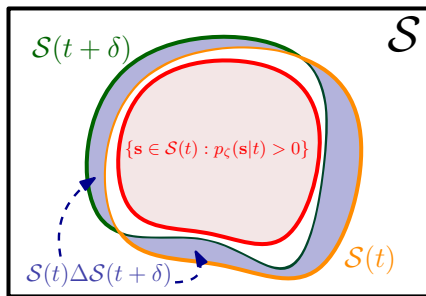
$$\mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right) p_{S|T}(\mathbf{S}|T)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|\mathbf{S}) \cdot p_S(\mathbf{S})} \right] = \bar{m}'(t) + O(h^2) + \text{“Non-vanishing Bias”}.$$

Bias-Corrected IPW Estimator of $\theta(t)$

$$\mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right) p_{S|T}(S|T)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S) \cdot p_S(S)} \right] = \bar{m}'(t) + O(h^2) + \text{“Non-vanishing Bias”}.$$

- 2 We replace $p_{S|T}(s|t)$ with a ζ -interior conditional density $p_\zeta(s|t)$ so that

$$\{s \in \mathcal{S}(t) : p_\zeta(s|t) > 0\} \subset \mathcal{S}(t + \delta) \quad \text{for any } \delta \in [-h, h].$$

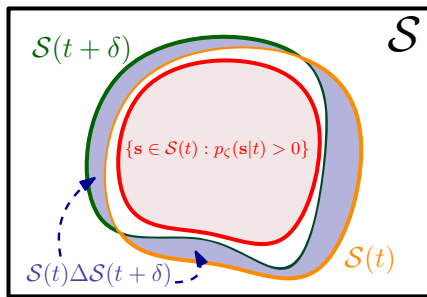


Bias-Corrected IPW Estimator of $\theta(t)$

$$\mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right) p_{S|T}(S|T)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S) \cdot p_S(S)} \right] = \bar{m}'(t) + O(h^2) + \text{“Non-vanishing Bias”}.$$

- 2 We replace $p_{S|T}(s|t)$ with a ζ -interior conditional density $p_{\zeta}(s|t)$ so that

$$\{s \in \mathcal{S}(t) : p_{\zeta}(s|t) > 0\} \subset \mathcal{S}(t + \delta) \quad \text{for any } \delta \in [-h, h].$$



- Now, we have that $\mathbb{E} \left[\frac{Y \left(\frac{T-t}{h} \right) K \left(\frac{T-t}{h} \right) p_{\zeta}(S|T)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S) \cdot p_S(S)} \right] = \bar{m}'(t) + O(h^2).$

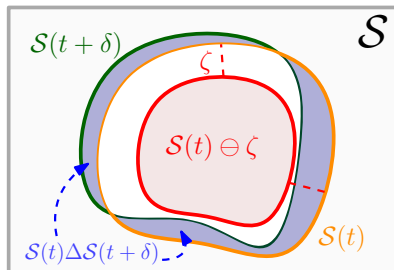
ζ -Interior Conditional Density

Question: How can we find a ζ -interior conditional density $p_{\zeta}(s|t)$?

ζ -Interior Conditional Density

Question: How can we find a ζ -interior conditional density $p_\zeta(\mathbf{s}|t)$?

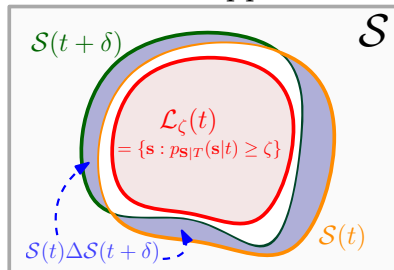
Support shrinking approach



$\mathcal{S}(t) \ominus \zeta = \left\{ \mathbf{s} \in \mathcal{S}(t) : \inf_{\mathbf{x} \in \partial \mathcal{S}(t)} \|\mathbf{s} - \mathbf{x}\|_2 \geq \zeta \right\}$
and define

$$p_\zeta(\mathbf{s}|t) = \frac{p_{\mathbf{S}|T}(\mathbf{s}|t) \cdot \mathbb{1}_{\{\mathbf{s} \in \mathcal{S}(t) \ominus \zeta\}}}{\int_{\mathcal{S}(t) \ominus \zeta} p_{\mathbf{S}|T}(\mathbf{s}_1|t) d\mathbf{s}_1}.$$

Level set approach



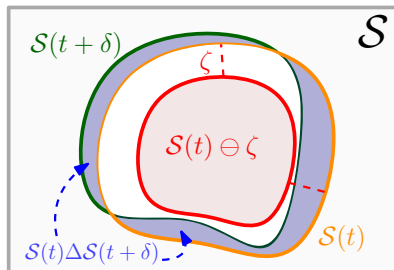
$\mathcal{L}_\zeta(t) = \left\{ \mathbf{s} \in \mathcal{S}(t) : p_{\mathbf{S}|T}(\mathbf{s}|t) \geq \zeta \right\}$
and define

$$p_\zeta(\mathbf{s}|t) = \frac{p_{\mathbf{S}|T}(\mathbf{s}|t) \cdot \mathbb{1}_{\{\mathbf{s} \in \mathcal{L}_\zeta(t)\}}}{\int_{\mathcal{L}_\zeta(t)} p_{\mathbf{S}|T}(\mathbf{s}_1|t) d\mathbf{s}_1}.$$

ζ -Interior Conditional Density

Question: How can we find a ζ -interior conditional density $p_\zeta(s|t)$?

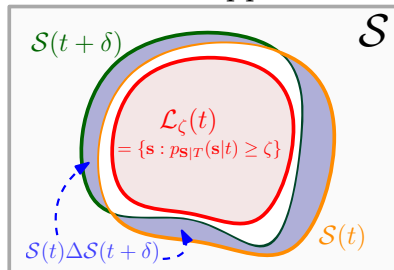
Support shrinking approach



$S(t) \ominus \zeta = \left\{ s \in S(t) : \inf_{x \in \partial S(t)} \|s - x\|_2 \geq \zeta \right\}$
and define

$$p_\zeta(s|t) = \frac{p_{S|T}(s|t) \cdot \mathbb{1}_{\{s \in S(t) \ominus \zeta\}}}{\int_{S(t) \ominus \zeta} p_{S|T}(s_1|t) ds_1}.$$

Level set approach



$\mathcal{L}_\zeta(t) = \{s \in S(t) : p_{S|T}(s|t) \geq \zeta\}$
and define

$$p_\zeta(s|t) = \frac{p_{S|T}(s|t) \cdot \mathbb{1}_{\{s \in \mathcal{L}_\zeta(t)\}}}{\int_{\mathcal{L}_\zeta(t)} p_{S|T}(s_1|t) ds_1}.$$

Remark: Practically, the level set approach is recommended, because we only need to choose $\zeta > 0$.

- **Bias-Corrected IPW Estimator:**

$$\hat{\theta}_{\text{C,IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h} \right) K \left(\frac{T_i - t}{h} \right) \hat{p}_{\zeta}(\mathbf{S}_i | t)}{\kappa_2 \cdot \hat{p}(T_i, \mathbf{S}_i)},$$

where

- $\hat{p}(t, \mathbf{s}), \hat{p}_{\zeta}(\mathbf{s} | t)$ are estimators of $p(t, \mathbf{s}), p_{\zeta}(\mathbf{s} | t)$.
- ζ can be set to, e.g., $\zeta = 0.5 \cdot \max \{ \hat{p}_{\mathbf{S}|T}(\mathbf{S}_i | t) : i = 1, \dots, n \}$.

- Bias-Corrected IPW Estimator:**

$$\hat{\theta}_{C,IPW}(t) = \frac{1}{nh^2} \sum_{i=1}^n \frac{Y_i \left(\frac{T_i-t}{h} \right) K \left(\frac{T_i-t}{h} \right) \hat{p}_{\zeta}(S_i|t)}{\kappa_2 \cdot \hat{p}(T_i, S_i)},$$

where

- $\hat{p}(t, s), \hat{p}_{\zeta}(s|t)$ are estimators of $p(t, s), p_{\zeta}(s|t)$.
- ζ can be set to, e.g., $\zeta = 0.5 \cdot \max \{ \hat{p}_{S|T}(S_i|t) : i = 1, \dots, n \}$.

- Bias-Corrected DR Estimator:**

$$\begin{aligned} & \hat{\theta}_{C,DR}(t) \\ &= \underbrace{\frac{1}{nh^2} \sum_{i=1}^n \frac{\left(\frac{T_i-t}{h} \right) K \left(\frac{T_i-t}{h} \right) \hat{p}_{\zeta}(S_i|t)}{\kappa_2 \cdot \hat{p}(T_i, S_i)} \left[Y_i - \hat{\mu}(t, S_i) - (T_i - t) \cdot \hat{\beta}(t, S_i) \right]}_{\text{IPW component}} \\ & \quad + \underbrace{\int \hat{\beta}(t, s) \cdot \hat{p}_{\zeta}(s|t) ds}_{\text{RA component}}. \end{aligned}$$

Theorem (Theorem 5 in Zhang and Chen 2025)

Under some regularity assumptions and

- ① $\hat{\mu}, \hat{\beta}, \hat{p}, \hat{p}_{\zeta}$ are estimated on a dataset independent of $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$;
- ② $\sqrt{nh^3} \|\hat{p}_{\zeta}(\mathbf{S}|t) - \bar{p}_{\zeta}(\mathbf{S}|t)\|_{L_2} = o_P(1)$, where $\hat{p}_{\zeta}(\mathbf{s}|t) \xrightarrow{P} \bar{p}_{\zeta}(\mathbf{s}|t)$;
- ③ at least one of the model specification conditions hold:
 - $\hat{p}(t, \mathbf{s}) \xrightarrow{P} \bar{p}(t, \mathbf{s}) = p(t, \mathbf{s})$ (joint density model),
 - $\hat{\mu}(t, \mathbf{s}) \xrightarrow{P} \bar{\mu}(t, \mathbf{s}) = \mu(t, \mathbf{s})$ and $\hat{\beta}(t, \mathbf{s}) \xrightarrow{P} \bar{\beta}(t, \mathbf{s}) = \beta(t, \mathbf{s})$ (outcome model);
- ④
$$\sup_{|u-t| \leq h} \|\hat{p}(u, \mathbf{S}) - p(u, \mathbf{S})\|_{L_2} \left[\|\hat{\mu}(t, \mathbf{S}) - \mu(t, \mathbf{S})\|_{L_2} + h \|\hat{\beta}(t, \mathbf{S}) - \beta(t, \mathbf{S})\|_{L_2} \right] = o_P\left(\frac{1}{\sqrt{nh}}\right)$$

we prove that

Asymptotic Properties of $\hat{\theta}_{C,DR}(t)$ Without Positivity

Theorem (Theorem 5 in Zhang and Chen 2025)

Under some regularity assumptions and

- ① $\hat{\mu}, \hat{\beta}, \hat{p}, \hat{p}_\zeta$ are estimated on a dataset independent of $\{(Y_i, T_i, S_i)\}_{i=1}^n$;
- ② $\sqrt{nh^3} \|\hat{p}_\zeta(S|t) - \bar{p}_\zeta(S|t)\|_{L_2} = o_P(1)$, where $\hat{p}_\zeta(s|t) \xrightarrow{P} \bar{p}_\zeta(s|t)$;
- ③ at least one of the model specification conditions hold:
 - $\hat{p}(t, s) \xrightarrow{P} \bar{p}(t, s) = p(t, s)$ (joint density model),
 - $\hat{\mu}(t, s) \xrightarrow{P} \bar{\mu}(t, s) = \mu(t, s)$ and $\hat{\beta}(t, s) \xrightarrow{P} \bar{\beta}(t, s) = \beta(t, s)$ (outcome model);
- ④
$$\sup_{|u-t| \leq h} \|\hat{p}(u, S) - p(u, S)\|_{L_2} \left[\|\hat{\mu}(t, S) - \mu(t, S)\|_{L_2} + h \|\hat{\beta}(t, S) - \beta(t, S)\|_{L_2} \right] = o_P\left(\frac{1}{\sqrt{nh}}\right)$$

we prove that

- $$\sqrt{nh^3} \left[\hat{\theta}_{C,DR}(t) - \theta(t) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{C,h,t}(Y_i, T_i, S_i; \bar{\mu}, \bar{\beta}, \bar{p}_{T|S}) + o_P(1).$$
- $$\sqrt{nh^3} \left[\hat{\theta}_{C,DR}(t) - \theta(t) - h^2 B_{C,\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{C,\theta}(t)).$$

Statistical Inference on $\theta(t)$ Without Positivity

Asymptotically valid inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ can be done via

$$\sqrt{nh^3} \left[\hat{\theta}_{C,DR}(t) - \theta(t) - h^2 B_{C,\theta}(t) \right] \xrightarrow{d} \mathcal{N}(0, V_{C,\theta}(t)).$$

- ① We estimate $V_{C,\theta}(t) = \mathbb{E} \left[\phi_{C,h,t}^2(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}, \bar{p}_\zeta) \right]$ with

$$\phi_{C,h,t}(Y, T, \mathbf{S}; \bar{\mu}, \bar{\beta}, \bar{p}, \bar{p}_\zeta) = \frac{\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right) \cdot \bar{p}_\zeta(\mathbf{S}|t)}{\sqrt{h} \cdot \kappa_2 \cdot \bar{p}(T, \mathbf{S})} \cdot [Y - \bar{\mu}(t, \mathbf{S}) - (T-t) \cdot \bar{\beta}(t, \mathbf{S})]$$

$$\text{by } \hat{V}_{C,\theta}(t) = \frac{1}{n} \sum_{i=1}^n \phi_{C,h,t}^2(Y, T, \mathbf{S}; \hat{\mu}, \hat{\beta}, \hat{p}, \hat{p}_\zeta).$$

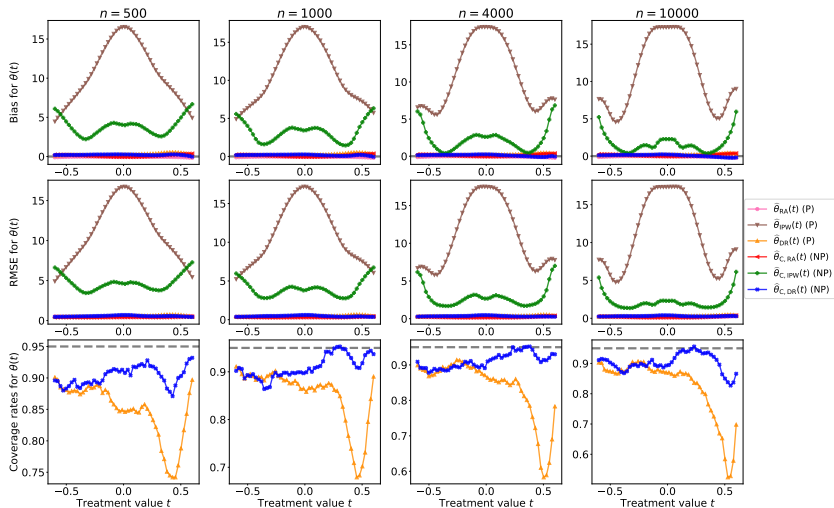
- ② $\hat{\mu}, \hat{\beta}, \hat{p}, \hat{p}_\zeta$ can be estimated via sample-splitting or cross-fitting.
- ③ We choose an implicit undersmoothing bandwidth $h = O\left(n^{-\frac{1}{5}}\right)$ to neglect the bias $h^2 B_{C,\theta}(t)$.

- ① Introduction
- ② Inference Theory for $\theta(t)$ Under Positivity
- ③ Inference Theory for $\theta(t)$ Without Positivity
- ④ Simulations and Case Study
- ⑤ Discussion



Simulations Without Positivity

$$Y = T^3 + T^2 + 10S + \epsilon, \quad T = \sin(\pi S) + E, \quad S \sim \text{Unif}[-1, 1], \quad E \sim \text{Unif}[-0.3, 0.3].$$



Note: $\beta(t, s) = \frac{\partial}{\partial t} \mu(t, s)$ is estimated via automatic differentiation of a well-trained neural net (inspired by [Luedtke 2024](#)).

A Case Study Under Positivity

We compare our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ under positivity with the finite-difference method (Colangelo and Lee 2020; CL20) on the U.S. Job Corps program (Schochet et al., 2001).

A Case Study Under Positivity

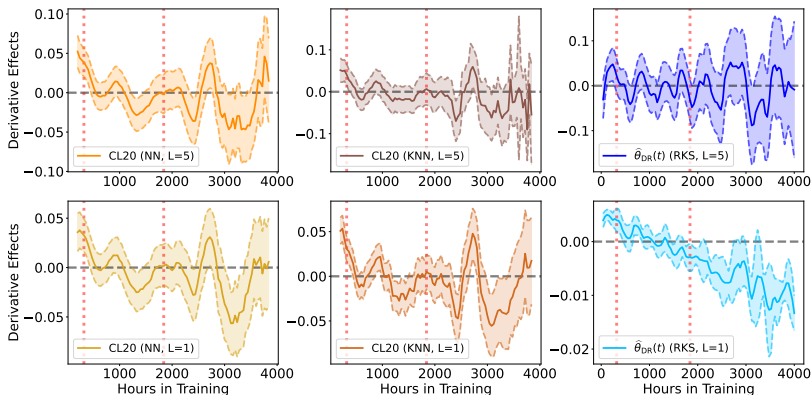
We compare our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ under positivity with the finite-difference method (Colangelo and Lee 2020; CL20) on the U.S. Job Corps program (Schochet et al., 2001).

- Y is the proportion of weeks employed in 2nd year after enrollment.
- T is the total hours of academic and vocational training received.
- S comprises 49 socioeconomic characteristics, and $n = 4024$.

A Case Study Under Positivity

We compare our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ under positivity with the finite-difference method (Colangelo and Lee 2020; CL20) on the U.S. Job Corps program (Schochet et al., 2001).

- Y is the proportion of weeks employed in 2nd year after enrollment.
- T is the total hours of academic and vocational training received.
- S comprises 49 socioeconomic characteristics, and $n = 4024$.



- ① Introduction
- ② Inference Theory for $\theta(t)$ Under Positivity
- ③ Inference Theory for $\theta(t)$ Without Positivity
- ④ Simulations and Case Study
- ⑤ Discussion



Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

- ① Under the positivity condition,
 - naive AIPW estimators are not doubly robust;
 - our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.

Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

① Under the positivity condition,

- naive AIPW estimators are not doubly robust;
- our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
- $\hat{\theta}_{\text{DR}}(t)$ admits an asymptotically linear form for pointwise (and uniform) inference when $h = O\left(n^{-\frac{1}{5}}\right)$.

Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

① Under the positivity condition,

- naive AIPW estimators are not doubly robust;
- our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
- $\hat{\theta}_{\text{DR}}(t)$ admits an asymptotically linear form for pointwise (and uniform) inference when $h = O\left(n^{-\frac{1}{5}}\right)$.

② Without the positivity condition,

- we prove the inconsistency of conventional IPW and DR estimators even when $Y(t) = \bar{m}(t) + \eta(\mathbf{S}) + \epsilon$.

Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

① Under the positivity condition,

- naive AIPW estimators are not doubly robust;
- our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
- $\hat{\theta}_{\text{DR}}(t)$ admits an asymptotically linear form for pointwise (and uniform) inference when $h = O\left(n^{-\frac{1}{5}}\right)$.

② Without the positivity condition,

- we prove the inconsistency of conventional IPW and DR estimators even when $Y(t) = \bar{m}(t) + \eta(S) + \epsilon$.
- our bias-corrected IPW and DR estimators reveal a novel connection to nonparametric set estimation problems ([Bonvini et al., 2023](#)).

Summary of Today's Talk

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt} \mathbb{E}[Y(t)]$ with and without the positivity condition.

① Under the positivity condition,

- naive AIPW estimators are not doubly robust;
- our proposed DR estimator $\hat{\theta}_{\text{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
- $\hat{\theta}_{\text{DR}}(t)$ admits an asymptotically linear form for pointwise (and uniform) inference when $h = O\left(n^{-\frac{1}{5}}\right)$.

② Without the positivity condition,

- we prove the inconsistency of conventional IPW and DR estimators even when $Y(t) = \bar{m}(t) + \eta(S) + \epsilon$.
- our bias-corrected IPW and DR estimators reveal a novel connection to nonparametric set estimation problems ([Bonvini et al., 2023](#)).

Causal Inference Meets Geometric Data Analysis
(<https://uwgeometry.github.io/>)!

Open Questions and Future Work

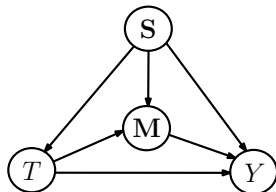
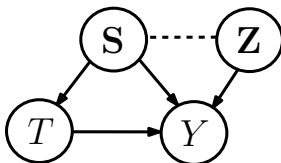
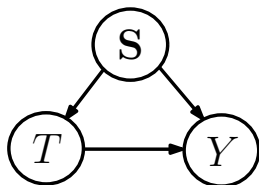
- ① **Efficiency Theory:** Can we derive efficient influence functions for our DR estimators through a sequence of kernel-smoothed parameters approximating $\theta(t)$ ([van der Laan et al., 2018](#))?

Open Questions and Future Work

- ① **Efficiency Theory:** Can we derive efficient influence functions for our DR estimators through a sequence of kernel-smoothed parameters approximating $\theta(t)$ ([van der Laan et al., 2018](#))?
- ② **Debiasing DR Estimators:** Can we debias our DR estimators through explicit bias estimation ([Calonico et al., 2018](#); [Cheng and Chen, 2019](#); [Takatsu and Westling, 2024](#)) or calibration ([van der Laan et al., 2024](#))?

Open Questions and Future Work

- 1 **Efficiency Theory:** Can we derive efficient influence functions for our DR estimators through a sequence of kernel-smoothed parameters approximating $\theta(t)$ (van der Laan et al., 2018)?
- 2 **Debiasing DR Estimators:** Can we debias our DR estimators through explicit bias estimation (Calonico et al., 2018; Cheng and Chen, 2019; Takatsu and Westling, 2024) or calibration (van der Laan et al., 2024)?
- 3 **Derivative Estimation in Other Causal Contexts:** Can we generalize our derivative estimators to other causal estimands:
 - instantaneous causal effect $\frac{d}{dt}\mathbb{E}[Y(t)|S=s]$ (Stolzenberg, 1980);
 - direct and indirect effects in mediation analysis (Huber et al., 2020)?



Thank you!

More details can be found in

[1] Y. Zhang, Y.-C. Chen, and A. Giessing. Nonparametric Inference on Dose-Response Curves Without the Positivity Condition. *arXiv preprint*, 2024.
<https://arxiv.org/abs/2405.09003>.

[2] Y. Zhang and Y.-C. Chen. Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments. *arXiv preprint*, 2025. <https://arxiv.org/abs/2501.06969>

All the code and data are available at
<https://github.com/zhangyk8/npDRDeriv>.

Python Package: [npDoseResponse](#).

Reference

- M. Bonvini, E. H. Kennedy, and L. J. Keele. Minimax optimal subgroup identification. *arXiv preprint arXiv:2306.17464*, 2023.
- S. Calonico, M. D. Cattaneo, and M. H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.
- G. Cheng and Y.-C. Chen. Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics*, 13(1):2194 – 2256, 2019.
- V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018.
- K. Colangelo and Y.-Y. Lee. Double debiased machine learning nonparametric inference with continuous treatments. *arXiv preprint arXiv:2004.03036*, 2020.
- W. Dai, T. Tong, and M. G. Genton. Optimal estimation of derivatives in nonparametric regression. *Journal of Machine Learning Research*, 17(164):1–25, 2016.
- K. Easwaran. Why physics uses second derivatives. *The British Journal for the Philosophy of Science*, 65(4): 845–862, 2014.
- R. D. Gill and J. M. Robins. Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 29(6):1785–1811, 2001.
- T. Haavelmo. Methods of measuring the marginal propensity to consume. *Journal of the American Statistical Association*, 42(237):105–122, 1947.
- K. Hirano and G. W. Imbens. *The Propensity Score with Continuous Treatments*, chapter 7, pages 73–84. John Wiley & Sons, Ltd, 2004.
- M. Huber, Y.-C. Hsu, Y.-Y. Lee, and L. Lettry. Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *Journal of Applied Econometrics*, 35(7):814–840, 2020.
- K. Imai and D. A. van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467):854–866, 2004.

Reference

- N. Kallus and A. Zhou. Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics*, pages 1243–1251. PMLR, 2018.
- E. H. Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.
- M. Lange. *An Introduction to the Philosophy of Physics: Locality, Fields, Energy, and Mass*. Blackwell, Oxford, 2002.
- A. Luedtke. Simplifying debiased inference via automatic differentiation and probabilistic programming. *arXiv preprint arXiv:2405.08675*, 2024.
- Y. Mack and H.-G. Müller. Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 59–72, 1989.
- W. K. Newey and T. M. Stoker. Efficiency of weighted average derivative estimators and index models. *Econometrica*, 61(5):1199–1223, 1993.
- J. Neyman. Optimal asymptotic tests of composite hypotheses. *Probability and Statistics*, pages 213–234, 1959.
- J. L. Powell, J. H. Stock, and T. M. Stoker. Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430, 1989.
- J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12): 1393–1512, 1986.
- D. Rothenhäusler and B. Yu. Incremental causal effects. *arXiv preprint arXiv:1907.13258*, 2019.
- P. Z. Schochet, J. Burghardt, and S. Glazerman. National job corps study: The impacts of job corps on participants’ employment and related outcomes. Mathematica policy research reports, Mathematica Policy Research, 2001.
- J. Shao. *Mathematical Statistics*. Springer Science & Business Media, 2003.

Reference

- R. M. Stolzenberg. The measurement and decomposition of causal effects in nonlinear and nonadditive models. *Sociological Methodology*, 11:459–488, 1980.
- K. Takatsu and T. Westling. Debiased inference for a covariate-adjusted regression function. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae041, 2024.
- L. van der Laan, A. Luedtke, and M. Carone. Automatic doubly robust inference for linear functionals via calibrated debiased machine learning. *arXiv preprint arXiv:2411.02771*, 2024.
- M. J. van der Laan, A. Bibaut, and A. R. Luedtke. Cv-tmle for nonpathwise differentiable target parameters. In M. J. van der Laan and S. Rose, editors, *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, pages 455–481. Springer, 2018.
- M. P. Wand and M. C. Jones. *Kernel Smoothing*. CRC press, 1994.
- L. Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.
- Y. Zhang and Y.-C. Chen. Doubly robust inference on causal derivative effects for continuous treatments. *arXiv preprint arXiv:2501.*, 2025.
- Y. Zhang, Y.-C. Chen, and A. Giessing. Nonparametric inference on dose-response curves without the positivity condition. *arXiv preprint arXiv:2405.09003*, 2024.

Detailed Regularity Assumptions

Assumption (Differentiability of the conditional mean outcome function)

For any $(t, s) \in \mathcal{T} \times \mathcal{S}$ and $\mu(t, s) = \mathbb{E}(Y|T = t, S = s)$, it holds that

- ① $\mu(t, s)$ is at least four times continuously differentiable with respect to t .*
- ② $\mu(t, s)$ and all of its partial derivatives are uniformly bounded on $\mathcal{T} \times \mathcal{S}$.*

Let \mathcal{J} be the support of the joint density $p(t, s)$.

Assumption (Differentiability of the density functions)

For any $(t, s) \in \mathcal{J}$, it holds that

- ① The joint density $p(t, s)$ and the conditional density $p_{T|S}(t|s)$ are at least three times continuously differentiable with respect to t .*
- ② $p(t, s)$, $p_{T|S}(t|s)$, $p_{S|T}(s|t)$, as well as all of the partial derivatives of $p(t, s)$ and $p_{T|S}(t|s)$ are bounded and continuous up to the boundary $\partial\mathcal{J}$.*
- ③ The support \mathcal{T} of the marginal density $p_T(t)$ is compact and $p_T(t)$ is uniformly bounded away from 0 within \mathcal{T} .*

Assumption (Regular kernel conditions)

A kernel function $K : \mathbb{R} \rightarrow [0, \infty)$ is bounded and compactly supported on $[-1, 1]$ with $\int_{\mathbb{R}} K(t) dt = 1$ and $K(t) = K(-t)$. In addition, it holds that

- ① $\kappa_j := \int_{\mathbb{R}} w^j K(u) du < \infty$ and $\nu_j := \int_{\mathbb{R}} w^j K^2(u) du < \infty$ for all $j = 1, 2, \dots$
- ② K is a second-order kernel, i.e., $\kappa_1 = 0$ and $\kappa_2 > 0$.
- ③ $\mathcal{K} = \left\{ t' \mapsto \left(\frac{t' - t}{h} \right)^{k_1} K \left(\frac{t' - t}{h} \right) : t \in \mathcal{T}, h > 0, k_1 = 0, 1 \right\}$ is a bounded VC-type class of measurable functions on \mathbb{R} .

Assumption (Smoothness condition on $\mathcal{S}(t)$)

For any $\delta \in \mathbb{R}$ and $t \in \mathcal{T}$, there exists an absolute constant $A_0 > 0$ such that either (i) " $\mathcal{S}(t) \ominus (A_0|\delta|) \subset \mathcal{S}(t + \delta)$ " for the support shrinking approach or (ii) " $\mathcal{L}_{A_0|\delta|}(t) \subset \mathcal{S}(t + \delta)$ " for the level set approach.

The self-normalizing technique can reduce the instability of IPW and DR estimators ([Kallus and Zhou, 2018](#)):

1 Self-Normalized Estimators Under Positivity:

$$\hat{\theta}_{\text{IPW}}^{\text{norm}}(t) = \frac{\hat{\theta}_{\text{IPW}}(t)}{\frac{1}{nh} \sum_{j=1}^n \frac{K\left(\frac{T_j - t}{h}\right)}{\hat{p}_{T|S}(T_j|S_j)}} = \frac{\sum_{i=1}^n \frac{Y_i \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\hat{p}_{T|S}(T_i|S_i)}}{\kappa_2 h \sum_{j=1}^n \frac{K\left(\frac{T_j - t}{h}\right)}{\hat{p}_{T|S}(T_j|S_j)}},$$

and

$$\hat{\theta}_{\text{DR}}^{\text{norm}}(t) = \frac{\sum_{i=1}^n \frac{[Y_i - \hat{\mu}(t, S_i) - (T_i - t) \cdot \hat{\beta}(t, S_i)] \left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\hat{p}_{T|S}(T_i|S_i)}}{\kappa_2 h \sum_{j=1}^n \frac{K\left(\frac{T_j - t}{h}\right)}{\hat{p}_{T|S}(T_j|S_j)}} + \frac{1}{n} \sum_{i=1}^n \hat{\beta}(t, S_i).$$

2 Self-Normalized Estimators Without Positivity:

$$\hat{\theta}_{C,IPW}^{\text{norm}}(t) = \frac{\hat{\theta}_{C,IPW}(t)}{\frac{1}{nh} \sum_{j=1}^n \frac{K\left(\frac{T_j-t}{h}\right) \cdot \hat{p}_{\zeta}(S_j|t)}{\hat{p}(T_j, S_j)}} = \frac{\sum_{i=1}^n \frac{Y_i \left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right) \cdot \hat{p}_{\zeta}(S_i|t)}{\hat{p}(T_i, S_i)}}{\kappa_2 h \sum_{j=1}^n \frac{K\left(\frac{T_j-t}{h}\right) \cdot \hat{p}_{\zeta}(S_j|t)}{\hat{p}(T_j, S_j)}},$$

and

$$\begin{aligned} \hat{\theta}_{C,DR}^{\text{norm}}(t) = & \frac{\sum_{i=1}^n \frac{[Y_i - \hat{\mu}(t, S_i) - (T_i - t) \cdot \hat{\beta}(t, S_i)] \left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right) \cdot \hat{p}_{\zeta}(S_i|t)}{\hat{p}(T_i, S_i)}}{\kappa_2 h \sum_{j=1}^n \frac{K\left(\frac{T_j-t}{h}\right) \cdot \hat{p}_{\zeta}(S_j|t)}{\hat{p}(T_j, S_j)}} \\ & + \int \hat{\beta}(t, s) \cdot \hat{p}_{\zeta}(s|t) ds. \end{aligned}$$

We generate i.i.d. observations $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ from the following data-generating model (Colangelo and Lee, 2020):

$$Y = 1.2T + T^2 + TS_1 + 1.2\boldsymbol{\xi}^T \mathbf{S} + \epsilon \sqrt{0.5 + F_{\mathcal{N}(0,1)}(S_1)}, \quad \epsilon \sim \mathcal{N}(0, 1),$$

$$T = F_{\mathcal{N}(0,1)}(3\boldsymbol{\xi}^T \mathbf{S}) - 0.5 + 0.75E, \quad \mathbf{S} = (S_1, \dots, S_d)^T \sim \mathcal{N}_d(\mathbf{0}, \Sigma), \quad E \sim \mathcal{N}(0, 1),$$

where

- $F_{\mathcal{N}(0,1)}$ is the CDF of $\mathcal{N}(0, 1)$ and $d = 20$.
- $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ has its entry $\xi_j = \frac{1}{j^2}$ for $j = 1, \dots, d$ and $\Sigma_{ii} = 1$, $\Sigma_{ij} = 0.5$ when $|i - j| = 1$, and $\Sigma_{ij} = 0$ when $|i - j| > 1$ for $i, j = 1, \dots, d$.
- The dose-response curve is given by $m(t) = 1.2t + t^2$, and our parameter of interest is the derivative effect curve $\theta(t) = 1.2 + 2t$.

Simulations Under the Positivity Condition

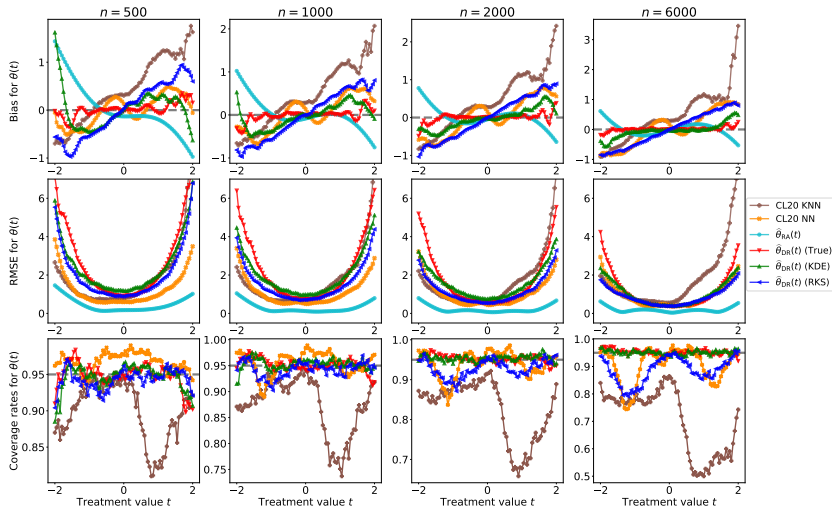


Figure: Comparisons between our proposed estimators and the finite-difference approaches by [Colangelo and Lee \(2020\)](#) (“CL20”) under positivity and with 5-fold cross-fitting across various sample sizes.

Simulations Under the Positivity Condition

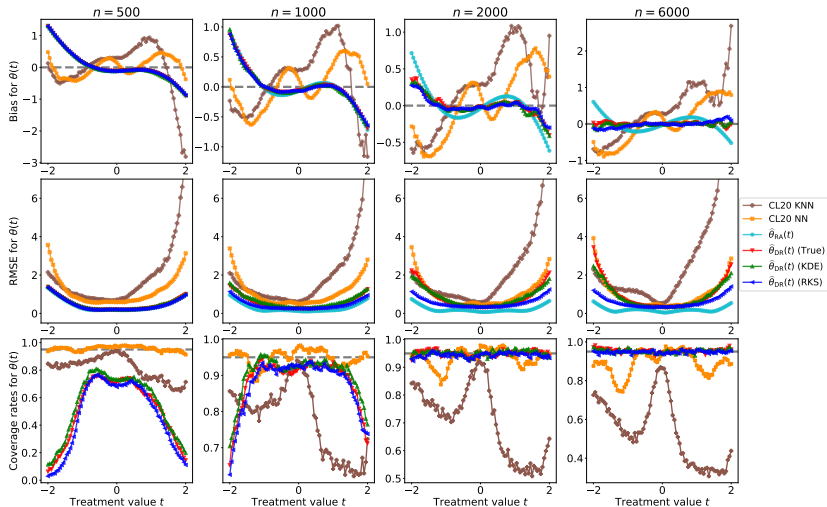


Figure: Comparisons between our proposed estimators and the finite-difference approaches by [Colangelo and Lee \(2020\)](#) (“CL20”) under positivity and **without cross-fitting** across various sample sizes.