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

Department of STATISTICS

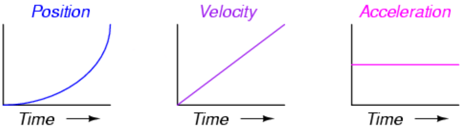
UNIVERSITY *of* WASHINGTON

The Notion of Derivative

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

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

- **Physics:**

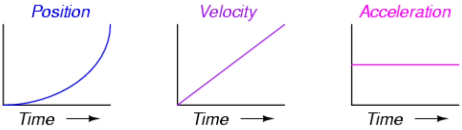


$$\text{Position } f(t) \overset{\text{derivative}}{\Longrightarrow} \text{Velocity } v(t) = f'(t)$$
$$\overset{\text{derivative}}{\Longrightarrow} \text{Acceleration } a(t) = v'(t)$$

The Notion of Derivative

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

- **Physics:**



$$\text{Position } \ f(t) \ \overset{\text{derivative}}{\Longrightarrow} \ \text{Velocity } \ v(t) = f'(t)$$

$$\overset{\text{derivative}}{\Longrightarrow} \ \text{Acceleration } \ a(t) = v'(t)$$

- **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.

$$\text{Position } f(t) \overset{\text{derivative}}{\Longrightarrow} \text{Velocity } v(t) = f'(t)$$

$$\overset{\text{derivative}}{\Longrightarrow} \text{Acceleration } a(t) = v'(t)$$

Derivative and Causation

Derivatives measure rates of change over infinitesimal neighborhoods.

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

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

$$\text{Acceleration } a(t) = v'(t) \overset{\text{"cause"}}{\Longrightarrow} \text{Velocity } v(t) \quad \text{over} \quad [t_0, t_1],$$
$$\text{Velocity } v(t) = f'(t) \overset{\text{"cause"}}{\Longrightarrow} \text{Position } f(t) \quad \text{over} \quad [t_0, t_1].$$

Derivative and Causation

Derivatives measure rates of change over infinitesimal neighborhoods.

$$\begin{aligned} \text{Position } f(t) &\overset{\text{derivative}}{\Longrightarrow} \text{Velocity } v(t) = f'(t) \\ &\overset{\text{derivative}}{\Longrightarrow} \text{Acceleration } a(t) = v'(t) \end{aligned}$$

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

$$\begin{aligned} \text{Acceleration } a(t) = v'(t) &\overset{\text{“cause”}}{\Longrightarrow} \text{Velocity } v(t) \quad \text{over} \quad [t_0, t_1], \\ \text{Velocity } v(t) = f'(t) &\overset{\text{“cause”}}{\Longrightarrow} \text{Position } f(t) \quad \text{over} \quad [t_0, t_1]. \end{aligned}$$

*“**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).

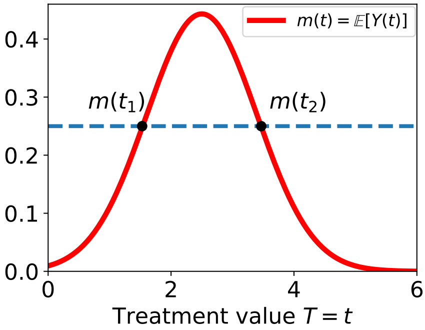
The Role of Derivatives in Causal Inference

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 $T = t$.

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

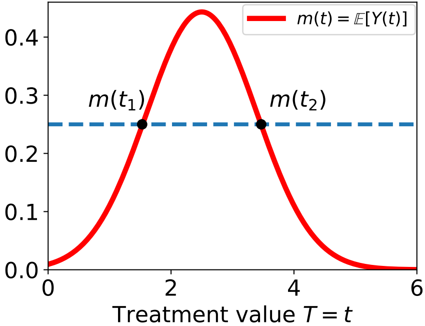
- $\mathbb{E}[Y(t)] =$ mean potential outcome under a static intervention $T = t$.
- When t varies in a continuous space, $t \mapsto \mathbb{E}[Y(t)] := m(t)$ is a curve!



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 $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$.

Estimand of Interest and its Alternatives

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

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

Estimand of Interest and its Alternatives

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

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

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**

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

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):

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

  for some small deterministic $\delta > 0$.

Estimand of Interest and its Alternatives

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

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

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):

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

  for some small deterministic $\delta > 0$.

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

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

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

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$, 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)* $Y_i(t) \perp\!\!\!\perp T_i \,\big|\, \boldsymbol{S}_i$ *for all* $t \in \mathcal{T}$.
3. ***(Positivity)*** $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ *for all* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$, 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)* $Y_i(t) \perp\!\!\!\perp T_i \,\big|\, \boldsymbol{S}_i$ *for all* $t \in \mathcal{T}$.
3. ***(Positivity)*** $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ *for all* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

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

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$, 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)* $Y_i(t) \perp\!\!\!\perp T_i \,\big|\, \boldsymbol{S}_i$ *for all* $t \in \mathcal{T}$.
3. ***(Positivity)*** $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ *for all* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

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

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

To identify and estimate $\theta(t)$ from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$, 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)* $Y_i(t) \perp\!\!\!\perp T_i \,\big|\, \boldsymbol{S}_i$ *for all* $t \in \mathcal{T}$.
3. ***(Positivity)*** $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ *for all* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

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

[1] Some 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* $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min}$ *for all* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

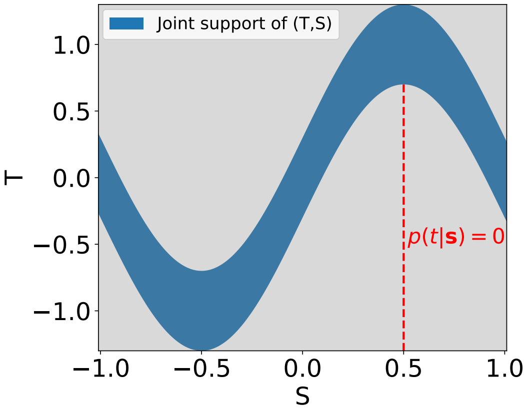
$$T = \sin(\pi S) + E, \quad E \sim \text{Unif}[-0.3, 0.3], \quad S \sim \text{Unif}[-1, 1], \quad \text{and} \quad E \perp\!\!\!\perp S.$$

An Example of the Positivity Violation

Assumption (Positivity Condition)

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

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



- **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 $m(t)$.

$$\text{Regression Adjustment (RA)} \; + \; \text{Inverse Probability Weighting (IPW)} \begin{cases} \Rightarrow \\ \not\Rightarrow \end{cases} \text{DR}$$

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 $m(t)$.

$$\text{Regression Adjustment (RA)} + \text{Inverse Probability Weighting (IPW)} \begin{cases} \Rightarrow \\ \not\Rightarrow \end{cases} \text{DR}$$

Without the positivity condition:

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

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

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 $m(t)$.

$$\text{Regression Adjustment (RA)} + \text{Inverse Probability Weighting (IPW)} \begin{cases} \Rightarrow \\ \not\Rightarrow \end{cases} \text{DR}$$

Without the positivity condition:

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

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

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 $m(t)$.

$$\text{Regression Adjustment (RA)} + \text{Inverse Probability Weighting (IPW)} \begin{cases} \Rightarrow \\ \not\Rightarrow \end{cases} \text{DR}$$

Without the positivity condition:

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

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

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.

Recap of the Identification Under Positivity

Assumption (Identification Conditions)

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

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

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

Recap of the Identification Under Positivity

Assumption (Identification Conditions)

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

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

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

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

Here, $K : \mathbb{R} \to [0, \infty)$ 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

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

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

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}(t, \boldsymbol{S}_i).$$

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i).$$

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

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

Dose-Response Curve Estimation Under Positivity

There are three major strategies for estimating

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i).$$

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

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

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

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

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

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

1. **RA Estimator:**

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

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

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

1. **RA Estimator:**

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

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

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

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

1. **RA Estimator:**

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

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

2. **IPW Estimator:** Inspired by the derivative estimator in Mack and Müller (1989), we propose

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(t, \boldsymbol{S}_i) \qquad \text{“+”} \qquad \widehat{m}_{\mathrm{IPW}}(t) = \frac{1}{nh}\sum_{i=1}^{n}\frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|\boldsymbol{S}}(T_i|\boldsymbol{S}_i)} \cdot Y_i$$

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}(t, \boldsymbol{S}_i) \qquad \text{“+”} \qquad \widehat{m}_{\mathrm{IPW}}(t) = \frac{1}{nh}\sum_{i=1}^{n} \frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|\boldsymbol{S}}(T_i|\boldsymbol{S}_i)} \cdot Y_i$$

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

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

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

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

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

$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(t, \boldsymbol{S}_i) \qquad \text{“+”} \qquad \widehat{m}_{\mathrm{IPW}}(t) = \frac{1}{nh}\sum_{i=1}^{n}\frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|\boldsymbol{S}}(T_i|\boldsymbol{S}_i)}\cdot Y_i$$

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

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

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

$\Longrightarrow$

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

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

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

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

$\Longrightarrow$

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

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

$$\Longrightarrow$$

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

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 $h \to 0$.

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

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

$$\Longrightarrow$$

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

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 $h \to 0$.
2. Different from $\widehat{m}_{\mathrm{IPW}}(t)$ and $\widehat{m}_{\mathrm{DR}}(t)$, we must compute the inverse probability weights as $\frac{1}{\widehat{p}_{T|\boldsymbol{S}}(T_i|\boldsymbol{S}_i)}$ but not $\frac{1}{\widehat{p}_{T|\boldsymbol{S}}(t|\boldsymbol{S}_i)}$ for $i = 1, ..., n$.

Theorem (Theorem 1 in Zhang and Chen 2025)

Under some regularity assumptions and

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

we prove that

Asymptotic Properties of $\widehat{\theta}_{\text{DR}}(t)$

Theorem (Theorem 1 in Zhang and Chen 2025)

Under some regularity assumptions and

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

we prove that

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

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

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

Statistical Inference on $\theta(t)$

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

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

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

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

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

Statistical Inference on $\theta(t)$

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

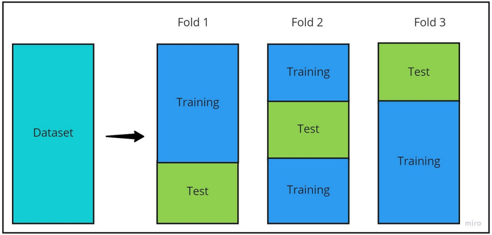
$$\sqrt{nh^3}\left[\widehat{\theta}_{\mathrm{DR}}(t) - \theta(t) - h^2 B_\theta(t)\right] \xrightarrow{d} \mathcal{N}\left(0, V_\theta(t)\right).$$

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

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

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

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



Statistical Inference on $\theta(t)$

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

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

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

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

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

2. $\widehat{\mu}, \widehat{\beta}, \widehat{p}_{T|\boldsymbol{S}}$ 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).

Why Do We Need Positivity?

Assumption (Identification Conditions)

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

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

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

The IPW approaches also rely on the following identities:

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

Why Do We Need Positivity?

Assumption (Identification Conditions)

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

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

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

The IPW approaches also rely on the following identities:

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

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

Identification Strategy Without Positivity

Assumption (Identification Conditions)

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *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)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *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* $\mathbb{E}\left[Y(t)|\boldsymbol{S} = \boldsymbol{s}\right]$ *is continuously differentiable with respect to* t *for any* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$ *with* $p_{S|T}(\boldsymbol{s}|t) > 0$ *and*

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

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

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

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

Identification Strategy Without Positivity

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

Identification Strategy Without Positivity

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

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

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

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

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

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

Identification Strategy Without Positivity

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

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

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

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

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

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

Validity of Our Identification Strategies Without Positivity

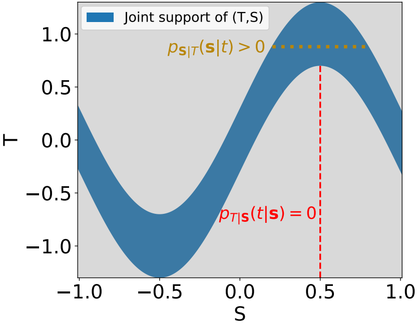
We identify $\theta(t)$ through

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

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

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

  to be well-defined when $p_{\boldsymbol{S}|T}(\boldsymbol{s}|t) > 0$.



Key Example: Additive Confounding Model

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

Key Example: Additive Confounding Model

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

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

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

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

Key Example: Additive Confounding Model

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

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

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

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

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

Key Example: Additive Confounding Model

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

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

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

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

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

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

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

Estimation Biases of IPW Estimators Without Positivity

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

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

Estimation Biases of IPW Estimators Without Positivity

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

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

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

Estimation Biases of IPW Estimators Without Positivity

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

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

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

- We show in Proposition 2 of Zhang and Chen (2025) that

$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{m}_{\text{IPW}}(t)\right] = \bar{m}(t)\cdot\rho(t) + \omega(t) \neq m(t),$$

$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{\theta}_{\text{IPW}}(t)\right] = \begin{cases} \bar{m}'(t)\cdot\rho(t) \\ \infty \end{cases} \neq \theta(t),$$

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

Estimation Biases of IPW Estimators Without Positivity

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

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

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

- We show in Proposition 2 of Zhang and Chen (2025) that

$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{m}_{\mathrm{IPW}}(t)\right] = \bar{m}(t) \cdot \rho(t) + \omega(t) \neq m(t),$$

$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{\theta}_{\mathrm{IPW}}(t)\right] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

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

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

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

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

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

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

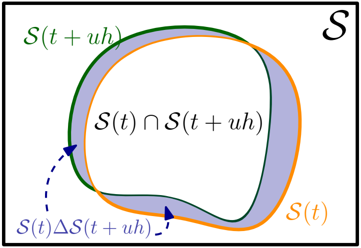
$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{\theta}_{\text{IPW}}(t)\right] = \lim_{h\to 0} \mathbb{E}\left[\frac{Y\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S)}\right] = \begin{cases} \bar{m}'(t)\cdot\rho(t) \\ \infty \end{cases} \neq \theta(t),$$

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

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

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

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



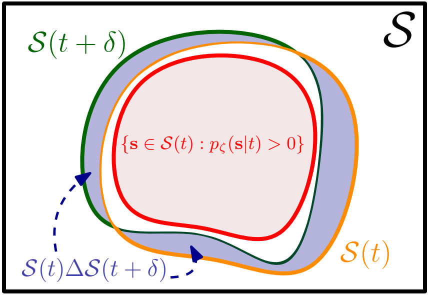
Bias-Corrected IPW Estimator of $\theta(t)$

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

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

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

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

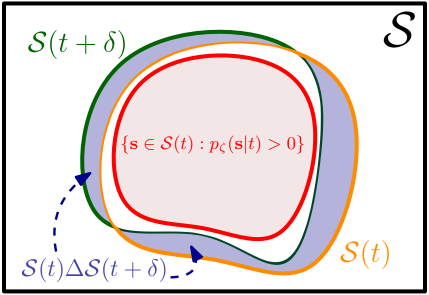


Bias-Corrected IPW Estimator of $\theta(t)$

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

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

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



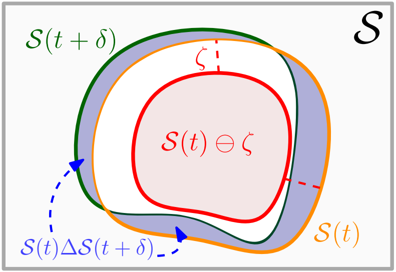
- Now, we have that $\mathbb{E}\left[\frac{Y\left(\frac{T-t}{h}\right)K\left(\frac{T-t}{h}\right)p_\zeta(\boldsymbol{S}|T)}{h^2\cdot\kappa_2\cdot p_{T|S}(T|\boldsymbol{S})\cdot p_S(\boldsymbol{S})}\right] = \bar{m}'(t) + O(h^2).$

ζ-Interior Conditional Density

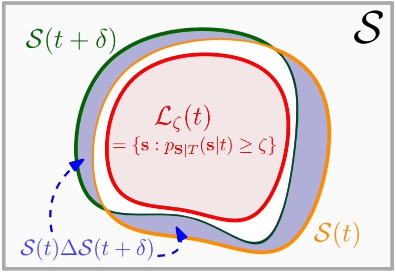
Question: How can we find a ζ-interior conditional density $p_\zeta(\boldsymbol{s}|t)$?

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

Support shrinking approach



Level set approach



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

and define

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

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

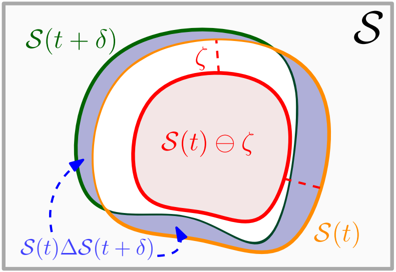
and define

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

ζ-Interior Conditional Density

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

Support shrinking approach

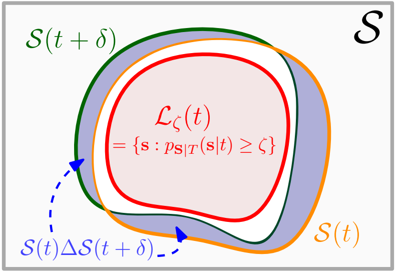


$$\mathcal{S}(t)\ominus\zeta = \left\{\boldsymbol{s}\in\mathcal{S}(t) : \inf_{\boldsymbol{x}\in\partial\mathcal{S}(t)} ||\boldsymbol{s}-\boldsymbol{x}||_2 \geq \zeta\right\}$$

and define

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

Level set approach



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

and define

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

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

Bias-Corrected IPW and DR Estimators of $\theta(t)$

- **Bias-Corrected IPW Estimator:**

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

where

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

Bias-Corrected IPW and DR Estimators of $\theta(t)$

- **Bias-Corrected IPW Estimator:**

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

where

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

- **Bias-Corrected DR Estimator:**

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

Theorem (Theorem 5 in Zhang and Chen 2025)

Under some regularity assumptions and

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

we prove that

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

Theorem (Theorem 5 in Zhang and Chen 2025)

Under some regularity assumptions and

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

we prove that

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

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

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

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

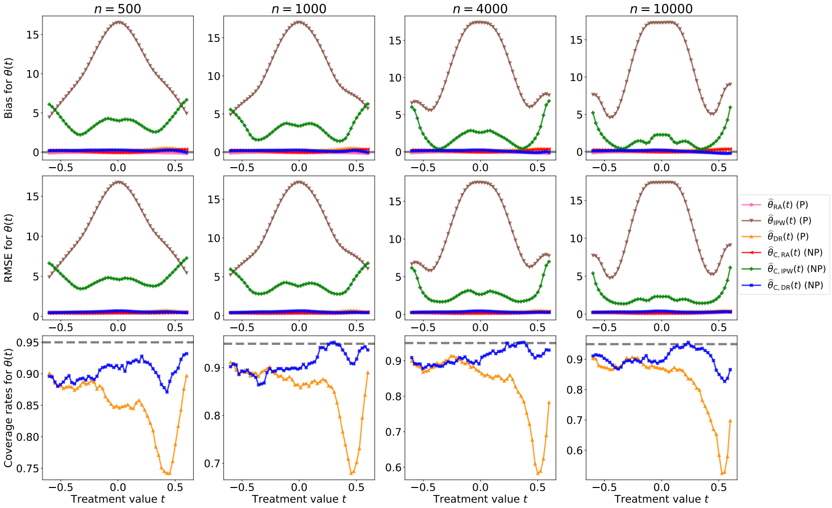
1. We estimate $V_{C,\theta}(t) = \mathbb{E}\left[\phi^2_{C,h,t}\left(Y, T, \boldsymbol{S}; \bar{\mu}, \bar{\beta}, \bar{p}, \bar{p}_\zeta\right)\right]$ with

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

   by $\widehat{V}_{C,\theta}(t) = \frac{1}{n}\sum_{i=1}^{n} \phi^2_{C,h,t}\left(Y, T, \boldsymbol{S}; \widehat{\mu}, \widehat{\beta}, \widehat{p}, \widehat{p}_\zeta\right)$.
2. $\widehat{\mu}, \widehat{\beta}, \widehat{p}, \widehat{p}_\zeta$ 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 $h^2 B_{C,\theta}(t)$.

Simulations Without Positivity

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



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

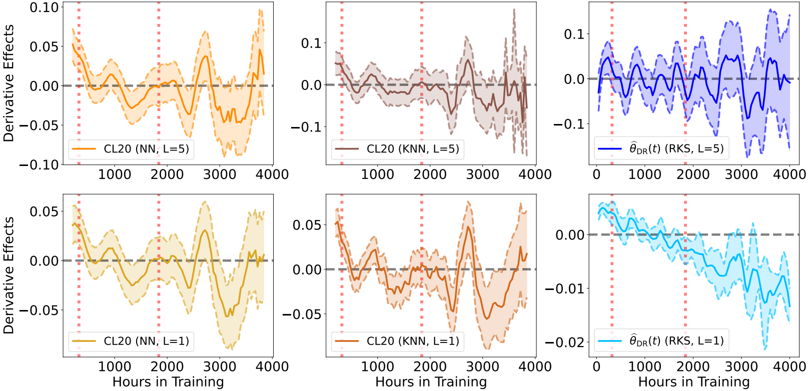
A Case Study Under Positivity

We compare our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 2^{nd} year after enrollment.
- T is the total hours of academic and vocational training received.
- $\boldsymbol{S}$ comprises 49 socioeconomic characteristics, and $n = 4024$.

We compare our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 2^{nd} year after enrollment.
- T is the total hours of academic and vocational training received.
- $\boldsymbol{S}$ comprises 49 socioeconomic characteristics, and $n = 4024$.



Summary of Today's Talk

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

Summary of Today's Talk

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

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

Summary of Today's Talk

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

1. Under the positivity condition,
   - naive AIPW estimators are not doubly robust;
   - our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
   - $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ with and without the positivity condition.

1. Under the positivity condition,
   - naive AIPW estimators are not doubly robust;
   - our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
   - $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 $Y(t) = \bar{m}(t) + \eta(\boldsymbol{S}) + \epsilon$.

Summary of Today's Talk

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

1. Under the positivity condition,
   - naive AIPW estimators are not doubly robust;
   - our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
   - $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 $Y(t) = \bar{m}(t) + \eta(\boldsymbol{S}) + \epsilon$.
   - 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 $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ with and without the positivity condition.

1. Under the positivity condition,
   - naive AIPW estimators are not doubly robust;
   - our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ achieves doubly robust consistency at the standard nonparametric rate.
   - $\widehat{\theta}_{\mathrm{DR}}(t)$ 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 $Y(t) = \bar{m}(t) + \eta(\boldsymbol{S}) + \epsilon$.
   - 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

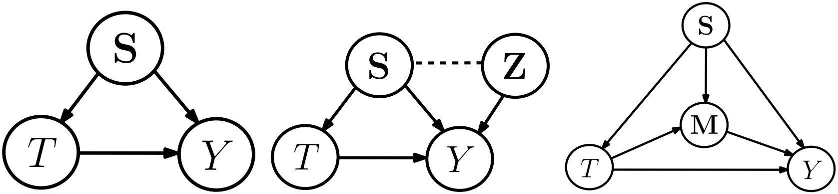
1. **Efficiency Theory:** Can we derive efficient influence functions for our DR estimators through a sequence of kernel-smoothed parameters approximating $\theta(t)$ (van der Laan et al., 2018)?

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)?

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 $\frac{d}{dt}\mathbb{E}\left[Y(t)|\boldsymbol{S}=\boldsymbol{s}\right]$ (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
hhttps://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.

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

1. $\mu(t, \boldsymbol{s})$ *is at least four times continuously differentiable with respect to* t.
2. $\mu(t, \boldsymbol{s})$ *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 $p(t, \boldsymbol{s})$.

Assumption (Differentiability of the density functions)

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

1. *The joint density* $p(t, \boldsymbol{s})$ *and the conditional density* $p_{T|\boldsymbol{S}}(t|\boldsymbol{s})$ *are at least three times continuously differentiable with respect to* t.
2. $p(t, \boldsymbol{s})$, $p_{T|\boldsymbol{S}}(t|\boldsymbol{s})$, $p_{\boldsymbol{S}|T}(\boldsymbol{s}|t)$, *as well as all of the partial derivatives of* $p(t, \boldsymbol{s})$ *and* $p_{T|\boldsymbol{S}}(t|\boldsymbol{s})$ *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* $\mathcal{T}$.

Detailed Regularity Assumptions

Assumption (Regular kernel conditions)

A kernel function $K : \mathbb{R} \to [0, \infty)$ 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. $\kappa_j := \int_{\mathbb{R}} u^j K(u)\, du < \infty$ *and* $\nu_j := \int_{\mathbb{R}} u^j K^2(u)\, du < \infty$ *for all* $j = 1, 2, ...$
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.

Self-Normalized IPW and DR Estimators

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

1. **Self-Normalized Estimators Under Positivity:**

$$\widehat{\theta}_{\mathrm{IPW}}^{\mathrm{norm}}(t) = \frac{\widehat{\theta}_{\mathrm{IPW}}(t)}{\frac{1}{nh}\sum_{j=1}^{n} \frac{K\left(\frac{T_j-t}{h}\right)}{\widehat{p}_{T|S}(T_j|\mathbf{S}_j)}} = \frac{\sum_{i=1}^{n} \frac{Y_i\left(\frac{T_i-t}{h}\right)K\left(\frac{T_i-t}{h}\right)}{\widehat{p}_{T|S}(T_i|\mathbf{S}_i)}}{\kappa_2 h \sum_{j=1}^{n} \frac{K\left(\frac{T_j-t}{h}\right)}{\widehat{p}_{T|S}(T_j|\mathbf{S}_j)}},$$

and

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

2. **Self-Normalized Estimators Without Positivity:**

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

and

$$\widehat{\theta}_{\text{C,DR}}^{\text{norm}}(t) = \frac{\sum_{i=1}^{n} \frac{\left[Y_i-\widehat{\mu}(t,\boldsymbol{S}_i)-(T_i-t)\cdot\widehat{\beta}(t,\boldsymbol{S}_i)\right]\left(\frac{T_i-t}{h}\right)K\left(\frac{T_i-t}{h}\right)\cdot \widehat{p}_{\zeta}(\boldsymbol{S}_i|t)}{\widehat{p}(T_i,\boldsymbol{S}_i)}}{\kappa_2 h \sum_{j=1}^{n} \frac{K\left(\frac{T_j-t}{h}\right)\cdot \widehat{p}_{\zeta}(\boldsymbol{S}_j|t)}{\widehat{p}(T_j,\boldsymbol{S}_j)}} + \int \widehat{\beta}(t,\boldsymbol{s})\cdot \widehat{p}_{\zeta}(\boldsymbol{s}|t)\,d\boldsymbol{s}.$$

Simulations Under the Positivity Condition

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

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

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

where

- $F_{\mathcal{N}(0,1)}$ is the CDF of $\mathcal{N}(0,1)$ and $d = 20$.
- $\boldsymbol{\xi} = (\xi_1, ..., \xi_d)^T \in \mathbb{R}^d$ has its entry $\xi_j = \frac{1}{j^2}$ for $j = 1, ..., d$ and $\Sigma_{ii} = 1$, $\Sigma_{ij} = 0.5$ when $|i-j| = 1$, and $\Sigma_{ij} = 0$ when $|i-j| > 1$ for $i, j = 1, ..., d$.
- 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

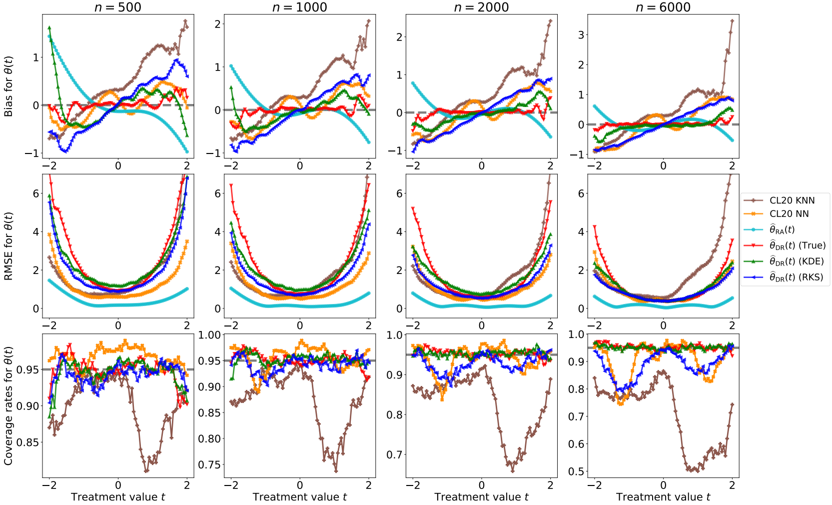


Figure: Comparisons between our proposed estimators and the finite-difference approaches by Colangelo and Lee (2020) ("CL20") under positivity and with 5-fold cross-fitting across various sample sizes.

Simulations Under the Positivity Condition

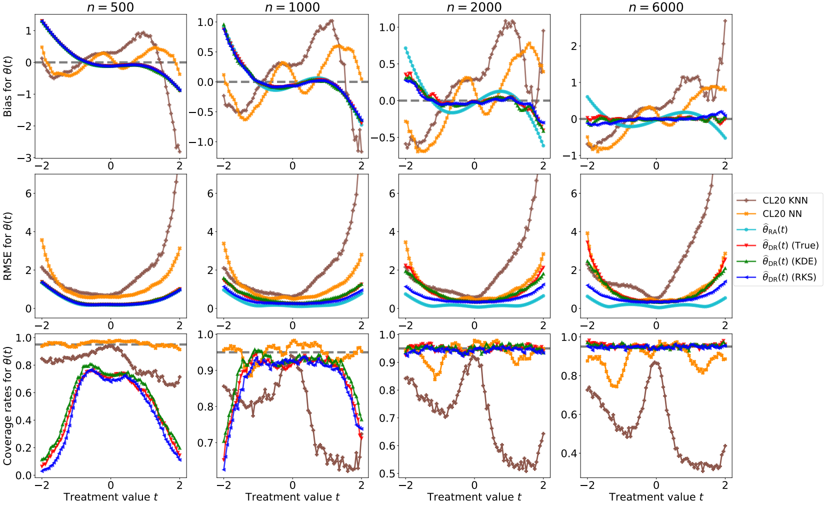


Figure: Comparisons between our proposed estimators and the finite-difference approaches by Colangelo and Lee (2020) ("CL20") under positivity and **without cross-fitting** across various sample sizes.