# Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments
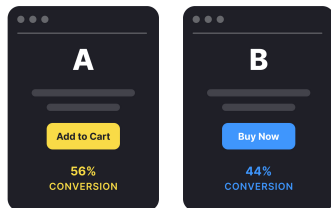
## Yikun Zhang

Joint work with *Professor Yen-Chi Chen*

*Department of Statistics,*
*University of Washington*

IMS Young Mathematical Scientists Forum (Statistics and Data Science)
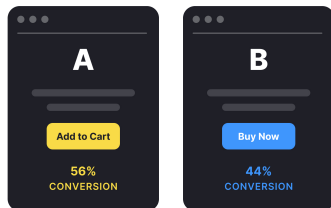November 24, 2025

Department of
STATISTICS

W UNIVERSITY *of* WASHINGTON

**Goal:** Study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.

**Goal:** Study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.



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

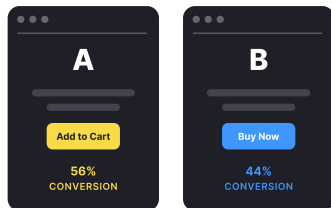- $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] =$ average treatment effect.

**Goal:** 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 $T = t$.

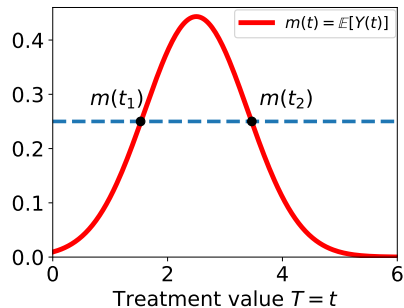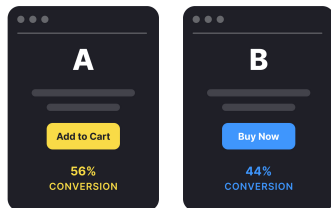- $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] =$ average treatment effect.

**Problem:** When $t$ varies in a continuous space, $t \mapsto \mathbb{E}[Y(t)] := m(t)$ is a curve!

**Goal:** Study the causal effect of a treatment $T \in \mathcal{T}$ on an outcome of interest $Y \in \mathcal{Y}$.



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

- $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] =$ average treatment effect.

**Problem:** When $t$ varies in a continuous space, $t \mapsto \mathbb{E}\left[Y(t)\right] := m(t)$ is a curve!



- $m'(t_1) \neq m'(t_2)$ even when $m(t_1) = m(t_2)$!

- The **derivative effect curve**

$$t \mapsto \theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$$

is a continuous generalization to the average treatment effect $\mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right]$.

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

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

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

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

**Problem:** $\theta(t)$ is non-regular and cannot be estimated in the rate $1/\sqrt{n}$.

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

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

**Problem:** $\theta(t)$ is non-regular and cannot be estimated in the rate $1/\sqrt{n}$.

There are some closely related but distinct estimands:

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

$$\mathbb{E}\left[Y(T + \delta)\right] - \mathbb{E}\left[Y(T)\right] \quad \text{for some deterministic} \quad \delta > 0.$$

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

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

**Problem:** $\theta(t)$ is non-regular and cannot be estimated in the rate $1/\sqrt{n}$.

There are some closely related but distinct estimands:

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

$$\mathbb{E}\left[Y(T + \delta)\right] - \mathbb{E}\left[Y(T)\right] \quad \text{for some deterministic} \quad \delta > 0.$$

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

$$\mathbb{E}\left[\theta(T)\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mathbb{E}\left(Y|T, S\right)\right], \quad \text{where } S \in \mathcal{S} \subset \mathbb{R}^d \text{ is a covariate vector.}$$

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

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

**Problem:** $\theta(t)$ is non-regular and cannot be estimated in the rate $1/\sqrt{n}$.

There are some closely related but distinct estimands:

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

$$\mathbb{E}\left[Y(T+\delta)\right] - \mathbb{E}\left[Y(T)\right] \quad \text{for some deterministic} \quad \delta > 0.$$

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

$$\mathbb{E}\left[\theta(T)\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mathbb{E}\left(Y|T, S\right)\right], \quad \text{where } S \in \mathcal{S} \subset \mathbb{R}^d \text{ is a covariate vector.}$$

Pros  These estimands may have more realistic interpretations in the actual context.

Cons  They quantify only the overall causal effects, not those at a specific level of interest.

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

## Assumption (Identification Conditions)

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.

2. *(Ignorability)* $Y(t)$ *is conditionally independent of $T$ given $\boldsymbol{S}$ for all $t \in \mathcal{T}$.*

3. *(**Positivity**) The conditional density satisfies $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ for all $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.*

---

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

**RCTs**

**Observational Studies**

## Assumption (Identification Conditions)

1. *(Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.*

2. *(Ignorability) $Y(t)$ is conditionally independent of $T$ given $\boldsymbol{S}$ for all $t \in \mathcal{T}$.*

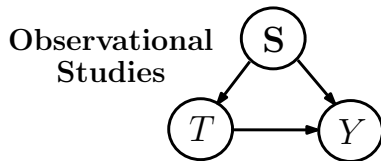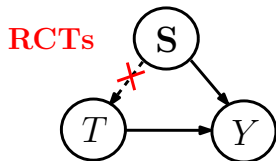3. *(**Positivity**) The conditional density satisfies $p_{T|\boldsymbol{S}}(t|\boldsymbol{s}) \geq p_{\min} > 0$ for all $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.*

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

- The positivity condition is required for $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s}) = \frac{\partial}{\partial t}\mathbb{E}\left(Y|T = t, \boldsymbol{S} = \boldsymbol{s}\right)$ to be well-defined on $\mathcal{T} \times \mathcal{S}$.

---

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

Assumption (Positivity Condition)

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

► Positivity is a very strong assumption with continuous treatments!

## Assumption (Positivity Condition)

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

▶ Positivity is a very strong assumption with continuous treatments!

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



$p_{T|S}(t|s) = 0$ in the gray regions, and the positivity condition fails!!

$$t \mapsto m(t) = \mathbb{E}\left[Y(t)\right] \quad \text{and} \quad t \mapsto \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \quad \text{for} \quad t \in \mathcal{T}.$$

**Under the positivity condition:**

1. Propose a doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.

$$t \mapsto m(t) = \mathbb{E}\left[Y(t)\right] \quad \text{and} \quad t \mapsto \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \quad \text{for} \quad t \in \mathcal{T}.$$

**Under the positivity condition:**

1. Propose a doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.

   Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{\begin{array}{l} \Rightarrow \\ \not\Rightarrow \end{array}\right.$ DR

$$t \mapsto m(t) = \mathbb{E}\left[Y(t)\right] \quad \text{and} \quad t \mapsto \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \quad \text{for} \quad t \in \mathcal{T}.$$

**Under the positivity condition:**

1 Propose a doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\begin{cases} \Rightarrow \\ \nRightarrow \end{cases}$ DR

**Without the positivity condition:**

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

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

$$t \mapsto m(t) = \mathbb{E}\left[Y(t)\right] \quad \text{and} \quad t \mapsto \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \quad \text{for} \quad t \in \mathcal{T}.$$

**Under the positivity condition:**

① Propose a doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.

Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\begin{cases} \Rightarrow \\ \nRightarrow \end{cases}$ DR

**Without the positivity condition:**

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

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

③ The usual IPW estimators of $m(t)$ and $\theta(t)$ are still *biased* even under model (1).

$$t \mapsto m(t) = \mathbb{E}\left[Y(t)\right] \quad \text{and} \quad t \mapsto \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \quad \text{for} \quad t \in \mathcal{T}.$$

**Under the positivity condition:**

1. Propose a doubly robust (DR) estimator of $\theta(t)$ via kernel smoothing.

   Regression Adjustment (RA) + Inverse Probability Weighting (IPW) $\left\{ \begin{array}{l} \Rightarrow \\ \not\Rightarrow \end{array} \right.$ DR

**Without the positivity condition:**

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

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

3. The usual IPW estimators of $m(t)$ and $\theta(t)$ are still *biased* even under model (1).

4. Propose our bias-corrected IPW and DR estimators for $m(t)$ and $\theta(t)$.
   - Has a novel connection to nonparametric support and level set estimation problems.

# Nonparametric Inference on $\theta(t)$ Under Positivity

## Assumption (Identification Conditions)

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.

2. *(Ignorability)* $Y(t)$ *is conditionally independent of $T$ given $S$ for all $t \in \mathcal{T}$.*

3. *(**Positivity**) The conditional density satisfies $p_{T|S}(t|s) \geq p_{\min} > 0$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.*

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

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

## Assumption (Identification Conditions)

1. *(Consistency)* $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
2. *(Ignorability)* $Y(t)$ is conditionally independent of $T$ given $S$ for all $t \in \mathcal{T}$.
3. *(**Positivity**) The conditional density satisfies $p_{T|S}(t|s) \geq p_{\min} > 0$ for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.*

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

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

$$\textbf{IPW:} \begin{cases} m(t) = \mathbb{E}[Y(t)] = \lim\limits_{h \to 0} \mathbb{E}\left[\frac{Y}{p_{T|S}(T|S)} \cdot \frac{1}{h}K\left(\frac{T-t}{h}\right)\right], \\ \theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)] = \textbf{???}. \end{cases}$$

- $K : \mathbb{R} \to [0, \infty)$ is a kernel function, *e.g.*, $K(u) = \begin{cases} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) & \text{(Gaussian)}, \\ \frac{3}{4}(1 - u^2) \cdot \mathbb{1}_{\{|u| \leq 1\}} & \text{(Parabolic)}. \end{cases}$
- $h > 0$ is a smoothing bandwidth parameter.

Given the data $\{(Y_i, T_i, S_i)\}_{i=1}^{n}$, there are three main strategies for estimating

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

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

$$\widehat{m}_{\text{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mu}(t, S_i).$$

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

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

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

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

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

① **RA Estimator:**

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

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

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

**1 RA Estimator:**

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

**Question:** How can we generalize the IPW form $m(t) = \lim\limits_{h \to 0} \mathbb{E}\left[\frac{Y \cdot K\left(\frac{T-t}{h}\right)}{h \cdot p_{T|S}(T|S)}\right]$ to identify and estimate $\theta(t)$?

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

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

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

$$\widehat{\theta}_{\mathrm{RA}}(t) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\beta}(t, S_i) \qquad \text{``+''} \qquad \widehat{\theta}_{\mathrm{IPW}}(t) = \frac{1}{nh^2} \sum_{i=1}^{n} \frac{\left(\frac{T_i - t}{h}\right) K\left(\frac{T_i - t}{h}\right)}{\kappa_2 \cdot \widehat{p}_{T|S}(T_i|S_i)} \cdot Y_i \qquad \Longrightarrow$$

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

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

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

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

The "New IPW component" leverages a local polynomial approximation to push the residual of the IPW component to (roughly) second order.

- Neyman orthogonality (Neyman, 1959; Chernozhukov et al., 2018) holds for this form of $\widehat{\theta}_{\mathrm{DR}}(t)$ as $h \to 0$.

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

*we prove that*

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

*we prove that*

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

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

An asymptotically valid inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ can be conducted through

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

An asymptotically valid inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ can be conducted through

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

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

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

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

An asymptotically valid inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ can be conducted through
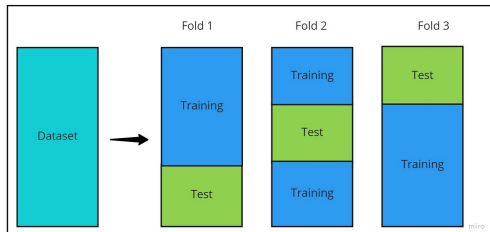
$$\sqrt{nh^3}\left[\widehat{\theta}_{\mathrm{DR}}(t) - \theta(t) - h^2 B_\theta(t)\right] \xrightarrow{d} \mathcal{N}\left(0, V_\theta(t)\right).$$

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

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

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

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

An asymptotically valid inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ can be conducted through

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

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

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

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

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

3. $h^2\,B_\theta(t)$ is asymptotically negligible when $h = O\left(n^{-\frac{1}{5}}\right)$, aligning with outputs from usual bandwidth selection methods (Wand and Jones, 1994; Wasserman, 2006).

**Question:**[2] Do we have a nonparametric efficiency lower bound for $\widehat{\theta}_{DR}(t)$?

---

[2]I acknowledge Ted Westling and Aaron Hudson for pointing out this direction.

**Question:**[2] Do we have a nonparametric efficiency lower bound for $\widehat{\theta}_{\mathrm{DR}}(t)$?

- $t \mapsto \theta(t) := \Psi(\mathrm{P}_0)(t)$ is *not* pathwise differentiable (Bickel et al., 1998; Hirano and Porter, 2012; Luedtke and van der Laan, 2016):

$$\forall t \in \mathcal{T}, \quad \exists \{\mathrm{P}_\epsilon : \epsilon \in \mathbb{R}\} \quad \text{s.t.} \quad \lim_{\epsilon \to 0} \frac{\Psi(\mathrm{P}_\epsilon)(t) - \Psi(\mathrm{P}_0)(t)}{\epsilon} \quad \text{does not exist.}$$

---

[2]I acknowledge Ted Westling and Aaron Hudson for pointing out this direction.

**Question:**[2] Do we have a nonparametric efficiency lower bound for $\widehat{\theta}_{\mathrm{DR}}(t)$?

- $t \mapsto \theta(t) := \Psi(\mathrm{P}_0)(t)$ is *not* pathwise differentiable (Bickel et al., 1998; Hirano and Porter, 2012; Luedtke and van der Laan, 2016):

$$\forall t \in \mathcal{T}, \quad \exists \{\mathrm{P}_\epsilon : \epsilon \in \mathbb{R}\} \quad \text{s.t.} \quad \lim_{\epsilon \to 0} \frac{\Psi(\mathrm{P}_\epsilon)(t) - \Psi(\mathrm{P}_0)(t)}{\epsilon} \quad \text{does not exist.}$$

- For a fixed $h > 0$, the smooth functional $\Phi(\mathrm{P}_0)(t) := \mathbb{E}\left[ \frac{Y \cdot \left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S)} \right]$ is pathwise differentiable (van der Laan et al., 2018; Takatsu and Westling, 2024).

---

[2]I acknowledge Ted Westling and Aaron Hudson for pointing out this direction.

**Question:**[2] Do we have a nonparametric efficiency lower bound for $\widehat{\theta}_{DR}(t)$?

- $t \mapsto \theta(t) := \Psi(P_0)(t)$ is *not* pathwise differentiable (Bickel et al., 1998; Hirano and Porter, 2012; Luedtke and van der Laan, 2016):

$$\forall t \in \mathcal{T}, \quad \exists \{P_\epsilon : \epsilon \in \mathbb{R}\} \quad \text{s.t.} \quad \lim_{\epsilon \to 0} \frac{\Psi(P_\epsilon)(t) - \Psi(P_0)(t)}{\epsilon} \quad \text{does not exist.}$$

- For a fixed $h > 0$, the smooth functional $\Phi(P_0)(t) := \mathbb{E}\left[\frac{Y \cdot \left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S)}\right]$ is pathwise differentiable (van der Laan et al., 2018; Takatsu and Westling, 2024).

- Up to a shrinking bias $O(h^2)$, the efficient influence function for $\Phi(P_0)(t)$ leads to

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

▶ The asymptotic variances of $\widehat{\theta}_{DR}(t)$ and $\widehat{\theta}_{EIF}(t)$ are the same (or differing by $O(h^2)$)!
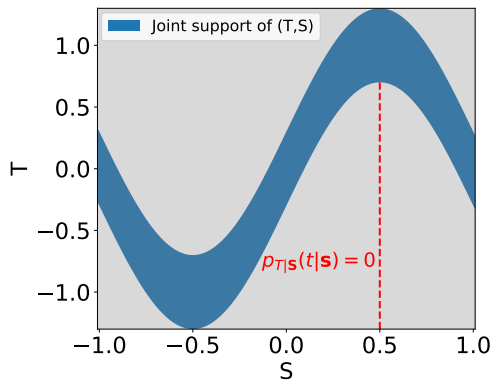
[2] I acknowledge Ted Westling and Aaron Hudson for pointing out this direction.

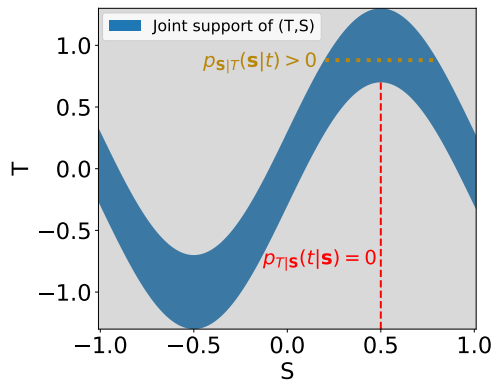# Nonparametric Inference on $\theta(t)$ Without Positivity

## Assumption (Identification Conditions)

1. *(Consistency) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.*

2. *(Ignorability) $Y(t)$ is conditionally independent of $T$ given $\boldsymbol{S}$ for all $t \in \mathcal{T}$.*

3. *(Treatment Variation) $\mathrm{Var}(T|\boldsymbol{S}=\boldsymbol{s}) > 0$ for all $\boldsymbol{s} \in \mathcal{S}$.*

## Assumption (Identification Conditions)

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.

2. *(Ignorability)* $Y(t)$ *is conditionally independent of* $T$ *given* $\boldsymbol{S}$ *for all* $t \in \mathcal{T}$.

3. *(Treatment Variation)* $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$ *for all* $\boldsymbol{s} \in \mathcal{S}$.



## Assumption (Extrapolation; Zhang et al. 2024)

*Assume* $(t, \boldsymbol{s}) \mapsto \mathbb{E}\left[Y(t)|\boldsymbol{S} = \boldsymbol{s}\right]$ *to be differentiable w.r.to* $t$ *for any* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$ *with* $p_{\boldsymbol{S}|T}(\boldsymbol{s}|t) > 0$ *and*

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

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

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

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

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

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

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{or} \quad Y = \bar{m}(T) + \eta(S) + \epsilon.$$

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

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

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{or} \quad Y = \bar{m}(T) + \eta(S) + \epsilon.$$

► We identify $m(t) = \mathbb{E}[Y(t)]$ and $\theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)]$ via two pillars in calculus:

$$\underbrace{\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,S)\bigg| T = t\right]}_{\textit{Differentiation}} \quad \text{and} \quad \underbrace{m(t) = \mathbb{E}\left[Y + \int_{u=T}^{u=t}\theta(u)\,du\right]}_{\textit{Integration}} \quad \text{with } \mu(t,s) = \mathbb{E}(Y|T=t,S).$$

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

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

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{or} \quad Y = \bar{m}(T) + \eta(S) + \epsilon.$$

▶ We identify $m(t) = \mathbb{E}[Y(t)]$ and $\theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)]$ via two pillars in calculus:

$$\underbrace{\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,S)\bigg| T = t\right]}_{\textit{Differentiation}} \quad \text{and} \quad \underbrace{m(t) = \mathbb{E}\left[Y + \int_{u=T}^{u=t} \theta(u)\, du\right]}_{\textit{Integration}} \text{ with } \mu(t,s) = \mathbb{E}(Y|T=t, S).$$

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

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

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

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

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

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

- Recall the standard (oracle) IPW estimators for $m(t)$ and $\theta(t)$:

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

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

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

- Recall the standard (oracle) IPW estimators for $m(t)$ and $\theta(t)$:

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

Proposition (Proposition 2 in Zhang and Chen 2025)

$$\lim_{h \to 0} \mathbb{E}\left[\tilde{m}_{\text{IPW}}(t)\right] = \bar{m}(t) \cdot \rho(t) + \omega(t) \neq m(t), \qquad \text{with} \quad \rho(t) = \mathbb{P}\left(S \in \mathcal{S}(t)\right),$$

$$\lim_{h \to 0} \mathbb{E}\left[\tilde{\theta}_{\text{IPW}}(t)\right] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t), \qquad \text{and} \quad \omega(t) = \mathbb{E}\left[\eta(S) \mathbb{1}_{\{S \in \mathcal{S}(t)\}}\right].$$

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

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

- Recall the standard (oracle) IPW estimators for $m(t)$ and $\theta(t)$:

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

**Proposition (Proposition 2 in Zhang and Chen 2025)**

$$\lim_{h \to 0} \mathbb{E}\left[\tilde{m}_{\mathrm{IPW}}(t)\right] = \bar{m}(t) \cdot \rho(t) + \omega(t) \neq m(t), \qquad \text{with} \quad \rho(t) = \mathbb{P}\left(S \in \mathcal{S}(t)\right),$$

$$\lim_{h \to 0} \mathbb{E}\left[\tilde{\theta}_{\mathrm{IPW}}(t)\right] = \begin{cases} \bar{m}'(t) \cdot \rho(t) \\ \infty \end{cases} \neq \theta(t), \qquad \text{and} \quad \omega(t) = \mathbb{E}\left[\eta(S)\mathbb{1}_{\{S \in \mathcal{S}(t)\}}\right].$$

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

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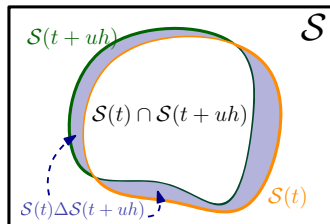
where $\rho(t) = \mathbb{P}\left(\boldsymbol{S} \in \mathcal{S}(t)\right)$.

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

where $\rho(t) = \mathbb{P}\left(S \in \mathcal{S}(t)\right)$.

1. We first want to disentangle $\theta(t) = \bar{m}'(t)$ from the bias term:
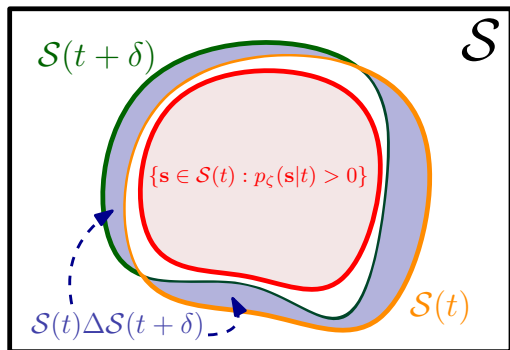
$$\mathbb{E}\left[\frac{Y \cdot \left(\frac{T-t}{h}\right)K\left(\frac{T-t}{h}\right) \cdot p_{S|T}(S|t)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S) \cdot p_S(S)}\right] = \bar{m}'(t) + O(h^2)$$

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

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

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

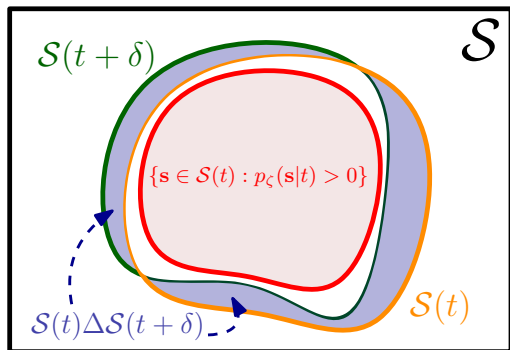2. We replace $p_{S|T}(s|t)$ with a $\zeta$-interior conditional density $p_\zeta(s|t)$ so that

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

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

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

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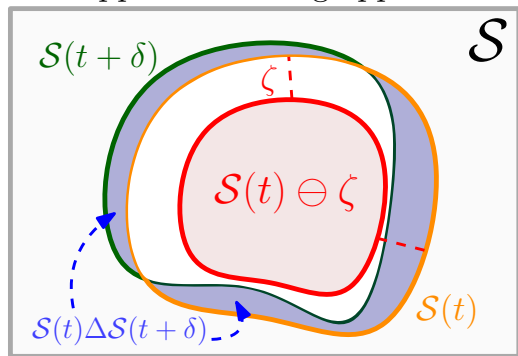


Now, we have that

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

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

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


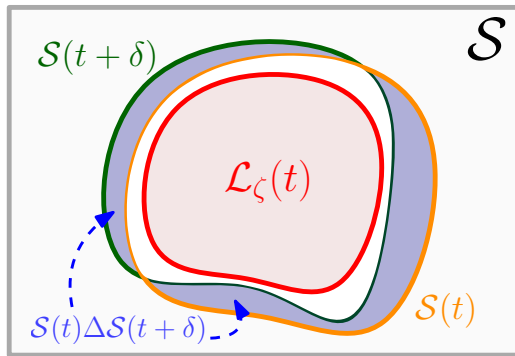
Support shrinking approach

Level set approach

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

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

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

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

▶ **Bias-Corrected IPW Estimator Without Positivity:**

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

• $\widehat{p}(t, s), \widehat{p}_\zeta(s|t)$ are estimators of $p(t, s), p_\zeta(s|t)$ and $\zeta = 0.5 \cdot \max \{ \widehat{p}_{S|T}(S_i | t) : i = 1, ..., n \}$.

▶ **Bias-Corrected IPW Estimator Without Positivity:**

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

• $\widehat{p}(t, s), \widehat{p}_{\zeta}(s|t)$ are estimators of $p(t, s), p_{\zeta}(s|t)$ and $\zeta = 0.5 \cdot \max\{\widehat{p}_{S|T}(S_i|t) : i = 1, ..., n\}$.

▶ **Bias-Corrected DR Estimator Without Positivity:**

$$\widehat{\theta}_{C,\text{DR}}(t) = \underbrace{\frac{1}{nh^2} \sum_{i=1}^{n} \frac{\left(\frac{T_i-t}{h}\right) K\left(\frac{T_i-t}{h}\right) \widehat{p}_{\zeta}(S_i|t)}{\kappa_2 \cdot \widehat{p}(T_i, S_i)} \left[Y_i - \widehat{\mu}(t, S_i) - (T_i - t) \cdot \widehat{\beta}(t, S_i)\right]}_{\text{IPW component}}$$

$$+ \underbrace{\int \widehat{\beta}(t, s) \cdot \widehat{p}_{\zeta}(s|t) \, ds}_{\text{RA component}}.$$

## 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 $\{(Y_i, T_i, S_i)\}_{i=1}^{n}$;*

2. *$\sqrt{nh} \, ||\widehat{p}_{\zeta}(S|t) - \bar{p}_{\zeta}(S|t)||_{L_2} = o_P(1)$ with $\widehat{p}_{\zeta}(s|t) \xrightarrow{P} \bar{p}_{\zeta}(s|t)$;*

3. *at least one of the model specification conditions hold:*

   - *$\widehat{p}(t, s) \xrightarrow{P} \bar{p}(t, s) = p(t, s)$ (joint density model),*

   - *$\widehat{\mu}(t, s) \xrightarrow{P} \bar{\mu}(t, s) = \mu(t, s)$ and $\widehat{\beta}(t, s) \xrightarrow{P} \bar{\beta}(t, s) = \beta(t, s)$ (outcome model);*

4. *$\displaystyle\sup_{|u-t| \leq h} ||\widehat{p}(u, S) - p(u, S)||_{L_2} \left[ ||\widehat{\mu}(t, S) - \mu(t, S)||_{L_2} + h \left|\left| \widehat{\beta}(t, S) - \beta(t, S) \right|\right|_{L_2} \right] = o_P\left(\frac{1}{\sqrt{nh}}\right)$,*
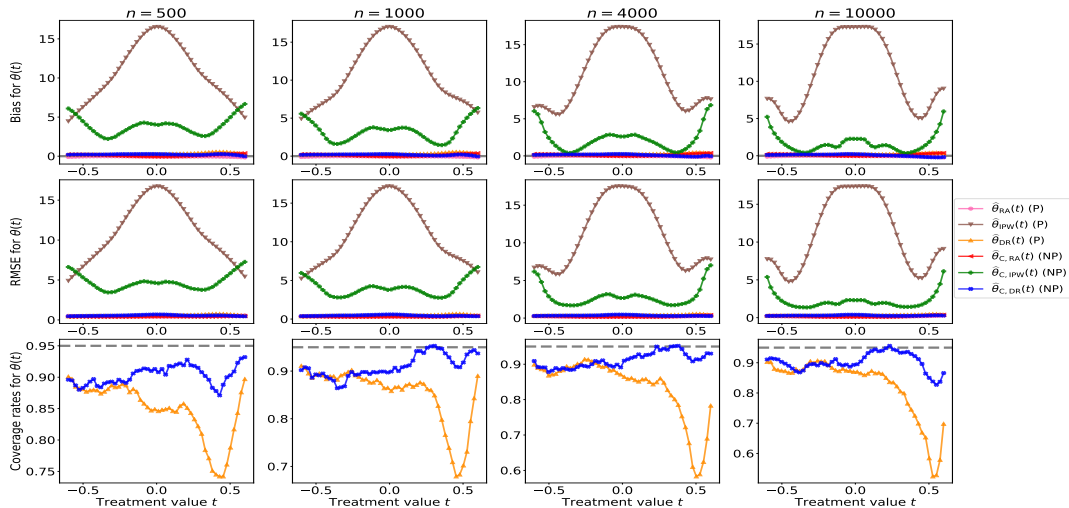
*we prove that*

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

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

# Experiments and Discussion

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



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

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.

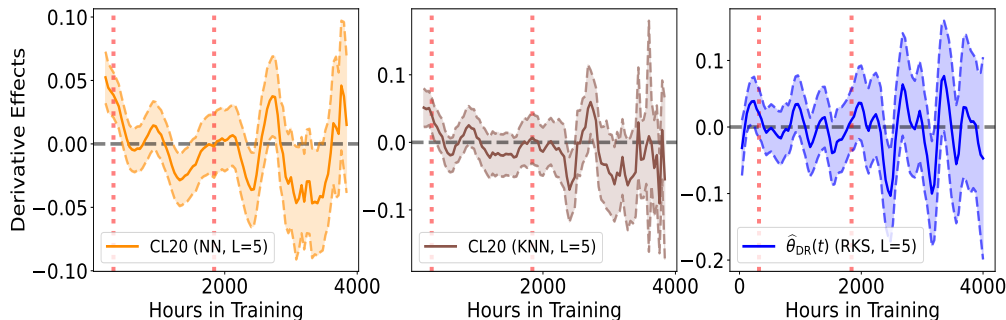- $S$ comprises 49 socioeconomic characteristics, and $n = 4024$.

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.
- $S$ comprises 49 socioeconomic characteristics, and $n = 4024$.

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right], t \in \mathcal{T} \subset \mathbb{R}$.

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right], t \in \mathcal{T} \subset \mathbb{R}$.

**❶ Under the positivity condition:**

- We propose $\widehat{\theta}_{\mathrm{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right], t \in \mathcal{T} \subset \mathbb{R}$.

**❶ Under the positivity condition:**

- We propose $\widehat{\theta}_{\text{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

**❷ Without the positivity condition:**

- Our bias-corrected IPW and DR estimators $\widehat{\theta}_{\text{C,IPW}}(t), \widehat{\theta}_{\text{C,DR}}(t)$ reveal interesting connections to nonparametric level set estimation problems (Bonvini et al., 2023):

$$\textit{Causal Inference} \quad \Longleftrightarrow \quad \textit{Geometric Data Analysis.}$$

We study (nonparametric) doubly robust inference on $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right], t \in \mathcal{T} \subset \mathbb{R}$.

**1 Under the positivity condition:**

- We propose $\widehat{\theta}_{\mathrm{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

**2 Without the positivity condition:**

- Our bias-corrected IPW and DR estimators $\widehat{\theta}_{\mathrm{C,IPW}}(t), \widehat{\theta}_{\mathrm{C,DR}}(t)$ reveal interesting connections to nonparametric level set estimation problems (Bonvini et al., 2023):

  *Causal Inference* $\iff$ *Geometric Data Analysis.*

**3 Future Works:**

- Sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022).

- Generalize our derivative estimators to other causal estimands:
  - instantaneous causal effect $\frac{d}{dt}\mathbb{E}\left[Y(t)|S = s\right]$ (Stolzenberg, 1980);
  - direct and indirect effects in mediation analysis (Huber et al., 2020; Xu et al., 2021)?

# **Thank you!**

More details can be found in

[1] 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.

[2] 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.

All the code and data are available at
hhttps://github.com/zhangyk8/npDRDeriv.

Python Package: npDoseResponse.

# Reference

P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer New York, 1998.

M. Bonvini, E. H. Kennedy, and L. J. Keele. Minimax optimal subgroup identification. *arXiv preprint arXiv:2306.17464*, 2023.

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.

V. Chernozhukov, C. Cinelli, W. Newey, A. Sharma, and V. Syrgkanis. Long story short: Omitted variable bias in causal machine learning. Technical report, National Bureau of Economic Research, 2022.

K. Colangelo and Y.-Y. Lee. Double debiased machine learning nonparametric inference with continuous treatments. *arXiv preprint arXiv:2004.03036*, 2020.

R. D. Gill and J. M. Robins. Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 29(6):1785–1811, 2001.

K. Hirano and G. W. Imbens. *The Propensity Score with Continuous Treatments*, chapter 7, pages 73–84. John Wiley & Sons, Ltd, 2004.

K. Hirano and J. R. Porter. Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790, 2012.

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.

A. Luedtke. Simplifying debiased inference via automatic differentiation and probabilistic programming. *arXiv preprint arXiv:2405.08675*, 2024.

A. R. Luedtke and M. J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713–742, 2016.

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.

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.

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. Xu, N. Sani, A. Ghassami, and I. Shpitser. Multiply robust causal mediation analysis with continuous treatments. *arXiv preprint arXiv:2105.09254*, 2021.

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 $K : \mathbb{R} \to [0, \infty)$ 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 $j = 1, 2, \ldots$.

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:**

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

and

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

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

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

and

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

- **Static Intervention:** Causal dose-response curve $t_0 \mapsto m(t_0) = \mathbb{E}\left[Y(t_0)\right] = \mathbb{E}\left[Y^{(t_0)}\right]$.

- **Static Intervention:** Causal dose-response curve $t_0 \mapsto m(t_0) = \mathbb{E}\left[Y(t_0)\right] = \mathbb{E}\left[Y^{(t_0)}\right]$.

- **Stochastic Intervention:** $Y^{q_h(t_0)}$ is the potential outcome had the treatment been randomly assigned from the conditional density

$$q_h(t, t_0|\boldsymbol{s}) = \frac{K\left(\frac{t-t_0}{h}\right) \cdot p_{T|S}(t|\boldsymbol{s})}{\int_{\mathcal{T}} K\left(\frac{u-t_0}{h}\right) \cdot p_{T|S}(u|\boldsymbol{s})\, du} \qquad \text{with } K : \mathbb{R} \to [0, \infty) \text{ being a kernel function.}$$
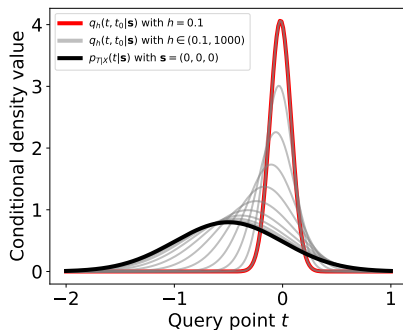


Figure: Kernel-smoothed tilted densities at $\boldsymbol{s} = \boldsymbol{0}$.

- **Static Intervention:** Causal dose-response curve $t_0 \mapsto m(t_0) = \mathbb{E}\left[Y(t_0)\right] = \mathbb{E}\left[Y^{(t_0)}\right]$.

- **Stochastic Intervention:** $Y^{q_h(t_0)}$ is the potential outcome had the treatment been randomly assigned from the conditional density

$$q_h(t, t_0|\boldsymbol{s}) = \frac{K\left(\frac{t-t_0}{h}\right) \cdot p_{T|S}(t|\boldsymbol{s})}{\int_{\mathcal{T}} K\left(\frac{u-t_0}{h}\right) \cdot p_{T|S}(u|\boldsymbol{s})\, du} \quad \text{with } K : \mathbb{R} \to [0, \infty) \text{ being a kernel function.}$$

▶ **Property:** The identification of the causal curve $t_0 \mapsto \bar{m}_h(t_0) = \mathbb{E}\left[Y^{q_h(t_0)}\right]$ **doesn't** require the positivity condition!!

# Kernel-Smoothed Tilted Intervention

- **Static Intervention:** Causal dose-response curve $t_0 \mapsto m(t_0) = \mathbb{E}\left[Y(t_0)\right] = \mathbb{E}\left[Y^{(t_0)}\right]$.

- **Stochastic Intervention:** $Y^{q_h(t_0)}$ is the potential outcome had the treatment been randomly assigned from the conditional density
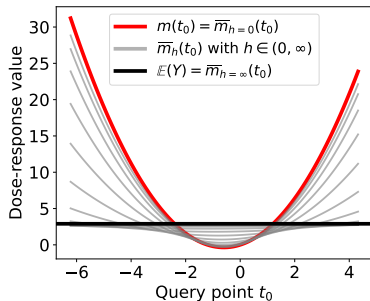
$$q_h(t, t_0 | \boldsymbol{s}) = \frac{K\left(\frac{t-t_0}{h}\right) \cdot p_{T|S}(t|\boldsymbol{s})}{\int_{\mathcal{T}} K\left(\frac{u-t_0}{h}\right) \cdot p_{T|S}(u|\boldsymbol{s})\, du} \quad \text{with } K : \mathbb{R} \to [0, \infty) \text{ being a kernel function.}$$

▶ **Property:** The identification of the causal curve $t_0 \mapsto \bar{m}_h(t_0) = \mathbb{E}\left[Y^{q_h(t_0)}\right]$ **doesn't** require the positivity condition!!

- When positivity is violated and $h \to 0$, $t_0 \mapsto \bar{m}_h(t_0)$ converges to

$$\int_{\mathcal{S}(t_0)} \mu(t_0, \boldsymbol{s})\, d\mathrm{P}(\boldsymbol{s}) + \int_{\mathcal{S} \backslash \mathcal{S}(t_0)} \mu(t_{0,\mathrm{proj}}, \boldsymbol{s})\, d\mathrm{P}(\boldsymbol{s})$$

that depends on the **geometry** of $\mathcal{S}(t_0)$.

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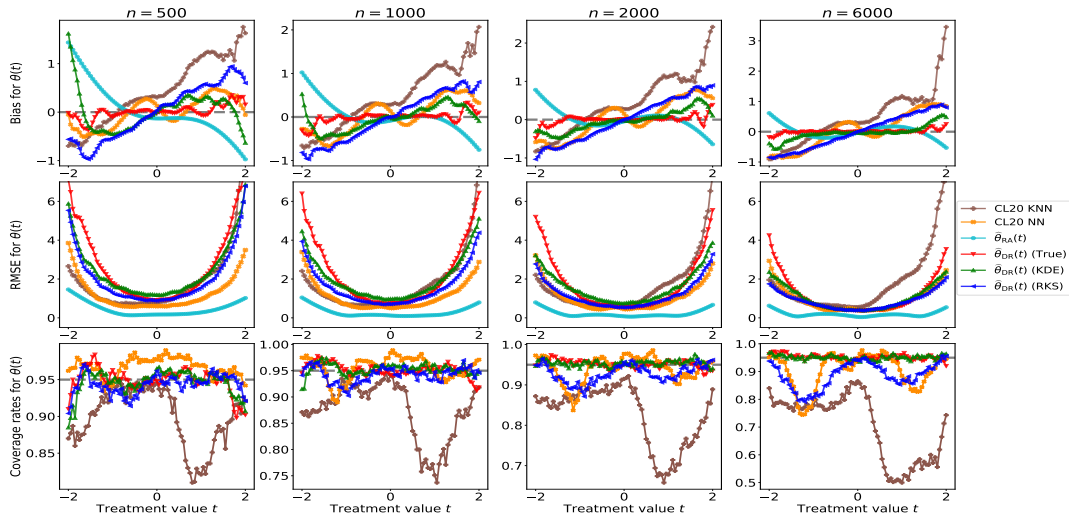
$$Y = 1.2\,T + T^2 + TS_1 + 1.2\,\boldsymbol{\xi}^T\boldsymbol{S} + \epsilon\sqrt{0.5 + F_{\mathcal{N}(0,1)}(S_1)}, \quad \epsilon \sim \mathcal{N}(0,1),$$

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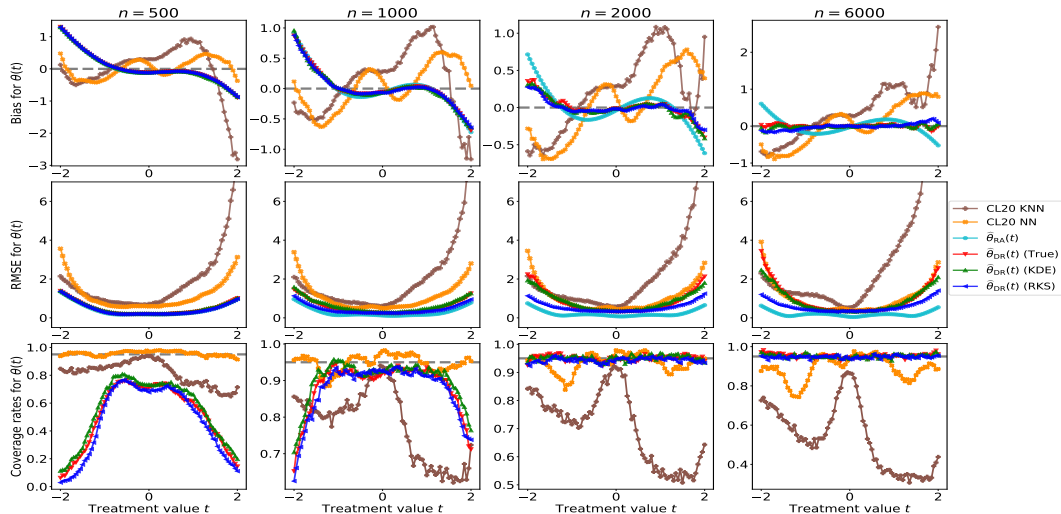
where

- $F_{\mathcal{N}(0,1)}$ 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 $\theta(t) = 1.2 + 2t$.

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.

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.