Nonparametric Inference on Causal Effects of Continuous Treatments: With and Without the Positivity Condition

Yikun Zhang

Joint work with *Professor Yen-Chi Chen*

Department of Statistics,
University of Washington

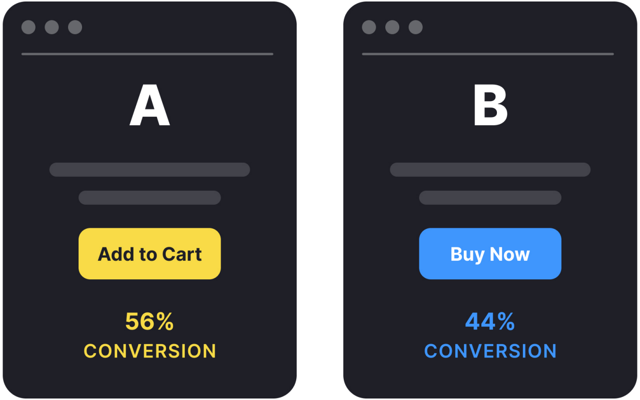
Pinterest
May 20, 2025

Department of
STATISTICS

UNIVERSITY *of* WASHINGTON

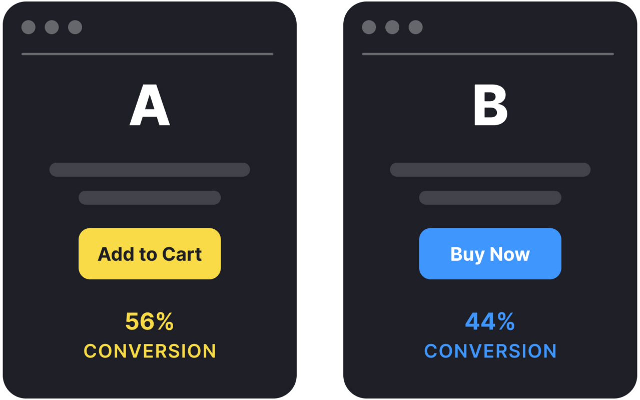
Fundamental Problem of Causal Inference

- Study the causal effect of a treatment $T \in \mathcal{T}$ on the outcome of interest $Y \in \mathcal{Y}$.



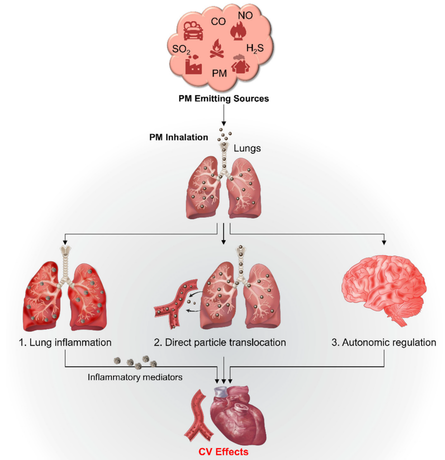
Fundamental Problem of Causal Inference

- Study the causal effect of a treatment $T \in \mathcal{T}$ on the outcome of interest $Y \in \mathcal{Y}$.



- The treatment variable T is *binary*, *i.e.*, $\mathcal{T} = \{0, 1\}$.
- Only one potential outcome, $Y(1)$ or $Y(0)$, can be observed for each individual.
- The common causal estimand is the average treatment effect $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$.

- We want to study the causal effects of $PM_{2.5}$ levels on Cardiovascular Mortality Rates (CMRs).



Biological pathways associated with particulate matter (PM) and cardiovascular disease (Miller and Newby, 2020; Basith et al., 2022).

FIPS	County name	Longitude	Latitude	PM2.5	CMR
1025	Clarke	-87.830772	31.676955	6.766443	379.421713
1061	Geneva	-85.839330	31.094869	8.254272	378.524698
1073	Jefferson	-86.896571	33.554343	10.825441	352.790427
1077	Lauderdale	-87.654117	34.901500	9.208783	332.594557
5085	Lonoke	-91.887917	34.754412	8.213144	365.061085
8045	Garfield	-107.903621	39.599420	2.601772	250.781477

The dataset contains the average annual cardiovascular mortality rates (CMRs) and $PM_{2.5}$ levels across $n = 2132$ U.S. counties from 1990 to 2010 (Wyatt et al., 2020a,b).

FIPS	County name	Longitude	Latitude	PM2.5	CMR
1025	Clarke	-87.830772	31.676955	6.766443	379.421713
1061	Geneva	-85.839330	31.094869	8.254272	378.524698
1073	Jefferson	-86.896571	33.554343	10.825441	352.790427
1077	Lauderdale	-87.654117	34.901500	9.208783	332.594557
5085	Lonoke	-91.887917	34.754412	8.213144	365.061085
8045	Garfield	-107.903621	39.599420	2.601772	250.781477

The dataset contains the average annual cardiovascular mortality rates (CMRs) and $PM_{2.5}$ levels across $n = 2132$ U.S. counties from 1990 to 2010 (Wyatt et al., 2020a,b).

- The treatment variable T, *i.e.*, the $PM_{2.5}$ level at each county, is a quantitative measure. In other words, it is *not a binary but continuous variable*!

The common causal estimands under a *binary* treatment are

- $\mathbb{E}[Y(t)]$ = mean counterfactual outcome when we set $T = t \in \{0, 1\}$.
- $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$ = average treatment effect.

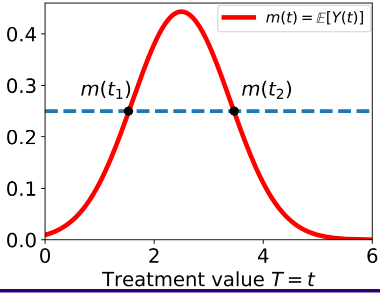
▶ **Question:** What are the counterparts of the above estimands under a *continuous* treatment $T \in \mathcal{T} \subset \mathbb{R}$?

The common causal estimands under a *binary* treatment are

- $\mathbb{E}\left[Y(t)\right]$ = mean counterfactual outcome when we set $T = t \in \{0, 1\}$.
- $\mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right]$ = average treatment effect.

▶ **Question:** What are the counterparts of the above estimands under a *continuous* treatment $T \in \mathcal{T} \subset \mathbb{R}$?

- $t \mapsto m(t) := \mathbb{E}\left[Y(t)\right]$ = (causal) dose-response curve.
- $t \mapsto \theta(t) := m'(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right]$ = (causal) derivative effect curve.



The goal of our study is to identify and estimate

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

[1] Credits to Marco Carone for this nice categorization of existing approaches.

Challenges and Existing Approaches

The goal of our study is to identify and estimate

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

Challenge: $m(t)$ and $\theta(t)$ are not pathwise differentiable (Bickel et al., 1998) and cannot be estimated in the rate $1/\sqrt{n}$.

[1]Credits to Marco Carone for this nice categorization of existing approaches.

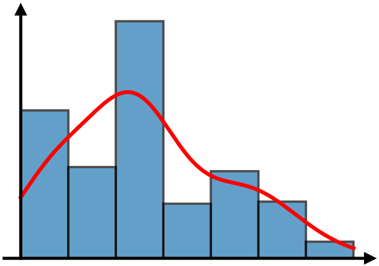
The goal of our study is to identify and estimate

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

Challenge: $m(t)$ and $\theta(t)$ are not pathwise differentiable (Bickel et al., 1998) and cannot be estimated in the rate $1/\sqrt{n}$.

Existing Approaches:[1]

1. *Discretization:* Divide the range of T into bins and assign observations accordingly.



[1]Credits to Marco Carone for this nice categorization of existing approaches.

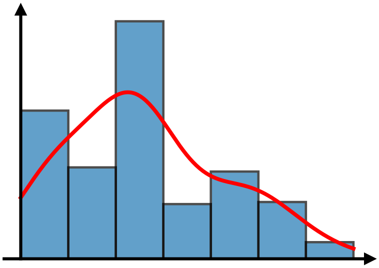
The goal of our study is to identify and estimate

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

Challenge: $m(t)$ and $\theta(t)$ are not pathwise differentiable (Bickel et al., 1998) and cannot be estimated in the rate $1/\sqrt{n}$.

Existing Approaches:[1]

1. *Discretization:* Divide the range of T into bins and assign observations accordingly.



Pros: Allow direct applications of standard methods for discrete treatments (*e.g.*, the block-based diagnostics by Hirano and Imbens 2004; Bia and Mattei 2008).

Cons: Difficult to choose the cutoff points for binning.

Cons: Potentially lose useful information.

[1] Credits to Marco Carone for this nice categorization of existing approaches.

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

2. *Marginal Structural Model*: Impose parametric structural assumptions on $\mathbb{E}\left[Y(t)\right]$ (Robins et al., 2000; van der Laan and Robins, 2003; Neugebauer and van der Laan, 2007), *e.g.*,

$$m(t) = \mathbb{E}\left[Y(t)\right] = \alpha_1 + \alpha_2 \cdot t.$$

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

2. *Marginal Structural Model*: Impose parametric structural assumptions on $\mathbb{E}\left[Y(t)\right]$ (Robins et al., 2000; van der Laan and Robins, 2003; Neugebauer and van der Laan, 2007), *e.g.*,

$$m(t) = \mathbb{E}\left[Y(t)\right] = \alpha_1 + \alpha_2 \cdot t.$$

Pros: Only need to estimate regression parameters α_1, α_2, which can achieve the parametric rate of convergence.

Cons: The parametric structural assumption could be violated!

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

3. *Stochastic Interventions* (Díaz-Muñoz and van der Laan, 2012; Haneuse and Rotnitzky, 2013; Schindl et al., 2024):

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

3. *Stochastic Interventions* (Díaz-Muñoz and van der Laan, 2012; Haneuse and Rotnitzky, 2013; Schindl et al., 2024):
   - Incremental causal 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 effect (Härdle and Stoker, 1989; Powell et al., 1989; Newey and Stoker, 1993; Hines et al., 2023):

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

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

3. *Stochastic Interventions* (Díaz-Muñoz and van der Laan, 2012; Haneuse and Rotnitzky, 2013; Schindl et al., 2024):
   - Incremental causal 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 effect (Härdle and Stoker, 1989; Powell et al., 1989; Newey and Stoker, 1993; Hines et al., 2023):

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

Pros These new 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.

The goal of our study is to identify and estimate

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

4. *Nonparametric Structural Assumptions:*
   - Shape constraint, *e.g.*, monotonicity (Westling et al., 2020; Westling and Carone, 2020).

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

4. *Nonparametric Structural Assumptions:*
   - Shape constraint, *e.g.*, monotonicity (Westling et al., 2020; Westling and Carone, 2020).
   - **Smoothness conditions**, *e.g.*, higher-order differentiability + localization techniques (Kennedy et al., 2017; Kallus and Zhou, 2018; Colangelo and Lee, 2020; Bonvini and Kennedy, 2022; Takatsu and Westling, 2024; Luedtke and Chung, 2024).

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

4. *Nonparametric Structural Assumptions:*
   - Shape constraint, *e.g.*, monotonicity (Westling et al., 2020; Westling and Carone, 2020).
   - **Smoothness conditions**, *e.g.*, higher-order differentiability + localization techniques (Kennedy et al., 2017; Kallus and Zhou, 2018; Colangelo and Lee, 2020; Bonvini and Kennedy, 2022; Takatsu and Westling, 2024; Luedtke and Chung, 2024).

Pros: Allow flexibility in estimating $m(t)$ and $\theta(t)$.

Cons: Require estimating nuisance functions and/or tuning (hyper)parameters.

Existing Approaches (Continued)

The goal of our study is to identify and estimate

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

4. *Nonparametric Structural Assumptions:*
    - Shape constraint, *e.g.*, monotonicity (Westling et al., 2020; Westling and Carone, 2020).
    - **Smoothness conditions**, *e.g.*, higher-order differentiability + localization techniques (Kennedy et al., 2017; Kallus and Zhou, 2018; Colangelo and Lee, 2020; Bonvini and Kennedy, 2022; Takatsu and Westling, 2024; Luedtke and Chung, 2024).

Pros Allow flexibility in estimating $m(t)$ and $\theta(t)$.

Cons Require estimating nuisance functions and/or tuning (hyper)parameters.

▶ Our works leverage **smoothness conditions** with **kernel smoothing** techniques.

The goal of our study is to identify and estimate

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

Assumption (Identification Condition)

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

In randomized controlled trials (RCTs),

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

The goal of our study is to identify and estimate

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

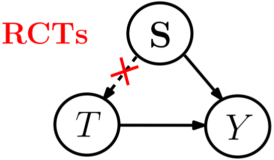
Assumption (Identification Condition)

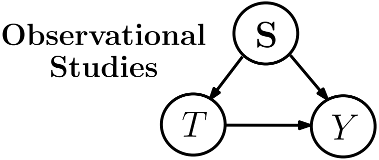
1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.

In randomized controlled trials (RCTs),

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

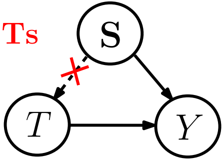
- Estimating $m(t)$ is to fit the regression function $t \mapsto \mathbb{E}(Y|T = t)$ on $\{(Y_i, T_i)\}_{i=1}^n$.
- Recovering $\theta(t)$ is a derivative estimation problem (Gasser and Müller, 1984).



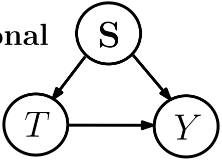


[2]Some mild interchangeability assumptions are needed; see Theorem 1.1 in Shao (2003).

RCTs



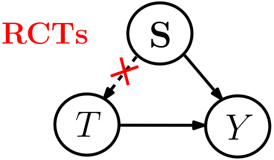
Observational Studies

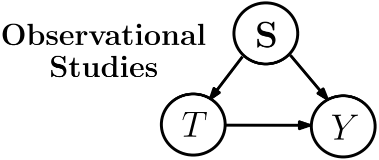


Assumption (Identification Conditions)

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* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

[2] 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* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

$$m(t) = \mathbb{E}\left[Y(t)\right] = \mathbb{E}\left[\mathbb{E}(Y|T = t, \boldsymbol{S})\right] \quad \text{and} \quad \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] \stackrel{(*)^2}{=} \mathbb{E}\left[\frac{\partial}{\partial t}\mathbb{E}(Y|T = t, \boldsymbol{S})\right].$$

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

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

An Example of the Positivity Violation

Assumption (Positivity Condition)

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

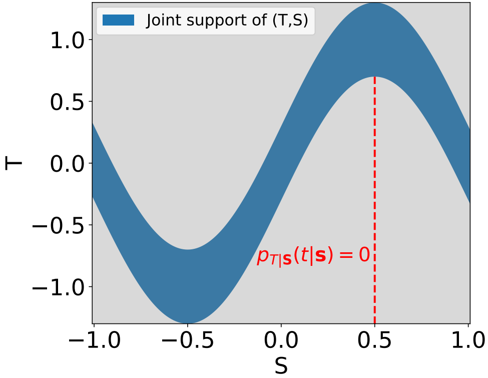
- Positivity is a very strong assumption with continuous treatments!

Assumption (Positivity Condition)

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

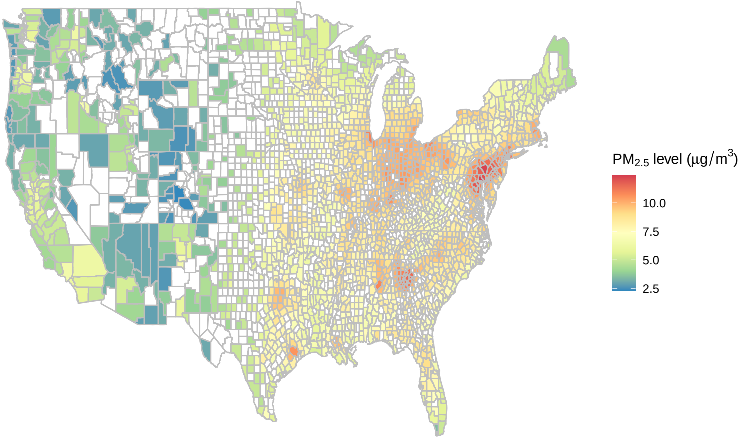
- Positivity is a very strong assumption with continuous treatments!

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



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

$PM_{2.5}$ Distribution at the County Level



Average $PM_{2.5}$ levels from 1990 to 2010 in $n = 2132$ counties.

- T is $PM_{2.5}$ level, and S consists of the county location and socioeconomic factors.
- Only one or several $PM_{2.5}$ levels are available per county in the dataset, and the positivity condition is violated!

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

Under the positivity condition:

1. Review the existing estimators for $m(t)$ 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. Review the existing estimators for $m(t)$ via kernel smoothing.
2. Propose our doubly robust (DR) estimator for $\theta(t)$.

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

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

Under the positivity condition:

1. Review the existing estimators for $m(t)$ via kernel smoothing.
2. Propose our doubly robust (DR) estimator for $\theta(t)$.

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

Without the positivity condition:

3. $m(t)$ and $\theta(t)$ are identifiable with a new extrapolation assumption satisfied by, *e.g.*,

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

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

Under the positivity condition:

1. Review the existing estimators for $m(t)$ via kernel smoothing.
2. Propose our doubly robust (DR) estimator for $\theta(t)$.

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

Without the positivity condition:

3. $m(t)$ and $\theta(t)$ are identifiable with a new extrapolation assumption satisfied by, *e.g.*,

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

4. The usual IPW estimators for $m(t)$ 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. Review the existing estimators for $m(t)$ via kernel smoothing.
2. Propose our doubly robust (DR) estimator for $\theta(t)$.

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

Without the positivity condition:

3. $m(t)$ and $\theta(t)$ are identifiable with a new extrapolation assumption satisfied by, *e.g.*,

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

4. The usual IPW estimators for $m(t)$ and $\theta(t)$ are still *biased* even under model (1).
5. 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.

Part I: Nonparametric Inference on $m(t)$ and $\theta(t)$ Under Positivity

This part is based on **Sections 2 and 3** in [1]:

[1] Y. Zhang and Y.-C. Chen (2025). **Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments**. *arXiv:2501.06969*. https://arxiv.org/abs/2501.06969.

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* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

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

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* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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

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

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

- $K: \mathbb{R} \to [0, \infty)$ 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 observed data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, there are three main strategies for estimating

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

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

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

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

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

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

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

To estimate $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,S)\right]$ from $\{(Y_i, T_i, S_i)\}_{i=1}^n$, we could also have 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 to generalize the IPW form $m(t) = \lim_{h\to 0}\mathbb{E}\left[\frac{Y\cdot K\left(\frac{T-t}{h}\right)}{h\cdot p_{T|S}(T|S)}\right]$ to estimate $\theta(t)$?

To estimate $\theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,S)\right]$ from $\{(Y_i, T_i, S_i)\}_{i=1}^n$, we could also have 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 to generalize the IPW form $m(t) = \lim_{h\to 0}\mathbb{E}\left[\frac{Y\cdot K\left(\frac{T-t}{h}\right)}{h\cdot p_{T|S}(T|S)}\right]$ to estimate $\theta(t)$?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

▶ **Remark:** All these AIPW estimators for $\theta(t)$ are, at best, **singly robust**!!

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

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

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

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

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

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

we prove that

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

we prove that

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

Statistical Inference on $\theta(t)$

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

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

Statistical Inference on $\theta(t)$

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

Statistical Inference on $\theta(t)$

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

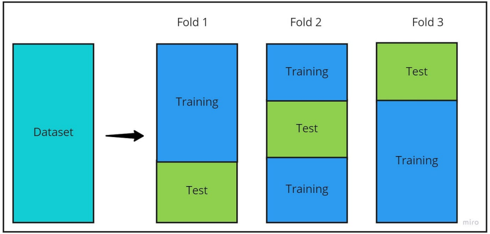
$$\sqrt{nh^3}\left[\widehat{\theta}_{\mathrm{DR}}(t) - \theta(t) - h^2 B_\theta(t)\right] \xrightarrow{d} \mathcal{N}\left(0, V_\theta(t)\right).$$

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.



Statistical Inference on $\theta(t)$

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. 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).

Nonparametric Efficiency Guarantee for $\widehat{\theta}_{\mathrm{DR}}(t)$

Question:[3] Do we have a nonparametric efficiency lower bound for $\widehat{\theta}_{\mathrm{DR}}(t)$?

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

Question:[3] 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):

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

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

Question:[3] 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).

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

Nonparametric Efficiency Guarantee for $\widehat{\theta}_{\mathrm{DR}}(t)$

Question:[3] 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).
- Up to a shrinking bias $O(h^2)$, the efficient influence function for $\Phi(\mathrm{P}_0)(t)$ leads to

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

► The asymptotic variances of $\widehat{\theta}_{\mathrm{DR}}(t)$ and $\widehat{\theta}_{\mathrm{EIF}}(t)$ are the same (or differing by $O(h^2)$)!

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

Part II: Nonparametric Inference on $m(t)$ and $\theta(t)$ Without Positivity

This part is based on **Sections 4 and 5** in [1] and **Sections 2, 3, and 4** in [2]:

[1] Y. Zhang and Y.-C. Chen (2025). **Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments**. *arXiv:2501.06969*. https://arxiv.org/abs/2501.06969.

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

Why Do We Need Positivity?

Assumption (Identification Conditions)

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

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

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

The IPW approaches also rely on the following identities:

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

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* $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$.

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:

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

▶ **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 $p(t, \boldsymbol{s})$.

Key Example: Additive Confounding Model

Consider the additive confounding model (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Gilbert et al., 2023):

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{with} \quad \mathbb{E}(\epsilon) = 0 \quad \text{and} \quad \mathrm{Var}(\epsilon) > 0. \tag{2}$$

- $\bar{m} : \mathcal{T} \to \mathbb{R}, \eta : \mathcal{S} \to \mathbb{R}$ are unknown functions, while $\epsilon \in \mathbb{R}$ is exogenous.
- $m(t) = \mathbb{E}\left[Y(t)\right] = \bar{m}(t) + \mathbb{E}\left[\eta(S)\right]$ and $\theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \bar{m}'(t)$.

Consider the additive confounding model (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Gilbert et al., 2023):

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{with} \quad \mathbb{E}(\epsilon) = 0 \quad \text{and} \quad \mathrm{Var}(\epsilon) > 0. \tag{2}$$

- $\bar{m} : \mathcal{T} \to \mathbb{R}, \eta : \mathcal{S} \to \mathbb{R}$ are unknown functions, while $\epsilon \in \mathbb{R}$ is exogenous.
- $m(t) = \mathbb{E}\left[Y(t)\right] = \bar{m}(t) + \mathbb{E}\left[\eta(S)\right]$ and $\theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \bar{m}'(t)$.

▶ **Identification of $\theta(t)$:** Under model (2) and consistency, we have

$$\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,S)\Big| T = t\right] := \theta_C(t) \quad \text{and} \quad \mathbb{E}(Y) = \mathbb{E}\left[m(T)\right].$$

Key Example: Additive Confounding Model

Consider the additive confounding model (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Gilbert et al., 2023):

$$Y(t) = \bar{m}(t) + \eta(S) + \epsilon \quad \text{with} \quad \mathbb{E}(\epsilon) = 0 \quad \text{and} \quad \text{Var}(\epsilon) > 0. \tag{2}$$

- $\bar{m} : \mathcal{T} \to \mathbb{R}, \eta : \mathcal{S} \to \mathbb{R}$ are unknown functions, while $\epsilon \in \mathbb{R}$ is exogenous.
- $m(t) = \mathbb{E}\left[Y(t)\right] = \bar{m}(t) + \mathbb{E}\left[\eta(S)\right]$ and $\theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \bar{m}'(t)$.

▶ **Identification of $\theta(t)$:** Under model (2) and consistency, we have

$$\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, S)\Big| T = t\right] := \theta_C(t) \quad \text{and} \quad \mathbb{E}(Y) = \mathbb{E}\left[m(T)\right].$$

▶ **Identification of $m(t)$:** By the fundamental theorem of calculus,

$$m(t) = \mathbb{E}\left[Y + \int_{u=T}^{u=t} \theta_C(u)\, du\right] = \mathbb{E}(Y) + \mathbb{E}\left\{\int_{u=T}^{u=t} \mathbb{E}\left[\frac{\partial}{\partial t}\mu(T, S)\Big| T = u\right] du\right\} \text{ for any } t \in \mathcal{T}.$$

▶ **Drawback of (2):** The treatment effect is homogeneous for any $S = s \in \mathcal{S}$.

Estimation of $m(t)$ and $\theta(t)$ Without Positivity

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

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

- **RA (Integral) Estimator Without Positivity**:

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

Estimation of $m(t)$ and $\theta(t)$ Without Positivity

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

- **RA (Integral) Estimator Without Positivity**:

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

- $\beta(t, s) = \frac{\partial}{\partial t}\mu(t, s) \overset{\text{fitted by}}{\longleftarrow}$ (partial) local polynomial regression (Fan and Gijbels, 1996) or neural networks (Paszke et al., 2017; Blondel and Roulet, 2024).

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

- **RA (Integral) Estimator Without Positivity**:

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

- $\beta(t,s) = \frac{\partial}{\partial t}\mu(t,s) \overset{\text{fitted by}}{\longleftarrow}$ (partial) local polynomial regression (Fan and Gijbels, 1996) or neural networks (Paszke et al., 2017; Blondel and Roulet, 2024).
- $\mathrm{F}_{S|T}(s|t) \overset{\text{fitted by}}{\longleftarrow}$ Nadaraya-Watson conditional CDF estimator (Hall et al., 1999).

Estimation of $m(t)$ and $\theta(t)$ Without Positivity

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

▶ **RA (Integral) Estimator Without Positivity**:

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

- $\beta(t, s) = \frac{\partial}{\partial t}\mu(t, s) \overset{\text{fitted by}}{\longleftarrow}$ (partial) local polynomial regression (Fan and Gijbels, 1996) or neural networks (Paszke et al., 2017; Blondel and Roulet, 2024).
- $\mathrm{F}_{S|T}(s|t) \overset{\text{fitted by}}{\longleftarrow}$ Nadaraya-Watson conditional CDF estimator (Hall et al., 1999).
- Compute the integral via a fast Riemann sum approximation (Zhang et al., 2024).

Estimation of $m(t)$ and $\theta(t)$ Without Positivity

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

- **RA (Integral) Estimator Without Positivity**:

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

- $\beta(t,s) = \frac{\partial}{\partial t}\mu(t,s) \overset{\text{fitted by}}{\longleftarrow}$ (partial) local polynomial regression (Fan and Gijbels, 1996) or neural networks (Paszke et al., 2017; Blondel and Roulet, 2024).
- $\mathrm{F}_{S|T}(s|t) \overset{\text{fitted by}}{\longleftarrow}$ Nadaraya-Watson conditional CDF estimator (Hall et al., 1999).
- Compute the integral via a fast Riemann sum approximation (Zhang et al., 2024).
- Establish the consistency of nonparametric bootstrap for $\widehat{m}_{\mathrm{C,RA}}(t)$ and $\widehat{\theta}_{\mathrm{C,RA}}(t)$.

Estimation Biases of IPW Estimators Without Positivity

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

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

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

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

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

Estimation Biases of IPW Estimators Without Positivity

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

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(S) + \epsilon$.
- Recall the standard (oracle) IPW estimators of $m(t)$ 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)

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

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

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

- For identification, we assume $Y(t) = \bar{m}(t) + \eta(S) + \epsilon$.
- Recall the standard (oracle) IPW estimators of $m(t)$ 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)

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

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

► **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!!

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

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

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

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

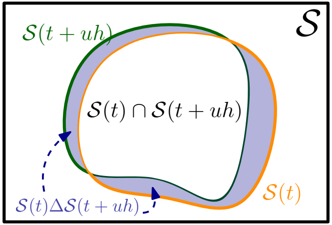
$$\lim_{h\to 0} \mathbb{E}\left[\widetilde{\theta}_{\mathrm{IPW}}(t)\right] = \lim_{h\to 0} \mathbb{E}\left[\frac{Y\left(\frac{T-t}{h}\right) K\left(\frac{T-t}{h}\right)}{h^2 \cdot \kappa_2 \cdot p_{T|S}(T|S)}\right] = \begin{cases} \bar{m}'(t)\cdot \rho(t) \\ \infty \end{cases} \neq \theta(t),$$

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:

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

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



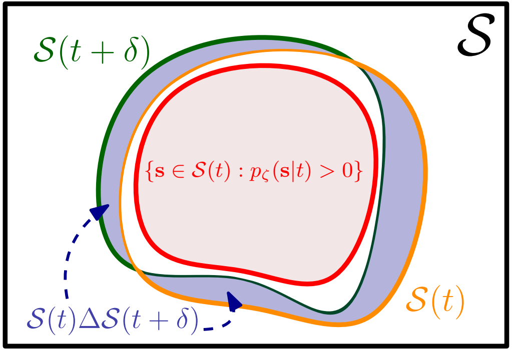
Bias-Corrected IPW Estimator for $\theta(t)$

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

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

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

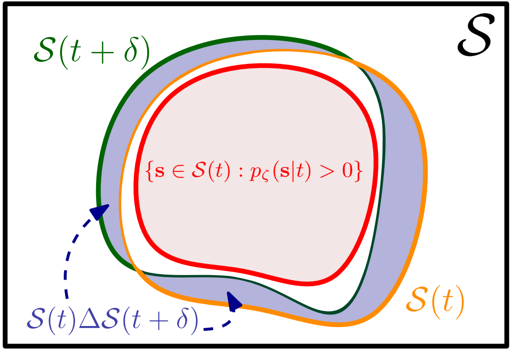
$$\{s\in\mathcal{S}(t) : p_\zeta(s|t)>0\}\subset\mathcal{S}(t+\delta) \quad \text{for any} \quad \delta\in[-h,h].$$



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

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

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



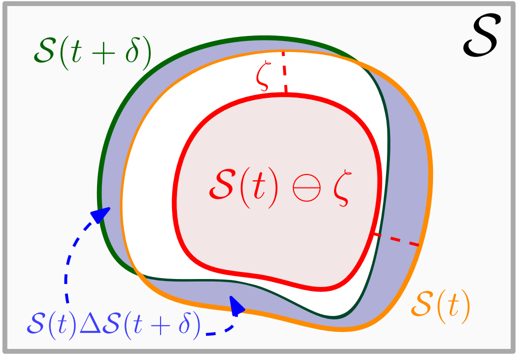
Now, we have that

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

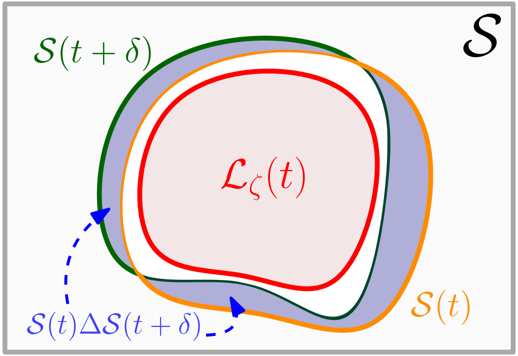
Question: How can we find a ζ-interior conditional density $p_\zeta(\boldsymbol{s}|t)$?

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

Support shrinking approach



Level set approach



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

▶ **Remark:** Practically, the RA estimators $\widehat{\theta}_{\mathrm{C,RA}}(t)$ and $\widehat{m}_{\mathrm{C,RA}}(t)$ are recommended!

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

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

Application: $PM_{2.5}$ on CMR

This part is based on **Section 5.3** in [2]:

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

All the code and data are available at

https://github.com/zhangyk8/npDoseResponse/tree/main.

Python Package: npDoseResponse and R Package: npDoseResponse.

$PM_{2.5}$ and CMRs Data Recap

FIPS	County name	Longitude	Latitude	PM2.5	CMR
1025	Clarke	-87.830772	31.676955	6.766443	379.421713
1061	Geneva	-85.839330	31.094869	8.254272	378.524698
1073	Jefferson	-86.896571	33.554343	10.825441	352.790427
1077	Lauderdale	-87.654117	34.901500	9.208783	332.594557
5085	Lonoke	-91.887917	34.754412	8.213144	365.061085
8045	Garfield	-107.903621	39.599420	2.601772	250.781477

1. The dataset (Wyatt et al., 2020a,b) contains the average annual CMRs (Y) and $PM_{2.5}$ levels (T) across $n = 2132$ U.S. counties over 1990-2010.

$PM_{2.5}$ and CMRs Data Recap

FIPS	County name	Longitude	Latitude	PM2.5	CMR
1025	Clarke	-87.830772	31.676955	6.766443	379.421713
1061	Geneva	-85.839330	31.094869	8.254272	378.524698
1073	Jefferson	-86.896571	33.554343	10.825441	352.790427
1077	Lauderdale	-87.654117	34.901500	9.208783	332.594557
5085	Lonoke	-91.887917	34.754412	8.213144	365.061085
8045	Garfield	-107.903621	39.599420	2.601772	250.781477

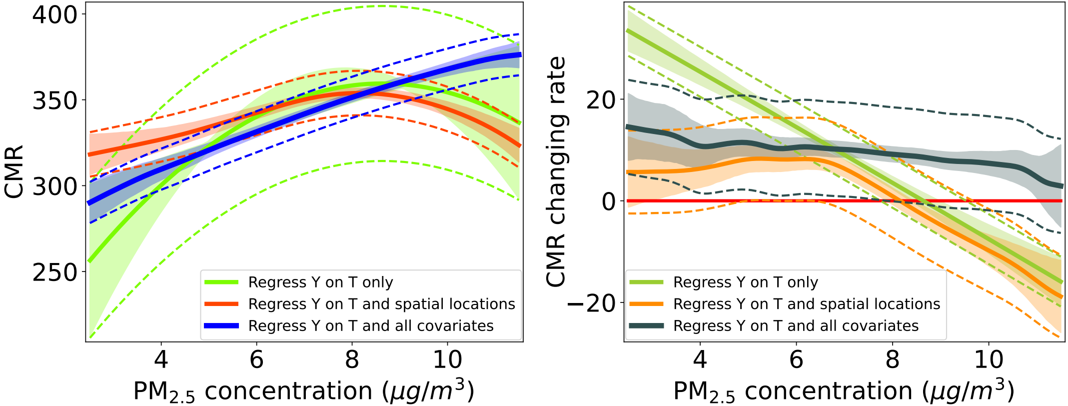
1. The dataset (Wyatt et al., 2020a,b) contains the average annual CMRs (Y) and $PM_{2.5}$ levels (T) across $n = 2132$ U.S. counties over 1990-2010.
2. The covariate vector $\boldsymbol{S} \in \mathbb{R}^{10}$ consists of two parts:
   - 2 spatial confounders: latitude and longitude of each county.
   - 8 county-level socioeconomic factors acquired from the US census.

$PM_{2.5}$ and CMRs Data Recap

FIPS	County name	Longitude	Latitude	PM2.5	CMR
1025	Clarke	-87.830772	31.676955	6.766443	379.421713
1061	Geneva	-85.839330	31.094869	8.254272	378.524698
1073	Jefferson	-86.896571	33.554343	10.825441	352.790427
1077	Lauderdale	-87.654117	34.901500	9.208783	332.594557
5085	Lonoke	-91.887917	34.754412	8.213144	365.061085
8045	Garfield	-107.903621	39.599420	2.601772	250.781477

1. The dataset (Wyatt et al., 2020a,b) contains the average annual CMRs (Y) and $PM_{2.5}$ levels (T) across $n = 2132$ U.S. counties over 1990-2010.
2. The covariate vector $\boldsymbol{S} \in \mathbb{R}^{10}$ consists of two parts:
   - 2 spatial confounders: latitude and longitude of each county.
   - 8 county-level socioeconomic factors acquired from the US census.
3. Focus on the values of $PM_{2.5}$ between 2.5 $\mu g/m^3$ and 11.5 $\mu g/m^3$ to avoid boundary effects (Takatsu and Westling, 2024).

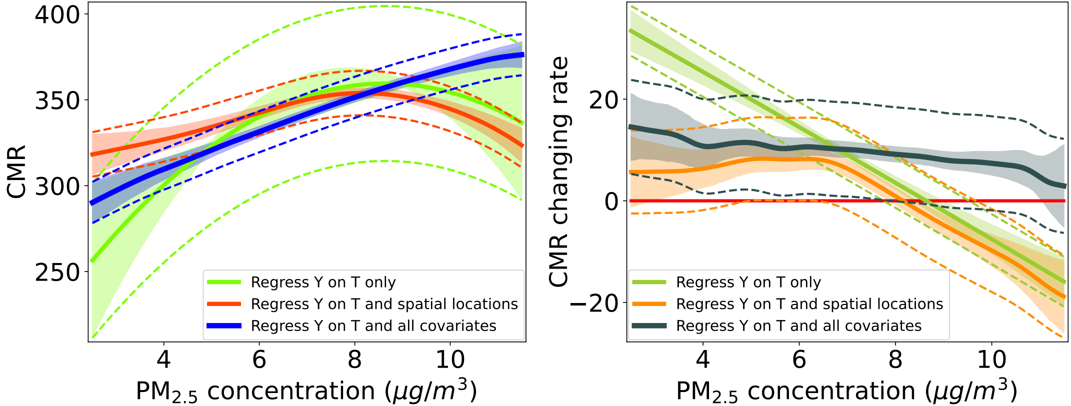
Effect of $PM_{2.5}$ on the Cardiovascular Mortality Rate (CMR)



Shaded areas: 95% pointwise confidence intervals; **Regions between dashed lines:** 95% uniform confidence bands.

- We compare three models:
  1. Regress Y on T alone via local quadratic regression.
  2. Regress Y on T with spatial locations.
  3. Regress Y on T with both spatial and socioeconomic covariates.

Effect of $PM_{2.5}$ on the Cardiovascular Mortality Rate (CMR)



Shaded areas: 95% pointwise confidence intervals; **Regions between dashed lines:** 95% uniform confidence bands.

- We compare three models:
  1. Regress Y on T alone via local quadratic regression.
  2. Regress Y on T with spatial locations.
  3. Regress Y on T with both spatial and socioeconomic covariates.
- For model 3, the increasing trends are **significant** when $PM_{2.5} < 8\,\mu g/m^3$.

Discussion

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

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

1. **Under the positivity condition:**
   - We propose $\widehat{\theta}_{\mathrm{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

Punchlines of Today's Talk

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

1. **Under the positivity condition:**
   - We propose $\widehat{\theta}_{\mathrm{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

2. **Without the positivity condition:**
   - Our key technique relies on two pillars in calculus:

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

Punchlines of Today's Talk

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

1. **Under the positivity condition:**
   - We propose $\widehat{\theta}_{\mathrm{DR}}(t)$ with standard nonparametric consistency and efficiency guarantee:

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

2. **Without the positivity condition:**
   - Our key technique relies on two pillars in calculus:

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

   - Our bias-corrected IPW and DR estimators reveal interesting connections to nonparametric level set estimation problems (Bonvini et al., 2023):

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

1. **Debiasing Doubly Robust Estimators:** Can we debias our DR estimators $\widehat{\theta}_{\mathrm{DR}}(t)$ and $\widehat{\theta}_{\mathrm{C,DR}}(t)$ through explicit bias estimation (Calonico et al., 2018; Cheng and Chen, 2019; Takatsu and Westling, 2024) or calibration (van der Laan et al., 2024)?

Ongoing Works and Future Directions

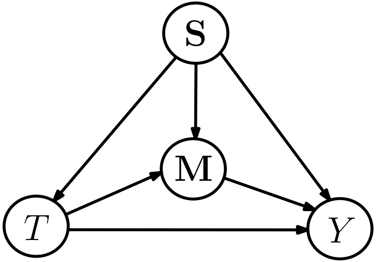
1. **Debiasing Doubly Robust Estimators:** Can we debias our DR estimators $\widehat{\theta}_{\mathrm{DR}}(t)$ and $\widehat{\theta}_{\mathrm{C,DR}}(t)$ through explicit bias estimation (Calonico et al., 2018; Cheng and Chen, 2019; Takatsu and Westling, 2024) or calibration (van der Laan et al., 2024)?

2. **Violation of Ignorability:** Can we conduct sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022a)?

1. **Debiasing Doubly Robust Estimators:** Can we debias our DR estimators $\widehat{\theta}_{\mathrm{DR}}(t)$ and $\widehat{\theta}_{\mathrm{C,DR}}(t)$ through explicit bias estimation (Calonico et al., 2018; Cheng and Chen, 2019; Takatsu and Westling, 2024) or calibration (van der Laan et al., 2024)?

2. **Violation of Ignorability:** Can we conduct sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022a)?

3. **Mediation Analysis:** Can we generalize our strategies for the estimation of direct and indirect causal effects (Huber et al., 2020; Xu et al., 2021)?



Thank you!

More details can be found in

[1] Y. Zhang, Y.-C. Chen, and A. Giessing. Nonparametric Inference on Dose-Response Curves Without the Positivity Condition. *arXiv preprint*, 2024. https://arxiv.org/abs/2405.09003.
[2] Y. Zhang and Y.-C. Chen. Doubly Robust Inference on Causal Derivative Effects for Continuous Treatments. *arXiv preprint*, 2025. https://arxiv.org/abs/2501.06969.

All the code and data are available at
https://github.com/zhangyk8/npDoseResponse/tree/main.

Python Package: npDoseResponse and R Package: npDoseResponse.

Check out my other works at https://zhangyk8.github.io/.

D. M. Bashtannyk and R. J. Hyndman. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298, 2001.

S. Basith, B. Manavalan, T. H. Shin, C. B. Park, W.-S. Lee, J. Kim, and G. Lee. The impact of fine particulate matter 2.5 on the cardiovascular system: a review of the invisible killer. *Nanomaterials*, 12(15):2656, 2022.

M. Bia and A. Mattei. A stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, 8(3):354–373, 2008.

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

M. Blondel and V. Roulet. The elements of differentiable programming. *arXiv preprint arXiv:2403.14606*, 2024.

M. Bonvini and E. H. Kennedy. Fast convergence rates for dose-response estimation. *arXiv preprint arXiv:2207.11825*, 2022.

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.

J. E. Chacón, T. Duong, and M. Wand. Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica*, pages 807–840, 2011.

Y.-C. Chen, C. R. Genovese, and L. Wasserman. A comprehensive approach to mode clustering. *Electronic Journal of Statistics*, 10(1):210 – 241, 2016.

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, and K. Kato. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597, 2014.

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, 2022a.

V. Chernozhukov, W. K. Newey, and R. Singh. Automatic debiased machine learning of causal and structural effects. *Econometrica*, 90(3):967–1027, 2022b.

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

I. Díaz-Muñoz and M. van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2012.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66. Chapman & Hall/CRC, 1996.

J. Fan, W. Härdle, and E. Mammen. Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26(3):943–971, 1998.

Q. Fan, Y.-C. Hsu, R. P. Lieli, and Y. Zhang. Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics*, 40(1):313–327, 2022.

T. Gasser and H.-G. Müller. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, pages 171–185, 1984.

B. Gilbert, A. Datta, J. A. Casey, and E. L. Ogburn. A causal inference framework for spatial confounding. *arXiv preprint arXiv:2112.14946*, 2023.

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

P. Hall, R. C. Wolff, and Q. Yao. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445):154–163, 1999.

S. Haneuse and A. Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30):5260–5277, 2013.

W. Härdle and T. M. Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995, 1989.

O. Hines, K. Diaz-Ordaz, and S. Vansteelandt. Optimally weighted average derivative effects. *arXiv preprint arXiv:2308.05456*, 2023.

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.

E. H. Kennedy, Z. Ma, M. D. McHugh, and D. S. Small. Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1229–1245, 2017.

Q. Li and J. Racine. Cross-validated local linear nonparametric regression. *Statistica Sinica*, pages 485–512, 2004.

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

A. Luedtke and I. Chung. One-step estimation of differentiable hilbert-valued parameters. *The Annals of Statistics*, 52 (4):1534–1563, 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.

M. R. Miller and D. E. Newby. Air pollution and cardiovascular disease: car sick. *Cardiovascular Research*, 116(2): 279–294, 2020.

R. Neugebauer and M. van der Laan. Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434, 2007.

Reference

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.

C. J. Paciorek. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1):107–125, 2010.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS 2017 Workshop on Autodiff*, 2017.

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.

J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

D. Rothenhäusler and B. Yu. Incremental causal effects. *arXiv preprint arXiv:1907.13258*, 2019.

K. Schindl, S. Shen, and E. H. Kennedy. Incremental effects for continuous exposures. *arXiv preprint arXiv:2409.11967*, 2024.

P. Schnell and G. Papadogeorgou. Mitigating unobserved spatial confounding when estimating the effect of supermarket access on cardiovascular disease deaths. *Annals of Applied Statistics*, 14:2069–2095, 12 2020.

Reference

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.

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 and J. M. Robins. *Unified methods for censored longitudinal data and causality*. Springer, 2003.

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.

T. Westling and M. Carone. A unified study of nonparametric inference for monotone functions. *Annals of Statistics*, 48(2):1001, 2020.

T. Westling, P. Gilbert, and M. Carone. Causal isotonic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):719–747, 2020.

Reference

L. H. Wyatt, G. C. Peterson, T. J. Wade, L. M. Neas, and A. G. Rappold. The contribution of improved air quality to reduced cardiovascular mortality: Declines in socioeconomic differences over time. *Environment international*, 136:105430, 2020a.

L. H. Wyatt, G. C. L. Peterson, T. J. Wade, L. M. Neas, and A. G. Rappold. Annual pm2.5 and cardiovascular mortality rate data: Trends modified by county socioeconomic status in 2,132 us counties. *Data in Brief*, 30:105318, 2020b.

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

L. Yang and R. Tschernig. Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(4):793–815, 1999.

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.

1. **Order q (Partial) Local Polynomial Regression** (Fan and Gijbels, 1996): Let $\widehat{\boldsymbol{\beta}}(t, \boldsymbol{s}) \in \mathbb{R}^{q+1}$ and $\widehat{\boldsymbol{\alpha}}(t, \boldsymbol{s}) \in \mathbb{R}^d$ be the minimizer of

$$\underset{(\boldsymbol{\beta},\boldsymbol{\alpha})^T \in \mathbb{R}^{q+1+d}}{\arg\min} \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{q} \beta_j (T_i - t)^q - \sum_{\ell=1}^{d} \alpha_\ell (S_{i,\ell} - s_\ell) \right]^2 K_T\left(\frac{T_i - t}{h}\right) K_S\left(\frac{\boldsymbol{S}_i - \boldsymbol{s}}{b}\right).$$

   - $K_T : \mathbb{R} \to [0, \infty)$, $K_S : \mathbb{R}^d \to [0, \infty)$ are two symmetric kernel functions, and $h, b > 0$ are smoothing bandwidth parameters.
   - The second component $\widehat{\beta}_2(t, \boldsymbol{s}) := \widehat{\beta}(t, \boldsymbol{s})$ is a consistent estimator of $\beta(t, \boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$.

2. **Nadaraya-Watson Conditional CDF Estimator for $F_{S|T}(\boldsymbol{s}|t)$** (Hall et al., 1999):

$$\widehat{F}_{S|T}(\boldsymbol{s}|t) = \widehat{P}_{\hbar}(\boldsymbol{s}|t) = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\boldsymbol{S}_i \leq \boldsymbol{s}\}} \cdot \bar{K}_T\left(\frac{T_i - t}{\hbar}\right)}{\sum_{j=1}^{n} \bar{K}_T\left(\frac{T_j - t}{\hbar}\right)}.$$

   - $\bar{K}_T : \mathbb{R} \to [0, \infty)$ is a kernel function and $\hbar > 0$ is its smoothing bandwidth parameter.

3 **Kernel Smoothing Methods for Estimating $p_{T|S}(t|s)$:**

- *Method 1 (Kernel density estimation on residuals):* If $T = g_S(S) + g_E(E)$ with $\mathbb{E}\left[g_E(E)|S\right] = 0$, then we can estimate $g_S(s) = \mathbb{E}\left(T|S = s\right)$ via any machine learning method to obtain

$$\widehat{p}_{T|S}(t|s) = \frac{1}{nh_e}\sum_{i=1}^{n} K_e\left[\frac{t - \widehat{g}_S(s) - (T_i - \widehat{g}_S(S_i))}{h_e}\right],$$

where $K_e : \mathbb{R} \to [0, \infty)$ is a kernel function and $h_e > 0$ is a bandwidth parameter.

- *Method 2 (Regression on kernel-smoothed outcomes (RKS) in Chernozhukov et al. 2022b):* Let $g(t, s) = \mathbb{E}\left[K_r\left(\frac{T-t}{h_r}\right)\Big|S = s\right]$. We obtain $\widehat{p}_{T|S}(t|s) = \widehat{g}(t, s)$ by regressing $\left\{K_r\left(\frac{T_i-t}{h_r}\right)\right\}_{i=1}^{n}$ against $\{S_i\}_{i=1}^{n}$ via any machine learning method. Note that

$$g(t, s) = \mathbb{E}\left[K_r\left(\frac{T-t}{h_r}\right)\Big|S = s\right] \to p_{T|S}(t|s) \quad \text{as} \quad h_r \to 0,$$

where $K_r : \mathbb{R} \to [0, \infty)$ is a kernel function and $h_r > 0$ is a bandwidth parameter.

Localized RA Derivative Estimator for $\theta(t)$ Without Positivity

Combining two nuisance function estimators $\widehat{\beta}(t, \boldsymbol{s})$ and $\widehat{F}_{\boldsymbol{S}|T}(\boldsymbol{s}|t)$, we derive our **localized RA derivative estimator** of $\theta(t)$ with kernel smoothing as:

$$\widehat{\theta}_{\mathrm{C,RA}}(t) = \int \widehat{\beta}(t, \boldsymbol{s})\, d\widehat{P}_{\hbar}(\boldsymbol{s}|t) = \frac{\sum_{i=1}^{n} \widehat{\beta}(t, \boldsymbol{S}_i) \cdot \bar{K}_T\left(\frac{T_i - t}{\hbar}\right)}{\sum_{j=1}^{n} \bar{K}_T\left(\frac{T_j - t}{\hbar}\right)}.$$

Our **RA integral estimator** takes the form

$$\widehat{m}_{\mathrm{C,RA}}(t) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_i + \int_{\tilde{t}=T_i}^{\tilde{t}=t} \widehat{\theta}_{\mathrm{C,RA}}(\tilde{t})\, d\tilde{t} \right].$$

- $\widehat{m}_{\mathrm{C,RA}}(t)$, under our kernel smoothing-based estimators, is a *linear smoother*.
- We can also fit $\mu(t, \boldsymbol{s})$ via neural networks and obtain an estimator for $\beta(t, \boldsymbol{s})$ via automatic differentiation in `PyTorch` (Paszke et al., 2017).

▶ **Issue:** The integral in $\widehat{m}_{\mathrm{C,RA}}(t)$ could be analytically difficult to compute.

Our integral estimator takes the form

$$\widehat{m}_{\mathrm{C,RA}}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i + \int_{\tilde{t}=T_i}^{\tilde{t}=t} \widehat{\theta}_{\mathrm{C,RA}}(\tilde{t})\, d\tilde{t}\right].$$

▶ **Riemann Sum Approximation:** Let $T_{(1)} \leq \cdots \leq T_{(n)}$ be the order statistics of $T_1, ..., T_n$ and $\Delta_j = T_{(j+1)} - T_{(j)}$ for $j = 1, ..., n-1$.

- Approximate $\widehat{m}_{\mathrm{C,RA}}(T_{(j)})$ for each $j = 1, ..., n$ as:

$$\widehat{m}_{\mathrm{C,RA}}(T_{(j)}) \approx \frac{1}{n}\sum_{i=1}^{n} Y_i + \frac{1}{n}\sum_{i=1}^{n-1}\Delta_i\left[i \cdot \widehat{\theta}_{\mathrm{C,RA}}(T_{(i)})\mathbb{1}_{\{i<j\}} - (n-i)\cdot \widehat{\theta}_{\mathrm{C,RA}}(T_{(i+1)})\mathbb{1}_{\{i\geq j\}}\right].$$

- Evaluate $\widehat{m}_{\mathrm{C,RA}}(t)$ at any $t \in \left[T_{(j)}, T_{(j+1)}\right]$ by a linear interpolation between $\widehat{m}_{\mathrm{C,RA}}(T_{(j)})$ and $\widehat{m}_{\mathrm{C,RA}}(T_{(j+1)})$.
- The approximation error $O_P\left(\frac{1}{n}\right)$ is asymptotically negligible.

1. Compute $\widehat{m}_{C,RA}(t)$ on the original data $\{(Y_i, T_i, S_i)\}_{i=1}^n$.
2. Generate B bootstrap samples $\left\{\left(Y_i^{*(b)}, T_i^{*(b)}, S_i^{*(b)}\right)\right\}_{i=1}^n$ by sampling with replacement and compute $\widehat{m}_{C,RA}^{*(b)}(t)$ for each $b = 1, \ldots, B$.
3. Let $\alpha \in (0,1)$ be a pre-specified significance level.
   - For pointwise inference at $t_0 \in \mathcal{T}$, calculate the $1-\alpha$ quantile $\zeta_{1-\alpha}^*(t_0)$ of $\{D_1(t_0), \ldots, D_B(t_0)\}$, where $D_b(t_0) = \left|\widehat{m}_{C,RA}^{*(b)}(t_0) - \widehat{m}_{C,RA}(t_0)\right|$ for $b = 1, \ldots, B$.
   - For uniform inference on $m(t)$, compute the $1-\alpha$ quantile $\xi_{1-\alpha}^*$ of $\{D_{\sup,1}, \ldots, D_{\sup,B}\}$, where $D_{\sup,b} = \sup_{t\in\mathcal{T}} \left|\widehat{m}_{C,RA}^{*(b)}(t) - \widehat{m}_{C,RA}(t)\right|$ for $b = 1, \ldots, B$.
4. Define the $1-\alpha$ confidence interval for $m(t_0)$ as:

$$\left[\widehat{m}_{C,RA}(t_0) - \zeta_{1-\alpha}^*(t_0),\ \widehat{m}_{C,RA}(t_0) + \zeta_{1-\alpha}^*(t_0)\right]$$

and the simultaneous $1-\alpha$ confidence band for every $t \in \mathcal{T}$ as:

$$\left[\widehat{m}_{C,RA}(t) - \xi_{1-\alpha}^*,\ \widehat{m}_{C,RA}(t) + \xi_{1-\alpha}^*\right].$$

Multiplier Bootstrap With $\widehat{\theta}_{\mathrm{DR}}(t)$ for Uniform Inference

Let $\{Z_i\}_{i=1}^n$ be a sequence of i.i.d. random variables independent of the observed data $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ with $\mathbb{E}(Z_i) = \mathrm{Var}(Z_i) = 1$ and sub-exponential tails (Fan et al., 2022; Colangelo and Lee, 2020).

1. Sample B different i.i.d. datasets $\left\{Z_i^{(b)}\right\}_{i=1}^n, b = 1, ..., B$.
2. Compute the bootstrap DR estimators for $\theta(t)$ for $b = 1, ..., B$ as:

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

3. Let $\widehat{Q}(1-\tau)$ be the $(1-\tau)$ quantile of the sequence $\left\{\sup_{t\in\mathcal{T}} \sqrt{nh^3} \left|\frac{\widehat{\theta}_{\mathrm{DR}}^{(b)*}(t) - \widehat{\theta}_{\mathrm{DR}}(t)}{\sqrt{\widehat{V}_{\theta}(t)}}\right|\right\}_{b=1}^B$.
4. The uniform $(1-\tau)$-level confidence band of $\theta(t)$ is given by

$$\left[\widehat{\theta}_{\mathrm{DR}}(t) \pm \widehat{Q}(1-\tau)\sqrt{\frac{\widehat{V}_{\theta}(t)}{nh^3}}\right].$$

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

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

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

and

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

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

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

and

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

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* $\mathcal{T} \times \mathcal{S}$*.*
3. *There exist absolute constants* $\sigma, A_0 > 0$ *such that* $\mathrm{Var}(Y|T = t, \boldsymbol{S} = \boldsymbol{s}) = \sigma^2$ *and* $\mathbb{E}|Y|^4 < A_0 < \infty$ *uniformly in* $\mathcal{J}$*.*

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 $\mathcal{T}$.*

Assumption (Boundary conditions)

1. *There exists some constants* $r_1, r_2 \in (0,1)$ *such that for any* $(t, \mathbf{s}) \in \mathcal{J}$ *and all* $\delta \in (0, r_1]$, *there is a point* $(t', \mathbf{s}') \in \mathcal{J}$ *satisfying*

$$\mathcal{B}\left((t', \mathbf{s}'),\, r_2\delta\right) \subset \mathcal{B}\left((t, \mathbf{s}),\, \delta\right) \cap \mathcal{J},$$

where $\mathcal{B}((t, \mathbf{s}),\, r) = \left\{(t_1, \mathbf{s}_1) \in \mathbb{R}^{d+1} : ||(t_1 - t, \mathbf{s}_1 - \mathbf{s})||_2 \leq r\right\}$ *with* $||\cdot||_2$ *being the standard Euclidean norm.*

2. *For any* $(t, \mathbf{s}) \in \partial\mathcal{J}$, *the boundary of* $\mathcal{J}$, *it satisfies that* $\frac{\partial}{\partial t} p(t, \mathbf{s}) = \frac{\partial}{\partial s_j} p(t, \mathbf{s}) = 0$ *and* $\frac{\partial^2}{\partial s_j^2} \mu(t, \mathbf{s}) = 0$ *for all* $j = 1, ..., d$.

3. *For any* $\delta > 0$, *the Lebesgue measure of the set* $\partial\mathcal{J} \oplus \delta$ *satisfies* $|\partial\mathcal{J} \oplus \delta| \leq A_1 \cdot \delta$ *for some absolute constant* $A_1 > 0$, *where* $\partial\mathcal{J} \oplus \delta = \left\{\mathbf{z} \in \mathbb{R}^{d+1} : \inf_{\mathbf{x} \in \partial\mathcal{J}} ||\mathbf{z} - \mathbf{x}||_2 \leq \delta\right\}$.

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* $j = 1, 2, \ldots$.
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 of the following holds true:*

(i) " $\mathcal{S}(t) \ominus (A_0|\delta|) \subset \mathcal{S}(t+\delta)$ *" for the support shrinking approach;*

(ii) " $\mathcal{L}_{A_0|\delta|}(t) \subset \mathcal{S}(t+\delta)$ *" for the level set approach.*

- The support $\mathcal{J}$ of $(T, \mathbf{S})$ may not cover $\mathcal{T} \times \mathcal{S}$ without positivity.
- The localized derivative estimator $\widehat{\theta}_C(t) = \int \widehat{\beta}(t, \mathbf{s})\, d\widehat{\mathrm{P}}_{\hbar}(\mathbf{s}|t)$ only requires $\widehat{\beta}(t, \mathbf{s})$ to be consistent in $\mathcal{J}$.

Lemma (Lemma 3 in Zhang et al. 2024)

Under some regularity conditions, as $h, b, \frac{\max\{h,b\}^4}{h} \to 0$ *and* $\frac{|\log(hb^d)|}{nh^3b^d} \to \infty$,

$$\sup_{(t,\mathbf{s})\in\mathcal{J}} \left|\widehat{\beta}(t, \mathbf{s}) - \frac{\partial}{\partial t}\mu(t, \mathbf{s})\right| = O\left(h^q + b^2 + \frac{\max\{h, b\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right).$$

Combining with the consistency of $\widehat{\mathrm{P}}_{\hbar}(\boldsymbol{s}|t)$ via the technique in Fan et al. (1998), we have the following results.

Theorem (Theorem 4 in Zhang et al. 2024)

Under some regularity conditions, when $q = 2$ *and* $h, b, \hbar, \frac{\max\{h,b\}^4}{h} \to 0$ *and* $\frac{n \max\{h,\hbar\} b^d}{\log n}, \frac{n\hbar}{\log n} \to \infty$,

$$\sup_{t \in \mathcal{T}} \left|\widehat{\theta}_{\mathrm{C,RA}}(t) - \theta_{\mathrm{C}}(t)\right| = \underbrace{O\left(h^2 + b^2 + \frac{\max\{b,h\}^4}{h}\right)}_{\text{Bias term}} + \underbrace{O_P\left(\sqrt{\frac{\log n}{nh^3}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right)}_{\text{Stochastic variation}},$$

$$\sup_{t \in \mathcal{T}} \left|\widehat{m}_{\mathrm{C,RA}}(t) - m(t)\right| = O\left(h^2 + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right).$$

$$\sup_{t\in\mathcal{T}} \left|\widehat{m}_{\mathrm{C,RA}}(t) - m(t)\right| = O\left(h^2 + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right).$$

- Blue term: the estimation bias of local polynomial estimator $\widehat{\beta}(t,\boldsymbol{s}) = \widehat{\beta}_2(t,\boldsymbol{s})$.
- Orange term: additional bias of $\widehat{\beta}_2(t,\boldsymbol{s})$ at the boundary $\partial\mathcal{J}$.
- Teal term: asymptotic rate from $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$.
- Red term: stochastic variation of $\widehat{\beta}_2(t,\boldsymbol{s})$.
- Cyan term: asymptotic rate from the Nadaraya-Watson conditional CDF estimator $\widehat{\mathrm{P}}_{\hbar}(\boldsymbol{s}|t)$.

Lemma (Lemma 5 in Zhang et al. 2024)

Under the same regularity conditions, if $h \asymp n^{-\frac{1}{\gamma}}$ *and* $\hbar \asymp n^{-\frac{1}{\varpi}}$ *for some* $\gamma \geq \varpi > 0$ *such that* $\frac{nh^5}{\log n} \to c_1$ *and* $\frac{n\hbar^5}{\log n} \to c_2$ *for some* $c_1, c_2 \geq 0$ *and* $\frac{n \max\{h,\hbar\}b^d}{\log n}, \frac{n\hbar}{\log n}, \frac{h^3 \log n}{\hbar}, \frac{nh^3\hbar^4}{\log n} \to \infty$ *as* $n \to \infty$, *then for any* $t \in \mathcal{T}$,

$$\sqrt{nh^3}\left[\widehat{\theta}_{\mathrm{C,RA}}(t) - \theta(t)\right] = \mathbb{G}_n \bar{\varphi}_t + o_P(1) \quad \text{and} \quad \sqrt{nh^3}\left[\widehat{m}_{\mathrm{C,RA}}(t) - m(t)\right] = \mathbb{G}_n \varphi_t + o_P(1),$$

where

$$\bar{\varphi}_t(Y, T, \mathbf{S}) = \frac{C_{K_T}\left[Y - \mu(T, \mathbf{S})\right]}{\sqrt{h} \cdot p_T(t)} \left(\frac{T-t}{h}\right) K_T\left(\frac{T-t}{h}\right)$$

and $\varphi_t(Y, T, \mathbf{S}) = \mathbb{E}_{T_1}\left[\int_{T_1}^{t} \bar{\varphi}_{\tilde{t}}(Y, T, \mathbf{S})\, d\tilde{t}\right]$ *with* $\mathbb{G}_n = \sqrt{n}\left(\mathbb{P}_n - \mathrm{P}\right)$, *where* $C_{K_T} > 0$ *is a constant that only depends on* K_T.

▶ **Note:** $\bar{\varphi}_t$ and φ_t are the IPW components of the *approximated* efficient influence functions.

Theorem (Theorems 6 and 7 in Zhang et al. 2024)

Under the same regularity conditions, if $h \asymp n^{-\frac{1}{\gamma}}$ *and* $b \lesssim \hbar \asymp n^{-\frac{1}{\varpi}}$ *for some* $\gamma \geq \varpi > 0$ *such that* $\frac{nh^{d+5}}{\log n} \to c_1$ *and* $\frac{n\hbar^5}{\log n} \to c_2$ *for some* $c_1, c_2 \geq 0$ *and* $\frac{\hbar}{h^3 \log n}, \hbar n^{\frac{1}{3}} \log n, \frac{\sqrt{n\hbar}}{\log n}, \frac{n \max\{h,\hbar\} b^d}{\log n} \to \infty$ *as* $n \to \infty$*,*

1. $$\left|\sqrt{nh^3} \sup_{t\in\mathcal{T}} \left|\widehat{m}_{C,RA}(t) - m(t)\right| - \sup_{t\in\mathcal{T}} \left|\mathbb{G}_n \varphi_t\right|\right| = O_P\left(\sqrt{nh^3 \max\{h,\hbar\}^4} + \sqrt{\frac{h^3 \log n}{\hbar}} + \frac{\log n}{\sqrt{n\hbar}} + \sqrt{\frac{\log n}{nb^d \hbar}}\right).$$

2. *there exists a mean-zero Gaussian process* $\mathbb{B}$ *such that*

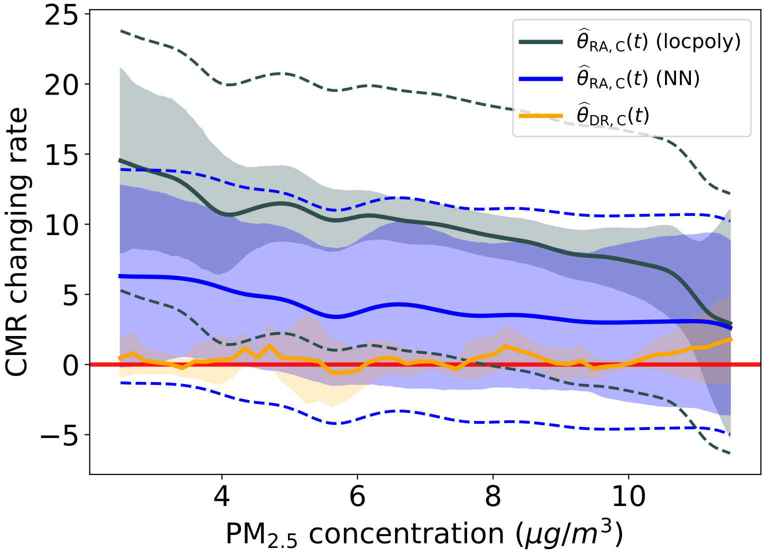
$$\sup_{u\geq 0}\left|P\left(\sqrt{nh^3} \sup_{t\in\mathcal{T}} \left|\widehat{m}_{C,RA}(t) - m(t)\right| \leq u\right) - P\left(\sup_{f\in\mathcal{F}} |\mathbb{B}(f)| \leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^3}\right)^{\frac{1}{8}} + \left(\frac{\log^2 n}{nb^d \hbar}\right)^{\frac{3}{8}}\right).$$

3. $$\sup_{u\geq 0}\left|P\left(\sqrt{nh^3} \sup_{t\in\mathcal{T}} \left|\widehat{m}^*_{C,RA}(t) - \widehat{m}_{C,RA}(t)\right| \leq u \,\middle|\, \mathbb{U}_n\right) - P\left(\sup_{f\in\mathcal{F}} |\mathbb{B}(f)| \leq u\right)\right| = O_P\left(\left(\frac{\log^5 n}{nh^3}\right)^{\frac{1}{8}} + \left(\frac{\log^2 n}{nb^d \hbar}\right)^{\frac{3}{8}}\right),$$

where $\mathcal{F} = \{(v, x, z) \mapsto \varphi_t(v, x, z) : t \in \mathcal{T}\}$.

1. $\mathcal{F}$ is not Donsker because φ_t is not uniformly bounded as $h \to 0$.
   - However, $\tilde{\mathcal{F}} = \left\{(v, x, z) \mapsto \sqrt{h^3} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}\right\}$ is of VC-type.
   - Gaussian approximation in Chernozhukov et al. (2014) can be applied to bound the difference between $\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|$ and $\sup_{f \in \mathcal{F}} |\mathbb{B}(f)|$.
2. As long as $\mathrm{Var}(Y|T = t, \boldsymbol{S} = \boldsymbol{s}) \geq \sigma^2 > 0$, $\mathrm{Var}\left[\varphi_t(Y, T, \boldsymbol{S})\right]$ is a positive finite number.
   - The asymptotic linearity (or V-statistic) is non-degenerate.
   - Pointwise bootstrap confidence intervals are asymptotically valid.
3. For the validity of uniform bootstrap confidence band, one can choose the bandwidths $h \asymp \hbar = O\left(n^{-\frac{1}{5}}\right)$ and $\left(\frac{\log n}{n}\right)^{\frac{4}{5d}} \lesssim b \lesssim n^{-\frac{1}{5}}$.
   - These orders align with the outputs from the usual bandwidth selection methods (Bashtannyk and Hyndman, 2001; Li and Racine, 2004).
   - No explicit undersmoothing is required!!

Additional Results for $PM_{2.5}$ on CMRs



Shaded areas: 95% pointwise confidence intervals; **Regions between dashed lines:** 95% uniform confidence bands.

Simulation Setup for Estimating $m(t)$ and $\theta(t)$ Without Positivity

- Use the Epanechnikov kernel for K_T and K_S (with the product kernel technique) and Gaussian kernel for $\bar{K}_T$.
- Select the bandwidth parameters $h, b > 0$ by modifying the rule-of-thumb method in Yang and Tschernig (1999).
- Set the bandwidth parameter $\hbar > 0$ to the normal reference rule in Chacón et al. (2011); Chen et al. (2016).
- Set the bootstrap resampling time $B = 1000$ and the nominal level for confidence intervals or bands to 95%.
- Compare our proposed estimators with the regression adjustment estimators under the same choices of bandwidth parameters:

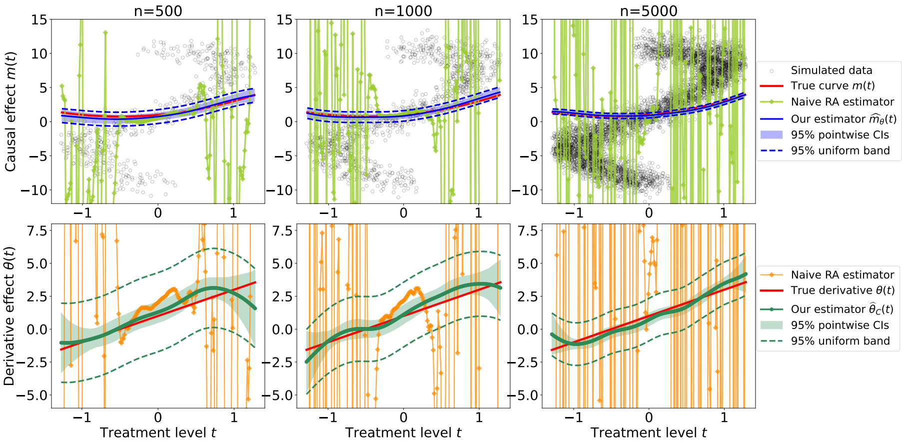
$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}(t, S_i) \quad \text{and} \quad \widehat{\theta}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\beta}(t, S_i).$$

Single Confounder Model Without Positivity

Generate i.i.d. observations $\{(Y_i, T_i, S_i)\}_{i=1}^{2000}$ from

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

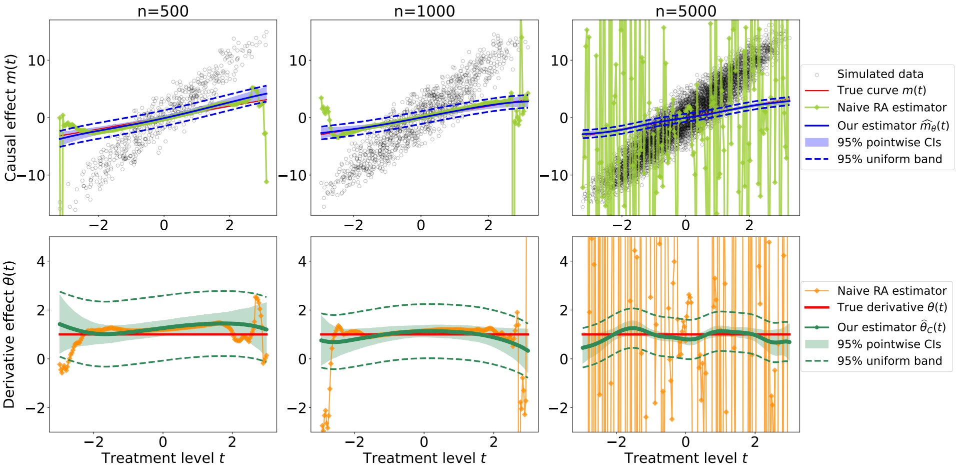
- $E \sim \text{Uniform}[-0.3, 0.3]$ is an independent treatment variation,
- $\epsilon \sim \mathcal{N}(0, 1)$ is an exogenous normal noise.



Generate i.i.d. observations $\{(Y_i, T_i, S_i)\}_{i=1}^{2000}$ from

$$Y = T + 6S_1 + 6S_2 + \epsilon, \quad T = 2S_1 + S_2 + E, \quad \text{and} \quad (S_1, S_2) \sim \text{Uniform}[-1,1]^2,$$

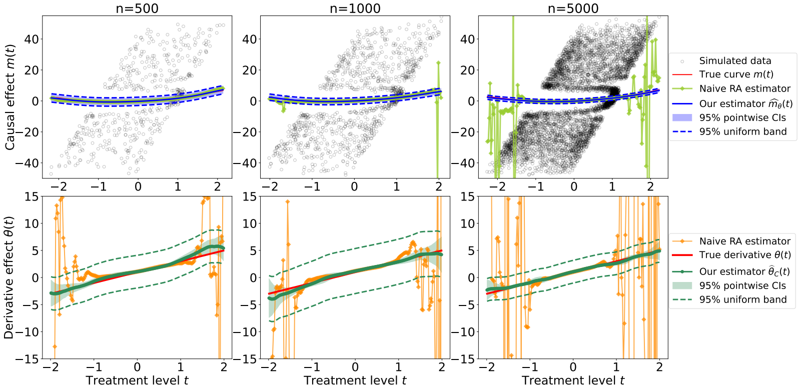
- $E \sim \text{Uniform}[-0.5, 0.5]$ and $\epsilon \sim \mathcal{N}(0,1)$.



Generate i.i.d. observations $\{(Y_i, T_i, S_i)\}_{i=1}^{2000}$ from

$$Y = T^2 + T + 10Z + \epsilon, \quad T = \cos\left(\pi Z^3\right) + \frac{Z}{4} + E, \quad \text{and} \quad Z = 4S_1 + S_2,$$

- $(S_1, S_2) \sim \text{Uniform}[-1, 1]^2$, $E \sim \text{Uniform}[-0.1, 0.1]$, and $\epsilon \sim \mathcal{N}(0, 1)$.
- Those doubly robust methods based on pseudo-outcomes (Kennedy et al., 2017; Takatsu and Westling, 2024) do not work in this example.



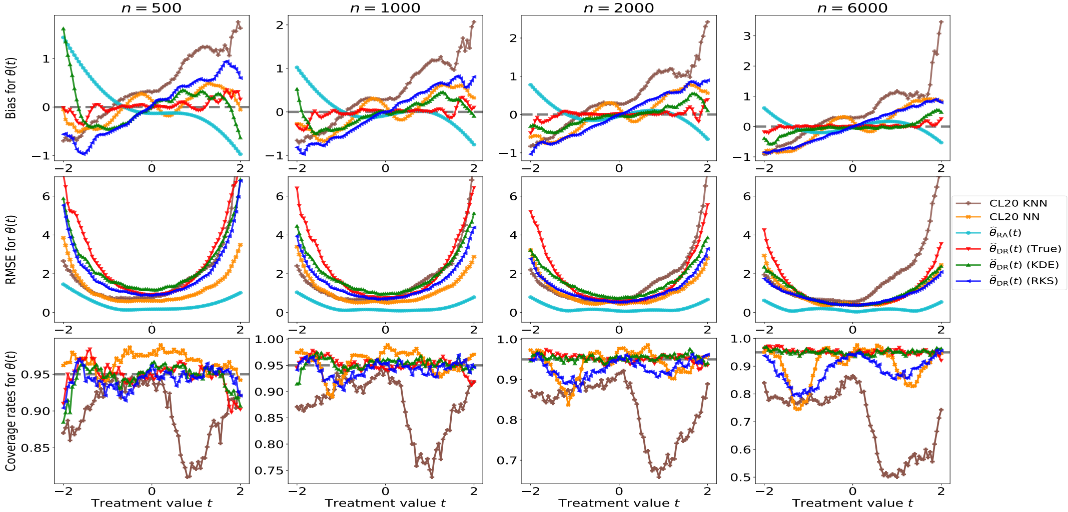
We generate i.i.d. observations $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$ from the following data-generating model (Colangelo and Lee, 2020):

$$Y = 1.2\,T + T^2 + TS_1 + 1.2\,\boldsymbol{\xi}^T\boldsymbol{S} + \epsilon\sqrt{0.5 + F_{\mathcal{N}(0,1)}(S_1)}, \quad \epsilon \sim \mathcal{N}(0,1),$$
$$T = F_{\mathcal{N}(0,1)}\left(3\boldsymbol{\xi}^T\boldsymbol{S}\right) - 0.5 + 0.75E, \quad \boldsymbol{S} = (S_1, ..., S_d)^T \sim \mathcal{N}_d\left(\boldsymbol{0}, \Sigma\right),\ E \sim \mathcal{N}(0,1),$$

where

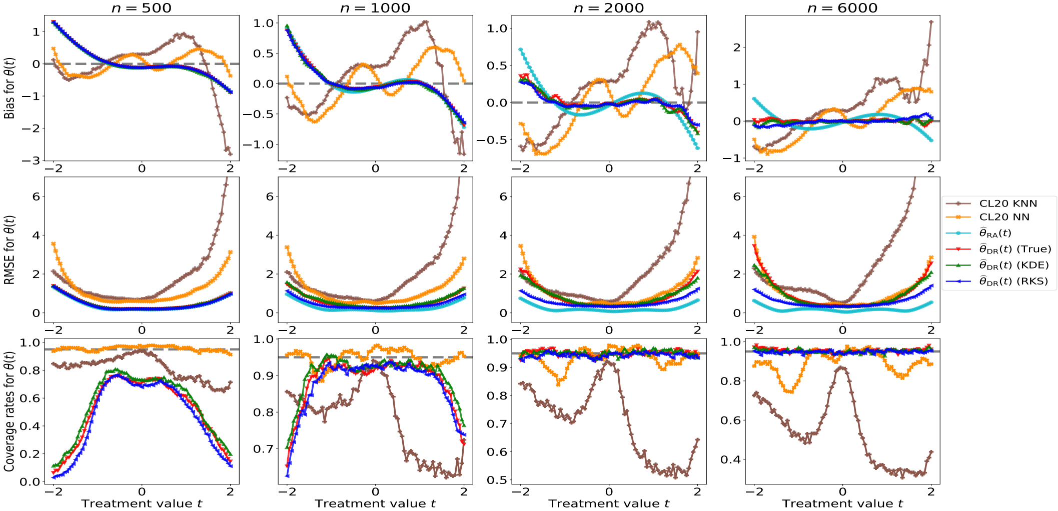
- $F_{\mathcal{N}(0,1)}$ is the CDF of $\mathcal{N}\left(0,1\right)$ and $d = 20$.
- $\boldsymbol{\xi} = (\xi_1, ..., \xi_d)^T \in \mathbb{R}^d$ has its entry $\xi_j = \frac{1}{j^2}$ for $j = 1, ..., d$ and $\Sigma_{ii} = 1$, $\Sigma_{ij} = 0.5$ when $|i-j| = 1$, and $\Sigma_{ij} = 0$ when $|i-j| > 1$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



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.

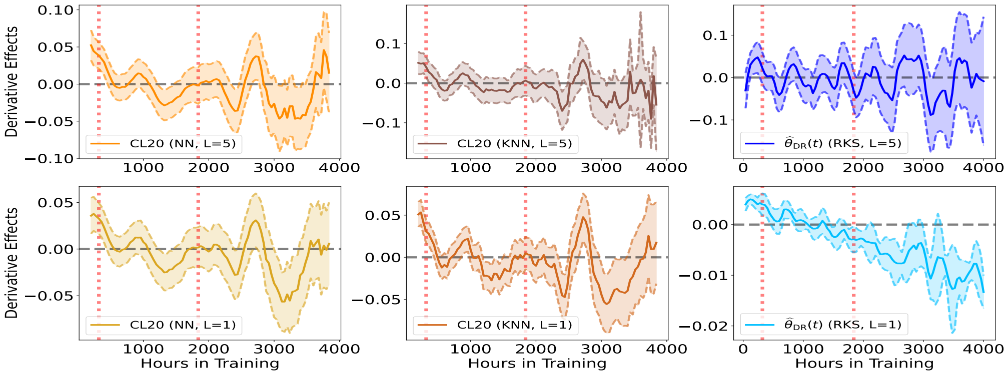
Simulations for Estimating $\theta(t)$ Under Positivity



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.

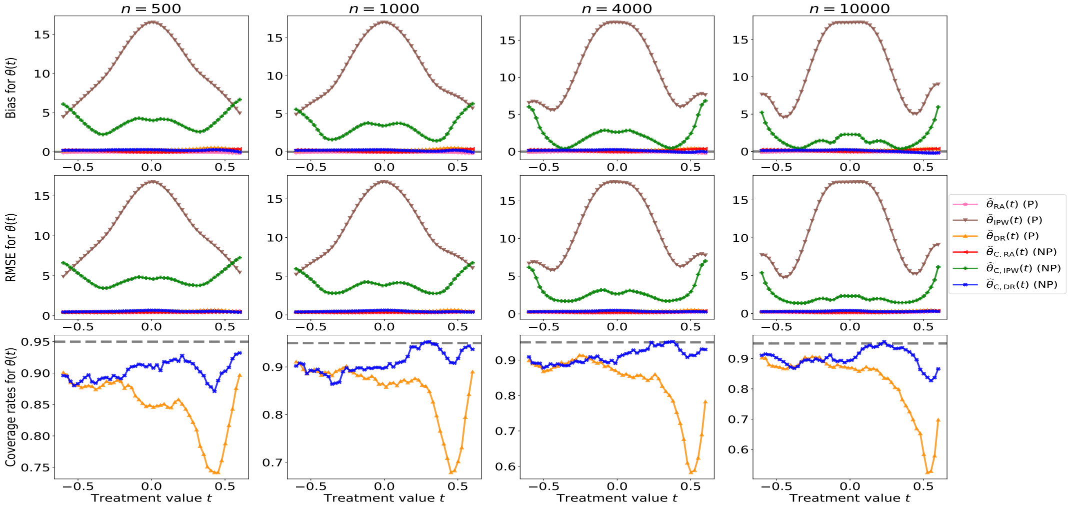
We compare our proposed DR estimator $\widehat{\theta}_{\mathrm{DR}}(t)$ under positivity with the finite-difference method (Colangelo and Lee 2020; CL20) on the U.S. Job Corps program (Schochet et al., 2001).

- 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 $n = 4024$.



Simulations for $\widehat{\theta}_{C,RA}(t)$, $\widehat{\theta}_{C,IPW}(t)$, $\widehat{\theta}_{C,DR}(t)$ 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 network (inspired by Luedtke 2024).