

Nonparametric Inference on Dose-Response Curves Without the Positivity Condition

Yikun Zhang

Joint work with *Professor Yen-Chi Chen*

*Department of Statistics,
University of Washington*

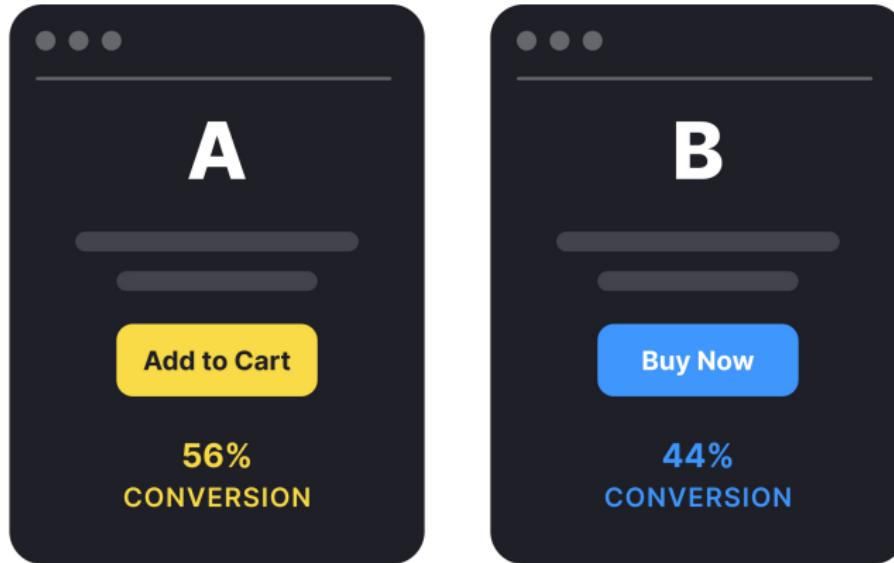
Causal Inference Working Group, Johns Hopkins University
March 11, 2025

Introduction



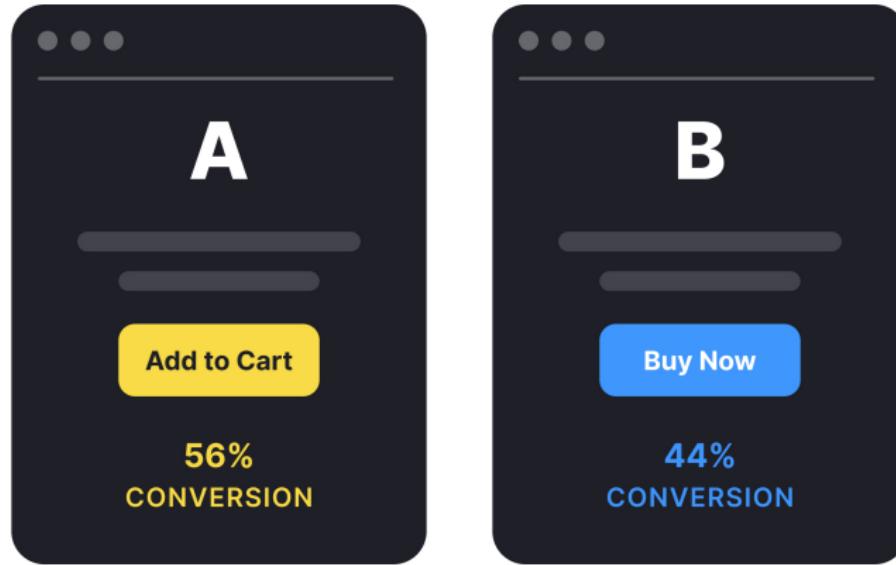
Fundamental Problem of Causal Inference

- ▶ 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

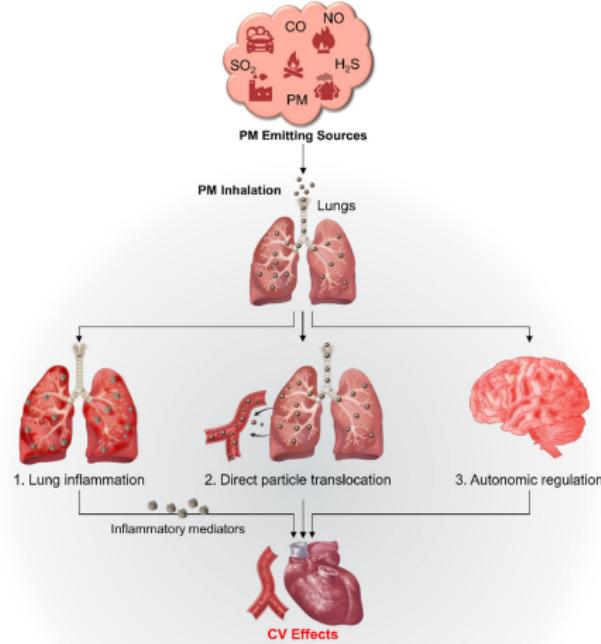
- ▶ Study the causal effect of a treatment $T \in \mathcal{T}$ on the outcome of interest $Y \in \mathcal{Y}$.



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

Motivation for Continuous Treatments

- 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](#)).

Motivation for Continuous Treatments

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](#)).

Motivation for Continuous Treatments

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!*

Causal Inference For Continuous Treatments

For *binary* treatment (*i.e.*, $\mathcal{T} = \{0, 1\}$), common causal estimands are

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

► **Question:** What are the counterparts of the above estimands under *continuous* treatment (*i.e.*, $\mathcal{T} \subset \mathbb{R}$)?

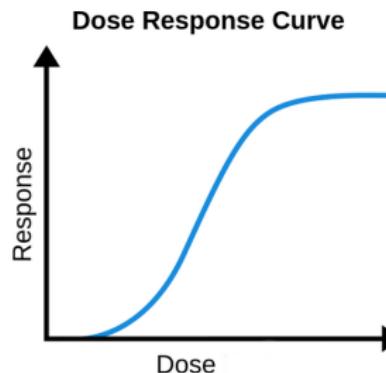
Causal Inference For Continuous Treatments

For *binary* treatment (*i.e.*, $\mathcal{T} = \{0, 1\}$), common causal estimands are

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

► **Question:** What are the counterparts of the above estimands under *continuous* treatment (*i.e.*, $\mathcal{T} \subset \mathbb{R}$)?

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



Identification of Dose-Response Curves in RCTs

In randomized controlled trials (RCTs),

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

Identification of Dose-Response Curves in RCTs

In randomized controlled trials (RCTs),

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

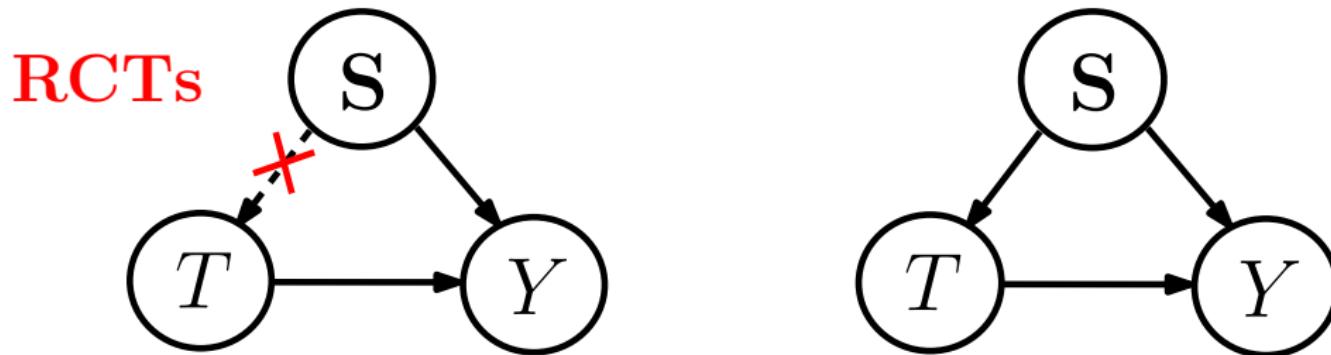
- We can estimate $m(t)$ by fitting a regression on $\{(Y_i, T_i)\}_{i=1}^n$.
- Recovering $\theta(t)$ is a derivative estimation problem ([Gasser and Müller, 1984](#)).

Identification of Dose-Response Curves in RCTs

In randomized controlled trials (RCTs),

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

- We can estimate $m(t)$ by fitting a regression on $\{(Y_i, T_i)\}_{i=1}^n$.
- Recovering $\theta(t)$ is a derivative estimation problem ([Gasser and Müller, 1984](#)).



- Some identification assumptions are required to estimate $m(t)$ and $\theta(t)$ from the observable data $\{(Y_i, T_i, S_i)\}_{i=1}^n$.

Assumption (Identification Conditions)

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

Assumption (Identification Conditions)

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

$$m(t) = \mathbb{E}[Y(t)] \stackrel{(*)}{=} \mathbb{E}\{\mathbb{E}[Y(t)|S]\}$$

(*) Law of total expectation

$$\stackrel{(**)}{=} \mathbb{E}\{\mathbb{E}[Y(t)|T=t, S]\}$$

(**) Ignorability

$$\stackrel{(***)}{=} \mathbb{E}[\mathbb{E}(Y|T=t, S)]$$

(***) Consistency & Positivity

Assumption (Identification Conditions)

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

$$\begin{aligned} m(t) &= \mathbb{E}[Y(t)] \stackrel{(*)}{=} \mathbb{E}\{\mathbb{E}[Y(t)|S]\} && (*) \text{ Law of total expectation} \\ &\stackrel{(**)}{=} \mathbb{E}\{\mathbb{E}[Y(t)|T=t, S]\} && (**)\text{ Ignorability} \\ &\stackrel{(***)}{=} \mathbb{E}[\mathbb{E}(Y|T=t, S)] && (***)\text{ Consistency \& Positivity} \end{aligned}$$

- The positivity condition is required in (***) for

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

to be well-defined on $\mathcal{T} \times \mathcal{S}$.

Estimation of Dose-Response Curves Under Positivity

There are three major strategies for estimating

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

from the data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, where $K : \mathbb{R} \rightarrow [0, \infty)$ is a kernel function.

Estimation of Dose-Response Curves Under Positivity

There are three major strategies for estimating

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

from the data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, where $K : \mathbb{R} \rightarrow [0, \infty)$ is a kernel function.

① **Regression Adjustment** ([Robins, 1986; Gill and Robins, 2001](#)):

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

② **Inverse Probability Weighting** ([Hirano and Imbens, 2004](#)):

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

③ **Doubly Robust** ([Kennedy et al., 2017; Colangelo and Lee, 2020](#)).

Estimation of Dose-Response Curves Under Positivity

There are three major strategies for estimating

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

from the data $\{(Y_i, T_i, S_i)\}_{i=1}^n$, where $K : \mathbb{R} \rightarrow [0, \infty)$ is a kernel function.

① **Regression Adjustment** ([Robins, 1986; Gill and Robins, 2001](#)):

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

② **Inverse Probability Weighting** ([Hirano and Imbens, 2004](#)):

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

③ **Doubly Robust** ([Kennedy et al., 2017; Colangelo and Lee, 2020](#)).

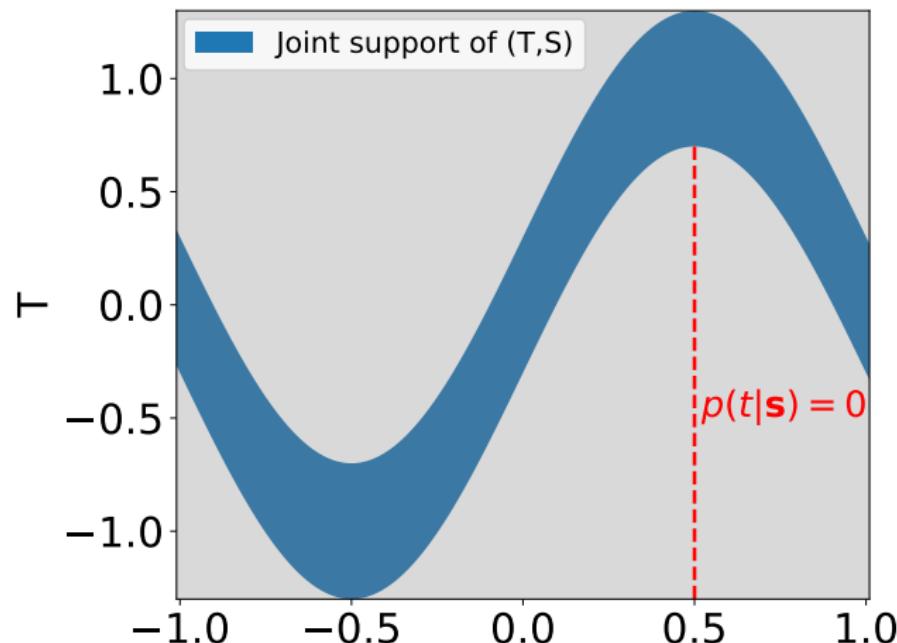
► **Issue:** Positivity is a strong assumption with continuous treatments!

Violation of the Positivity Condition

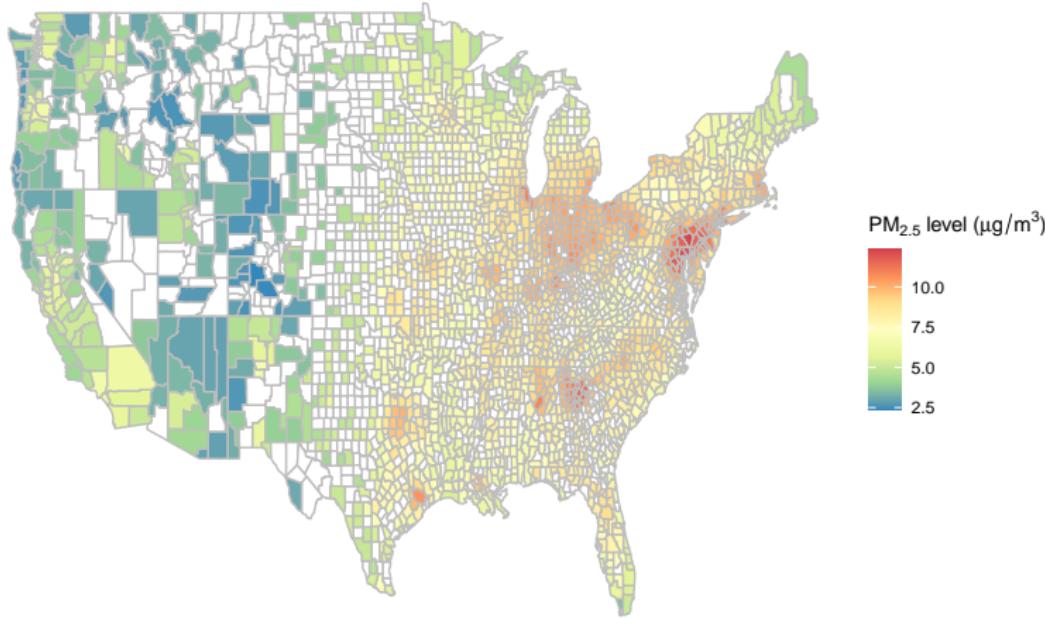
Assumption (Positivity Condition)

The conditional density $p(t|s)$ is uniformly bounded away from zero for all $(t, s) \in \mathcal{T} \times \mathcal{S}$.

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



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!

Highlight of Today's Talk

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

- ① The positivity condition may fail in some regions of $\mathcal{T} \times \mathcal{S}$.
 - We propose a new identification strategy for $m(t)$ and $\theta(t)$.

Highlight of Today's Talk

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

- ① The positivity condition may fail in some regions of $\mathcal{T} \times \mathcal{S}$.
 - We propose a new identification strategy for $m(t)$ and $\theta(t)$.
- ② We derive a novel integral estimator $\hat{m}_\theta(t)$ of $m(t)$ for all $t \in \mathcal{T}$.

Highlight of Today's Talk

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

- ① The positivity condition may fail in some regions of $\mathcal{T} \times \mathcal{S}$.
 - We propose a new identification strategy for $m(t)$ and $\theta(t)$.
- ② We derive a novel integral estimator $\hat{m}_\theta(t)$ of $m(t)$ for all $t \in \mathcal{T}$.
 - Construct a localized derivative estimator $\hat{\theta}_C(t)$ of $\theta(t) = m'(t)$ around the observations $T_i, i = 1, \dots, n$.
 - Extrapolate $\hat{\theta}_C(t)$ to any treatment level of interest via the integration.

Highlight of Today's Talk

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

- ① The positivity condition may fail in some regions of $\mathcal{T} \times \mathcal{S}$.
 - We propose a new identification strategy for $m(t)$ and $\theta(t)$.
- ② We derive a novel integral estimator $\widehat{m}_\theta(t)$ of $m(t)$ for all $t \in \mathcal{T}$.
 - Construct a localized derivative estimator $\widehat{\theta}_C(t)$ of $\theta(t) = m'(t)$ around the observations $T_i, i = 1, \dots, n$.
 - Extrapolate $\widehat{\theta}_C(t)$ to any treatment level of interest via the integration.
 - Both $\widehat{m}_\theta(t)$ and $\widehat{\theta}_C(t)$ are consistent in \mathcal{T} even when the positivity condition fails.

Highlight of Today's Talk

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

- ① The positivity condition may fail in some regions of $\mathcal{T} \times \mathcal{S}$.
 - We propose a new identification strategy for $m(t)$ and $\theta(t)$.
- ② We derive a novel integral estimator $\hat{m}_\theta(t)$ of $m(t)$ for all $t \in \mathcal{T}$.
 - Construct a localized derivative estimator $\hat{\theta}_C(t)$ of $\theta(t) = m'(t)$ around the observations $T_i, i = 1, \dots, n$.
 - Extrapolate $\hat{\theta}_C(t)$ to any treatment level of interest via the integration.
 - Both $\hat{m}_\theta(t)$ and $\hat{\theta}_C(t)$ are consistent in \mathcal{T} even when the positivity condition fails.
- ③ Nonparametric bootstrap inference with our proposed estimators $\hat{m}_\theta(t)$ and $\hat{\theta}_C(t)$ for $m(t)$ and $\theta(t)$ is asymptotically valid.

Identification



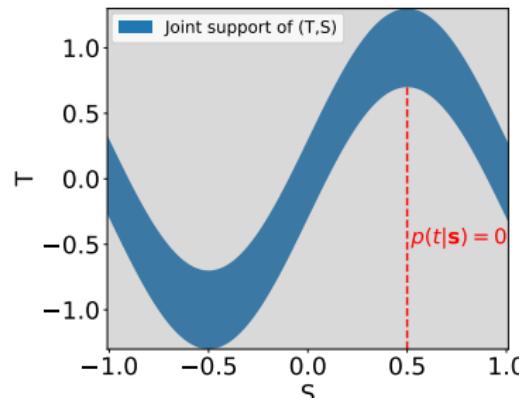
Why Do We Need Positivity?

Assumption (Identification Conditions)

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

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



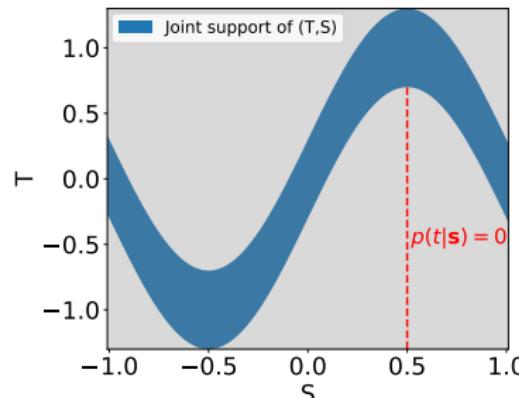
Why Do We Need Positivity?

Assumption (Identification Conditions)

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

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



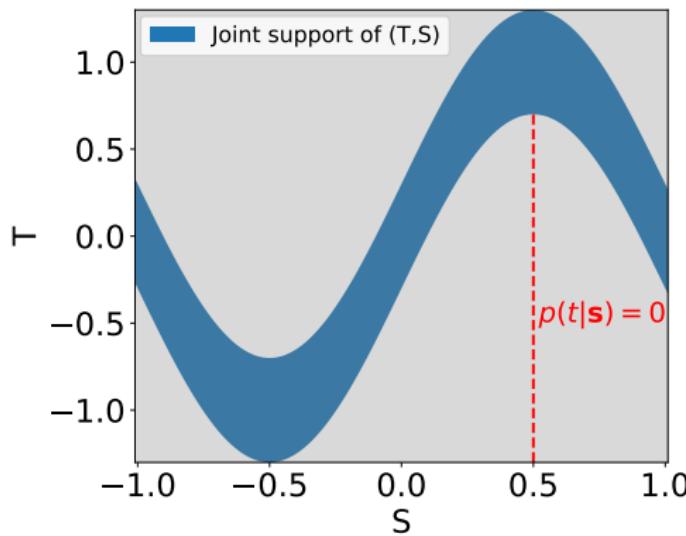
► **Identification Issue:** Without positivity,

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

is *not well-defined* outside the support $\mathcal{E} \subset \mathcal{T} \times \mathcal{S}$ of the joint density $p(t, s)$.

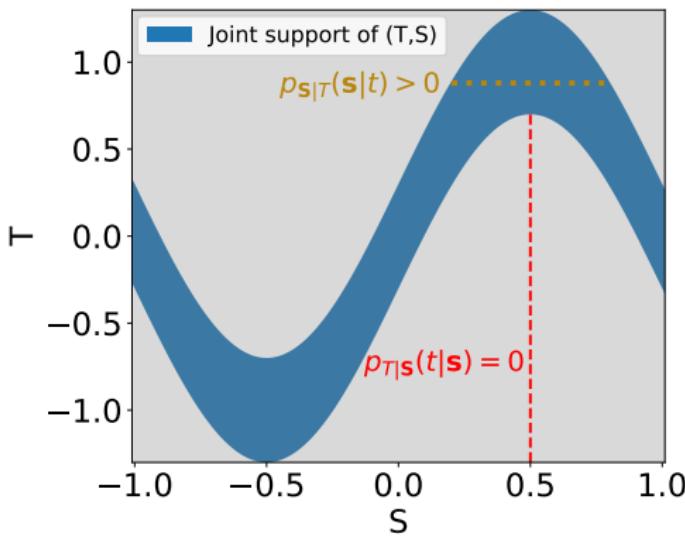
Assumption (Identification Conditions)

- ① (*Consistency*) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- ② (*Ignorability*) $Y(t)$ is conditionally independent of T given S for all $t \in \mathcal{T}$.
- ③ (*Treatment Variation*) $\text{Var}(T|S = s) > 0$ for all $s \in \mathcal{S}$.



Assumption (Identification Conditions)

- ① (*Consistency*) $Y = Y(t)$ whenever $T = t \in \mathcal{T}$.
- ② (*Ignorability*) $Y(t)$ is conditionally independent of T given S for all $t \in \mathcal{T}$.
- ③ (*Treatment Variation*) $\text{Var}(T|S = s) > 0$ for all $s \in \mathcal{S}$.



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

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

$$\begin{aligned}\theta(t) &= \frac{d}{dt} \mathbb{E}[Y(t)] = \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \right] \\ &\stackrel{*}{=} \mathbb{E} \left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \mid T = t \right].\end{aligned}$$

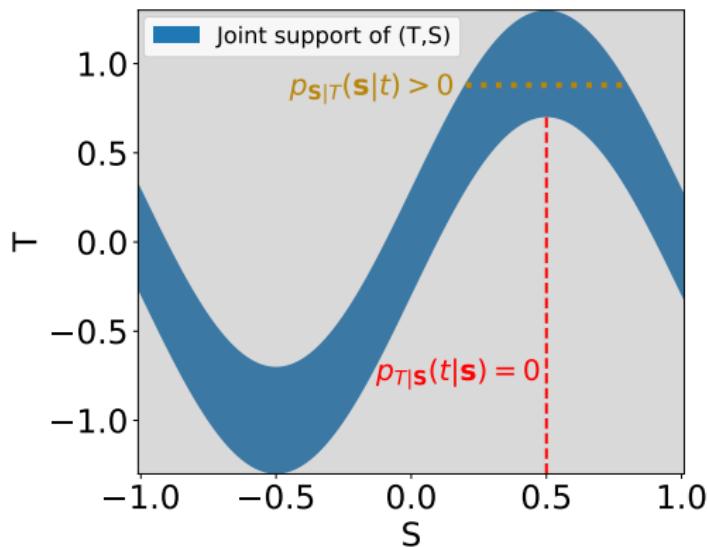
Additionally, it holds true that $\mathbb{E}(Y) = \mathbb{E}[m(T)]$.

Identification Strategy Without Positivity

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

Identification Strategy Without Positivity

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

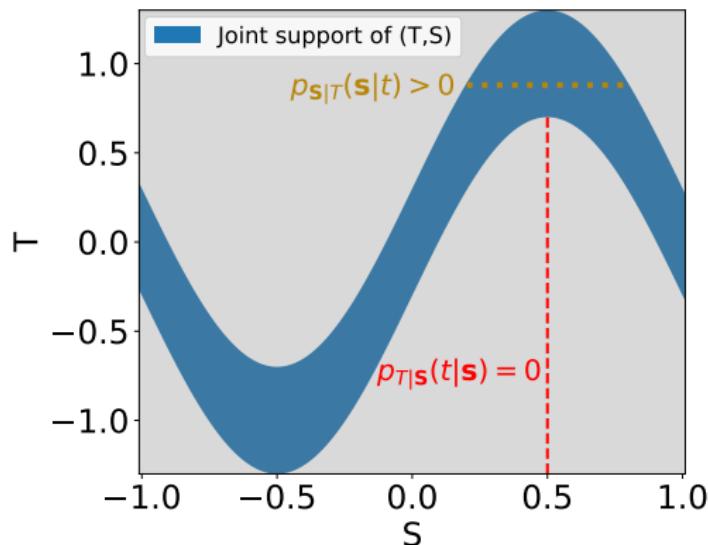


$$\begin{aligned}\theta(t) &= \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T=t\right] \\ &\stackrel{(*)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|T=t, S] \middle| T=t\right] \\ &\stackrel{(**)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y|T=t,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

(*) Ignorability; (** Consistency.

Identification Strategy Without Positivity

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



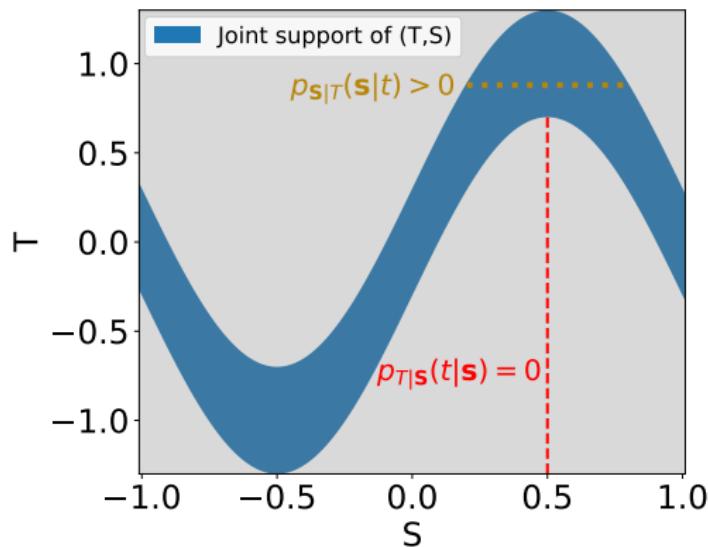
$$\begin{aligned}\theta(t) &= \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|S] \middle| T=t\right] \\ &\stackrel{(*)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y(t)|T=t, S] \middle| T=t\right] \\ &\stackrel{(**)}{=} \mathbb{E}\left[\frac{\partial}{\partial t} \mathbb{E}[Y|T=t,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

(*) Ignorability; (** Consistency.

- By the fundamental theorem of calculus, $m(t) = m(T) + \int_T^t \overbrace{m'(u)}^{\theta_C(u)} du$ so that

Identification Strategy Without Positivity

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



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

(*) Ignorability; (** Consistency.

- By the fundamental theorem of calculus, $m(t) = m(T) + \int_T^t \overbrace{m'(u)}^{=\theta_C(u)} du$ so that

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

Example: Additive Confounding Model

Consider the additive confounding model, which is commonly assumed in spatial statistics (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Gilbert et al., 2023)¹:

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

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

¹Indeed, our paper is inspired by Professor Elizabeth Ogburn's talk at UW in February 2023.

Example: Additive Confounding Model

Consider the additive confounding model, which is commonly assumed in spatial statistics ([Paciorek, 2010](#); [Schnell and Papadogeorgou, 2020](#); [Gilbert et al., 2023](#))¹:

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

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

Proposition (Proposition 2 in [Zhang et al. 2024](#))

Under the additive confounding model (1), the extrapolation condition holds:

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

¹Indeed, our paper is inspired by Professor Elizabeth Ogburn's talk at UW in February 2023.

Example: Additive Confounding Model

Consider the additive confounding model, which is commonly assumed in spatial statistics (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Gilbert et al., 2023)¹:

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

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

Proposition (Proposition 2 in Zhang et al. 2024)

Under the additive confounding model (1), the extrapolation condition holds:

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

► **Drawback of (1):** The treatment effect is homogeneous for any $S = s \in \mathcal{S}$.

¹Indeed, our paper is inspired by Professor Elizabeth Ogburn's talk at UW in February 2023.

Estimation and Inference



Proposed Integral Estimator of $m(t)$

Recall our identification formulae

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

Proposed Integral Estimator of $m(t)$

Recall our identification formulae

$$m(t) = \mathbb{E} \left[Y + \int_{\tilde{t}=T}^{\tilde{t}=t} \theta_C(\tilde{t}) d\tilde{t} \right] \quad \text{and} \quad \theta_C(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, s) \middle| T = t \right] = \int \frac{\partial}{\partial t} \mu(t, s) dP(s|t).$$

Our **integral estimator** of $m(t)$ is given by

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

and our **localized derivative** estimator is

$$\hat{\theta}_C(t) = \int \hat{\beta}_2(t, s) d\hat{P}(s|t).$$

Proposed Integral Estimator of $m(t)$

Recall our identification formulae

$$m(t) = \mathbb{E} \left[Y + \int_{\tilde{t}=T}^{\tilde{t}=t} \theta_C(\tilde{t}) d\tilde{t} \right] \quad \text{and} \quad \theta_C(t) = \mathbb{E} \left[\frac{\partial}{\partial t} \mu(t, s) \middle| T = t \right] = \int \frac{\partial}{\partial t} \mu(t, s) dP(s|t).$$

Our **integral estimator** of $m(t)$ is given by

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

and our **localized derivative** estimator is

$$\hat{\theta}_C(t) = \int \hat{\beta}_2(t, s) d\hat{P}(s|t).$$

- $\beta_2(t, s) := \frac{\partial}{\partial t} \mu(t, s)$ is fitted by (partial) local polynomial regression.
- $P(s|t)$ is estimated by Nadaraya-Watson conditional cumulative distribution function (CDF) estimator.

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

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

- $K_T : \mathbb{R} \rightarrow [0, \infty)$, $K_S : \mathbb{R}^d \rightarrow [0, \infty)$ are two symmetric kernel functions, and $h, b > 0$ are smoothing bandwidth parameters.
- The second component $\hat{\beta}_2(t, s)$ is a consistent estimator of $\beta_2(t, s) = \frac{\partial}{\partial t} \mu(t, s)$.

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

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

- $K_T : \mathbb{R} \rightarrow [0, \infty)$, $K_S : \mathbb{R}^d \rightarrow [0, \infty)$ are two symmetric kernel functions, and $h, b > 0$ are smoothing bandwidth parameters.
- The second component $\hat{\beta}_2(t, s)$ is a consistent estimator of $\beta_2(t, s) = \frac{\partial}{\partial t} \mu(t, s)$.

- ② **Nadaraya-Watson conditional CDF Estimator (Hall et al., 1999):**

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

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

Proposed Localized Derivative Estimator of $\theta(t)$

Combining two nuisance function estimators $\widehat{\beta}_2(t, s)$ and $\widehat{P}(s|t)$, we derive our **localized derivative estimator** of $\theta(t)$ as:

$$\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t, s) d\widehat{P}_{\hbar}(s|t) = \frac{\sum_{i=1}^n \widehat{\beta}_2(t, S_i) \cdot \bar{K}_T\left(\frac{T_i-t}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j-t}{\hbar}\right)}.$$

Our integral estimator takes the form

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

Proposed Localized Derivative Estimator of $\theta(t)$

Combining two nuisance function estimators $\widehat{\beta}_2(t, s)$ and $\widehat{P}(s|t)$, we derive our **localized derivative estimator** of $\theta(t)$ as:

$$\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t, s) d\widehat{P}_{\hbar}(s|t) = \frac{\sum_{i=1}^n \widehat{\beta}_2(t, S_i) \cdot \bar{K}_T\left(\frac{T_i-t}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j-t}{\hbar}\right)}.$$

Our integral estimator takes the form

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

- Other methods can be applied to estimate $\frac{\partial}{\partial t} \mu(t, s)$ and $P(s|t)$.
- $\widehat{m}_\theta(t)$, under our kernel-based estimators, is a *linear smoother*.

Proposed Localized Derivative Estimator of $\theta(t)$

Combining two nuisance function estimators $\widehat{\beta}_2(t, s)$ and $\widehat{P}(s|t)$, we derive our **localized derivative estimator** of $\theta(t)$ as:

$$\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t, s) d\widehat{P}_{\hbar}(s|t) = \frac{\sum_{i=1}^n \widehat{\beta}_2(t, S_i) \cdot \bar{K}_T\left(\frac{T_i-t}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j-t}{\hbar}\right)}.$$

Our integral estimator takes the form

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

- Other methods can be applied to estimate $\frac{\partial}{\partial t} \mu(t, s)$ and $P(s|t)$.
 - $\widehat{m}_\theta(t)$, under our kernel-based estimators, is a *linear smoother*.
- **Issue:** The integral could be analytically difficult to compute.

Fast Computing Algorithm for the Integral Estimator

Our integral estimator takes the form

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

- **Riemann Sum Approximation:** Let $T_{(1)} \leq \dots \leq T_{(n)}$ be the order statistics of T_1, \dots, T_n and $\Delta_j = T_{(j+1)} - T_{(j)}$ for $j = 1, \dots, n - 1$.
- Approximate $\hat{m}_\theta(T_{(j)})$ for each $j = 1, \dots, n$ as:

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

Fast Computing Algorithm for the Integral Estimator

Our integral estimator takes the form

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

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

- Approximate $\hat{m}_\theta(T_{(j)})$ for each $j = 1, \dots, n$ as:

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

- Evaluate $\hat{m}_\theta(t)$ at any $t \in [T_{(j)}, T_{(j+1)}]$ by a linear interpolation between $\hat{m}_\theta(T_{(j)})$ and $\hat{m}_\theta(T_{(j+1)})$.
- The approximation error is at most $O_P(\frac{1}{n})$, which is *asymptotically negligible*.

Nonparametric Bootstrap Inference

- ① Compute $\hat{m}_\theta(t)$ on the original data $\{(Y_i, T_i, S_i)\}_{i=1}^n$.

Nonparametric Bootstrap Inference

- ① Compute $\hat{m}_\theta(t)$ 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 $\hat{m}_\theta^{*(b)}(t)$ for each $b = 1, \dots, B$.

Nonparametric Bootstrap Inference

- ① Compute $\hat{m}_\theta(t)$ 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 $\hat{m}_\theta^{*(b)}(t)$ for each $b = 1, \dots, B$.
- 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 $\{D_1(t_0), \dots, D_B(t_0)\}$, where $D_b(t_0) = \left| \hat{m}_\theta^{*(b)}(t_0) - \hat{m}_\theta(t_0) \right|$ for $b = 1, \dots, B$.
 - For uniform inference on $m(t)$, compute the $1 - \alpha$ quantile $\xi_{1-\alpha}^*$ of $\{D_{\text{sup},1}, \dots, D_{\text{sup},B}\}$, where $D_{\text{sup},b} = \sup_{t \in \mathcal{T}} \left| \hat{m}_\theta^{*(b)}(t) - \hat{m}_\theta(t) \right|$ for $b = 1, \dots, B$.

Nonparametric Bootstrap Inference

- ① Compute $\hat{m}_\theta(t)$ 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 $\hat{m}_\theta^{*(b)}(t)$ for each $b = 1, \dots, B$.
- 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 $\{D_1(t_0), \dots, D_B(t_0)\}$, where $D_b(t_0) = \left|\hat{m}_\theta^{*(b)}(t_0) - \hat{m}_\theta(t_0)\right|$ for $b = 1, \dots, B$.
 - For uniform inference on $m(t)$, compute the $1 - \alpha$ quantile $\xi_{1-\alpha}^*$ of $\{D_{\text{sup},1}, \dots, D_{\text{sup},B}\}$, where $D_{\text{sup},b} = \sup_{t \in \mathcal{T}} \left|\hat{m}_\theta^{*(b)}(t) - \hat{m}_\theta(t)\right|$ for $b = 1, \dots, B$.
- ④ Define the $1 - \alpha$ confidence interval for $m(t_0)$ as:

$$[\hat{m}_\theta(t_0) - \zeta_{1-\alpha}^*(t_0), \hat{m}_\theta(t_0) + \zeta_{1-\alpha}^*(t_0)]$$

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

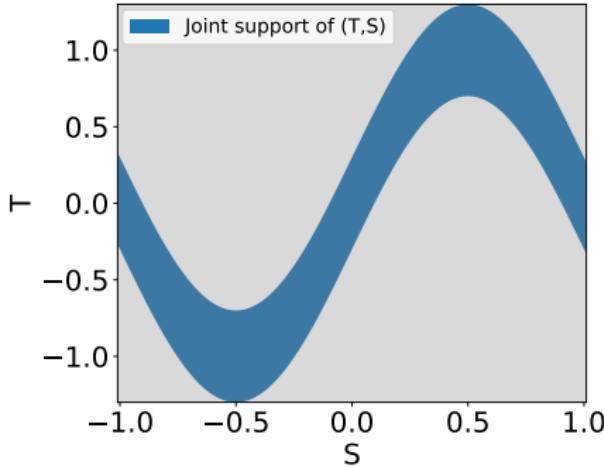
$$[\hat{m}_\theta(t) - \xi_{1-\alpha}^*, \hat{m}_\theta(t) + \xi_{1-\alpha}^*].$$

Asymptotic Theory



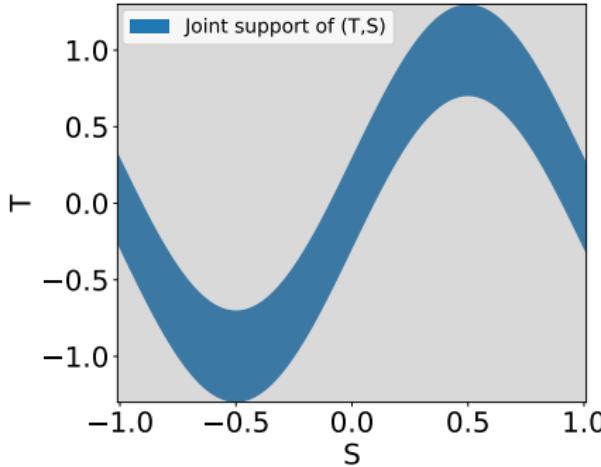
Uniform Consistency of Local Polynomial Regression

- The support \mathcal{E} of (T, S) may not cover $\mathcal{T} \times \mathcal{S}$ without positivity.



Uniform Consistency of Local Polynomial Regression

- The support \mathcal{E} of (T, S) may not cover $\mathcal{T} \times \mathcal{S}$ without positivity.



- $\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t, s) d\widehat{P}_h(s|t)$ only requires $\widehat{\beta}_2(t, s)$ to be consistent in \mathcal{E} .

Lemma (Lemma 3 in Zhang et al. 2024)

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

$$\sup_{(t,s) \in \mathcal{E}} \left| \widehat{\beta}_2(t, s) - \frac{\partial}{\partial t} \mu(t, s) \right| = O \left(h^q + b^2 + \frac{\max\{h, b\}^4}{h} \right) + O_P \left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}} \right).$$

Uniform Consistencies of Proposed Estimators

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

Theorem (Theorem 4 in [Zhang et al. 2024](#))

Let $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t) \geq p_{T,\min} > 0$ for all $t \in \mathcal{T}'$. When $q = 2$ and $h, b, \hbar, \frac{\max\{h,b\}^4}{h} \rightarrow 0$ and $\frac{n \max\{h, \hbar\} b^d}{\log n}, \frac{n \hbar}{\log n} \rightarrow \infty$,

$$\sup_{t \in \mathcal{T}'} |\widehat{\theta}_C(t) - \theta_C(t)| = \underbrace{O\left(h^2 + b^2 + \frac{\max\{b, h\}^4}{h}\right)}_{\text{Bias term}} + \underbrace{O_P\left(\sqrt{\frac{\log n}{nh^3}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right)}_{\text{Stochastic variation}},$$
$$\sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| = O\left(h^2 + b^2 + \frac{\max\{b, h\}^4}{h}\right) + O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right).$$

Uniform Rate of Convergence For the Integral Estimator

$$\sup_{t \in T'} |\hat{m}_\theta(t) - m(t)| = O \left(h^2 + b^2 + \frac{\max\{b, h\}^4}{h} \right) + O_P \left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \frac{h^2}{n} + \sqrt{\frac{\log n}{nh}} \right).$$

- **Blue term:** the estimation bias of local polynomial estimator $\hat{\beta}_2(t, s)$.
- **Orange term:** additional bias of $\hat{\beta}_2(t, s)$ at the boundary $\partial\mathcal{E}$.

Uniform Rate of Convergence For the Integral Estimator

$$\sup_{t \in T'} |\hat{m}_\theta(t) - m(t)| = O \left(\textcolor{blue}{h^2} + b^2 + \frac{\max\{b, h\}^4}{h} \right) + O_P \left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \textcolor{red}{h^2} + \sqrt{\frac{\log \textcolor{teal}{n}}{n h}} \right).$$

- **Blue term:** the estimation bias of local polynomial estimator $\hat{\beta}_2(t, s)$.
- **Orange term:** additional bias of $\hat{\beta}_2(t, s)$ at the boundary $\partial \mathcal{E}$.
- **Teal term:** asymptotic rate from $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$.
- **Red term:** stochastic variation of $\hat{\beta}_2(t, s)$.

Uniform Rate of Convergence For the Integral Estimator

$$\sup_{t \in T'} |\hat{m}_\theta(t) - m(t)| = O \left(h^2 + b^2 + \frac{\max\{b, h\}^4}{h} \right) + O_P \left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\log n}{nh^3}} + \frac{h^2}{n} + \sqrt{\frac{\log n}{nh}} \right).$$

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

Asymptotic Linearity of Proposed Estimators

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 $\frac{nh^5}{\log n} \rightarrow c_1$ and $\frac{n\hbar^5}{\log n} \rightarrow c_2$ for some $c_1, c_2 \geq 0$ and $\frac{n \max\{h, \hbar\} b^d}{\log n}, \frac{n\hbar}{\log n}, \frac{h^3 \log n}{\hbar}, \frac{nh^3\hbar^4}{\log n} \rightarrow \infty$ as $n \rightarrow \infty$, then for any $t \in \mathcal{T}'$,

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

where

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

and $\varphi_t(Y, T, S) = \mathbb{E}_{T_1} \left[\int_{T_1}^t \bar{\varphi}_t(Y, T, S) d\tilde{t} \right]$ with $\mathbb{G}_n = \sqrt{n} (\mathbb{P}_n - P)$, 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.

Nonparametric Bootstrap Consistency

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

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

1

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

Nonparametric Bootstrap Consistency

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

Under the same regularity conditions, if $h \asymp n^{-\frac{1}{\gamma}}$ and $b \lesssim \bar{h} \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \rightarrow c_1$ and $\frac{n\bar{h}^5}{\log n} \rightarrow c_2$ for some $c_1, c_2 \geq 0$ and

$$\frac{\bar{h}}{\bar{h}^3 \log n}, \bar{h}n^{\frac{1}{3}} \log n, \frac{\sqrt{n}\bar{h}}{\log n}, \frac{n \max\{h, \bar{h}\} b^d}{\log n} \rightarrow \infty \text{ as } n \rightarrow \infty,$$

1

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

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

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

Nonparametric Bootstrap Consistency

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

Under the same regularity conditions, if $h \asymp n^{-\frac{1}{\gamma}}$ and $b \lesssim \bar{h} \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \rightarrow c_1$ and $\frac{n\bar{h}^5}{\log n} \rightarrow c_2$ for some $c_1, c_2 \geq 0$ and

$$\frac{\bar{h}}{\bar{h}^3 \log n}, \bar{h}n^{\frac{1}{3}} \log n, \frac{\sqrt{n}\bar{h}}{\log n}, \frac{n \max\{h, \bar{h}\} b^d}{\log n} \rightarrow \infty \text{ as } n \rightarrow \infty,$$

1

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

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

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

3

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

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

Remarks on Our Asymptotic Results

- ① \mathcal{F} is not Donsker because φ_t is not uniformly bounded as $h \rightarrow 0$.
 - 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 $\sup_{f \in \mathcal{F}} |\mathbb{B}(f)|$.

Remarks on Our Asymptotic Results

- ① \mathcal{F} is not Donsker because φ_t is not uniformly bounded as $h \rightarrow 0$.
 - 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 $\text{Var}(Y|T = t, S = s) \geq \sigma^2 > 0$, $\text{Var}[\varphi_t(Y, T, S)]$ is a positive finite number.
 - The asymptotic linearity (or V-statistic) is non-degenerate.
 - Pointwise bootstrap confidence intervals are asymptotically valid.

Remarks on Our Asymptotic Results

- ① \mathcal{F} is not Donsker because φ_t is not uniformly bounded as $h \rightarrow 0$.
 - 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 $\text{Var}(Y|T = t, S = s) \geq \sigma^2 > 0$, $\text{Var}[\varphi_t(Y, T, S)]$ 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!!

Case Study: PM_{2.5} on CMR



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

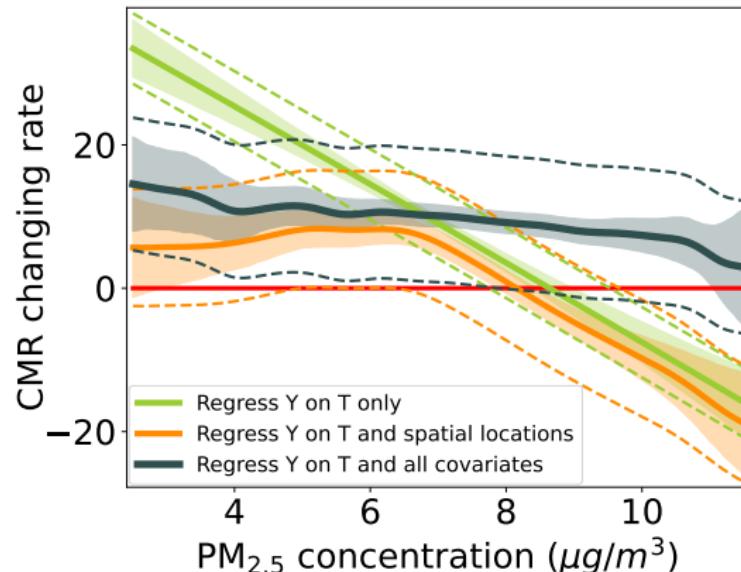
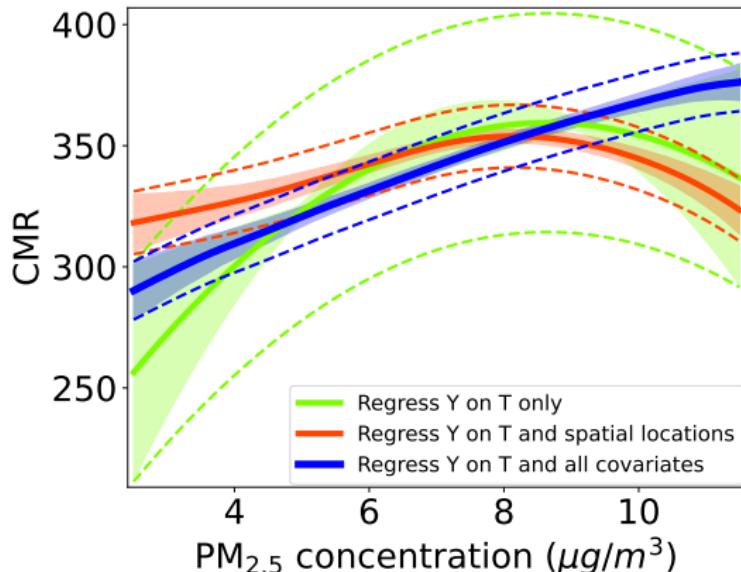
- ① The dataset ([Wyatt et al., 2020a,b](#)) contains the average annual CMRs (Y) and PM_{2.5} levels (T) across $n = 2132$ U.S. counties over 1990-2010.
- ② The covariate vector $S \in \mathbb{R}^{10}$ consists of two parts:
 - 2 spatial confounders: latitude and longitude of each county.
 - 8 county-level 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 $S \in \mathbb{R}^{10}$ 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 $\mu\text{g}/\text{m}^3$ and 11.5 $\mu\text{g}/\text{m}^3$ to avoid boundary effects ([Takatsu and Westling, 2022](#)).

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

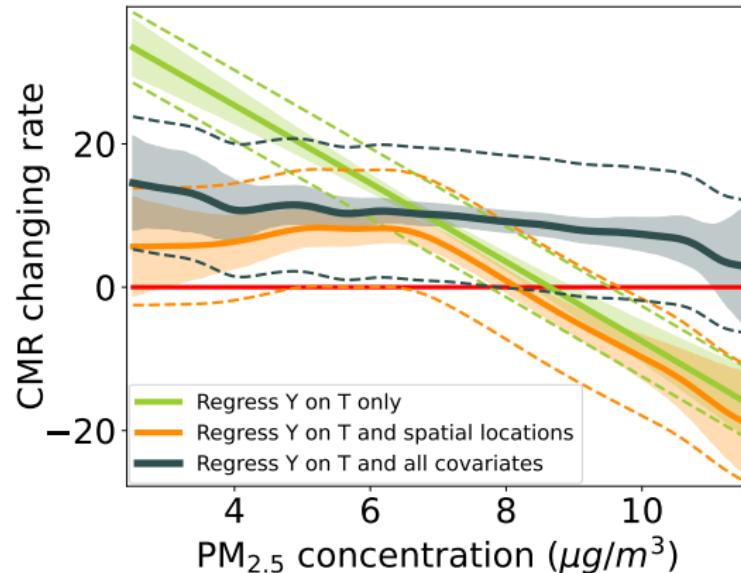
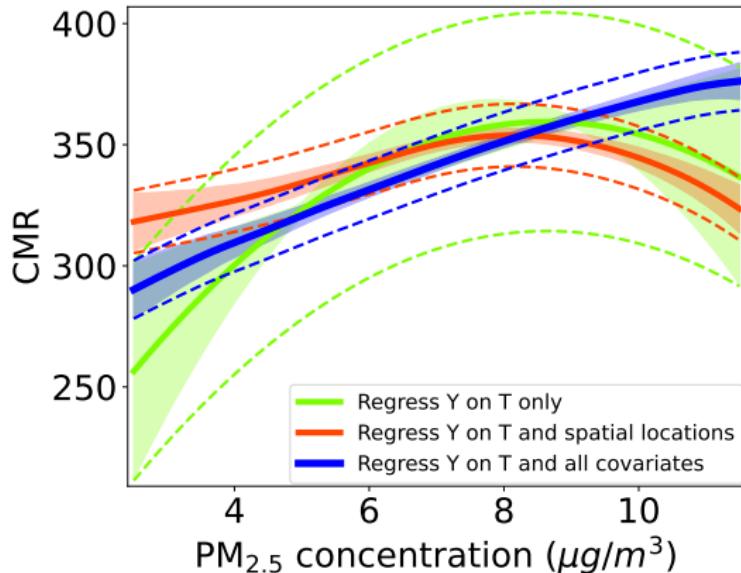


Shaded areas: 95% pointwise confidence intervals; Regions between dashed lines: 95% uniform confidence bands.

- 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 $\text{PM}_{2.5} < 8 \mu\text{g}/\text{m}^3$.

Discussion



Summary of Today's Talk

We study nonparametric inference on the dose-response curve $m(t) = \mathbb{E}[Y(t)]$ and its derivative $\theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)]$ without the **positivity** condition.

- Our key technique relies on two pillars in calculus:

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

Summary of Today's Talk

We study nonparametric inference on the dose-response curve $m(t) = \mathbb{E}[Y(t)]$ and its derivative $\theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)]$ without the **positivity** condition.

- Our key technique relies on two pillars in calculus:

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

- Both estimators are consistent without the positivity condition.

Summary of Today's Talk

We study nonparametric inference on the dose-response curve $m(t) = \mathbb{E}[Y(t)]$ and its derivative $\theta(t) = \frac{d}{dt}\mathbb{E}[Y(t)]$ without the **positivity** condition.

- Our key technique relies on two pillars in calculus:

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

- Both estimators are consistent without the positivity condition.
- Our integration idea opens a new direction for causal inference with continuous treatments under violations of positivity!

Ongoing Works and Future Directions

- 1 **IPW and Doubly Robust Estimation:** Our estimators are based on regression adjustment forms. Can we generalize to IPW and doubly robust forms ([Zhang and Chen, 2025](#))?

Ongoing Works and Future Directions

- ① **IPW and Doubly Robust Estimation:** Our estimators are based on regression adjustment forms. Can we generalize to IPW and doubly robust forms ([Zhang and Chen, 2025](#))?
- ② **Violation of Ignorability:** Can we conduct sensitivity analysis on unmeasured confounding ([Chernozhukov et al., 2022](#))?
- ③ **High-Dimensional Confounders:** $\left(\frac{\log n}{n}\right)^{\frac{4}{5d}} \lesssim b \lesssim n^{-\frac{1}{5}}$ only works when $d < 5$.
Can we use additive models ([Guo et al., 2019](#)) to resolve the dimensionality issue?

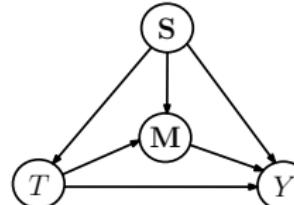
$$\mathbb{E}(Y|T=t, S=s, Z=z) = m(t) + \eta(s) + \sum_{j=1}^{d'} g_j(z_j) \quad \text{with} \quad z \in \mathbb{R}^{d'}, d' \gg d.$$

Ongoing Works and Future Directions

- ① **IPW and Doubly Robust Estimation:** Our estimators are based on regression adjustment forms. Can we generalize to IPW and doubly robust forms ([Zhang and Chen, 2025](#))?
- ② **Violation of Ignorability:** Can we conduct sensitivity analysis on unmeasured confounding ([Chernozhukov et al., 2022](#))?
- ③ **High-Dimensional Confounders:** $\left(\frac{\log n}{n}\right)^{\frac{4}{5d}} \lesssim b \lesssim n^{-\frac{1}{5}}$ only works when $d < 5$.
Can we use additive models ([Guo et al., 2019](#)) to resolve the dimensionality issue?

$$\mathbb{E}(Y|T=t, S=s, Z=z) = m(t) + \eta(s) + \sum_{j=1}^{d'} g_j(z_j) \quad \text{with} \quad z \in \mathbb{R}^{d'}, d' \gg d.$$

- ④ **Mediation Analysis:** Can we generalize our strategies for the estimation of direct and indirect causal effects ([Huber et al., 2020; Xu et al., 2021](#)) without positivity?



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>.

All the code and data are available at

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

Python Package: [npDoseResponse](#) and R Package: [npDoseResponse](#).

- [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>.

Reference

- 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.
- 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.
- 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, C. Cinelli, W. Newey, A. Sharma, and V. Syrgkanis. Long story short: Omitted variable bias in causal machine learning. Technical report, National Bureau of Economic Research, 2022.
- K. Colangelo and Y.-Y. Lee. Double debiased machine learning nonparametric inference with continuous treatments. *arXiv preprint arXiv:2004.03036*, 2020.
- 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.
- 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.

Reference

- R. D. Gill and J. M. Robins. Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 29(6):1785–1811, 2001.
- Z. Guo, W. Yuan, and C.-H. Zhang. Decorrelated local linear estimator: Inference for non-linear effects in high-dimensional additive models. *arXiv preprint arXiv:1907.12732*, 2019.
- 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.
- K. Hirano and G. W. Imbens. *The Propensity Score with Continuous Treatments*, chapter 7, pages 73–84. John Wiley & Sons, Ltd, 2004.
- M. Huber, Y.-C. Hsu, Y.-Y. Lee, and L. Lettry. Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *Journal of Applied Econometrics*, 35(7):814–840, 2020.
- E. Kammann and M. P. Wand. Geoadditive models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 52(1):1–18, 2003.
- 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.
- M. R. Miller and D. E. Newby. Air pollution and cardiovascular disease: car sick. *Cardiovascular Research*, 116(2):279–294, 2020.
- C. J. Paciorek. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1):107–125, 2010.

Reference

- 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.
- 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.
- K. Takatsu and T. Westling. Debiased inference for a covariate-adjusted regression function. *arXiv preprint arXiv:2210.06448*, 2022.
- H. Thaden and T. Kneib. Structural equation models for dealing with spatial confounding. *The American Statistician*, 72(3):239–252, 2018.
- 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:2501.06969*, 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.

Regularity Assumptions (Smoothness Conditions)

Let $\mathcal{E} \subset \mathcal{T} \times \mathcal{S}$ be the support of $p(t, s)$, \mathcal{E}° be the interior of \mathcal{E} , and $\partial\mathcal{E}$ be the boundary of \mathcal{E} .

- ① For any $(t, s) \in \mathcal{E}^\circ$, $\mu(t, s)$ is at least $(q + 1)$ times continuously differentiable with respect to t and at least four times continuously differentiable with respect to s . All these partial derivatives of $\mu(t, s)$ are continuous up to the boundary $\partial\mathcal{E}$. Furthermore, $\mu(t, s)$ and the partial derivatives are uniformly bounded on \mathcal{E} . Finally, there exist absolute constants $\sigma, A_0 > 0$ such that $\text{Var}(Y|T = t, S = s) = \sigma^2$ and $\mathbb{E}|Y|^4 < A_0 < \infty$ uniformly in \mathcal{E} .
- ② $p(t, s)$ is bounded and at least twice continuously differentiable with bounded partial derivatives up to the second order on \mathcal{E}° . All these partial derivatives of $p(t, s)$ are continuous up to the boundary $\partial\mathcal{E}$. Furthermore, \mathcal{E} is compact and $p(t, s)$ is uniformly bounded away from 0 on \mathcal{E} . Finally, the marginal density $p_T(t)$ of T is non-degenerate, *i.e.*, its support \mathcal{T} has a nonempty interior.

Regularity Assumptions (Boundary Conditions)

- ③ There exists some constants $r_1, r_2 \in (0, 1)$ such that for any $(t, s) \in \mathcal{E}$ and all $\delta \in (0, r_1]$, there is a point $(t', s') \in \mathcal{E}$ satisfying

$$\mathcal{B}((t', s'), r_2\delta) \subset \mathcal{B}((t, s), \delta) \cap \mathcal{E},$$

where

$$\mathcal{B}((t, s), r) = \left\{ (t_1, s_1) \in \mathbb{R}^{d+1} : \|(t_1 - t, s_1 - s)\|_2 \leq r \right\}$$

with $\|\cdot\|_2$ being the standard Euclidean norm.

- ④ For any $(t, s) \in \partial\mathcal{E}$, the boundary of \mathcal{E} , it satisfies that $\frac{\partial}{\partial t} p(t, s) = \frac{\partial}{\partial s_j} p(t, s) = 0$ and $\frac{\partial^2}{\partial s_j^2} \mu(t, s) = 0$ for all $j = 1, \dots, d$.
- ⑤ For any $\delta > 0$, the Lebesgue measure of the set $\partial\mathcal{E} \oplus \delta$ satisfies $|\partial\mathcal{E} \oplus \delta| \leq A_1 \cdot \delta$ for some absolute constant $A_1 > 0$, where

$$\partial\mathcal{E} \oplus \delta = \left\{ z \in \mathbb{R}^{d+1} : \inf_{x \in \partial\mathcal{E}} \|z - x\|_2 \leq \delta \right\}.$$

Regularity Assumptions (Kernel Conditions)

- ⑥ $K_T : \mathbb{R} \rightarrow [0, \infty)$ and $K_S : \mathbb{R}^d \rightarrow [0, \infty)$ are compactly supported and Lipschitz continuous kernels such that $\int_{\mathbb{R}} K_T(t) dt = \int_{\mathbb{R}^d} K_S(s) ds = 1$, $K_T(t) = K_T(-t)$, and K_S is radially symmetric with $\int s \cdot K_S(s) ds = 0$. In addition, for all $j = 1, 2, \dots$, and $\ell = 1, \dots, d$,

$$\begin{aligned}\kappa_j^{(T)} &:= \int_{\mathbb{R}} u^j K_T(u) du < \infty, & \nu_j^{(T)} &:= \int_{\mathbb{R}} u^j K_T^2(u) du < \infty, \\ \kappa_{j,\ell}^{(S)} &:= \int_{\mathbb{R}^d} u_\ell^j K_S(u) du < \infty, & \text{and} & \nu_{j,k}^{(S)} := \int_{\mathbb{R}^d} u_\ell^j K_S^2(u) du < \infty.\end{aligned}$$

Finally, both K_T and K_S are second-order kernels, i.e., $\kappa_2^{(T)} > 0$ and $\kappa_{2,\ell}^{(S)} > 0$ for all $\ell = 1, \dots, d$.

- ⑦ Let $\mathcal{K}_{q,d} = \left\{ (y, z) \mapsto \left(\frac{y-t}{h}\right)^\ell \left(\frac{z_i-s_i}{b}\right)^{k_1} \left(\frac{z_j-s_j}{b}\right)^{k_2} K_T\left(\frac{y-t}{h}\right) K_S\left(\frac{z-s}{b}\right) : (t, s) \in \mathcal{T} \times \mathcal{S}; i, j = 1, \dots, d; \ell = 0, \dots, 2q; k_1, k_2 = 0, 1; h, b > 0 \right\}$. It holds that $\mathcal{K}_{q,d}$ is a bounded VC-type class of measurable functions on \mathbb{R}^{d+1} .

Regularity Assumptions (Kernel Conditions)

- ⑧ The function $\bar{K}_T : \mathbb{R} \rightarrow [0, \infty)$ is a second-order, Lipschitz continuous, and symmetric kernel with a compact support, i.e., $\int_{\mathbb{R}} \bar{K}_T(t) dt = 1$, $\bar{K}_T(t) = \bar{K}_T(-t)$, and $\int_{\mathbb{R}} t^2 \bar{K}_T(t) dt \in (0, \infty)$.
- ⑨ Let $\bar{\mathcal{K}} = \left\{ y \mapsto \bar{K}_T \left(\frac{y-t}{\hbar} \right) : t \in \mathcal{T}, \hbar > 0 \right\}$. It holds that $\bar{\mathcal{K}}$ is a bounded VC-type class of measurable functions on \mathbb{R} .

Recall that the class \mathcal{G} of measurable functions on \mathbb{R}^{d+1} is VC-type if there exist constants $A_2, v_2 > 0$ such that for any $0 < \epsilon < 1$,

$$\sup_Q N \left(\mathcal{G}, L_2(Q), \epsilon \|G\|_{L_2(Q)} \right) \leq \left(\frac{A_2}{\epsilon} \right)^{v_2},$$

where $N \left(\mathcal{G}, L_2(Q), \epsilon \|G\|_{L_2(Q)} \right)$ is the $\epsilon \|G\|_{L_2(Q)}$ -covering number of the (semi-)metric space $(\mathcal{G}, \|\cdot\|_{L_2(Q)})$, Q is any probability measure on \mathbb{R}^{d+1} , G is an envelope function of \mathcal{G} , and $\|G\|_{L_2(Q)}$ is defined as $\left[\int_{\mathbb{R}^{d+1}} [G(x)]^2 dQ(x) \right]^{\frac{1}{2}}$.

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:

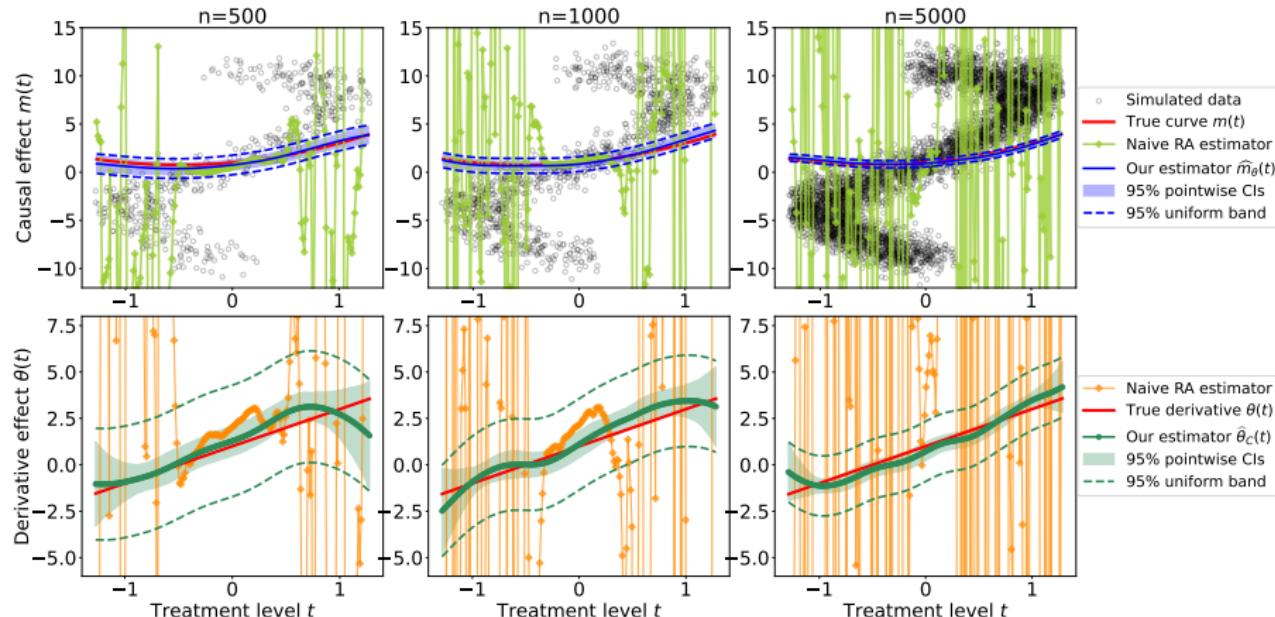
$$\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, S_i) \quad \text{and} \quad \hat{\theta}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_2(t, S_i).$$

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.

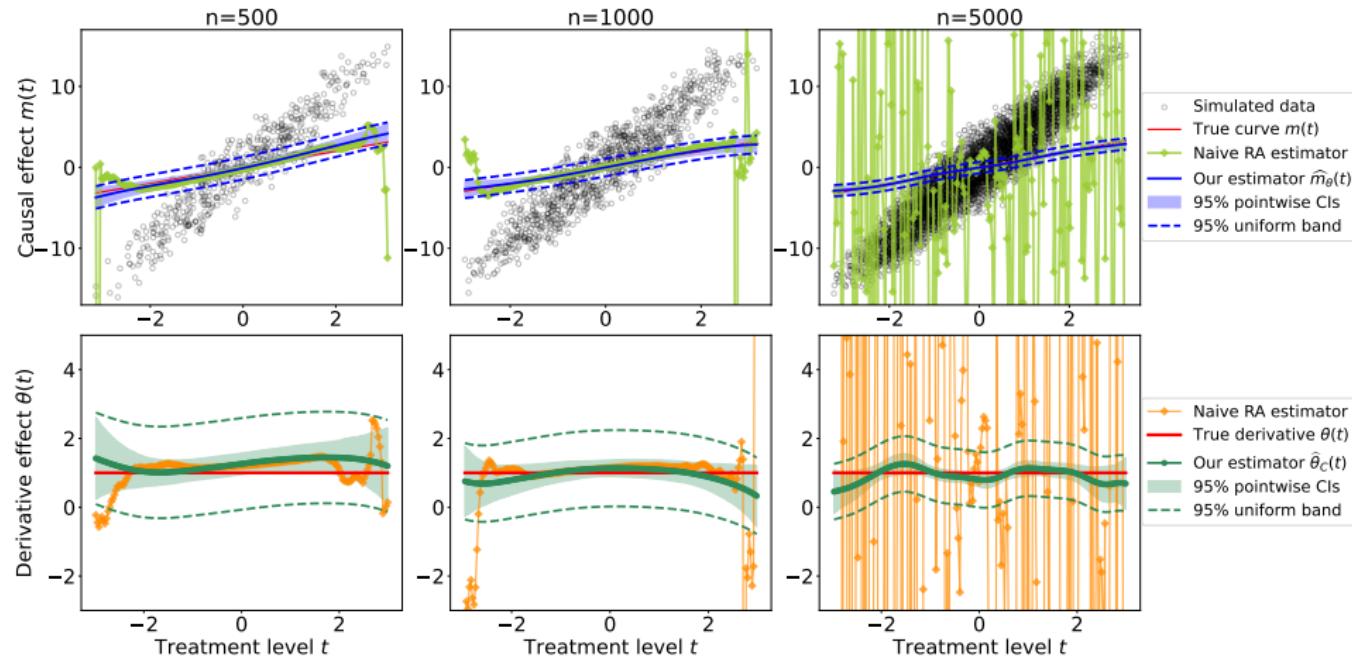


Linear Confounding Model Without Positivity

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 $\epsilon \sim \mathcal{N}(0, 1)$.

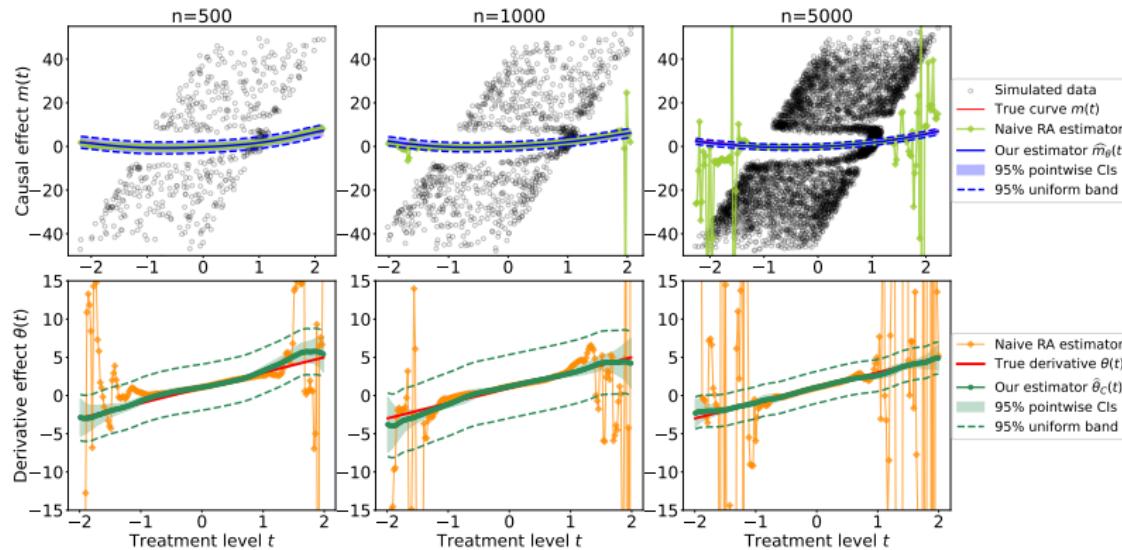


Nonlinear Confounding Model Without Positivity

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

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

- $(S_1, S_2) \sim \text{Uniform}[-1, 1]^2$, $E \sim \text{Uniform}[-0.1, 0.1]$, and $\epsilon \sim \mathcal{N}(0, 1)$.
- Those doubly robust methods based on pseudo-outcomes ([Kennedy et al., 2017](#); [Takatsu and Westling, 2022](#)) do not work in this example.



Nonparametric Bound on $m(t)$ When $\text{Var}(E) = 0$

For simplicity, we assume the additive confounding model

$$Y = \bar{m}(T) + \eta(S) + \epsilon, \quad T = f(S) + E \quad \text{with} \quad \mathbb{E}[\eta(S)] = 0 \quad \text{and} \quad \mathbb{E}(E) = 0.$$

When $\text{Var}(E) = 0$,

- $\mu(t, s)$ can be identified only on a lower-dimensional surface $\{(t, s) \in \mathcal{T} \times \mathcal{S} : t = f(s)\}$ so that

$$\mu(f(s), s) = \bar{m}(f(s)) + \eta(s) = m(f(s)) + \eta(s). \quad (2)$$

- The relation $T = f(S)$ can be recovered from the data $\{(T_i, S_i)\}_{i=1}^n$.

Assumption (Bounded random effect)

Let $L_f(t) = \{s \in \mathcal{S} : f(s) = t\}$ be a level set of the function $f : \mathcal{S} \rightarrow \mathbb{R}$ at $t \in \mathcal{T}$. There exists a constant $\rho_1 > 0$ such that

$$\rho_1 \geq \max \left\{ \sup_{t \in \mathcal{T}} \sup_{s \in L_f(t)} |\eta(s)|, \frac{\sup_{t \in \mathcal{T}} \sup_{s \in L_f(t)} \mu(f(s), s) - \inf_{t \in \mathcal{T}} \inf_{s \in L_f(t)} \mu(f(s), s)}{2} \right\}.$$

Nonparametric Bound on $m(t)$ When $\text{Var}(E) = 0$

By (2) and the first lower bound on $\rho_1 \geq \sup_{t \in \mathcal{T}} \sup_{\mathbf{s} \in L_f(t)} |\eta(\mathbf{s})|$ in the previous assumption, we know that

$$|\mu(f(\mathbf{s}), \mathbf{s}) - m(t)| = |\eta(\mathbf{s})| \leq \rho_1$$

for any $\mathbf{s} \in L_f(t)$. It also implies that

$$\begin{aligned} m(t) &\in \bigcap_{\mathbf{s} \in L_f(t)} [\mu(f(\mathbf{s}), \mathbf{s}) - \rho_1, \mu(f(\mathbf{s}), \mathbf{s}) + \rho_1] \\ &= \left[\sup_{\mathbf{s} \in L_f(t)} \mu(f(\mathbf{s}), \mathbf{s}) - \rho_1, \inf_{\mathbf{s} \in L_f(t)} \mu(f(\mathbf{s}), \mathbf{s}) + \rho_1 \right], \end{aligned}$$

which is the nonparametric bound on $m(t)$ that contains all the possible values of $m(t)$ for any fixed $t \in \mathcal{T}$ when $\text{Var}(E) = 0$.

- This bound is well-defined and nonempty under the second lower bound on ρ_1 in the previous assumption.