

# 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

School of Mathematics, Sun Yat-Sen University  
September 12, 2024

# Introduction



## A Central Problem in Causal Inference:

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

---

<sup>1</sup>Here,  $Y(t)$  is the potential outcome that would have been observed under treatment level  $T = t$ .

## A Central Problem in Causal Inference:

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

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

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

---

<sup>1</sup>Here,  $Y(t)$  is the potential outcome that would have been observed under treatment level  $T = t$ .

## A Central Problem in Causal Inference:

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

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

- $\mathbb{E}[Y(t)]$  = mean counterfactual outcome<sup>1</sup> 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.,  $\mathcal{T} \subset \mathbb{R}$ )?

---

<sup>1</sup>Here,  $Y(t)$  is the potential outcome that would have been observed under treatment level  $T = t$ .

## A Central Problem in Causal Inference:

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

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

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

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

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

---

<sup>1</sup>Here,  $Y(t)$  is the potential outcome that would have been observed under treatment level  $T = t$ .

# Identification of a Causal Dose-Response Curve

Without confounding,  $m(t) = \mathbb{E}[Y(t)] = \mathbb{E}(Y|T = t)$ .

- Fitting  $m(t)$  is to regress  $\{Y_i\}_{i=1}^n$  with respect to  $\{T_i\}_{i=1}^n$ .
- Recovering  $\theta(t)$  is a classical derivative estimation problem (Gasser and Müller, 1984).

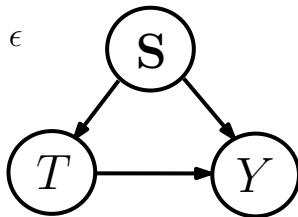
# Identification of a Causal Dose-Response Curve

Without confounding,  $m(t) = \mathbb{E}[Y(t)] = \mathbb{E}(Y|T = t)$ .

- Fitting  $m(t)$  is to regress  $\{Y_i\}_{i=1}^n$  with respect to  $\{T_i\}_{i=1}^n$ .
- Recovering  $\theta(t)$  is a classical derivative estimation problem ([Gasser and Müller, 1984](#)).

$$Y = \mu(T, \mathbf{S}) + \epsilon$$

$$T = f(\mathbf{S}) + E$$



- $E$  is an independent treatment variation with  $\mathbb{E}(E) = 0$ ,
- $\epsilon$  is an exogenous noise with  $\mathbb{E}(\epsilon) = 0$ ,  $\text{Var}(\epsilon) = \sigma^2 > 0$ , and  $\mathbb{E}(\epsilon^4) < \infty$ .



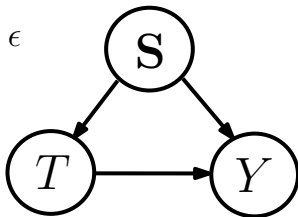
# Identification of a Causal Dose-Response Curve

Without confounding,  $m(t) = \mathbb{E}[Y(t)] = \mathbb{E}(Y|T = t)$ .

- Fitting  $m(t)$  is to regress  $\{Y_i\}_{i=1}^n$  with respect to  $\{T_i\}_{i=1}^n$ .
- Recovering  $\theta(t)$  is a classical derivative estimation problem (Gasser and Müller, 1984).

$$Y = \mu(T, \mathbf{S}) + \epsilon$$

$$T = f(\mathbf{S}) + E$$



- $E$  is an independent treatment variation with  $\mathbb{E}(E) = 0$ ,
- $\epsilon$  is an exogenous noise with  $\mathbb{E}(\epsilon) = 0$ ,  $\text{Var}(\epsilon) = \sigma^2 > 0$ , and  $\mathbb{E}(\epsilon^4) < \infty$ .

► **Solution:** Some identification assumptions are required to estimate  $m(t) = \mathbb{E}[Y(t)]$  and  $\theta(t) = m'(t)$  from  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .

## Assumption

- 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 (Treatment Variation)  $E$  has nonzero variance, i.e.,  $\text{Var}(E) > 0$ .

## Assumption

- 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 (Treatment Variation)  $E$  has nonzero variance, i.e.,  $\text{Var}(E) > 0$ .

Thus,  $m(t)$  and  $\theta(t)$  can be identified through

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

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

# Estimation of Dose-Response Curves Under Positivity

To estimate

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})],$$

we only need to recover  $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$  from  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .

# Estimation of Dose-Response Curves Under Positivity

To estimate

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})],$$

we only need to recover  $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$  from  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .

- ① **Regression Adjustment:**  $\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i)$ , where  $\hat{\mu}$  is any consistent estimator of  $\mu$  (Robins, 1986; Gill and Robins, 2001).
- ② **Inverse Probability Weighting (IPW):**  $\hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i - t}{h}\right)}{\hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot Y_i$  (Hirano and Imbens, 2004; Imai and van Dyk, 2004).
- ③ **Doubly Robust:** Kennedy et al. (2017); Westling et al. (2020); Colangelo and Lee (2020); Semenova and Chernozhukov (2021); Bonvini and Kennedy (2022); Takatsu and Westling (2022).

# Estimation of Dose-Response Curves Under Positivity

To estimate

$$m(t) = \mathbb{E}[Y(t)] = \mathbb{E}[\mu(t, \mathbf{S})],$$

we only need to recover  $\mu(t, \mathbf{s}) = \mathbb{E}(Y|T = t, \mathbf{S} = \mathbf{s})$  from  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .

- 1 **Regression Adjustment:**  $\hat{m}_{\text{RA}}(t) = \frac{1}{n} \sum_{i=1}^n \hat{\mu}(t, \mathbf{S}_i)$ , where  $\hat{\mu}$  is any consistent estimator of  $\mu$  (Robins, 1986; Gill and Robins, 2001).
- 2 **Inverse Probability Weighting (IPW):**  $\hat{m}_{\text{IPW}}(t) = \frac{1}{nh} \sum_{i=1}^n \frac{K\left(\frac{T_i - t}{h}\right)}{\hat{p}_{T|\mathbf{S}}(T_i|\mathbf{S}_i)} \cdot Y_i$  (Hirano and Imbens, 2004; Imai and van Dyk, 2004).
- 3 **Doubly Robust:** Kennedy et al. (2017); Westling et al. (2020); Colangelo and Lee (2020); Semenova and Chernozhukov (2021); Bonvini and Kennedy (2022); Takatsu and Westling (2022).

## Assumption (Positivity or Overlap Condition)

*The conditional density  $p_{T|\mathbf{S}}(t|\mathbf{s})$  is bounded above and away from zero almost surely for all  $t \in \mathcal{T}$  and  $\mathbf{s} \in \mathcal{S}$ .*

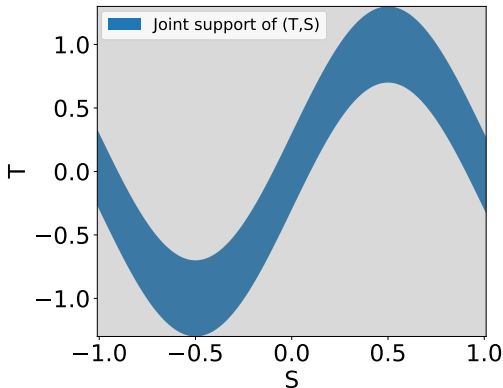
► **Issue:** Positivity is a strong condition with continuous treatments!

# Violation of the Positivity Condition

Consider a single confounder model:

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

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



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

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

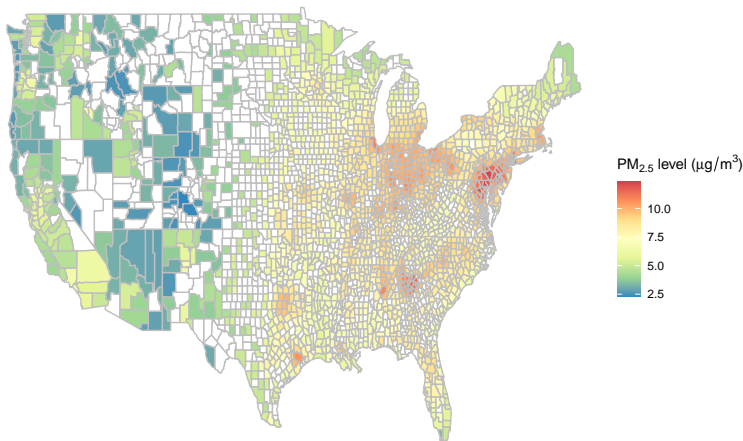


Figure: Average  $PM_{2.5}$  levels over the years 1990-2010 in  $n = 2132$  counties.  $T$  is  $PM_{2.5}$  level, while  $S$  consists of county location and demographic features.

► **Problem:** Only one  $PM_{2.5}$  level is available per county, but causal effects of different  $PM_{2.5}$  levels on county-level CMRs are of interest.



# Highlight of Today's Talk

- 1 The positivity condition may fail to hold in some regions of  $\mathcal{T} \times \mathcal{S}$ .

# Highlight of Today's Talk

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

# Highlight of Today's Talk

- ① The positivity condition may fail to hold in some regions of  $\mathcal{T} \times \mathcal{S}$ .
- ② We propose 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 fundamental theorem of calculus.

# Highlight of Today's Talk

- ① The positivity condition may fail to hold in some regions of  $\mathcal{T} \times \mathcal{S}$ .
- ② We propose 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 fundamental theorem of calculus.
  - Compute the integration via an efficient Riemann sum approximation.
  - $\hat{m}_\theta(t)$  is consistent within any compact set of  $\mathcal{T}$  even when the positivity condition fails in some regions of  $\mathcal{T} \times \mathcal{S}$ .

# Highlight of Today's Talk

- ① The positivity condition may fail to hold in some regions of  $\mathcal{T} \times \mathcal{S}$ .
- ② We propose 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 fundamental theorem of calculus.
  - Compute the integration via an efficient Riemann sum approximation.
  - $\hat{m}_\theta(t)$  is consistent within any compact set of  $\mathcal{T}$  even when the positivity condition fails in some regions of  $\mathcal{T} \times \mathcal{S}$ .
- ③ Nonparametric bootstrap inferences with our estimators on  $m(t)$  and  $\theta(t)$  are asymptotically valid.

# Methodology



# Interchangeability Assumption

Recall our model setup

$$Y = \mu(T, S) + \epsilon \quad \text{and} \quad T = f(S) + E \quad \text{with} \quad S \perp\!\!\!\perp E \quad \text{and} \quad \mathbb{E}(E) = 0.$$

## Assumption (Interchangeability)

*$\mu(t, s)$  is continuously differentiable with respect to  $t$  for any  $(t, s) \in \mathcal{T} \times \mathcal{S}$ , and the following two equalities hold true:*

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

# Interchangeability Assumption

Recall our model setup

$$Y = \mu(T, S) + \epsilon \quad \text{and} \quad T = f(S) + E \quad \text{with} \quad S \perp\!\!\!\perp E \quad \text{and} \quad \mathbb{E}(E) = 0.$$

## Assumption (Interchangeability)

$\mu(t, s)$  is continuously differentiable with respect to  $t$  for any  $(t, s) \in \mathcal{T} \times \mathcal{S}$ , and the following two equalities hold true:

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

► **Note:** Estimating  $\theta(t)$  by the form  $\theta_C(t) = \mathbb{E} \left[ \frac{\partial}{\partial t} \mu(t, S) \middle| T = t \right]$  is our key technique to bypass the positivity condition.



## Example: Additive Confounding Model

Consider the following additive confounding model

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

- This is a common working model in spatial confounding problems (Paciorek, 2010; Schnell and Papadogeorgou, 2020).
- It is also known as the geoaddivitive structural equation model (Kammann and Wand, 2003; Thaden and Kneib, 2018; Wiecha and Reich, 2024).

## Example: Additive Confounding Model

Consider the following additive confounding model

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

- This is a common working model in spatial confounding problems (Paciorek, 2010; Schnell and Papadogeorgou, 2020).
- It is also known as the geoaddivitive structural equation model (Kammann and Wand, 2003; Thaden and Kneib, 2018; Wiecha and Reich, 2024).

### Proposition (Proposition 1 in Zhang et al. 2024)

*Under the additive confounding model,*

- 1  $\bar{m}(t) = m(t).$
- 2  $\theta(t) = \theta_M(t) = \theta_C(t).$
- 3  $\mathbb{E}[\mu(T, S)] = \mathbb{E}[m(T)]$  even when  $\mathbb{E}[\eta(S)] \neq 0.$

## Three Critical Insights

- ①  $\mu(t, s)$  and  $\frac{\partial}{\partial t}\mu(t, s)$  can be consistently estimated at each observed data point  $(T_i, S_i)$ .
  - The positivity condition holds at  $(T_i, S_i)$  for  $i = 1, \dots, n$ .

## Three Critical Insights

- ①  $\mu(t, \mathbf{s})$  and  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  can be consistently estimated at each observed data point  $(T_i, \mathbf{S}_i)$ .
  - The positivity condition holds at  $(T_i, \mathbf{S}_i)$  for  $i = 1, \dots, n$ .
- ②  $\theta(t)$  can be consistently estimated via  $\theta_C(t) = \mathbb{E} \left[ \frac{\partial}{\partial t}\mu(t, \mathbf{S}) \mid T = t \right]$ .
  - Only require an accurate estimator of  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  at the covariate  $\mathbf{s}$  when the conditional density  $p(\mathbf{s}|t)$  is high.

# Three Critical Insights

- ①  $\mu(t, \mathbf{s})$  and  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  can be consistently estimated at each observed data point  $(T_i, S_i)$ .
  - The positivity condition holds at  $(T_i, S_i)$  for  $i = 1, \dots, n$ .
- ②  $\theta(t)$  can be consistently estimated via  $\theta_C(t) = \mathbb{E} \left[ \frac{\partial}{\partial t}\mu(t, \mathbf{S}) \mid T = t \right]$ .
  - Only require an accurate estimator of  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  at the covariate  $\mathbf{s}$  when the conditional density  $p(\mathbf{s}|t)$  is high.
- ③ By the fundamental theorem of calculus,

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

# Three Critical Insights

- ①  $\mu(t, \mathbf{s})$  and  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  can be consistently estimated at each observed data point  $(T_i, S_i)$ .
  - The positivity condition holds at  $(T_i, S_i)$  for  $i = 1, \dots, n$ .
- ②  $\theta(t)$  can be consistently estimated via  $\theta_C(t) = \mathbb{E} \left[ \frac{\partial}{\partial t}\mu(t, \mathbf{S}) \mid T = t \right]$ .
  - Only require an accurate estimator of  $\frac{\partial}{\partial t}\mu(t, \mathbf{s})$  at the covariate  $\mathbf{s}$  when the conditional density  $p(\mathbf{s}|t)$  is high.
- ③ By the fundamental theorem of calculus,

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

$\Rightarrow$  Under our interchangeability assumption,

$$\begin{aligned} m(t) &= \mathbb{E} \left[ m(T) + \int_{\tilde{t}=T}^{\tilde{t}=t} \theta(\tilde{t}) d\tilde{t} \right] = \mathbb{E} [\mu(T, \mathbf{S})] + \mathbb{E} \left[ \int_{\tilde{t}=T}^{\tilde{t}=t} \theta_C(\tilde{t}) d\tilde{t} \right] \\ &= \mathbb{E}(Y) + \mathbb{E} \left[ \int_{\tilde{t}=T}^{\tilde{t}=t} \theta_C(\tilde{t}) d\tilde{t} \right]. \end{aligned}$$

The form  $m(t) = \mathbb{E}(Y) + \mathbb{E} \left[ \int_T^t \theta_C(\tilde{t}) d\tilde{t} \right]$  leads to our proposed *integral estimator* of  $m(t)$  as:

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

where  $\hat{\theta}_C(t)$  is a consistent estimator of

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

The form  $m(t) = \mathbb{E}(Y) + \mathbb{E} \left[ \int_T^t \theta_C(\tilde{t}) d\tilde{t} \right]$  leads to our proposed *integral estimator* of  $m(t)$  as:

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

where  $\hat{\theta}_C(t)$  is a consistent estimator of

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

- Estimate  $\beta_2(t, \mathbf{s}) := \frac{\partial}{\partial t} \mu(t, \mathbf{s})$  by (partial) local polynomial regression (Fan and Gijbels, 1996).
- Fit  $P(\mathbf{s}|t)$  by Nadaraya-Watson conditional cumulative distribution function (CDF) estimator (Hall et al., 1999).



# (Partial) Order $q$ Local Polynomial Regression

- ① Let  $K_T : \mathbb{R} \rightarrow [0, \infty)$ ,  $K_S : \mathbb{R}^d \rightarrow [0, \infty)$  be two symmetric kernel functions and  $h, b > 0$  be their smoothing bandwidth parameters.
- Epanechnikov kernel  $K(u) = \frac{3}{4} (1 - u^2) \cdot \mathbb{1}_{\{|u| \leq 1\}}$  and Gaussian kernel  $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ .
  - Product kernel technique  $K_S(\mathbf{u}) = \prod_{i=1}^d K(u_i)$  for  $\mathbf{u} \in \mathbb{R}^d$ .
- ② Let  $\mathbf{X}_i(t, \mathbf{s}) = (1, (T_i - t), \dots, (T_i - t)^q, (S_{i,1} - s_1), \dots, (S_{i,d} - s_d)) \in \mathbb{R}^{q+1+d}$ ,

$$\mathbf{X}(t, \mathbf{s}) = \begin{pmatrix} \mathbf{X}_1(t, \mathbf{s}) \\ \vdots \\ \mathbf{X}_n(t, \mathbf{s}) \end{pmatrix} \text{ and } \mathbf{W}(t, \mathbf{s}) = \begin{pmatrix} K_T\left(\frac{T_1 - t}{h}\right) K_S\left(\frac{\mathbf{S}_1 - \mathbf{s}}{b}\right) & & \\ & \ddots & \\ & & K_T\left(\frac{T_n - t}{h}\right) K_S\left(\frac{\mathbf{S}_n - \mathbf{s}}{b}\right) \end{pmatrix}.$$

- ③ Solve a weighted least-square problem

$$\begin{aligned} \left(\widehat{\boldsymbol{\beta}}(t, \mathbf{s}), \widehat{\boldsymbol{\alpha}}(t, \mathbf{s})\right)^T &= \arg \min_{(\boldsymbol{\beta}, \boldsymbol{\alpha})^T \in \mathbb{R}^{q+1+d}} \left[ \mathbf{Y} - \mathbf{X}(t, \mathbf{s}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} \right]^T \mathbf{W}(t, \mathbf{s}) \left[ \mathbf{Y} - \mathbf{X}(t, \mathbf{s}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} \right] \\ &= \arg \min_{(\boldsymbol{\beta}, \boldsymbol{\alpha})^T \in \mathbb{R}^{q+1+d}} \sum_{i=1}^n \left[ Y_i - \sum_{j=0}^q \beta_j (T_i - t)^j - \sum_{\ell=1}^d \alpha_\ell (S_{i,\ell} - s_\ell) \right]^2 K_T\left(\frac{T_i - t}{h}\right) K_S\left(\frac{\mathbf{S}_i - \mathbf{s}}{b}\right). \end{aligned}$$

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

With  $\mathbf{Y} = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$ ,

$$\left( \hat{\beta}(t, \mathbf{s}), \hat{\alpha}(t, \mathbf{s}) \right)^T = \left[ \mathbf{X}^T(t, \mathbf{s}) \mathbf{W}(t, \mathbf{s}) \mathbf{X}(t, \mathbf{s}) \right]^{-1} \mathbf{X}(t, \mathbf{s})^T \mathbf{W}(t, \mathbf{s}) \mathbf{Y}.$$

We estimate  $\beta_2(t, \mathbf{s}) := \frac{\partial}{\partial t} \mu(t, \mathbf{s})$  by the second component  $\hat{\beta}_2(t, \mathbf{s})$  of  $\hat{\beta}(t, \mathbf{s}) \in \mathbb{R}^{q+1}$ , where  $q = 2$  is recommended.

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

With  $\mathbf{Y} = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$ ,

$$\left( \hat{\beta}(t, \mathbf{s}), \hat{\alpha}(t, \mathbf{s}) \right)^T = \left[ \mathbf{X}^T(t, \mathbf{s}) \mathbf{W}(t, \mathbf{s}) \mathbf{X}(t, \mathbf{s}) \right]^{-1} \mathbf{X}(t, \mathbf{s})^T \mathbf{W}(t, \mathbf{s}) \mathbf{Y}.$$

We estimate  $\beta_2(t, \mathbf{s}) := \frac{\partial}{\partial t} \mu(t, \mathbf{s})$  by the second component  $\hat{\beta}_2(t, \mathbf{s})$  of  $\hat{\beta}(t, \mathbf{s}) \in \mathbb{R}^{q+1}$ , where  $q = 2$  is recommended.

We fit  $P(\mathbf{s}|t)$  by Nadaraya-Watson conditional CDF estimator

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

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

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

With  $\mathbf{Y} = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$ ,

$$\left( \hat{\beta}(t, s), \hat{\alpha}(t, s) \right)^T = \left[ \mathbf{X}^T(t, s) \mathbf{W}(t, s) \mathbf{X}(t, s) \right]^{-1} \mathbf{X}(t, s)^T \mathbf{W}(t, s) \mathbf{Y}.$$

We estimate  $\beta_2(t, s) := \frac{\partial}{\partial t} \mu(t, s)$  by the second component  $\hat{\beta}_2(t, s)$  of  $\hat{\beta}(t, s) \in \mathbb{R}^{q+1}$ , where  $q = 2$  is recommended.

We fit  $P(s|t)$  by Nadaraya-Watson conditional CDF estimator

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

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

## ► Proposed Localized Derivative Estimator:

$$\hat{\theta}_C(t) = \int \hat{\beta}_2(t, s) d\hat{P}_h(s|t) = \frac{\sum_{i=1}^n \hat{\beta}_2(t, \mathbf{s}_i) \cdot \bar{K}_T\left(\frac{T_i - t}{h}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j - t}{h}\right)}.$$

# A Fast Computing Algorithm for Proposed Integral Estimator

Our *integral estimator* takes the form

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

► **Issue:** The integral could be analytically difficult to compute.

# A Fast Computing Algorithm for Proposed Integral Estimator

Our *integral estimator* takes the form

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

► **Issue:** The integral could be analytically difficult to compute.

► **Solution:** 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:

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

# A Fast Computing Algorithm for Proposed Integral Estimator

Our *integral estimator* takes the form

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

► **Issue:** The integral could be analytically difficult to compute.

► **Solution:** 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  $O_P\left(\frac{1}{n}\right)$ .

# Nonparametric Bootstrap Inference

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



# 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  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .
- ② Generate  $B$  bootstrap samples  $\left\{ \left( Y_i^{*(b)}, T_i^{*(b)}, \mathbf{S}_i^{*(b)} \right) \right\}_{i=1}^n$  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  $\{(Y_i, T_i, \mathbf{S}_i)\}_{i=1}^n$ .
- ② Generate  $B$  bootstrap samples  $\left\{ \left( Y_i^{*(b)}, T_i^{*(b)}, \mathbf{S}_i^{*(b)} \right) \right\}_{i=1}^n$  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:

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

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

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

# Asymptotic Theory



## (Uniform) Consistencies of Proposed Estimators

Let  $\mathcal{T}' \subset \mathcal{T}$  be a compact set so that  $p_T(t) \geq p_{T,\min} > 0$  for all  $t \in \mathcal{T}'$ .

Assume

- smoothness conditions on  $p(t, s)$  and  $\mu(t, s)$ ,
- boundary conditions on  $\mathcal{E} \subset \mathcal{T} \times \mathcal{S}$ , which is the support of  $p(t, s)$ ,
- regular and VC-type conditions on the kernel functions  $K_T, K_S, \bar{K}_T$ .

# (Uniform) Consistencies of Proposed Estimators

Let  $\mathcal{T}' \subset \mathcal{T}$  be a compact set so that  $p_T(t) \geq p_{T,\min} > 0$  for all  $t \in \mathcal{T}'$ .

Assume

- smoothness conditions on  $p(t, s)$  and  $\mu(t, s)$ ,
- boundary conditions on  $\mathcal{E} \subset \mathcal{T} \times \mathcal{S}$ , which is the support of  $p(t, s)$ ,
- regular and VC-type conditions on the kernel functions  $K_T, K_S, \bar{K}_T$ .

Then, as  $h, b, \hbar, \frac{\max\{h, b\}^4}{h} \rightarrow 0$  and  $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log \log n}, \frac{n\hbar}{|\log \hbar|}, \frac{|\log \hbar|}{\log \log n} \rightarrow \infty$ ,

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

and

$$\begin{aligned} \sup_{t \in \mathcal{T}'} |\hat{m}_\theta(t) - m(t)| &= O_P\left(\frac{1}{\sqrt{n}}\right) + O\left(h^q + b^2 + \frac{\max\{h, b\}^4}{h}\right) \\ &\quad + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}} + \hbar^2 + \sqrt{\frac{|\log \hbar|}{n\hbar}}\right). \end{aligned}$$

# Asymptotic Linearity of Proposed Estimators

Under the same regularity conditions, if  $h \asymp b \asymp n^{-\frac{1}{\gamma}}$  and  $\hbar \asymp n^{-\frac{1}{\varpi}}$  for some  $\gamma \geq \varpi > 0$  such that  $\frac{nh^{d+5}}{\log n} \rightarrow c_1$  and  $\frac{n\hbar^5}{\log n} \rightarrow c_2$  for some  $c_1, c_2 \geq 0$  and  $\frac{\log n}{nh^2}, \frac{h^{d+3} \log n}{\hbar}, \frac{h^{d+3}}{\hbar^2} \rightarrow 0$  as  $n \rightarrow \infty$ , then for any  $t \in \mathcal{T}'$ ,

$$\sqrt{nh^3b^d} \left[ \widehat{\theta}_C(t) - \theta_C(t) \right] = \mathbb{G}_n \bar{\varphi}_t + o_P(1),$$

$$\sqrt{nh^3b^d} \left[ \widehat{m}_\theta(t) - m(t) \right] = \mathbb{G}_n \varphi_t + o_P(1),$$

where

$$\bar{\varphi}_t(Y, T, \mathbf{S}) = \mathbb{E}_{(T_{i_3}, \mathbf{S}_{i_3})} \left[ \frac{\mathbf{e}_2^T \mathbf{M}_q^{-1} \boldsymbol{\Psi}_{t, \mathbf{S}_{i_3}}(Y, T, \mathbf{S})}{\sqrt{hb^d} \cdot p(t, \mathbf{S}_{i_3}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T \left( \frac{t - T_{i_3}}{\hbar} \right) \right]$$

and  $\varphi_t(Y, T, \mathbf{S}) = \mathbb{E}_{T_{i_2}} \left[ \int_{T_{i_2}}^t \bar{\varphi}_{\tilde{t}}(Y, T, \mathbf{S}) d\tilde{t} \right]$  with  $\mathbb{G}_n = \sqrt{n} (\mathbb{P}_n - \mathbb{P})$ .

- Note that  $\bar{\varphi}_t$  and  $\varphi_t$  may not be efficient influence functions.

# High-Level Proof of Asymptotic Linearity

Define

$$\mathbf{M}_q = \begin{pmatrix} \left( \kappa_{i+j-2}^{(T)} \right)_{1 \leq i, j \leq q+1} & \mathbf{0} \\ \mathbf{0} & \left( \kappa_{2, i-q-1}^{(S)} \mathbb{1}_{\{i=j\}} \right)_{q+1 < i, j \leq q+1+d} \end{pmatrix} \in \mathbb{R}^{(q+1+d) \times (q+1+d)}$$

and the function  $\Psi_{t,s}, \psi_{t,s} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^{q+1+d}$  as:

$$\Psi_{t,s}(y, z, v) = \begin{bmatrix} \left( y \cdot \left( \frac{z-t}{h} \right)^{j-1} K_T \left( \frac{z-t}{h} \right) K_S \left( \frac{v-s}{b} \right) \right)_{1 \leq j \leq q+1} \\ \left( y \cdot \left( \frac{v_{j-q-1}-s_{j-q-1}}{b} \right) K_T \left( \frac{z-t}{h} \right) K_S \left( \frac{v-s}{b} \right) \right)_{q+1 < j \leq q+1+d} \end{bmatrix}.$$

► **Key Argument:** Write  $\hat{m}_\theta(t) - m(t)$  into a V-statistic ([Shieh, 2014](#))

$$\begin{aligned} & \hat{m}_\theta(t) - m(t) \\ &= \frac{1}{n^3} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \int_{T_{i_1}} \frac{e_2^T \mathbf{M}_q^{-1} \Psi_{\tilde{t}, s_{i_2}}(Y_{i_3}, T_{i_3}, \mathbf{S}_{i_3})}{h^2 b^d \cdot p(\tilde{t}, \mathbf{S}_{i_2}) \cdot p_T(\tilde{t})} \cdot \frac{1}{h} \bar{K}_T \left( \frac{\tilde{t} - T_{i_2}}{h} \right) d\tilde{t} - \mathbb{E} \left[ \int_T \theta_C(\tilde{t}) d\tilde{t} \right] \\ &+ O_P \left( \frac{1}{\sqrt{n}} + h^2 + \sqrt{\frac{\log n}{nh}} \right). \end{aligned}$$



# Bootstrap Consistency

Under the same regularity conditions, if  $h \asymp b \asymp n^{-\frac{1}{\gamma}}$  and  $\hbar \asymp n^{-\frac{1}{\varpi}}$  for some  $\gamma \geq \varpi > 0$  such that  $\frac{nh^{d+5}}{\log n} \rightarrow c_1$  and  $\frac{n\hbar^5}{\log n} \rightarrow c_2$  for some  $c_1, c_2 \geq 0$  and  $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3} \log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \rightarrow \infty$  as  $n \rightarrow \infty$ ,

1

$$\begin{aligned} & \left| \sqrt{nh^3 b^d} \sup_{t \in \mathcal{T}'} |\hat{m}_\theta(t) - m(t)| - \sup_{t \in \mathcal{T}'} |\mathbb{G}_n \varphi_t| \right| \\ &= O_P \left( \sqrt{nh^{d+7}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{\frac{h^{d+3} \log n}{\hbar}} + \sqrt{\frac{h^{d+3}}{\hbar^2}} \right). \end{aligned}$$

# Bootstrap Consistency

Under the same regularity conditions, if  $h \asymp b \asymp n^{-\frac{1}{\gamma}}$  and  $\hbar \asymp n^{-\frac{1}{\varpi}}$  for some  $\gamma \geq \varpi > 0$  such that  $\frac{nh^{d+5}}{\log n} \rightarrow c_1$  and  $\frac{n\hbar^5}{\log n} \rightarrow c_2$  for some  $c_1, c_2 \geq 0$  and  $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3} \log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \rightarrow \infty$  as  $n \rightarrow \infty$ ,

1

$$\begin{aligned} & \left| \sqrt{nh^3 b^d} \sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| - \sup_{t \in \mathcal{T}'} |\mathbb{G}_n \varphi_t| \right| \\ &= O_P \left( \sqrt{nh^{d+7}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{\frac{h^{d+3} \log n}{\hbar}} + \sqrt{\frac{h^{d+3}}{\hbar^2}} \right). \end{aligned}$$

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

$$\sup_{u \geq 0} \left| \mathbb{P} \left( \sqrt{nh^3 b^d} \sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| \leq u \right) - \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\mathbb{B}(f)| \leq u \right) \right| = O \left( \left( \frac{\log^5 n}{nh^{d+3}} \right)^{\frac{1}{8}} \right).$$

# Bootstrap Consistency

Under the same regularity conditions, if  $h \asymp b \asymp n^{-\frac{1}{\gamma}}$  and  $\hbar \asymp n^{-\frac{1}{\varpi}}$  for some  $\gamma \geq \varpi > 0$  such that  $\frac{nh^{d+5}}{\log n} \rightarrow c_1$  and  $\frac{n\hbar^5}{\log n} \rightarrow c_2$  for some  $c_1, c_2 \geq 0$  and  $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3} \log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \rightarrow \infty$  as  $n \rightarrow \infty$ ,

1

$$\begin{aligned} & \left| \sqrt{nh^3 b^d} \sup_{t \in \mathcal{T}'} |\hat{m}_\theta(t) - m(t)| - \sup_{t \in \mathcal{T}'} |\mathbb{G}_n \varphi_t| \right| \\ &= O_P \left( \sqrt{nh^{d+7}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{\frac{h^{d+3} \log n}{\hbar}} + \sqrt{\frac{h^{d+3}}{\hbar^2}} \right). \end{aligned}$$

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

$$\sup_{u \geq 0} \left| \mathbb{P} \left( \sqrt{nh^3 b^d} \sup_{t \in \mathcal{T}'} |\hat{m}_\theta(t) - m(t)| \leq u \right) - \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\mathbb{B}(f)| \leq u \right) \right| = O \left( \left( \frac{\log^5 n}{nh^{d+3}} \right)^{\frac{1}{8}} \right).$$

3

$$\sup_{u \geq 0} \left| \mathbb{P} \left( \sqrt{nh^3 b^d} \cdot \sup_{t \in \mathcal{T}'} |\hat{m}_\theta^*(t) - \hat{m}_\theta(t)| \leq u \mid \mathbb{U}_n \right) - \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\mathbb{B}(f)| \leq u \right) \right| = O_P \left( \left( \frac{\log^5 n}{nh^{d+3}} \right)^{\frac{1}{8}} \right)$$

with

$$\mathcal{F} = \{(v, x, z) \mapsto \varphi_t(v, x, z) : t \in \mathcal{T}'\}.$$

- ①  $\mathcal{F}$  is not Donsker because  $\varphi_t$  is not uniformly bounded as  $h \rightarrow 0$ .
  - However,  $\tilde{\mathcal{F}} = \left\{ (v, x, z) \mapsto \sqrt{h^3 b^d} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}' \right\}$  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  $\varphi_t$  is not uniformly bounded as  $h \rightarrow 0$ .
  - However,  $\tilde{\mathcal{F}} = \left\{ (v, x, z) \mapsto \sqrt{h^3 b^d} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}' \right\}$  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  $\text{Var}(\epsilon) = \sigma^2 > 0$ ,  $\text{Var} [\varphi_t(Y, T, S)]$  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  $\varphi_t$  is not uniformly bounded as  $h \rightarrow 0$ .
  - However,  $\tilde{\mathcal{F}} = \left\{ (v, x, z) \mapsto \sqrt{h^3 b^d} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}' \right\}$  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  $\text{Var}(\epsilon) = \sigma^2 > 0$ ,  $\text{Var} [\varphi_t(Y, T, S)]$  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  $h \asymp b = O\left(n^{-\frac{1}{d+5}}\right)$  and  $\tilde{h} = O\left(n^{-\frac{1}{5}}\right)$ .
  - They match the outputs by the usual bandwidth selection methods ([Bashtannyk and Hyndman, 2001](#); [Li and Racine, 2004](#)).
  - No explicit undersmoothing is required!!

# Simulations and Case Study



- 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  $\bar{h} > 0$  to the normal reference rule in [Chacón et al. \(2011\)](#); [Chen et al. \(2016\)](#).
- Set the bootstrap resampling time  $B = 1000$  and the significance level  $\alpha = 0.05$ .
- Compare our proposed estimators with the regression adjustment estimators under the same choices of bandwidth parameters:

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

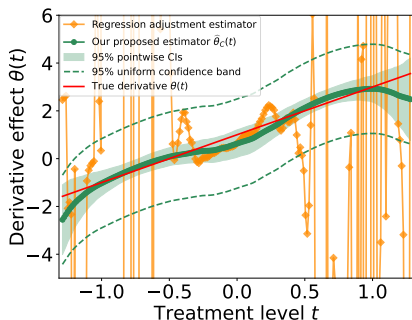
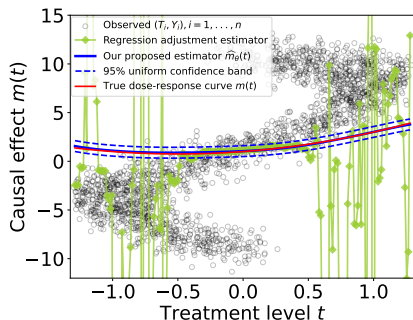


# Single Confounder Model

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.

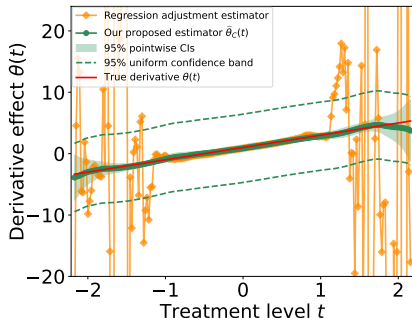
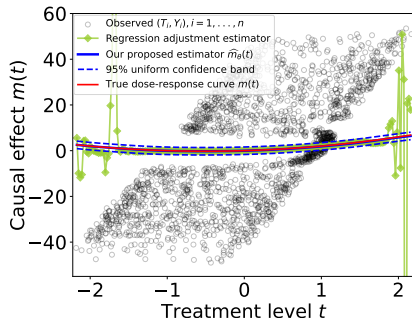


# Nonlinear Confounding Model

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

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

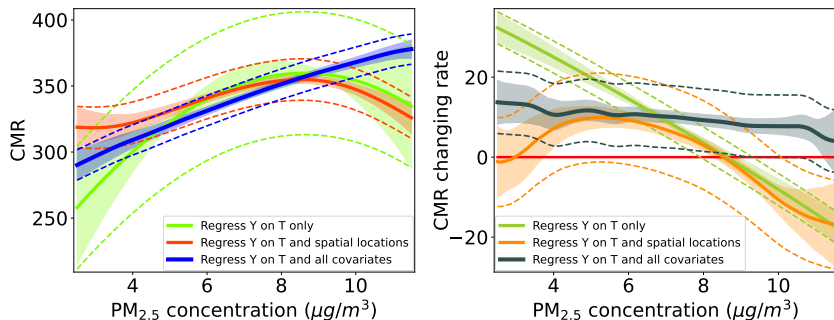
- $(S_1, S_2) \sim \text{Uniform}[-1, 1]^2$ ,  $E \sim \text{Uniform}[-0.1, 0.1]$ , and  $\epsilon \sim \mathcal{N}(0, 1)$ .
- Methods based on pseudo-outcomes ([Kennedy et al., 2017](#); [Takatsu and Westling, 2022](#)) does not work in this example.



# Effect of $\text{PM}_{2.5}$ on the Cardiovascular Mortality Rate (CMR)

- ① Recent studies identify a positive association between  $\text{PM}_{2.5}$  level ( $\mu\text{g}/\text{m}^3$ ) and county-level CMR (deaths/100,000 person-years) in the U.S. after controlling for socioeconomic factors (Wyatt et al., 2020a).
- ② Obtain the average annual CMR as  $Y$  and  $\text{PM}_{2.5}$  concentration as  $T$  over years 1990-2010 within  $n = 2132$  U.S. counties from Wyatt et al. (2020b).
- ③ The covariate vector  $S \in \mathbb{R}^{10}$  consists of two parts:
  - Two spatial confounding variables, *i.e.*, latitude and longitude of each county.
  - Eight county-level socioeconomic factors acquired from the US census.
- ④ Focus on the values of  $\text{PM}_{2.5}$  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)



After adjusting for all the available confounding variables,

- the estimated relationship between  $PM_{2.5}$  and CMR becomes monotonically increasing;
- the 95% confidence band of the estimated changing rate of CMR is unanimously above 0 when the  $PM_{2.5}$  level is below  $9 \mu g/m^3$ .

# Discussion



## Summary and Future Works

We study nonparametric inference on dose-response curves and their derivative functions.

- Propose an integral estimator of  $m(t)$  and a localized derivative estimator of  $\theta(t)$ .
- Both estimators are consistent without the positivity condition.

# Summary and Future Works

We study nonparametric inference on dose-response curves and their derivative functions.

- Propose an integral estimator of  $m(t)$  and a localized derivative estimator of  $\theta(t)$ .
- Both estimators are consistent without the positivity condition.

## ► Future Directions:

- 1 Better estimates of the nuisance functions  $\frac{\partial}{\partial t}\mu(t, s)$  and  $P(s|t)$ :
  - Bandwidth selection via the plug-in rule (Ruppert et al., 1995) or cross-validation (Li and Racine, 2004).
  - Regression splines for  $\frac{\partial}{\partial t}\mu(t, s)$  (Friedman, 1991; Zhou and Wolfe, 2000) and local logistic approaches for  $P(s|t)$  (Hall et al., 1999).

# Summary and Future Works

We study nonparametric inference on dose-response curves and their derivative functions.

- Propose an integral estimator of  $m(t)$  and a localized derivative estimator of  $\theta(t)$ .
- Both estimators are consistent without the positivity condition.

## ► Future Directions:

- 1 Better estimates of the nuisance functions  $\frac{\partial}{\partial t}\mu(t, s)$  and  $P(s|t)$ :
  - Bandwidth selection via the plug-in rule (Ruppert et al., 1995) or cross-validation (Li and Racine, 2004).
  - Regression splines for  $\frac{\partial}{\partial t}\mu(t, s)$  (Friedman, 1991; Zhou and Wolfe, 2000) and local logistic approaches for  $P(s|t)$  (Hall et al., 1999).
- 2 Generalize our proposed integral estimators to the IPW and doubly robust variants.
- 3 Sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022) and the interchangeability assumption.



- ④ Study the semi-parametric efficiency of the influence functions from our proposed estimators:

$$\bar{\varphi}_t(Y, T, S) = \mathbb{E}_{(T_{i_3}, S_{i_3})} \left[ \frac{\mathbf{e}_2^T \mathbf{M}_q^{-1} \Psi_{t, S_{i_3}}(Y, T, S)}{\sqrt{h b^d} \cdot p(t, S_{i_3}) \cdot p_T(t)} \cdot \frac{1}{h} \bar{K}_T \left( \frac{t - T_{i_3}}{h} \right) \right]$$

$$\text{and } \varphi_t(Y, T, S) = \mathbb{E}_{T_{i_2}} \left[ \int_{T_{i_2}}^t \bar{\varphi}_t(Y, T, S) d\tilde{t} \right].$$

- 4 Study the semi-parametric efficiency of the influence functions from our proposed estimators:

$$\bar{\varphi}_t(Y, T, \mathbf{S}) = \mathbb{E}_{(T_{i_3}, \mathbf{S}_{i_3})} \left[ \frac{\mathbf{e}_2^T \mathbf{M}_q^{-1} \boldsymbol{\Psi}_{t, \mathbf{S}_{i_3}}(Y, T, \mathbf{S})}{\sqrt{h b^d} \cdot p(t, \mathbf{S}_{i_3}) \cdot p_T(t)} \cdot \frac{1}{h} \bar{K}_T \left( \frac{t - T_{i_3}}{h} \right) \right]$$

and  $\varphi_t(Y, T, \mathbf{S}) = \mathbb{E}_{T_{i_2}} \left[ \int_{T_{i_2}}^t \bar{\varphi}_{\tilde{t}}(Y, T, \mathbf{S}) d\tilde{t} \right]$ .

- 5 Our proposed nonparametric estimators suffer from the curse of dimensionality.
- Impose a semi-parametric additive model ([Guo et al., 2019](#)) as:

$$\mathbb{E}(Y|T=t, \mathbf{S}=\mathbf{s}, \mathbf{Z}=\mathbf{z}) = m(t) + \eta(\mathbf{s}) + \sum_{j=1}^{d'} g_j(\mathbf{z}_j),$$

where  $\mathbf{Z} \in \mathbb{R}^{d'}$  is a high-dimensional covariate vector.

# 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>.

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

# Reference

- D. M. Bashtannyk and R. J. Hyndman. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298, 2001.
- M. Bonvini and E. H. Kennedy. Fast convergence rates for dose-response estimation. *arXiv preprint arXiv:2207.11825*, 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. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67, 1991.
- T. Gasser and H.-G. Müller. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, pages 171–185, 1984.
- 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.

# Reference

- 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.
- 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.
- 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.
- C. J. Paciorek. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1):107–125, 2010.
- J. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12): 1393–1512, 1986.
- D. Ruppert, S. J. Sheather, and M. P. Wand. An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270, 1995.
- 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.

- V. Semenova and V. Chernozhukov. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2021.
- G. S. Shieh. U-and V-statistics. *Wiley StatsRef: Statistics Reference Online*, 2014.
- 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.
- 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.
- N. Wiecha and B. J. Reich. Two-stage spatial regression models for spatial confounding. *arXiv preprint arXiv:2404.09358*, 2024.
- 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.
- 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, Y.-C. Chen, and A. Giessing. Nonparametric inference on dose-response curves without the positivity condition. *arXiv preprint arXiv:2405.09003*, 2024.
- S. Zhou and D. A. Wolfe. On derivative estimation in spline regression. *Statistica Sinica*, 10(1):93–108, 2000.

► **Question:** Why is it necessary for  $\text{Var}(E) > 0$ ?

- Suppose that  $\text{Var}(E) = 0$ , and we take

$$T = f(\mathbf{S}) + E \stackrel{\text{a.s.}}{=} S_1 \quad \text{with} \quad \mathbb{E}(S_1) = 0.$$

- Consider two equal conditional mean outcome functions

$$\mu_1(T, \mathbf{S}) = T + 2S_1 \stackrel{\text{a.s.}}{=} 3S_1 \quad \text{and} \quad \mu_2(T, \mathbf{S}) = 2T + S_1 \stackrel{\text{a.s.}}{=} 3S_1.$$

- However,  $\mu_1, \mu_2$  lead to two distinct treatment effects:

$$m_1(t) = \mathbb{E}[\mu_1(t, \mathbf{S})] = t \quad \text{and} \quad m_2(t) = \mathbb{E}[\mu_2(t, \mathbf{S})] = 2t.$$

# Regularity Assumptions (Smoothness Conditions)

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

- ① For any  $(t, s) \in \mathcal{T} \times \mathcal{S}$ ,  $\mu(t, s)$  is at least  $(q + 1)$  times continuously differentiable with respect to  $t$  and at least four times continuously differentiable with respect to  $s$ . Furthermore,  $\mu(t, s)$  and all of its partial derivatives are uniformly bounded on  $\mathcal{T} \times \mathcal{S}$ .
- ②  $p(t, s)$  is bounded and at least twice continuously differentiable with bounded partial derivatives up to the second order on  $\mathcal{E}^\circ$ . 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  $p_T(t)$  is non-degenerate.



## Regularity Assumptions (Boundary Conditions)

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

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

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.

- 4 For any  $(t, \mathbf{s}) \in \partial\mathcal{E}$ , the boundary of  $\mathcal{E}$ , 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  $j = 1, \dots, d$ .
- 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

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

# 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  $\int s \cdot K_S(s) ds = \mathbf{0}$ . In addition, for all  $j = 1, 2, \dots$ , and  $\ell = 1, \dots, d$ ,

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

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  $\bar{\mathcal{K}} = \left\{ y \mapsto \bar{K}_T\left(\frac{y-t}{h}\right) : t \in \mathcal{T}, h > 0 \right\}$ . 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

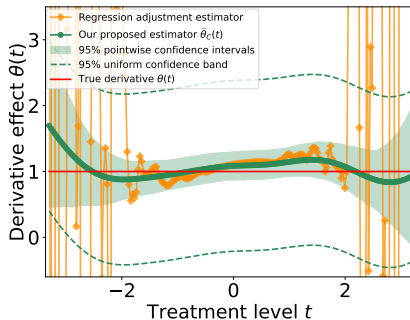
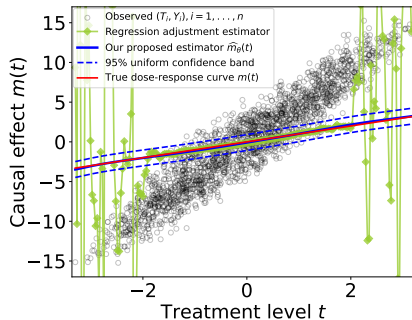
$$\left[ \int_{\mathbb{R}^{d+1}} [G(x)]^2 dQ(x) \right]^{\frac{1}{2}}.$$

# Linear Confounding Model

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



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

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

$$\mu(f(s), s) = \bar{m}(f(s)) + \eta(s) = m(f(s)) + \eta(s). \quad (1)$$

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

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

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

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

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  $\rho_1$  in the previous assumption.