Nonparametric Inference on Dose-Response Curves Without the Positivity Condition

Yikun Zhang[1]

Joint work with *Yen-Chi Chen*[1] and *Alexander Giessing*[2]

[1]Department of Statistics, University of Washington
[2]Department of Statistics and Data Science, National University of Singapore

Causal Inference and Missing Data Reading Group
November 4, 2024

Department of
STATISTICS

UNIVERSITY of WASHINGTON

Introduction

A Central Problem in Causal Inference:

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

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

[1] 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[1] 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.*, $\mathcal{T} \subset \mathbb{R}$)?

[1] 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[1] 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.

[1] Here, $Y(t)$ is the potential outcome that would have been observed under treatment level $T = t$.

Identification of Dose-Response Curves

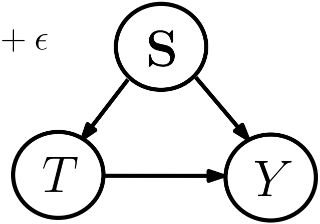
Without confounding, $m(t) = \mathbb{E}\left[Y(t)\right] = \mathbb{E}\left(Y|T = t\right)$.

- 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).

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

- 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).

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

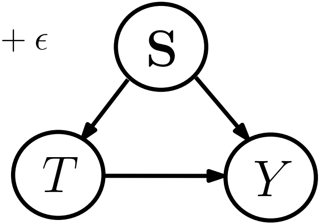


- E is an independent treatment variation with $\mathbb{E}(E) = 0$,
- ϵ is an exogenous noise with $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 > 0$, and $\mathbb{E}(\epsilon^4) < \infty$.

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

- 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 $\mathbb{E}(E) = 0$,
- ϵ 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 $m(t) = \mathbb{E}\left[Y(t)\right]$ and $\theta(t) = m'(t)$ from $\{(Y_i, T_i, S_i)\}_{i=1}^n$.

Identification of Dose-Response Curves

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given any* $\boldsymbol{S} = \boldsymbol{s} \in \mathcal{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$.

Identification of Dose-Response Curves

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given any* $\boldsymbol{S} = \boldsymbol{s} \in \mathcal{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$.

- **Question:** Why is it necessary for $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$ for all $\boldsymbol{s} \in \mathcal{S}$?

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given any* $\boldsymbol{S} = \boldsymbol{s} \in \mathcal{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$.

- **Question:** Why is it necessary for $\mathrm{Var}(T|\boldsymbol{S} = \boldsymbol{s}) > 0$ for all $\boldsymbol{s} \in \mathcal{S}$?

- Consider the following example with $\mathrm{Var}(T|\boldsymbol{S}) = 0$ as:

$$T = f(\boldsymbol{S}, E) = S_1 \quad \text{and} \quad \mathbb{E}(S_1) = 0.$$

- Let $Y = T + 2S_1 + \epsilon = 3S_1 + \epsilon$ and $\widetilde{Y} = 2T + S_1 + \widetilde{\epsilon} = 3S_1 + \widetilde{\epsilon}$. Then,

$$\mathbb{E}\left(Y|T = t, \boldsymbol{S} = \boldsymbol{s}\right) = 3s_1 = \mathbb{E}\left(\widetilde{Y}\middle|T = t, \boldsymbol{S} = \boldsymbol{s}\right).$$

- However,

$$m(t) = \mathbb{E}\left[Y(t)\right] = t \quad \text{and} \quad \widetilde{m}(t) = \mathbb{E}\left[\widetilde{Y}(t)\right] = 2t.$$

Identification of Dose-Response Curves Under Positivity

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given* $\boldsymbol{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S}) > 0$.

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

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given* $\boldsymbol{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S}) > 0$.

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

However, in order for $\mu(t, \boldsymbol{s}) = \mathbb{E}\left(Y|T = t, \boldsymbol{S} = \boldsymbol{s}\right)$ to be well-defined on $\mathcal{T} \times \mathcal{S}$, we need the positivity condition.

Assumption (Positivity or Overlap Condition)

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

Assumption

1. *(Consistency)* $Y = Y(t)$ *whenever* $T = t \in \mathcal{T}$.
2. *(Ignorability or Unconfoundedness)* $Y(t) \perp\!\!\!\perp T \,\big|\, \boldsymbol{S}$ *for all* $t \in \mathcal{T}$.
3. *(Treatment Variation) The conditional variance of* T *given* $\boldsymbol{S}$ *is strictly positive, i.e.,* $\mathrm{Var}(T|\boldsymbol{S}) > 0$.
4. *(Positivity) The conditional density* $p(t|\boldsymbol{s})$ *is bounded away from zero almost surely for all* $t \in \mathcal{T}$ *and* $\boldsymbol{s} \in \mathcal{S}$.

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

$$\begin{cases} m(t) = \mathbb{E}\left[Y(t)\right] = \mathbb{E}\left[\mu(t,\boldsymbol{S})\right], \\ \theta(t) = \frac{d}{dt}\mathbb{E}\left[Y(t)\right] = \frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})\right] \overset{(\star)^2}{=} \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right], \end{cases}$$

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

[2] For $(\star)$, we only need some mild assumption; see Theorem 1.1 in Shao (2003).

To estimate
$$m(t) = \mathbb{E}\left[Y(t)\right] = \mathbb{E}\left[\mu(t, \boldsymbol{S})\right],$$
we only need to recover $\mu(t, \boldsymbol{s}) = \mathbb{E}\left(Y | T = t, \boldsymbol{S} = \boldsymbol{s}\right)$ from $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$.

To estimate

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

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

1. **Regression Adjustment:** $\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i)$, where $\widehat{\mu}$ is any consistent estimator of μ (Robins, 1986; Gill and Robins, 2001).

To estimate

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

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

1. **Regression Adjustment:** $\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i)$, where $\widehat{\mu}$ is any consistent estimator of μ (Robins, 1986; Gill and Robins, 2001).
2. **Inverse Probability Weighting (IPW):** $\widehat{m}_{\mathrm{IPW}}(t) = \frac{1}{nh}\sum_{i=1}^n \frac{K\left(\frac{T_i - t}{h}\right)}{\widehat{p}_{T|S}(T_i|\boldsymbol{S}_i)} \cdot Y_i$ (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).

To estimate

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

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

1. **Regression Adjustment:** $\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i)$, where $\widehat{\mu}$ is any consistent estimator of μ (Robins, 1986; Gill and Robins, 2001).

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

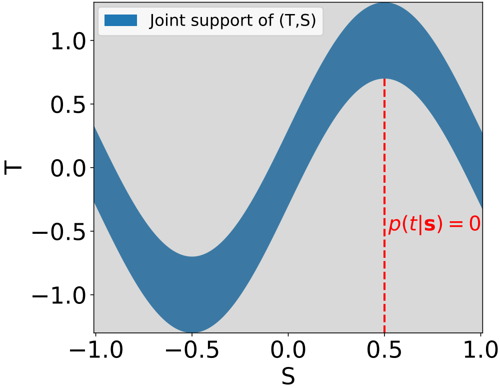
► **Issue:** Positivity is a very strong assumption with continuous treatments!

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)

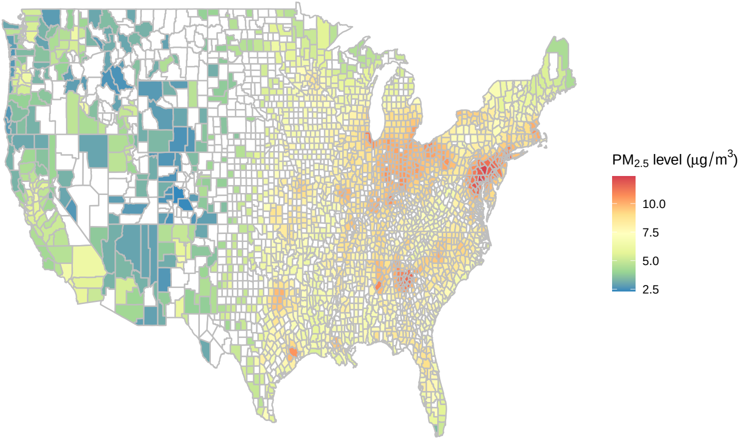


Figure: Average $PM_{2.5}$ levels from 1990 to 2010 in $n = 2132$ counties. T is $PM_{2.5}$ level, while S consists of the county location and some socioeconomic factors.

▶ **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.

1. The positivity condition may fail to hold in some regions of $\mathcal{T} \times \mathcal{S}$.
   - Identify $m(t)$ through an identification assumption on $\theta(t) = m'(t)$.

Highlight of Today's Talk

1. The positivity condition may fail to hold in some regions of $\mathcal{T} \times \mathcal{S}$.
   - Identify $m(t)$ through an identification assumption on $\theta(t) = m'(t)$.
2. We propose a novel integral estimator $\widehat{m}_\theta(t)$ 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 $\mathcal{T} \times \mathcal{S}$.
   - Identify $m(t)$ through an identification assumption on $\theta(t) = m'(t)$.
2. We propose 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 $T_i, i = 1, ..., n$.
   - Extrapolate $\widehat{\theta}_C(t)$ 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 $\mathcal{T} \times \mathcal{S}$.
   - Identify $m(t)$ through an identification assumption on $\theta(t) = m'(t)$.
2. We propose 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 $T_i, i = 1, ..., n$.
   - Extrapolate $\widehat{\theta}_C(t)$ to any treatment level of interest via the fundamental theorem of calculus.
   - $\widehat{m}_\theta(t)$ 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 $\mathcal{T} \times \mathcal{S}$.
   - Identify $m(t)$ through an identification assumption on $\theta(t) = m'(t)$.
2. We propose 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 $T_i, i = 1, ..., n$.
   - Extrapolate $\widehat{\theta}_C(t)$ to any treatment level of interest via the fundamental theorem of calculus.
   - $\widehat{m}_\theta(t)$ 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

Identification Condition for $\theta(t)$

Assumption (Interchangeability)

$\mathbb{E}[Y(t)|\boldsymbol{S}=\boldsymbol{s}]$ is continuously differentiable with respect to t for any $(t, \boldsymbol{s})$ such that $p(\boldsymbol{s}|t) > 0$, and the following two equalities hold true:

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

Identification Condition for $\theta(t)$

Assumption (Interchangeability)

$\mathbb{E}\left[Y(t)|\boldsymbol{S}=\boldsymbol{s}\right]$ is continuously differentiable with respect to t for any $(t,\boldsymbol{s})$ such that $p(\boldsymbol{s}|t)>0$, and the following two equalities hold true:

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

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

Identification Condition for $\theta(t)$

Assumption (Interchangeability)

$\mathbb{E}\left[Y(t)|\boldsymbol{S}=\boldsymbol{s}\right]$ is continuously differentiable with respect to t for any $(t,\boldsymbol{s})$ such that $p(\boldsymbol{s}|t)>0$, and the following two equalities hold true:

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

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

- Estimating $\theta(t)$ by $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\Big|T=t\right]$ is our key technique to bypass the positivity condition.

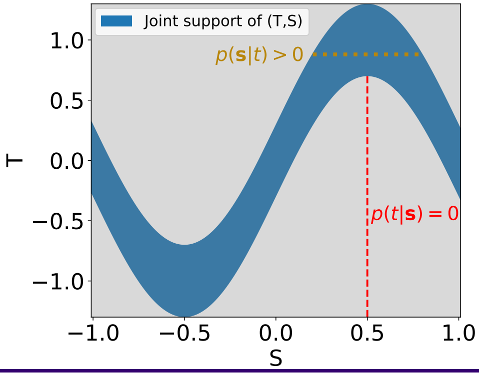
Identification Condition for $\theta(t)$

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

- Different from the identification via $\theta(t) = \frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right]$ under the positivity condition, we only need

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

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



Example: Additive Confounding Model

Consider the following additive confounding model

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

- This is a common working model in spatial confounding problems (Paciorek, 2010; Schnell and Papadogeorgou, 2020).
- It is also known as the geoadditive 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

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

- This is a common working model in spatial confounding problems (Paciorek, 2010; Schnell and Papadogeorgou, 2020).
- It is also known as the geoadditive 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. $\mathbb{E}\left[\mu(T,\boldsymbol{S})\right] = \mathbb{E}\left[m(T)\right]$ *even when* $\mathbb{E}\left[\eta(\boldsymbol{S})\right] \neq 0$.

Three Critical Insights

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

Three Critical Insights

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

2. $\theta(t)$ can be consistently estimated via $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S}) \middle| T = t\right]$.
   - Only require an accurate estimator of $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ at the covariate $\boldsymbol{s}$ when the conditional density $p(\boldsymbol{s}|t)$ is high.

Three Critical Insights

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

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

Three Critical Insights

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

2. $\theta(t)$ can be consistently estimated via $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S}) \middle| T = t\right]$.
   - Only require an accurate estimator of $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ at the covariate $\boldsymbol{s}$ when the conditional density $p(\boldsymbol{s}|t)$ is high.

3. By the fundamental theorem of calculus,

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

$\Longrightarrow$ Under our identification assumption for $\theta(t)$,

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

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:

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

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

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

Proposed Integral Estimator of Dose-Response Curve

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:

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

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

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

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

(Partial) Order q Local Polynomial Regression

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

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

3. Solve a weighted least-square problem

$$\left(\widehat{\boldsymbol{\beta}}(t, \boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t, \boldsymbol{s})\right)^T = \underset{(\boldsymbol{\beta}, \boldsymbol{\alpha})^T \in \mathbb{R}^{q+1+d}}{\arg\min} \left[\boldsymbol{Y} - \boldsymbol{X}(t, \boldsymbol{s}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix}\right]^T \boldsymbol{W}(t, \boldsymbol{s}) \left[\boldsymbol{Y} - \boldsymbol{X}(t, \boldsymbol{s}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix}\right]$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- **Proposed Localized Derivative Estimator of $\theta(t)$:**

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

Fast Computing Algorithm for Our Integral Estimator

Our *integral estimator* takes the form

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

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

Fast Computing Algorithm for Our Integral Estimator

Our *integral estimator* takes the form

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

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

► **Solution:** Let $T_{(1)} \le \cdots \le T_{(n)}$ be the order statistics of $T_1, ..., T_n$ and $\Delta_j = T_{(j+1)} - T_{(j)}$ for $j = 1, ..., n-1$.

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

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

Fast Computing Algorithm for Our Integral Estimator

Our *integral estimator* takes the form

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

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

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

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

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

Nonparametric Bootstrap Inference

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

Nonparametric Bootstrap Inference

1. Compute $\widehat{m}_\theta(t)$ on the original data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$.
2. Generate B bootstrap samples $\left\{\left(Y_i^{*(b)}, T_i^{*(b)}, \boldsymbol{S}_i^{*(b)}\right)\right\}_{i=1}^n$ by sampling with replacement and compute $\widehat{m}_\theta^{*(b)}(t)$ for each $b = 1, ..., B$.

Nonparametric Bootstrap Inference

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

2. Generate B bootstrap samples $\left\{\left(Y_i^{*(b)}, T_i^{*(b)}, \boldsymbol{S}_i^{*(b)}\right)\right\}_{i=1}^n$ by sampling with replacement and compute $\widehat{m}_\theta^{*(b)}(t)$ for each $b = 1, ..., B$.

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

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

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

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

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

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

Then, when $q = 2$, as $h, b, \hbar, \frac{\max\{h,b\}^4}{h} \to 0$ and $\frac{n\max\{h,\hbar\}b^d}{\log n}, \frac{n\hbar}{\log n} \to \infty$,

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

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

[3] We thank Alex Luedtke for pointing out an unexpected dimension dependence of our previous rate $o_P\left(\sqrt{\frac{\log n}{nh^3b^d}} + \hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right)$. Our new proof is inspired by Fan et al. (1998).

Under the same regularity conditions, if $h \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^5}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some $c_1, c_2 \geq 0$ and $\frac{n \max\{h,\hbar\} b^d}{\log n}, \frac{n\hbar}{\log n}, \frac{h^3 \log n}{\hbar}, \frac{nh^3\hbar^4}{\log n} \to \infty$ as $n \to \infty$, then for any $t \in \mathcal{T}'$,

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

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

where[4]

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

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

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

[4] The key of our previous proof is to write $\widehat{m}_\theta(t) - m(t)$ into a V-statistic (Shieh, 2014).

Bootstrap Consistency

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

1.
$$\left| \sqrt{nh^3} \sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| - \sup_{t \in \mathcal{T}'} |\mathbb{G}_n \varphi_t| \right| = O_P\left(\sqrt{nh^3 \max\{h, \hbar\}^4} + \sqrt{\frac{h^3 \log n}{\hbar}} + \frac{\log n}{\sqrt{n\hbar}} + \sqrt{\frac{\log n}{nb^d \hbar}} \right).$$

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

1. $$\left| \sqrt{nh^3} \sup_{t \in \mathcal{T}'} \left|\widehat{m}_\theta(t) - m(t)\right| - \sup_{t \in \mathcal{T}'} \left|\mathbb{G}_n \varphi_t\right| \right| = O_P\left(\sqrt{nh^3 \max\{h, \hbar\}^4} + \sqrt{\frac{h^3 \log n}{\hbar}} + \frac{\log n}{\sqrt{n\hbar}} + \sqrt{\frac{\log n}{nb^d \hbar}} \right).$$

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

$$\sup_{u \geq 0} \left| \mathrm{P}\left(\sqrt{nh^3} \sup_{t \in \mathcal{T}'} \left|\widehat{m}_\theta(t) - m(t)\right| \leq u \right) - \mathrm{P}\left(\sup_{f \in \mathcal{F}} \left|\mathbb{B}(f)\right| \leq u \right) \right| = O\left(\left(\frac{\log^5 n}{nh^3} \right)^{\frac{1}{8}} + \left(\frac{\log^2 n}{nb^d \hbar} \right)^{\frac{3}{8}} \right).$$

Bootstrap Consistency

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

1. $$\left| \sqrt{nh^3} \sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| - \sup_{t \in \mathcal{T}'} |\mathbb{G}_n \varphi_t| \right| = O_P\left(\sqrt{nh^3 \max\{h,\hbar\}^4} + \sqrt{\frac{h^3 \log n}{\hbar}} + \frac{\log n}{\sqrt{n\hbar}} + \sqrt{\frac{\log n}{nb^d \hbar}} \right).$$

2. there exists a mean-zero Gaussian process $\mathbb{B}$ such that
$$\sup_{u \geq 0} \left| \mathrm{P}\left(\sqrt{nh^3} \sup_{t \in \mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| \leq u \right) - \mathrm{P}\left(\sup_{f \in \mathcal{F}} |\mathbb{B}(f)| \leq u \right) \right| = O\left(\left(\frac{\log^5 n}{nh^3} \right)^{\frac{1}{8}} + \left(\frac{\log^2 n}{nb^d \hbar} \right)^{\frac{3}{8}} \right).$$

3. $$\sup_{u \geq 0} \left| \mathrm{P}\left(\sqrt{nh^3} \sup_{t \in \mathcal{T}'} |\widehat{m}^*_\theta(t) - \widehat{m}_\theta(t)| \leq u \middle| \mathbb{U}_n \right) - \mathrm{P}\left(\sup_{f \in \mathcal{F}} |\mathbb{B}(f)| \leq u \right) \right| = O_P\left(\left(\frac{\log^5 n}{nh^3} \right)^{\frac{1}{8}} + \left(\frac{\log^2 n}{nb^d \hbar} \right)^{\frac{3}{8}} \right)$$
   where
$$\mathcal{F} = \left\{ (v, x, z) \mapsto \varphi_t(v, x, z) : t \in \mathcal{T}' \right\}.$$

Remarks on Our Asymptotic Results

1. $\mathcal{F}$ is not Donsker because φ_t is not uniformly bounded as $h \to 0$.
   - However, $\widetilde{\mathcal{F}} = \left\{(v, x, z) \mapsto \sqrt{h^3} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}'\right\}$ 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 $h \to 0$.
   - However, $\widetilde{\mathcal{F}} = \left\{(v, x, \boldsymbol{z}) \mapsto \sqrt{h^3} \cdot \varphi_t(v, x, \boldsymbol{z}) : t \in \mathcal{T}'\right\}$ 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 $\text{Var}(\epsilon) = \sigma^2 > 0$, $\text{Var}\left[\varphi_t(Y, T, \boldsymbol{S})\right]$ is a positive finite number.
   - The asymptotic linearity (or V-statistic) is non-degenerate.
   - Pointwise bootstrap confidence intervals are asymptotically valid.

Remarks on Our Asymptotic Results

1. $\mathcal{F}$ is not Donsker because φ_t is not uniformly bounded as $h \to 0$.
   - However, $\widetilde{\mathcal{F}} = \left\{(v, x, z) \mapsto \sqrt{h^3} \cdot \varphi_t(v, x, z) : t \in \mathcal{T}'\right\}$ is of VC-type.
   - Gaussian approximation in Chernozhukov et al. (2014) can be applied to bound the difference between $\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|$ and $\sup_{f \in \mathcal{F}} |\mathbb{B}(f)|$.
2. As long as $\mathrm{Var}(\epsilon) = \sigma^2 > 0$, $\mathrm{Var}\left[\varphi_t(Y, T, S)\right]$ is a positive finite number.
   - The asymptotic linearity (or V-statistic) is non-degenerate.
   - Pointwise bootstrap confidence intervals are asymptotically valid.
3. For the validity of uniform bootstrap confidence band, one can choose the bandwidths $h \asymp \hbar = O\left(n^{-\frac{1}{5}}\right)$ and $\left(\frac{\log n}{n}\right)^{\frac{4}{5d}} \lesssim b \lesssim n^{-\frac{1}{5}}$.
   - They match up with 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

Simulation Setup

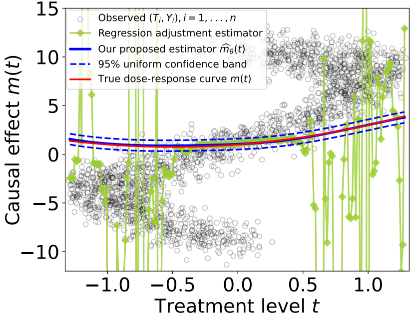
- Use the Epanechnikov kernel for K_T and K_S (with the product kernel technique) and Gaussian kernel for $\bar{K}_T$.
- Select the bandwidth parameters $h, b > 0$ by modifying the rule-of-thumb method in Yang and Tschernig (1999).
- Set the bandwidth parameter $\hbar > 0$ to the normal reference rule in Chacón et al. (2011); Chen et al. (2016).
- Set the bootstrap resampling time $B = 1000$ and the significance level $\alpha = 0.05$.
- Compare our proposed estimators with the regression adjustment estimators under the same choices of bandwidth parameters:

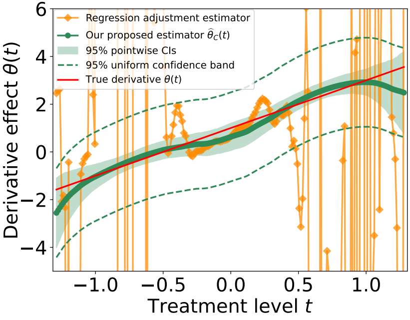
$$\widehat{m}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}(t, \boldsymbol{S}_i) \quad \text{and} \quad \widehat{\theta}_{\mathrm{RA}}(t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\beta}_2(t, \boldsymbol{S}_i).$$

Generate i.i.d. observations $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^{2000}$ 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.



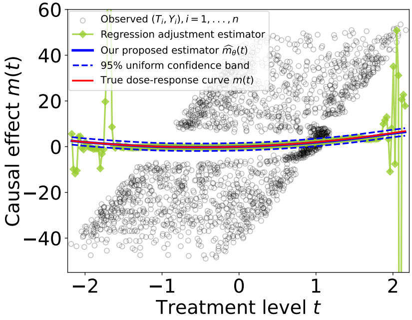


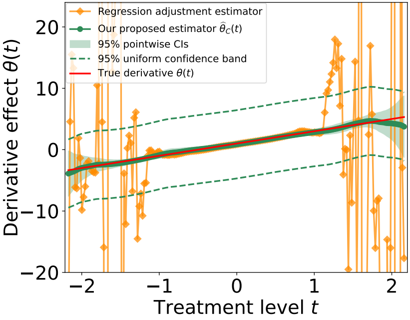
Nonlinear Confounding Model

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

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

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





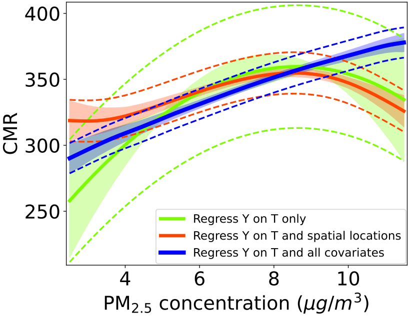
Effect of $PM_{2.5}$ on the Cardiovascular Mortality Rate (CMR)

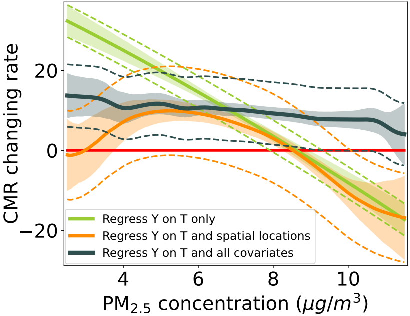
1. Recent studies identify a positive association between $PM_{2.5}$ level ($\mu g/m^3$) and county-level CMR (deaths/100,000 person-years) in the U.S. after controlling for socioeconomic factors (Wyatt et al., 2020a).
2. Obtain the average annual CMR as Y and $PM_{2.5}$ concentration as T over years 1990-2010 within $n = 2132$ U.S. counties from Wyatt et al. (2020b).

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

1. Recent studies identify a positive association between $PM_{2.5}$ level ($\mu g/m^3$) 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 $PM_{2.5}$ concentration as T over years 1990-2010 within $n = 2132$ U.S. counties from Wyatt et al. (2020b).
3. The covariate vector $\boldsymbol{S} \in \mathbb{R}^{10}$ 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 $PM_{2.5}$ between 2.5 $\mu g/m^3$ and 11.5 $\mu g/m^3$ 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.

- We identify $m(t)$ through the identification of $\theta(t)$ when the positivity condition fails to hold.
- We 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.

- We identify $m(t)$ through the identification of $\theta(t)$ when the positivity condition fails to hold.
- We 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 $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ and $\mathrm{P}(\boldsymbol{s}|t)$:
    - Bandwidth selection via the plug-in rule (Ruppert et al., 1995) or cross-validation (Li and Racine, 2004).
    - Regression splines for $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ (Friedman, 1991; Zhou and Wolfe, 2000) and local logistic approaches for $P(\boldsymbol{s}|t)$ (Hall et al., 1999).

Summary and Future Works

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

- We identify $m(t)$ through the identification of $\theta(t)$ when the positivity condition fails to hold.
- We 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 $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ and $\mathrm{P}(\boldsymbol{s}|t)$:
   - Bandwidth selection via the plug-in rule (Ruppert et al., 1995) or cross-validation (Li and Racine, 2004).
   - Regression splines for $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ (Friedman, 1991; Zhou and Wolfe, 2000) and local logistic approaches for $P(\boldsymbol{s}|t)$ (Hall et al., 1999).
2. Generalize our proposed integral estimators to the IPW and doubly robust variants.

Semi-parametric Inference With High-Dimensional Covariates

3. Sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022) and the interchangeability assumption.

Semi-parametric Inference With High-Dimensional Covariates

3. Sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022) and the interchangeability assumption.

4. Study the semi-parametric efficiency (Kennedy, 2016) of the influence functions from our proposed estimators:

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

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

3. Sensitivity analysis on unmeasured confounding (Chernozhukov et al., 2022) and the interchangeability assumption.

4. Study the semi-parametric efficiency (Kennedy, 2016) of the influence functions from our proposed estimators:

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

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

5. Our proposed nonparametric estimators suffer from the curse of dimensionality.

   - $\left(\frac{\log n}{n}\right)^{\frac{4}{5d}} \lesssim b \lesssim n^{-\frac{1}{5}}$ only works when $d<5$.
   - Impose a semi-parametric additive model (Guo et al., 2019) as:

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

   where $\boldsymbol{Z}\in\mathbb{R}^{d'}$ 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`.

We thank Alex Luedtke, Andrea Rotnitzky, Marco Carone, Zhichao Jiang, Pawel Morzywolek, and Daniel Suen for their insightful comments on the earlier version of this presentation.

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

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.

Reference

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.

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. Semiparametric theory and empirical processes in causal inference. *Statistical causal inferences and their applications in public health research*, pages 141–167, 2016.

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.

Reference

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.

J. Shao. *Mathematical Statistics*. Springer Science & Business Media, 2003.

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.

Reference

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.

Regularity Assumptions (Smoothness Conditions)

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

1. For any $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$, $\mu(t, \boldsymbol{s})$ is at least $(q + 1)$ times continuously differentiable with respect to t and at least four times continuously differentiable with respect to $\boldsymbol{s}$. Furthermore, $\mu(t, \boldsymbol{s})$ and all of its partial derivatives are uniformly bounded on $\mathcal{T} \times \mathcal{S}$.
2. $p(t, \boldsymbol{s})$ 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 $p(t, \boldsymbol{s})$ are continuous up to the boundary $\partial\mathcal{E}$. Furthermore, $\mathcal{E}$ is compact and $p(t, \boldsymbol{s})$ is uniformly bounded away from 0 on $\mathcal{E}$. Finally, the marginal density $p_T(t)$ is non-degenerate.

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

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

where

$$\mathcal{B}((t, \boldsymbol{s}),\, r) = \left\{(t_1, \boldsymbol{s}_1) \in \mathbb{R}^{d+1} : ||(t_1 - t, \boldsymbol{s}_1 - \boldsymbol{s})||_2 \leq r\right\}$$

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

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

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

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

Regularity Assumptions (Kernel Conditions)

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

$$\kappa_j^{(T)} := \int_{\mathbb{R}} u^j K_T(u)\, du < \infty, \quad \nu_j^{(T)} := \int_{\mathbb{R}} u^j K_T^2(u)\, du < \infty,$$

$$\kappa_{j,\ell}^{(S)} := \int_{\mathbb{R}^d} u_\ell^j K_S(\boldsymbol{u})\, d\boldsymbol{u} < \infty, \quad \text{and} \quad \nu_{j,k}^{(S)} := \int_{\mathbb{R}^d} u_\ell^j K_S^2(\boldsymbol{u})\, d\boldsymbol{u} < \infty.$$

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 $\ell = 1, ..., d$.

7. Let $\mathcal{K}_{q,d} = \Big\{ (y, \boldsymbol{z}) \mapsto \left(\frac{y-t}{h}\right)^\ell \left(\frac{z_i - s_i}{b}\right)^{k_1} \left(\frac{z_j - s_j}{b}\right)^{k_2} K_T\left(\frac{y-t}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{s}}{b}\right) : (t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}; i, j = 1, ..., d; \ell = 0, ..., 2q; k_1, k_2 = 0, 1; h, b > 0 \Big\}$. 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 $\bar{K}_T : \mathbb{R} \to [0, \infty)$ 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 $A_2, \upsilon_2 > 0$ such that for any $0 < \epsilon < 1$,

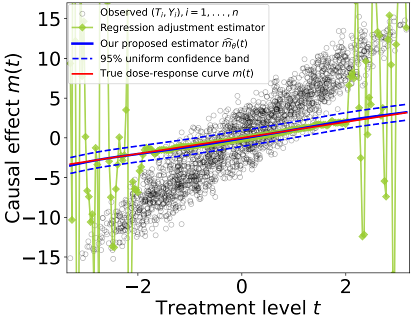
$$\sup_Q N\left(\mathcal{G}, L_2(Q), \epsilon \,||G||_{L_2(Q)}\right) \leq \left(\frac{A_2}{\epsilon}\right)^{\upsilon_2},$$

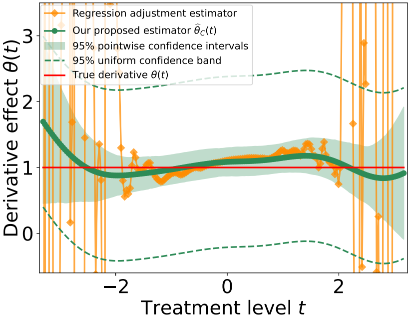
where $N\left(\mathcal{G}, L_2(Q), \epsilon \,||G||_{L_2(Q)}\right)$ is the $\epsilon \,||G||_{L_2(Q)}$-covering number of the (semi-)metric space $\left(\mathcal{G}, ||\cdot||_{L_2(Q)}\right)$, Q is any probability measure on $\mathbb{R}^{d+1}$, G is an envelope function of $\mathcal{G}$, and $||G||_{L_2(Q)}$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

Generate i.i.d. observations $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^{2000}$ 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 $\mathrm{Var}(E) = 0$

For simplicity, we assume the additive confounding model

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

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

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

$$\mu\left(f(\boldsymbol{s}), \boldsymbol{s}\right) = \bar{m}\left(f(\boldsymbol{s})\right) + \eta(\boldsymbol{s}) = m\left(f(\boldsymbol{s})\right) + \eta(\boldsymbol{s}). \tag{1}$$

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

Assumption (Bounded random effect)

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

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

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

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

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

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

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

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 $\mathrm{Var}(E) = 0$.

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