



NONPARAMETRIC INFERENCE ON DOSE-RESPONSE CURVES WITHOUT THE POSITIVITY CONDITION

Yikun Zhang[†] and Yen-Chi Chen

Department of Statistics, University of Washington

[†] yikun@uw.edu

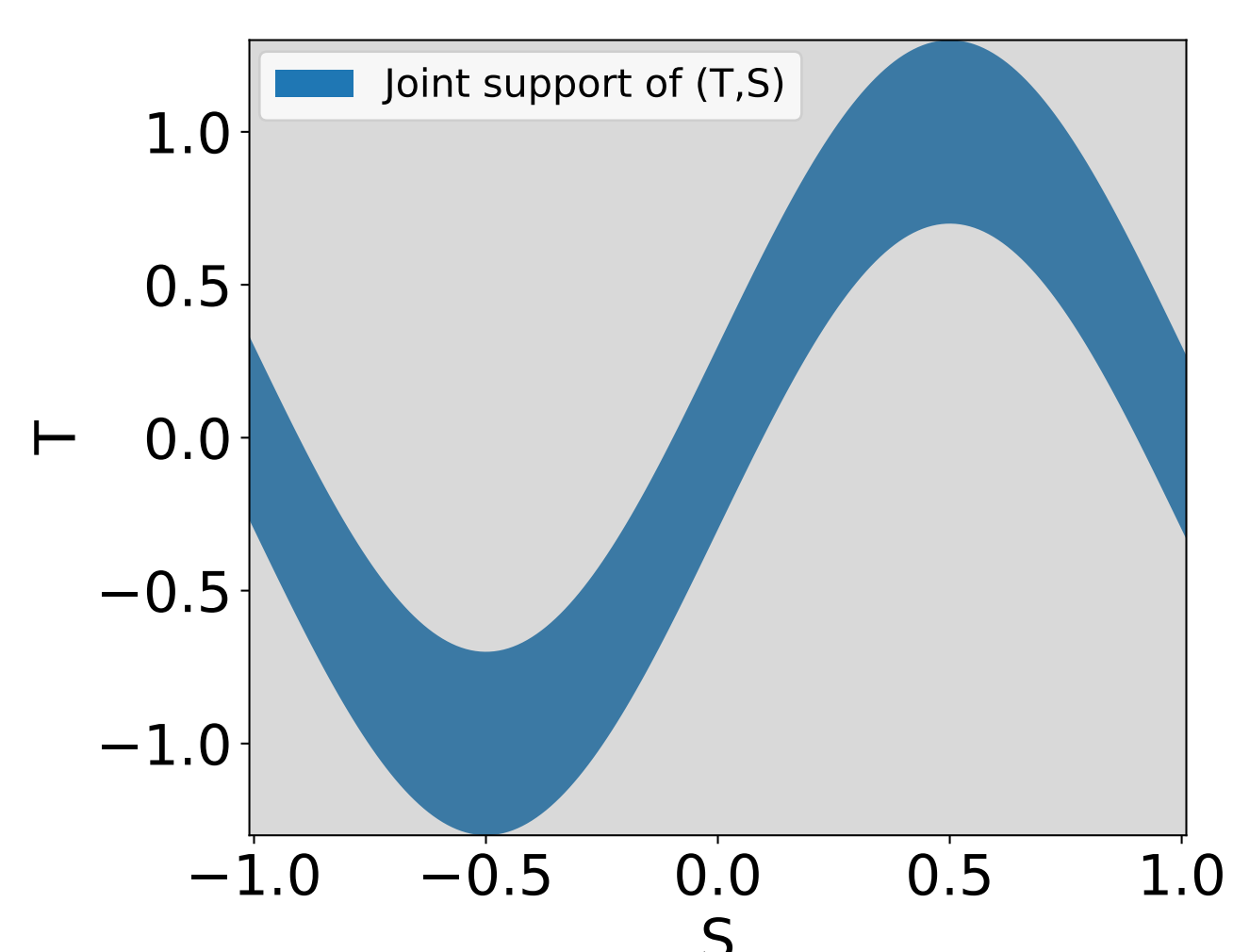


INTRODUCTION

Estimating the causal effects for continuous treatments (*i.e.*, the dose-response curves) often relies on the **positivity condition**:

Every subject has some chance of receiving any treatment level $T = t$ regardless of its covariates $S = s \in \mathbb{R}^d$.

- This condition **could fail** in observational studies with continuous treatments.



- We propose a novel integral estimator of the dose-response curve without assuming the positivity condition.

- It is based on a localized derivative estimator and the fundamental theorem of calculus.
- It can be efficiently computed in practice via Riemann sum approximations.
- It can be combined with bootstrap methods for valid inference on the dose-response curve and its derivative.

IDENTIFICATION CONDITIONS

Assume that $\{(Y_i, T_i, S_i)\}_{i=1}^n$ are IID from the model:

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

where $E \perp\!\!\!\perp S, \epsilon, \epsilon \perp\!\!\!\perp S, \mathbb{E}(E) = \mathbb{E}(\epsilon) = 0, \mathbb{E}(E^2) > 0$, and $\mathbb{E}(\epsilon^4) < \infty$.

Dose-response curve and its **derivative function** can be identified with observed data as:

$$m(t) = \mathbb{E}[\mu(t, S)] \quad \text{and} \quad \theta(t) = m'(t) = \frac{d}{dt} \mathbb{E}[\mu(t, S)]$$

under *consistency* and *ignorability* assumptions.

Interchangability Assumption: The function $\mu(t, s)$ is continuously differentiable with respect to t and

$$\mathbb{E}[\mu(T, S)] = \mathbb{E}[m(T)],$$

$$\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, S)\right] = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, S) \mid T = t\right].$$

MOTIVATING EXAMPLE

Consider the following additive confounding model:

$$Y = m(T) + \eta(S) + \epsilon \quad \text{and} \quad T = f(S) + E$$

with $\mathbb{E}[\eta(S)] = 0$. This model satisfies our interchangability assumption and is known as the geoaditive structural equation in spatial statistics.

THREE KEY INSIGHTS

- $\mu(t, s)$ and $\frac{\partial}{\partial t} \mu(t, s)$ can be consistently estimated at each observation (T_i, S_i) .
- $\theta(t)$ can be consistently estimated by the localized form $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t} \mu(t, S) \mid T = t\right]$.
- 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 Taking the expectation on both sides yield that

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

PROPOSED ESTIMATORS

Proposed Integral Estimator of $m(t)$:

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

where $\hat{\theta}_C(t)$ is a consistent estimator of $\theta_C(t) = \int \beta_2(t, s) dP(s|t)$ with $\beta_2(t, s) \equiv \frac{\partial}{\partial t} \mu(t, s)$.

- Fit $\beta_2(t, s)$ by local polynomial regression;
- Estimate $P(s|t)$ by Nadaraya-Watson conditional CDF estimator.

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

$$\hat{\theta}_C(t) = \frac{\sum_{i=1}^n \hat{\beta}_2(t, S_i) \cdot \bar{K}_T\left(\frac{T_i - t}{h}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j - t}{h}\right)}.$$

FAST COMPUTING ALGORITHM

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 $j = 1, \dots, n$ as:

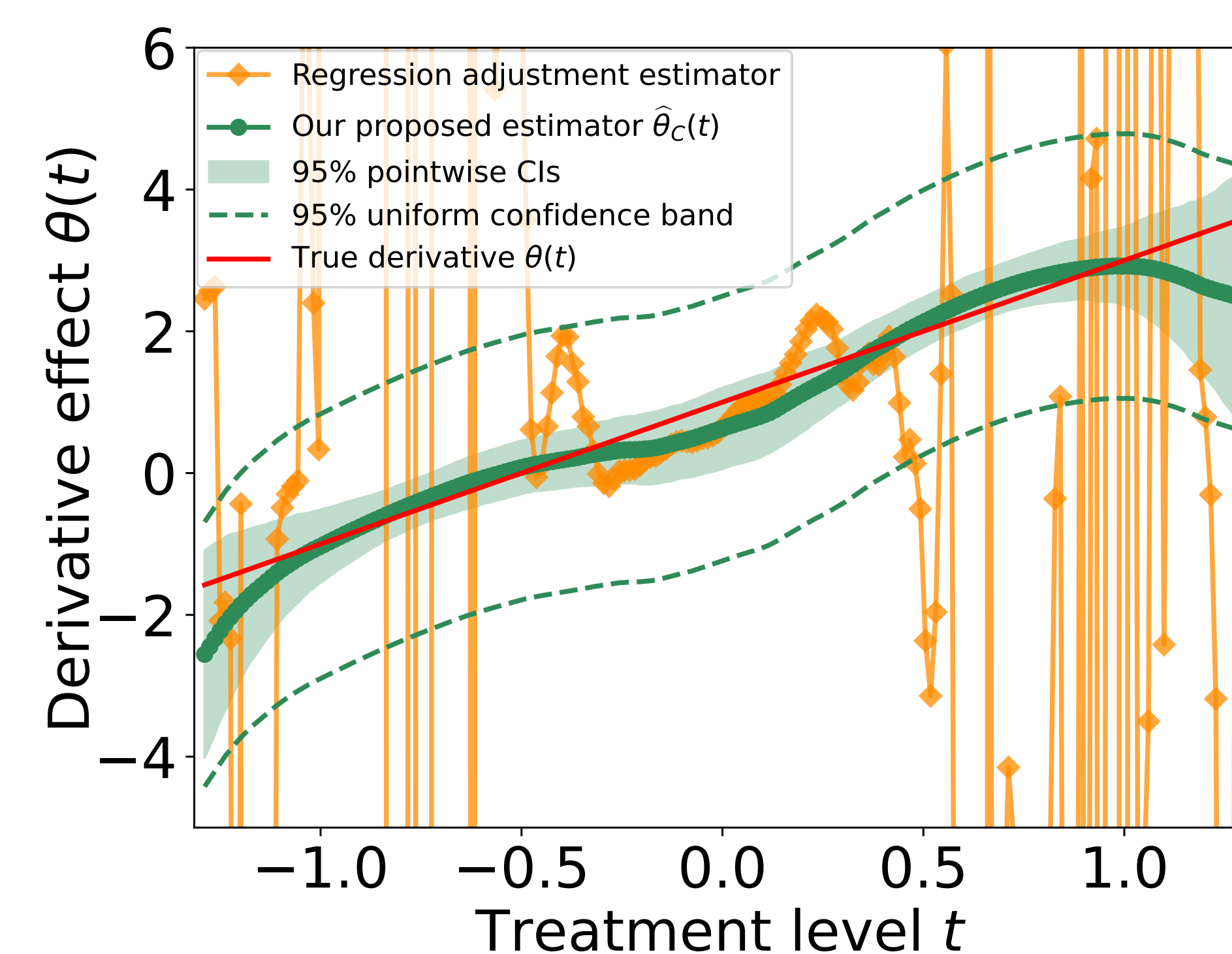
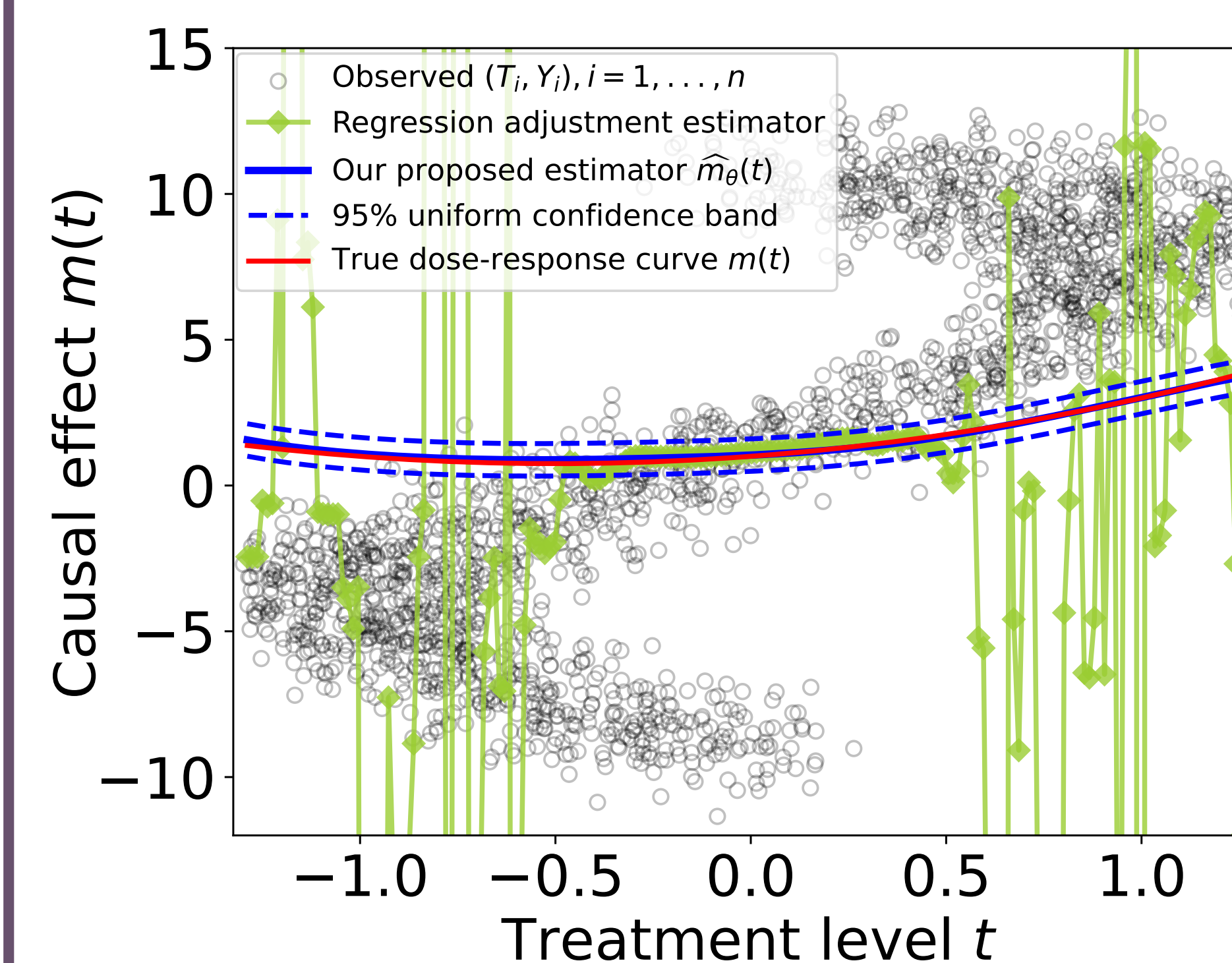
$$\begin{aligned} \hat{m}_\theta(T_{(j)}) &\approx \frac{1}{n} \sum_{i=1}^n Y_i + \frac{1}{n} \sum_{i=1}^{n-1} \Delta_i \left[i \cdot \hat{\theta}_C(T_{(i)}) \mathbb{1}_{\{i < j\}} \right. \\ &\quad \left. - (n-i) \cdot \hat{\theta}_C(T_{(i+1)}) \mathbb{1}_{\{i \geq j\}} \right]. \end{aligned}$$

- 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 $\hat{m}_\theta(T_{(j+1)})$.

SIMULATION STUDIES

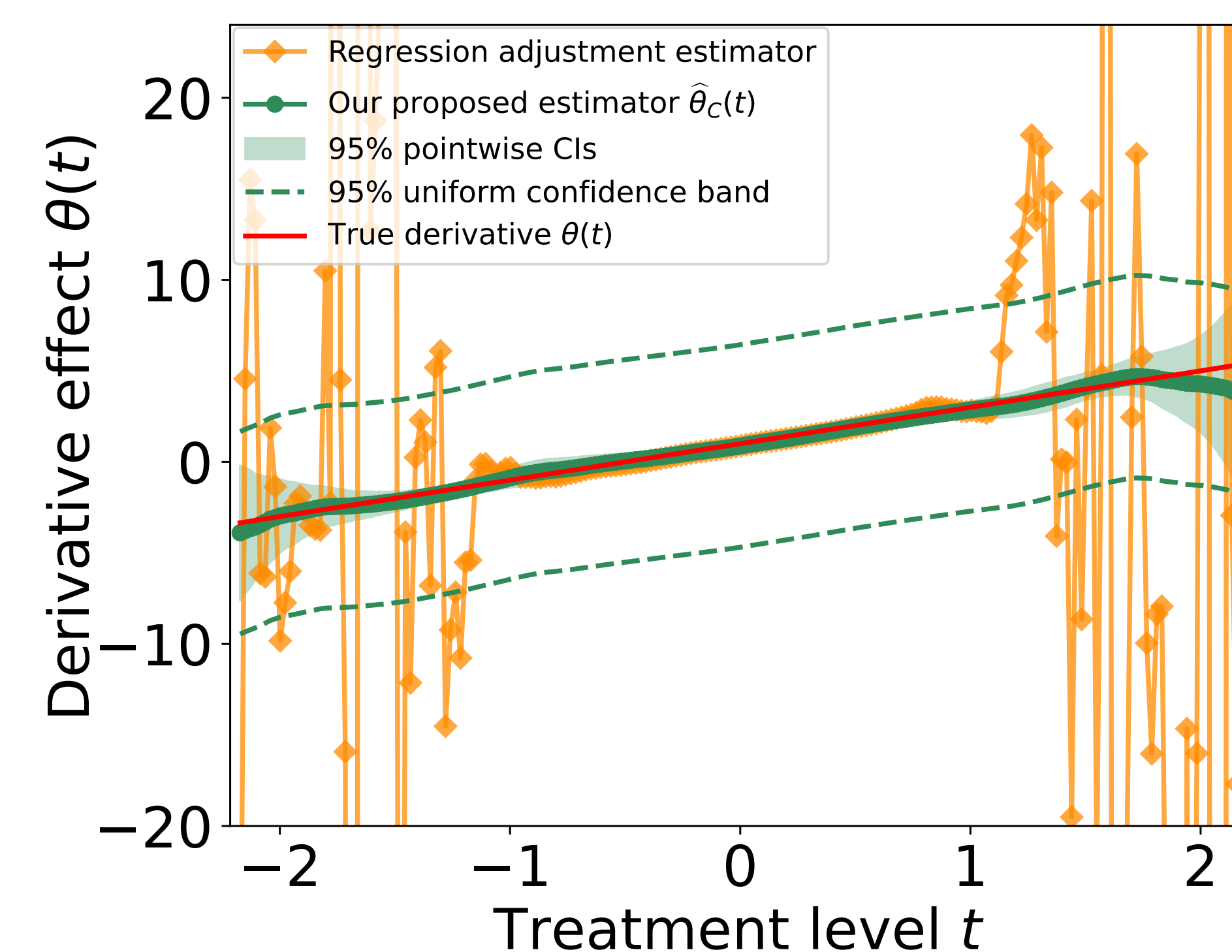
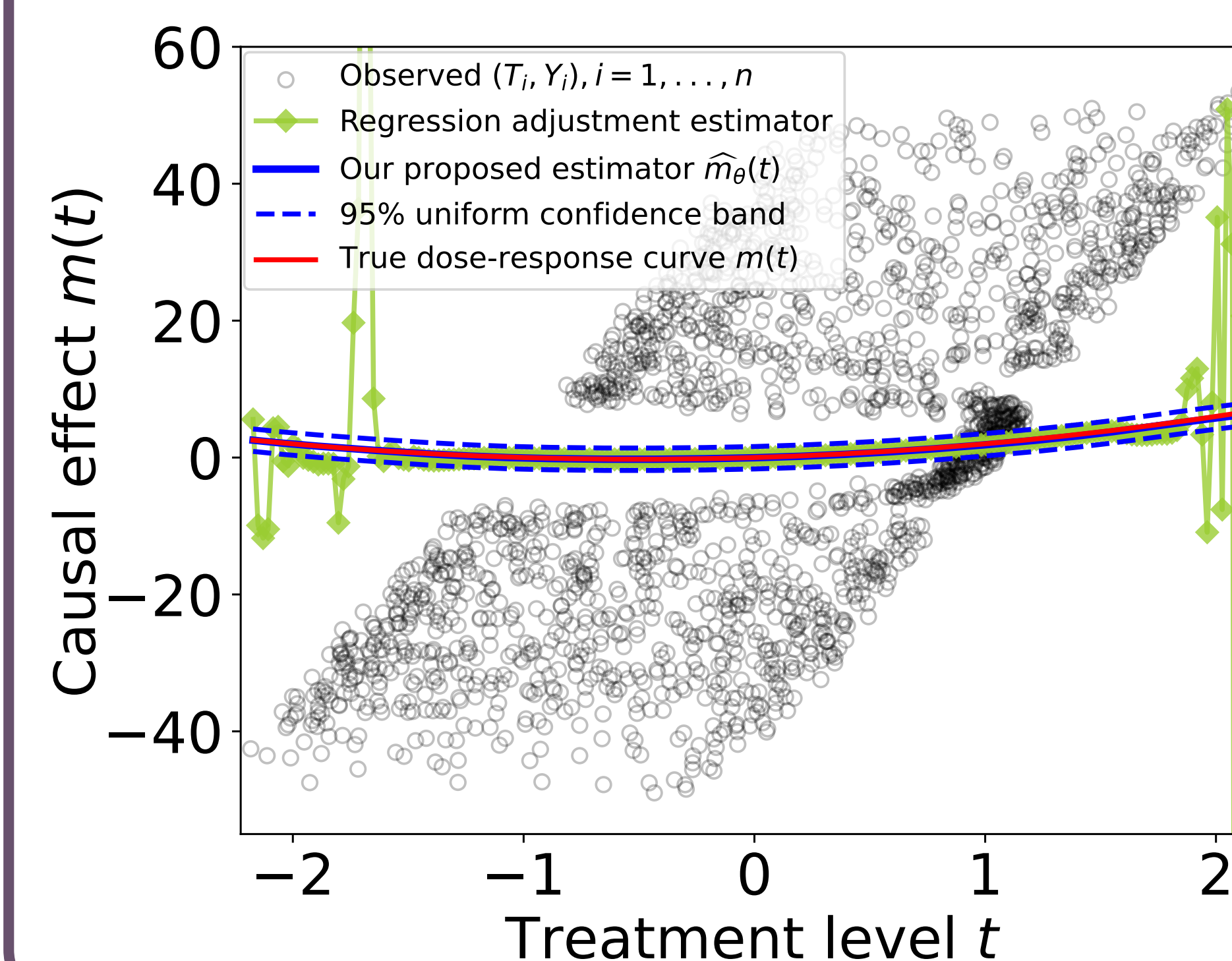
• Single Confounder Model:

$$Y = T^2 + T + 1 + 10S + \epsilon, \quad T = \sin(\pi S) + E, \quad S \sim \text{Unif}[-1, 1] \subset \mathbb{R}, \quad E \sim \text{Unif}[-0.3, 0.3], \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, 1).$$



• Nonlinear Confounding Model:

$$\begin{aligned} Y &= T^2 + T + 10Z + \epsilon, \quad T = \cos(\pi Z^3) + Z/4 + E, \quad Z = 4S_1 + S_2, \\ S &= (S_1, S_2) \sim \text{Unif}[-1, 1]^2 \subset \mathbb{R}^2, \quad E \sim \text{Unif}[-0.1, 0.1], \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, 1). \end{aligned}$$



EFFECT OF PM_{2.5} ON CARDIOVASCULAR MORTALITY RATE (CMR)

The covariate vector $S \in \mathbb{R}^{10}$ includes spatical locations (longitude, latitude) and eight socioeconomic factors.

