

Sharp Bounds on Continuous Treatment Effects under Unobserved Confounding

Junzhe Zhang
Syracuse University
jzhan403@syr.edu

Abstract

We study partial identification of conditional treatment effects with a continuous treatment under unobserved confounding. We target the conditional average potential outcome (CAPO) $\mathbb{E}[Y_x | z]$ and the conditional average treatment effect (CATE) between two treatment levels. We assume a bounded outcome and Lipschitz continuity of the response functions and of the propensity scores given the latent confounder, and show that these conditions imply the smooth treatment response and smooth treatment selection conditions of Kim et al. (2018). The combined bound of Kim et al. uses only two reference points on the identified diagonal and is generally loose. Bounding through every diagonal reference point instead gives explicit bounds whose average is the sharp identified set of the CAPO at every treatment, inside or outside the observational support. Capping the difference between the upper and lower bounds gives sharp causal bounds on the CATE. These bounds are informative without monotonicity or overlap assumptions.

1 INTRODUCTION

Evaluating treatment effects from observational data is a central task in causal inference, with applications to policy evaluation, healthcare, and reinforcement learning. The challenge is that observed treatment–outcome correlations generally do not imply causal effects: unobserved confounding can induce spurious relations, leaving the target effect non-identifiable, i.e., underdetermined by the observational data together with causal knowledge about the underlying system (Pearl, 2000). Several conditions and algorithms have been developed to decide whether a causal effect is identifiable (Pearl, 2000; Bareinboim and Pearl, 2016).

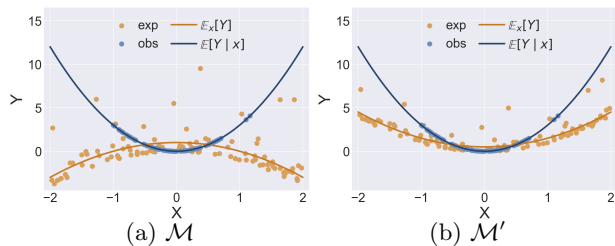


Figure 1: Two confounded systems that induce the *same* observational law $P(X, Y)$ (blue; regression $\mathbb{E}[Y | x]$ in dark blue) yet *different* interventional laws $P_x(Y)$ (orange), illustrating non-identifiability: (a) $\mathcal{M}$ with $Y \leftarrow -X^2 + U^2$ has a causal effect $\mathbb{E}_x[Y] = 1 - x^2$ (dark orange curve); (b) $\mathcal{M}'$ with $Y \leftarrow X^2 + \frac{1}{2}U^2$ has a causal effect $\mathbb{E}_x[Y] = x^2 + \frac{1}{2}$ (dark orange curve).

For example, under the *back-door* condition (Pearl, 2000, Ch. 3.3.1), or equivalently *ignorability* (Rosenbaum and Rubin, 1983), the effect is identified by covariate adjustment and can be estimated at standard rates (Rosenbaum and Rubin, 1983; Bang and Robins, 2005). In practice, however, the available causal assumptions are often insufficient to determine the target effect from observational data, in which case the effect is said to be *non-identifiable*. The following example illustrates this challenge.

Example 1. Consider a system $\mathcal{M}$ with treatment $X \leftarrow -U/2$ and outcome $Y \leftarrow -X^2 + U^2$, where $U \sim \mathcal{N}(0, 1)$ is unobserved. Fig. 1a shows samples (blue) from the resulting observational distribution $P(X, Y)$ alongside samples (orange) from the interventional distribution $P_x(Y)$ obtained by randomly assigning X over $[-2, 2]$. The two deviate significantly: the unobserved confounder U induces a spurious correlation between X and Y , making some treatments appear more effective than they actually are. Treatment effects $\mathbb{E}_x[Y]$ are not identifiable here (Huang and Valtorta, 2006; Shpitser and Pearl, 2006). To witness, consider the alternative system $\mathcal{M}'$ with the same treatment $X \leftarrow -U/2$ but outcome $Y \leftarrow X^2 + \frac{1}{2}U^2$.

Both models share the causal relationships ($X \rightarrow Y$ and $X \leftarrow U \rightarrow Y$) and induce the *identical* observational law $P(X, Y)$ (substituting $U = -2X$ gives $Y = 3X^2$ in either model), yet they disagree on the effect: intervening $\text{do}(X \leftarrow x)$ yields $\mathbb{E}_x[Y] = 1 - x^2$ under $\mathcal{M}$ but $\mathbb{E}_x[Y] = x^2 + \frac{1}{2}$ under $\mathcal{M}'$ (Fig. 1b). Both models thus agree that X affects Y and are confounded in the same way; they disagree on *which treatment is preferable*, with $x = 0$ optimal under $\mathcal{M}$ but the worst choice under $\mathcal{M}'$. The observational data alone cannot tell them apart. ■

When identifiability fails, a common alternative approach is *partial identification*: bounding the effect over all confounding patterns compatible with the data (Manski, 1990; Balke and Pearl, 1997). For discrete treatments and outcomes, sharp bounds follow from programs over a discretized latent domain (Balke and Pearl, 1997; Zhang et al., 2022) or from generative-model approximations (Xia et al., 2021, 2022). However, this enumeration does not carry over to continuous treatments, where three lines of work exist.

Sensitivity models bound how far hidden confounding may distort the treatment density: the continuous marginal sensitivity model (CMSM) of Jesson et al. (2022), its sharp and generalized versions (Frauen et al., 2023; Baitairian et al., 2025), the δ -MSM (Marmarelis et al., 2023), and exposure-dependent and marginal-structural variants (Bonvini et al., 2022; Dalal and Tchetgen Tchetgen, 2025). Several bound the conditional average potential outcome (CAPO) pointwise, but all are stated relative to the observed treatment density and require it to be positive at the query. *Optimization-based* bounds search a response-function class consistent with the data, usually given an instrument (Kilbertus et al., 2020; Gunsilius, 2020; Padh et al., 2023; Balazadeh Meresht et al., 2022). *Smoothness* bounds restrict how fast counterfactual means vary (Tan et al., 2024). Following Hausman and Newey (2016), Kim et al. (2018) bound the effect of treatment on the treated assuming it is Lipschitz in the intervened treatment (smooth treatment response, STR) and in the received treatment (smooth treatment selection, STS), and Khan et al. (2024) use Lipschitz smoothness for off-policy evaluation without overlap. App. A provides more details on related work.

We work in the smoothness line, in the setting of Kim et al. (2018) extended to covariates. Our targets are the CAPO $\tau(\mathbf{x}, \mathbf{z}) = \mathbb{E}[Y_{\mathbf{x}} | \mathbf{z}]$ and the conditional average treatment effect (CATE) $\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z})$ between two treatment levels. The bounds of Kim et al. (2018) are sharp under STR together with a monotonicity assumption. Under STR and STS, however, sharp bounds remain an open challenge: we show that their combined bound is not sharp in general.

This paper aims to address this challenge. Our contributions are threefold. (1) *Structural conditions for STR and STS*. We derive both conditions from Lipschitz continuity of the response functions and of the propensity scores given the latent confounder (Prop. 2), with a selection cost that depends only on the observed treatment density (Lem. 1). (2) *The bounds of Kim et al. (2018) are loose*. Their combined STR+STS bound uses only two reference points on the identified diagonal (Prop. 8) and is not sharp in general, even within their own model class (Prop. 9). (3) *Sharp bounds on the CAPO and the CATE*. Combining the STR and STS bounds through every diagonal reference point gives a valid bound (Thm. 6) that is the sharp identified set of the CAPO at every treatment, inside or outside the observational support (Thm. 10). Capping the difference of the upper and lower bounds gives the sharp identified set of the CATE, without any monotonicity assumption (Thm. 11). All proofs and experimental details are in the appendix.

Notations. We denote variables by capital letters (X) and their values by lowercase (x); $|\mathbf{X}|$ is the cardinality of a set $\mathbf{X}$. $P(\mathbf{X})$ is the distribution of $\mathbf{X}$, $P(\mathbf{x})$ abbreviates probability $P(\mathbf{X} = \mathbf{x})$, and $P(\mathbb{X})$ stands for the probability $P(\mathbf{X} \in \mathbb{X})$ that variables $\mathbf{X}$ take a value in a set $\mathbb{X}$. For reals a, b we write $a \vee b \triangleq \max\{a, b\}$ and $a \wedge b \triangleq \min\{a, b\}$.

2 LIPSCHITZ CONTINUITY FOR PARTIAL IDENTIFICATION

Our analysis focuses on the structural causal models (SCMs) (Pearl, 2000; Bareinboim et al., 2022) of a contextual bandit, with a treatment $\mathbf{X}$, an outcome Y , covariates $\mathbf{Z}$, and an unobserved confounder U affecting all three. In a model $\mathcal{M}$, structural functions set $\mathbf{Z} \leftarrow f_{\mathbf{Z}}(U)$, $\mathbf{X} \leftarrow f_{\mathbf{X}}(\mathbf{Z}, U, \varepsilon_{\mathbf{X}})$, and $Y \leftarrow f_Y(\mathbf{X}, \mathbf{Z}, U, \varepsilon_Y)$, where U , $\varepsilon_{\mathbf{X}}$, and ε_Y are mutually independent exogenous variables. The assignment noise $\varepsilon_{\mathbf{X}}$ makes the treatment random given $(\mathbf{Z}, U)$, so that it can have a conditional density $p(\mathbf{x} | \mathbf{z}, \mathbf{u})$; this is needed only for the structural condition (A3) below. The learner does not know $\mathcal{M}$; it only observes samples of $(\mathbf{X}, \mathbf{Z}, Y)$, summarized by the *observational distribution* $P(\mathbf{X}, Y, \mathbf{Z}) \triangleq P(\mathbf{X}, Y, \mathbf{Z}; \mathcal{M})$. An intervention $\text{do}(\mathbf{X} \leftarrow \mathbf{x})$ (for short, $\text{do}(\mathbf{x})$) replaces $f_{\mathbf{X}}$ by the constant $\mathbf{x}$, inducing the submodel $\mathcal{M}_{\mathbf{x}}$. The *potential outcome* $Y_{\mathbf{x}}(\mathbf{u})$ is the solution for Y in $\mathcal{M}_{\mathbf{x}}$, and the *interventional distribution* is the joint distribution $P_{\mathbf{x}}(Y, \mathbf{Z}) \triangleq P(Y_{\mathbf{x}}, \mathbf{Z}_{\mathbf{x}}; \mathcal{M})$ over potential outcomes.

Our goal is to infer the *conditional average potential outcome* (CAPO) $\tau(\mathbf{x}, \mathbf{z}) \triangleq \mathbb{E}_{\mathbf{x}}[Y | \mathbf{z}]$ at every treatment $\mathbf{x}$ and covariate value $\mathbf{z}$ (**output**) from the observational distribution $P(\mathbf{X}, \mathbf{Z}, Y)$ (**input**). We

also bound the *conditional average treatment effect* (CATE) $\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z}) = \mathbb{E}[Y_{\mathbf{x}_1} - Y_{\mathbf{x}_0} \mid \mathbf{z}]$ between two treatment levels (Thm. 11). Under unobserved confounding, τ is not identifiable: many distinct SCMs reproduce the same $P(\mathbf{X}, \mathbf{Z}, Y)$ yet disagree on the target query. Consequently, we aim at partial identification: deriving informative bounds over the CAPO and CATO from the observational distribution.

Without restrictions on the unobserved mechanisms, the data leave $Y_{\mathbf{x}}$ free at treatments that are rarely or never observed at $\mathbf{z}$, and no estimator can improve on the trivial range bound there. An informative bound requires restricting how fast the unobserved mechanisms may vary (Kim et al., 2018; Tan et al., 2024). We posit one for each unobserved mechanism: the outcome and the treatment assignment.

Each latent stratum induces a *response function* $h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) \triangleq \mathbb{E}[f_Y(\mathbf{x}, \mathbf{z}, \mathbf{u}, \varepsilon_Y)]$, the expected outcome of $\text{do}(\mathbf{X} \leftarrow \mathbf{x})$ at covariates $\mathbf{z}$ within stratum $\mathbf{u}$. Treatments and covariates take values in Borel sets $\mathbb{X} \subseteq \mathbb{R}^n$ and $\mathbb{Z} \subseteq \mathbb{R}^l$, which need not be bounded. With constants $L_Y, L_X > 0$, we assume the following.

(A1) (*Compact outcome domain.*) The outcome takes values in a compact interval $\mathbb{Y} = [\underline{y}, \bar{y}] \subset \mathbb{R}$.

(A2) (*Lipschitz response functions.*) For all $\mathbf{u} \in \mathbb{U}$, $\mathbf{z} \in \mathbb{Z}$, and $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$,

$$|h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) - h_{\mathbf{u}}(\mathbf{x}', \mathbf{z})| \leq L_Y \|\mathbf{x} - \mathbf{x}'\|. \quad (1)$$

(A3) (*Lipschitz propensity density.*) For all $\mathbf{u} \in \mathbb{U}$ and $\mathbf{z} \in \mathbb{Z}$, the treatment has a probability density $p(\mathbf{x} \mid \mathbf{z}, \mathbf{u})$ satisfying, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$,

$$|p(\mathbf{x} \mid \mathbf{z}, \mathbf{u}) - p(\mathbf{x}' \mid \mathbf{z}, \mathbf{u})| \leq L_X \|\mathbf{x} - \mathbf{x}'\|. \quad (2)$$

We write $\text{diam}(\mathbb{Y}) \triangleq \bar{y} - \underline{y}$ for the outcome range. Every response function output $f_Y(\mathbf{x}, \mathbf{z}, \mathbf{u}, \varepsilon_Y)$ lies in $[\underline{y}, \bar{y}]$. The constants L_Y and L_X are *sensitivity parameters* the analyst must prespecify.

2.1 Smoothness of Treatment on the Treated

Assumptions (A2) and (A3) restrict mechanisms that the data never reveal. To connect them to the data, we rewrite the target through a function whose values we partially identify. Specifically, let $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') \triangleq \mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', \mathbf{z}]$ denote the *effect of treatment on the treated* (ETT) at treatment $\mathbf{x}'$: the mean outcome under $\text{do}(\mathbf{x})$ of the units that observationally received $\mathbf{x}'$ (Pearl, 2000). It is the function studied by Kim et al. (2018), here conditional on $\mathbf{z}$. We call $\mathbf{x}$ the *intervened treatment* and $\mathbf{x}'$ the *received treatment*, and regard $g_{\mathbf{z}}$ as a function on $\mathbb{X} \times S_{\mathbf{z}}$, where $S_{\mathbf{z}} \triangleq \text{supp } P(\mathbf{X} \mid \mathbf{z})$ is the observational support at $\mathbf{z}$. Conditioning on the

received treatment $\mathbf{X} = \mathbf{x}'$ and applying the law of total expectation expands the target over the observational treatment distribution,

$$\tau(\mathbf{x}, \mathbf{z}) = \int g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') dP(\mathbf{x}' \mid \mathbf{z}). \quad (3)$$

The *nominal propensity score* $P(\mathbf{x}' \mid \mathbf{z}) = \int P(\mathbf{x}' \mid \mathbf{z}, \mathbf{u}) dP(\mathbf{u} \mid \mathbf{z})$ averages the latent-conditional assignments over the strata, and we write $p(\mathbf{x}' \mid \mathbf{z})$ for its density. As the nominal propensity is identified, bounding τ reduces to bounding $g_{\mathbf{z}}$. On the diagonal $\mathbf{x}' = \mathbf{x}$ the ETT is point-identified by consistency (Pearl, 2000, Ch. 7): $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}) = \mu(\mathbf{x}, \mathbf{z})$ for $\mathbf{x} \in S_{\mathbf{z}}$, where $\mu(\mathbf{x}, \mathbf{z}) \triangleq \mathbb{E}[Y \mid \mathbf{x}, \mathbf{z}]$ is the observed regression. Off the diagonal, where $\mathbf{x} \neq \mathbf{x}'$, it is unidentified.

Assumptions (A2) and (A3) restrict the unobserved mechanisms, but the bounds use them only through two consequences for the ETT function, which, following Kim et al. (2018), limit how fast it may vary in each coordinate. Note that $Y_{\mathbf{x}} \perp \mathbf{X} \mid \mathbf{U}, \mathbf{Z}$ in every SCM of $\mathcal{M}$, and the received treatment enters the ETT only through the latent posterior $P(\mathbf{U} \mid \mathbf{x}', \mathbf{z})$. Expanding the total expectation over $\mathbf{U}$ gives

$$g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') = \int h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) dP(\mathbf{u} \mid \mathbf{x}', \mathbf{z}). \quad (4)$$

This splits each ETT value into a response function, which varies with the intervened treatment and is controlled by (A2), and a latent posterior, which varies with the received treatment and is controlled by (A3). The key step bounds how fast the latent posterior moves with the received treatment, using only the density $p(\mathbf{x} \mid \mathbf{z})$ of the nominal propensity.

Lemma 1. *Under (A3), for every $\mathbf{z} \in \mathbb{Z}$ and every pair $\mathbf{x}'_0, \mathbf{x}'_1 \in \mathbb{X}$ with $p(\mathbf{x}'_0 \mid \mathbf{z}) > 0$ and $p(\mathbf{x}'_1 \mid \mathbf{z}) > 0$,*

$$\text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \leq \frac{L_X \|\mathbf{x}'_0 - \mathbf{x}'_1\|}{\max\{p(\mathbf{x}'_0 \mid \mathbf{z}), p(\mathbf{x}'_1 \mid \mathbf{z})\}}, \quad (5)$$

where $\text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \triangleq \text{TV}(P(\mathbf{U} \mid \mathbf{x}'_0, \mathbf{z}), P(\mathbf{U} \mid \mathbf{x}'_1, \mathbf{z}))$ is the total variation between latent posteriors.

Two received treatments share the response functions of Eq. 4 and differ only through the latent posterior, and the response functions take values in $\mathbb{Y}$. So, for every $\mathbf{x} \in \mathbb{X}$ and every such pair,

$$\begin{aligned} |g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_0) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_1)| &\leq \text{diam}(\mathbb{Y}) \text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \\ &\leq \kappa_{\mathbf{z}}(\mathbf{x}'_0, \mathbf{x}'_1) \triangleq \frac{L_X \text{diam}(\mathbb{Y}) \|\mathbf{x}'_0 - \mathbf{x}'_1\|}{\max\{p(\mathbf{x}'_0 \mid \mathbf{z}), p(\mathbf{x}'_1 \mid \mathbf{z})\}}. \end{aligned} \quad (6)$$

The coefficient $\kappa_{\mathbf{z}}$ need not satisfy the triangle inequality, so we pass to its chain closure $\kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{b}) \triangleq \inf \sum_i \kappa_{\mathbf{z}}(\mathbf{a}_i, \mathbf{a}_{i+1})$ over finite chains $\mathbf{a} = \mathbf{a}_0, \dots, \mathbf{a}_k = \mathbf{b}$ of points with positive density. This coefficient is

density-adaptive: it is small where many units are observed and grows where the observed density thins out, and it evaluates the density only at support points. A function satisfying $|f(\mathbf{a}) - f(\mathbf{b})| \leq \kappa_{\mathbf{z}}(\mathbf{a}, \mathbf{b})$ for all pairs also satisfies it with $\kappa_{\mathbf{z}}^*$, so the constraint translates.

Proposition 2. *Under (A1), (A2), and (A3), for every $\mathbf{z} \in \mathbb{Z}$ the ETT function satisfies the following two conditions with the costs $\rho_{\mathbf{T}} \triangleq L_Y \|\cdot\|$ on $\mathbb{X}$ and $\rho_{\mathbf{S}} \triangleq \kappa_{\mathbf{z}}^*$ on $S_{\mathbf{z}}^+ \triangleq \{\mathbf{x}' \in S_{\mathbf{z}} : p(\mathbf{x}' | \mathbf{z}) > 0\}$.*

(STR) (Smooth treatment response.) *For all $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{X}$ and $\mathbf{x}' \in S_{\mathbf{z}}^+$,*

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - g_{\mathbf{z}}(\tilde{\mathbf{x}}, \mathbf{x}')| \leq \rho_{\mathbf{T}}(\mathbf{x}, \tilde{\mathbf{x}}). \quad (7)$$

(STS) (Smooth treatment selection.) *For all $\mathbf{x} \in \mathbb{X}$ and $\mathbf{x}'_0, \mathbf{x}'_1 \in S_{\mathbf{z}}^+$,*

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_0) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_1)| \leq \rho_{\mathbf{S}}(\mathbf{x}'_0, \mathbf{x}'_1). \quad (8)$$

(STR) bounds how fast the counterfactual mean of a fixed group of units changes with the intervened treatment. (STS) bounds how fast it changes across the groups selected by different observed treatments, which is where confounding enters: without confounding $g_{\mathbf{z}}$ is constant in $\mathbf{x}'$, and (STS) holds with $\rho_{\mathbf{S}} = 0$. These are the smooth treatment response and smooth treatment selection conditions of Kim et al. (2018), stated conditionally on $\mathbf{z}$.

For each covariate value $\mathbf{z}$, let $\rho_{\mathbf{T}}$ be a pseudometric on $\mathbb{X}$ and $\rho_{\mathbf{S}}$ a pseudometric on $S_{\mathbf{z}}$, both continuous in the Euclidean topology (we suppress their dependence on $\mathbf{z}$), and say that the ETT function satisfies (STR) and (STS) with these costs when the two inequalities of Prop. 2 hold with them for all $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{X}$ and $\mathbf{x}', \mathbf{x}'_0, \mathbf{x}'_1 \in S_{\mathbf{z}}$. We call $\rho_{\mathbf{T}}$ the *response cost* and $\rho_{\mathbf{S}}$ the *selection cost*. The default choices $\rho_{\mathbf{T}} = L_Y \|\cdot\|$ and $\rho_{\mathbf{S}} = C \|\cdot\|$, with constants $L_Y, C > 0$, are the constants b and a of Kim et al. (2018); a general $\rho_{\mathbf{T}}$ extends their modulus of continuity, and a general $\rho_{\mathbf{S}}$ admits the density-adaptive $\kappa_{\mathbf{z}}^*$. Like L_Y and L_X , the costs are sensitivity parameters.

The rest of the paper proceeds as follows. Sec. 3 derives the STR and STS bounds and combines them through every diagonal reference (Thm. 6). Sec. 4 shows that the combined bound of Kim et al. (2018) is a special case of ours and is not sharp in general, whereas ours is the sharp identified set of the CAPO and of the CATE (Sec. 4.1). The bounding results of Sec. 3 are derived under (A1)–(A3) with the costs of Prop. 2, but extend to (A1), (STR), and (STS) with any costs $\rho_{\mathbf{T}}, \rho_{\mathbf{S}}$ as defined above (App. B).

3 LIPSCHITZ BOUNDS ON TREATMENT RESPONSES

By Eq. 3, bounding $\tau(\mathbf{x}, \mathbf{z})$ reduces to bounding the ETT $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}')$, of which only the diagonal is identifiable. We first derive the STR and STS bounds of Kim et al. (2018), conditional on $\mathbf{z}$, under structural conditions; each moves one argument of $g_{\mathbf{z}}$ onto the diagonal. Sec. 3.1 combines them to derive the sharp bound, over all possible pathways to transport an unidentified argument to an identified one on the diagonal.

By consistency, (STR) places each $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}')$ within $\rho_{\mathbf{T}}(\mathbf{x}, \mathbf{x}')$ of $\mu(\mathbf{x}', \mathbf{z}) = g_{\mathbf{z}}(\mathbf{x}', \mathbf{x}')$ (Fig. 2(a)).

Lemma 3. *Under (A1)–(A3), for every $(\mathbf{x}, \mathbf{z}) \in \mathbb{X} \times \mathbb{Z}$, the following inequalities hold:*

$$\underline{w}(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[Y_{\mathbf{x}} | \mathbf{z}] \leq \overline{w}(\mathbf{x}, \mathbf{z}), \quad (9)$$

where, for a draw $\mathbf{X}' \sim P(\mathbf{X} | \mathbf{z})$ of the received treatment, the bounds are

$$\begin{aligned} \underline{w}(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{X}', \mathbf{z}) - \rho_{\mathbf{T}}(\mathbf{x}, \mathbf{X}')) \vee \underline{y} | \mathbf{z}], \\ \overline{w}(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{X}', \mathbf{z}) + \rho_{\mathbf{T}}(\mathbf{x}, \mathbf{X}')) \wedge \overline{y} | \mathbf{z}]. \end{aligned} \quad (10)$$

Lem. 3 is Kim et al. (2018, Prop. 3.1), conditional on $\mathbf{z}$ and intersected with $\mathbb{Y}$. It uses no information at the query, so it holds at every $\mathbf{x} \in \mathbb{X}$, inside or outside $S_{\mathbf{z}}$; without confounding it sharpens to the bound of Khan et al. (2024) (Cor. 12).

(STS) instead moves the received treatment, placing $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}')$ within $\rho_{\mathbf{S}}(\mathbf{x}', \mathbf{x})$ of $\mu(\mathbf{x}, \mathbf{z})$ (Fig. 2(b)), which requires the query to be observed.

Lemma 4. *Under (A1)–(A3), for every $\mathbf{z} \in \mathbb{Z}$ and $\mathbf{x} \in S_{\mathbf{z}}$, the following inequalities hold:*

$$\underline{w}_X(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[Y_{\mathbf{x}} | \mathbf{z}] \leq \overline{w}_X(\mathbf{x}, \mathbf{z}), \quad (11)$$

where, for a draw $\mathbf{X}' \sim P(\mathbf{X} | \mathbf{z})$ of the received treatment, the bounds are

$$\begin{aligned} \underline{w}_X(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{x}, \mathbf{z}) - \rho_{\mathbf{S}}(\mathbf{X}', \mathbf{x})) \vee \underline{y} | \mathbf{z}], \\ \overline{w}_X(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{x}, \mathbf{z}) + \rho_{\mathbf{S}}(\mathbf{X}', \mathbf{x})) \wedge \overline{y} | \mathbf{z}]. \end{aligned} \quad (12)$$

Lem. 4 is Kim et al. (2018, Prop. D.1), conditional on a context $\mathbf{z}$. With costs measured by Euclidean distance between treatments, it beats the STR bound wherever $C < L_Y$. With the density-adaptive cost of Prop. 2 it needs no overlap floor: its averaged band is at most $\mathbb{E}[\kappa_{\mathbf{z}}(\mathbf{x}, \mathbf{X}') | \mathbf{z}] \leq L_X \text{diam}(\mathbb{Y}) \int_{S_{\mathbf{z}}} \|\mathbf{x} - \mathbf{x}'\| d\mathbf{x}'$, since $\max\{p(\mathbf{x} | \mathbf{z}), p(\mathbf{x}' | \mathbf{z})\} \geq p(\mathbf{x}' | \mathbf{z})$, which is finite for bounded $S_{\mathbf{z}}$ at any $\mathbf{x} \in S_{\mathbf{z}}^+$, however small $p(\mathbf{x} | \mathbf{z})$ is.

Example 2 (STR and STS bounds). For the SCM of Ex. 1 ($X \leftarrow -U/2$, $Y \leftarrow -X^2 + U^2$), (A1) requires bounded outcomes, so we truncate U to $[-4, 4]$,

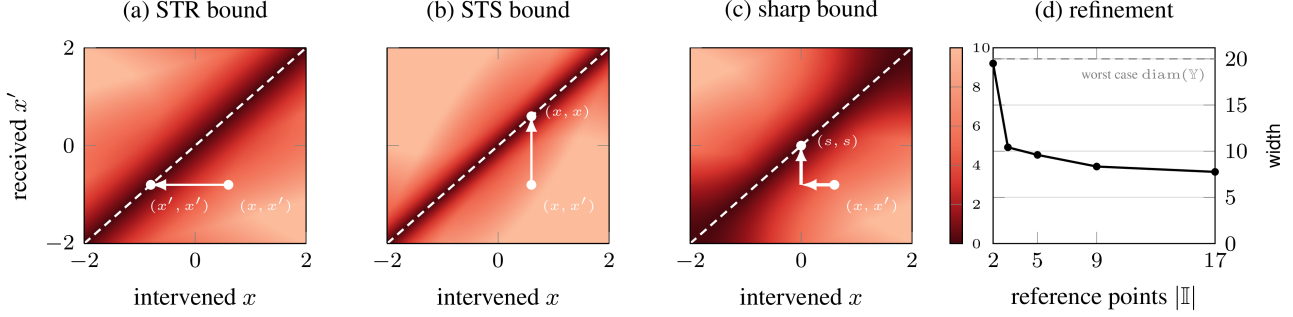


Figure 2: Half-width of the bound on the ETT over the (x, x') plane for Exs. 2 and 3. The identified diagonal $x = x'$ is dashed, and arrows point from each query to the diagonal point bounding it. (a) STR bound ((STR)); (b) STS bound ((STS)); (c) sharp bound over all diagonal reference points (Thm. 10); (d) enlarging a finite reference set $\mathbb{I}$ shrinks the averaged width toward the sharp bound. Panels (a)–(c) saturate at $\frac{1}{2} \text{diam}(\mathbb{Y}) = 10$.

which puts the treatments in $\mathbb{X} = [-2, 2]$ and the outcomes in $\mathbb{Y} = [-4, 16]$. The response functions are $h_u(x) = -x^2 + u^2$, which sweep exactly $\mathbb{Y}$ over that region, so $|h_u(x) - h_u(x')| = |x + x'| |x - x'| \leq 4|x - x'|$ on $[-2, 2]$: the outcome mechanism is 4-Lipschitz, and (STR) holds with $\rho_T = 4|\cdot|$ (Prop. 2). The observed treatments fill $\mathbb{X}$, and along that manifold $\mu(x) = 3x^2 \in [0, 12] \subset \mathbb{Y}$, so the diagonal $\mathbb{E}[Y_x | x] = \mu(x)$ is identified by consistency. The STR bound (Lem. 3) carries these identified diagonal values to a query x , averaging over $P(X | z)$ the interval $[3X'^2 - 4|x - X'|, 3X'^2 + 4|x - X'|] \cap [-4, 16]$. Assumption (A3) cannot hold in this system, since $X = -U/2$ is a deterministic function of U , so we posit (STS) directly, with $\rho_S = C|\cdot|$ and $C = 8$. The STS bound (Lem. 4) instead centers the ETT at the diagonal value at the *query* treatment, $\mu(x) = 3x^2$, averaging the interval $[3x^2 - 8|x - X'|, 3x^2 + 8|x - X'|] \cap [-4, 16]$. Since $C = 8 > L_Y = 4$, here the STR bound is tighter. ■

3.1 Combined Bounds over CAPOs

Each of the STR and STS bounds moves a single argument of the unidentified ETT $g_z(\mathbf{x}, \mathbf{x}')$ onto the identified diagonal: the intervened treatment $\mathbf{x} \mapsto \mathbf{x}'$ in the first case, the received treatment $\mathbf{x}' \mapsto \mathbf{x}$ in the second. Applying both, through the intermediate point $(\mathbf{s}, \mathbf{x}')$, bounds the ETT by the identified value $\mu(\mathbf{s}, \mathbf{z})$ at *any* diagonal reference point $\mathbf{s} \in S_z$: a step $\mathbf{x} \rightarrow \mathbf{s}$ in the intervened treatment at cost $\rho_T(\mathbf{x}, \mathbf{s})$, then a step $\mathbf{x}' \rightarrow \mathbf{s}$ in the received treatment at cost $\rho_S(\mathbf{x}', \mathbf{s})$. Formally, the two smoothness conditions are one Lipschitz condition for the product cost $d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{a}, \mathbf{b})) \triangleq \rho_T(\mathbf{x}, \mathbf{a}) + \rho_S(\mathbf{x}', \mathbf{b})$ on $\mathbb{X} \times S_z$.

Lemma 5. *Under (A1)–(A3), for every $\mathbf{z} \in \mathbb{Z}$ the ETT function is Lipschitz for the product cost: $|g_z(\mathbf{p}) - g_z(\mathbf{q})| \leq d_z(\mathbf{p}, \mathbf{q})$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{X} \times S_z$. In*

particular, for every $\mathbf{x} \in \mathbb{X}$ and $\mathbf{x}', \mathbf{s} \in S_z$,

$$|\mathbb{E}[Y_{\mathbf{x}} | \mathbf{x}', \mathbf{z}] - \mu(\mathbf{s}, \mathbf{z})| \leq \rho(\mathbf{s}, \mathbf{x}') \quad (13)$$

$$\text{with } \rho(\mathbf{s}, \mathbf{x}') \triangleq \rho_T(\mathbf{x}, \mathbf{s}) + \rho_S(\mathbf{x}', \mathbf{s}). \quad (14)$$

Lem. 5 contains both STR and STS bounds: $\mathbf{s} = \mathbf{x}'$ removes the STS step and recovers the pointwise STR bound of Lem. 3, while $\mathbf{s} = \mathbf{x}$ removes the STR step and recovers the pointwise STS bound of Lem. 4. Taking the better of these two at each $\mathbf{x}'$ is the combined STR+STS bound of Kim et al. (2018, Prop. D.3).

Lem. 5 reaches the reference point along an L-shaped path, one step in each argument (Fig. 3, blue). By the triangle inequality for each cost, no path that alternates the two arguments, and no path through several diagonal points, is cheaper (App. B), and Thm. 10 shows more: no argument at all can improve the resulting bound, because it is realizable by a candidate SCM compatible with the observational distribution.

Taking every diagonal point as a reference point gives, for each covariate value $\mathbf{z}$, the *upper and lower bounds*

$$\begin{aligned} U_z(\mathbf{x}, \mathbf{x}') &\triangleq \left(\inf_{\mathbf{s} \in S_z} [\mu(\mathbf{s}, \mathbf{z}) + \rho(\mathbf{s}, \mathbf{x}')] \right) \wedge \bar{y}, \\ L_z(\mathbf{x}, \mathbf{x}') &\triangleq \left(\sup_{\mathbf{s} \in S_z} [\mu(\mathbf{s}, \mathbf{z}) - \rho(\mathbf{s}, \mathbf{x}')] \right) \vee \underline{y}, \end{aligned} \quad (15)$$

with ρ as in Eq. 14. Since $\rho(\mathbf{s}, \mathbf{x}')$ is the product cost $d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{s}, \mathbf{s}))$ from the query to the diagonal point $(\mathbf{s}, \mathbf{s})$, U_z and L_z are the McShane–Whitney extensions of the diagonal values $\mu(\mathbf{s}, \mathbf{z})$ to $\mathbb{X} \times S_z$ (McShane, 1934; Whitney, 1934), truncated to $\mathbb{Y}$; Sec. 4.1 shows that these bounds are extremal.

Theorem 6. *Under (A1)–(A3), for every $\mathbf{z} \in \mathbb{Z}$ the ETT function satisfies $L_z \leq g_z \leq U_z$ on $\mathbb{X} \times S_z$. Consequently, for every $\mathbf{x} \in \mathbb{X}$, including $\mathbf{x} \notin S_z$, and a draw $\mathbf{X}' \sim P(\mathbf{X} | \mathbf{z})$ of the received treatment,*

$$\mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') | \mathbf{z}] \leq \tau(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') | \mathbf{z}]. \quad (16)$$

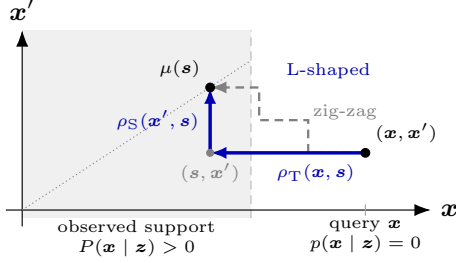


Figure 3: The L-shaped path of Lem. 5 (blue) from (x, x') through (s, x') to the diagonal point (s, s) , at cost $\rho(s, x')$ (Eq. 14). A zig-zag path (grey) costs at least as much; query x may lie outside the support.

For a query $x \notin S_z$, the regression $\mu(x, z)$ is undefined, so the STS bound of Lem. 4 is unavailable, and the combined bound of Kim et al. (2018, Prop. D.3) reduces to STR alone. The bound of Eq. 15 still uses both steps, through reference points $s \in S_z$ near x that trade $\rho_T(x, s)$ against $\rho_S(x', s)$. With the density-adaptive cost of Prop. 2, the selection step runs only between support points, so it stays informative even when little mass lies near the query, i.e., the bound needs no positivity assumption at the query.

Thm. 6 is Lem. 5 applied at every reference point, together with the range constraint. It holds for every SCM satisfying (A1)–(A3), and more generally for every SCM whose ETT function satisfies (STR) and (STS) with arbitrary costs. Over that class, these bounds cannot be improved (Sec. 4.1).

In practice, we compute the bounds over a finite set of reference points in the support, where μ is identified; restricting the references keeps the bounds valid.

Corollary 7. *Under (A1)–(A3), let $\mathbb{I} \subseteq S_z$ be a nonempty reference set, and let $U_{z, \mathbb{I}}$ and $L_{z, \mathbb{I}}$ be the bounds of Eq. 15 with the infimum and supremum restricted to $\mathbb{I}$. Then $L_{z, \mathbb{I}} \leq L_z \leq U_z \leq U_{z, \mathbb{I}}$, so that*

$$\mathbb{E}[L_{z, \mathbb{I}}(x, X') \mid z] \leq \tau(x, z) \leq \mathbb{E}[U_{z, \mathbb{I}}(x, X') \mid z] \quad (17)$$

for every $x \in \mathbb{X}$. The bounds tighten as $\mathbb{I}$ grows. For finite $\mathbb{I}$, they cost $O(n|\mathbb{I}|)$ per query when $P(X \mid z)$ is an empirical law on n points.

Since the STR and STS bounds use the references $s = x'$ and $s = x$, the combined bound is never looser than either; Cor. 14 in App. B states when it is tighter.

Example 3 (Combined bound). Continuing Ex. 2 ($\rho_T = 4|\cdot|$, $\rho_S = 8|\cdot|$, $\mu(s) = 3s^2$ on $S = [-2, 2]$). The bound of Thm. 6 takes the best diagonal reference for each pair (x, x') (Fig. 2(c)), and an intermediate point can beat both the STR and the STS bound. At $(x, x') = (-1, 1)$, the STR bound gives the upper value $\mu(1) + 4 \cdot 2 = 11$ and the STS bound

$\mu(-1) + 8 \cdot 2 = 19$, whereas the reference $s = 2/3$ gives $\frac{4}{3} + \frac{20}{3} + \frac{8}{3} = \frac{32}{3} \approx 10.67$. Averaging over $P(X \mid z)$ gives the sharp bound over true response $\tau(x) = \mathbb{E}[U^2] - x^2 \approx 1 - x^2$; finite references (Cor. 7) give widths of Fig. 2(d), which shrink as $\mathbb{I}$ grows. ■

4 SHARPNESS ANALYSIS

Thm. 6 shows that the bound is valid. This section shows that it cannot be improved. Sharpness is relative to a model class, and we establish it for the class defined by (STR) and (STS) with the general costs of Sec. 2.1. Let $\mathcal{M}(\rho_T, \rho_S)$ be the class of SCMs satisfying (A1) whose ETT functions satisfy (STR) and (STS); by Prop. 2, it contains every SCM satisfying (A1)–(A3) with the costs of that proposition. Given the observational law, the *identified set* of $\tau(x, z)$ is

$$\Theta(x, z) \triangleq \{ \mathbb{E}_{\mathcal{M}}[Y_x \mid z] : \mathcal{M} \in \mathcal{M}(\rho_T, \rho_S), \\ P_{\mathcal{M}}(X, Z, Y) = P(X, Z, Y) \}. \quad (18)$$

We first compare our bound with the combined bound of Kim et al. (2018), which uses only two of its reference points and can be tightened. We then show that our bound cannot: it gives $\Theta(x, z)$ exactly, and it yields the sharp identified set of the CATE (Sec. 4.1).

Kim et al. (2018) combine the two conditions by taking, at each received treatment, the better of the STR and STS bounds of Sec. 3. This restricts the bounds in Eq. 15 to two reference points.

Proposition 8. *Let $\mathbb{X} \subset \mathbb{R}$, without covariates, and $\rho_T = b|\cdot|$, $\rho_S = a|\cdot|$. For any query $x \in S$, the combined STR+STS upper bound of Kim et al. (2018, Prop. D.3) gives the following*

$$u_2(x) \triangleq \mathbb{E}[\min\{\mu(x) + a|X' - x|, \mu(X') + b|X' - x|\}],$$

which is the bound of Cor. 7 with the two-point reference set $\{x, x'\}$ chosen separately at each received treatment x' (without the intersection with $\mathbb{Y}$).

Dropping the other reference points costs sharpness.

Proposition 9. *The bound u_2 of Prop. 8 is not sharp in general: for some costs, observational distribution, and query $x \in S$, $u_2(x)$ strictly exceeds $\mathbb{E}[Y_x]$ in every SCM satisfying the STR and STS conditions with these costs and reproducing the observational distribution.*

For example, let $a = b = 1$, let the treatment be uniform on $S = \{0, 1, 2\}$, let $Y = \mu(X)$ with $\mu = (0, -1, 0)$, and take the query $x = 2$. Table 1 lists the terms $\mu(s) + |2 - s| + |x' - s|$. The two-point bound uses only the columns $s \in \{2, x'\}$, which give the values $(2, 0, 0)$ across $x' = 0, 1, 2$ and hence $u_2(x = 2) = 2/3$. The reference point $s = 1$ lowers the value at $x' = 0$

x'	$s = 0$	$s = 1$	$s = 2$	two-point	sharp
0	2	1	2	2	1
1	3	0	1	0	0
2	4	1	0	0	0

Table 1: The counterexample of Prop. 9 at the query $x = 2$: terms $\mu(s) + |2 - s| + |x' - s|$, the two-point bound of Kim et al. (2018) (columns $s \in \{2, x'\}$), and the sharp bound. Averaging over P gives $2/3$ and $1/3$.

from 2 to 1: every such model has $\mathbb{E}[Y_{x=2}] \leq 1/3$, and App. B constructs one that attains this upper bound. The sharpness argument for Prop. D.3 in Kim et al. (2018) constructs a data-generating process whose regression μ is constant. With a constant regression and these costs, $\inf_s [b|x - s| + a|x' - s|] = \min\{a, b\}|x - x'|$ is attained at $s \in \{x, x'\}$, so the two-point bound coincides with our bound; for a nonconstant regression, as here, it does not and can be generally improved.

4.1 Sharp Identified Set

The diagonal of any function satisfying (STR) and (STS) is Lipschitz for $\rho_T + \rho_S$: for all $\mathbf{s}, \mathbf{s}' \in S_z$,

$$|\mu(\mathbf{s}, \mathbf{z}) - \mu(\mathbf{s}', \mathbf{z})| \leq \rho_T(\mathbf{s}, \mathbf{s}') + \rho_S(\mathbf{s}, \mathbf{s}'), \quad (19)$$

and in particular it is continuous. This is a *testable implication* of the model, and it is the only restriction the model places on the observational law (Cor. 17 in App. B). When it holds, Thm. 10 characterizes $\Theta(\mathbf{x}, \mathbf{z})$ exactly at every query, including treatments $\mathbf{x}$ outside the observational support at $\mathbf{z}$.

Theorem 10. *Under (A1), (STR), and (STS), fix $\mathbf{z} \in \mathbb{Z}$ and suppose $\mu(\cdot, \mathbf{z})$ satisfies Eq. 19. Then:*

- (a) (Attainment) U_z and L_z take values in $\mathbb{Y}$, satisfy the inequalities of (STR) and (STS), and equal $\mu(\mathbf{s}, \mathbf{z})$ at every diagonal point $(\mathbf{s}, \mathbf{s})$;
- (b) (Sharp set) for every $\mathbf{x} \in \mathbb{X}$, including $\mathbf{x} \notin S_z$,

$$\Theta(\mathbf{x}, \mathbf{z}) = \left[\mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}], \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] \right]. \quad (20)$$

If Eq. 19 fails on a set of covariate values of positive probability, no SCM in $\mathcal{M}(\rho_T, \rho_S)$ reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$, and the model is refuted.

Part (a) holds because an infimum of d_z -Lipschitz functions is d_z -Lipschitz, while Eq. 19 pins the bounds to the data on the diagonal. So the McShane–Whitney extensions U_z and L_z of Sec. 3.1 are the largest and the smallest functions with values in $\mathbb{Y}$ that satisfy (STR) and (STS) and agree with μ on the diagonal. Part (b) constructs, for every admissible ETT function, an SCM that realizes it while reproducing the observa-

tional distribution; mixing the two bounds then sweeps the whole interval (App. B).

The CATE between two treatment levels, $\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z}) = \mathbb{E}[Y_{\mathbf{x}_1} - Y_{\mathbf{x}_0} \mid \mathbf{z}]$, is the difference of two CAPOs, and the bounds U_z and L_z combine with (STR) into its sharp identified set.

Theorem 11. *Let $\mathbf{x}_1, \mathbf{x}_0 \in \mathbb{X}$, including treatments outside S_z , let $r \triangleq \rho_T(\mathbf{x}_1, \mathbf{x}_0)$, and let $\mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$ denote the interval*

$$\left[-\mathbb{E}[\min\{U_z(\mathbf{x}_0, \mathbf{X}') - L_z(\mathbf{x}_1, \mathbf{X}'), r\} \mid \mathbf{z}], \mathbb{E}[\min\{U_z(\mathbf{x}_1, \mathbf{X}') - L_z(\mathbf{x}_0, \mathbf{X}'), r\} \mid \mathbf{z}] \right]. \quad (21)$$

- (a) (Validity) Under (A1)–(A3), $\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z}) \in \mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$.
- (b) (Sharpness) Under (A1), (STR), and (STS), if $\mu(\cdot, \mathbf{z})$ satisfies Eq. 19, then $\mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$ is exactly the identified set of the CATE over the SCMs in $\mathcal{M}(\rho_T, \rho_S)$ that reproduce $P(\mathbf{X}, \mathbf{Z}, Y)$.

Differencing the two CAPO bounds of Thm. 6 gives the upper value $\mathbb{E}[U_z(\mathbf{x}_1, \mathbf{X}') - L_z(\mathbf{x}_0, \mathbf{X}') \mid \mathbf{z}]$, which is valid but loose, because the upper and lower bounds are attained by different ETT functions. The cap at r is exactly the coupling that differencing misses: (STR) ties the two treatment levels together within each group of units. (STR) alone gives only the data-free interval $\pm r$ (Kim et al., 2018, Prop. 3.2), and Kim et al. (2018) obtain informative CATE bounds by adding monotonicity. Thm. 11 is informative wherever $U_z(\mathbf{x}_1, \cdot) - L_z(\mathbf{x}_0, \cdot) < r$ on a set of received treatments of positive probability, and needs no monotonicity.

5 EXPERIMENTS

We compare our finite-reference bound (Cor. 7) with the combined STR+STS bound of Kim et al. (2018), the continuous marginal sensitivity model (CMSM; Jesson et al., 2022), and the Lipschitz-envelope adaptation of Khan et al. (2024). Table 2 reports coverage of known model means and interval width divided by the outcome range, averaged over queries within each setting. Ours and Kim receive identical regressions, Lipschitz costs, and outcome ranges. Kim uses the per-pair endpoints; ours searches the supported reference points before clipping. CMSM uses one target-based coverage oracle per SCM or IST; Khan is evaluated under violated unconfoundedness. Experimental setup and calibration details are provided in App. C.

Discrete SCMs. We randomly draw 500 SCMs with $U \in \{-1, +1\}$ and $K \in \{5, 7, 9\}$ equally spaced treatments on $[-1, 1]$. Assignment follows $P(X = x_j \mid U = u) = (1 + \rho u x_j)/K$, and outcomes are affine rescalings

Table 2: Query-averaged coverage and interval width normalized by the outcome range. Reduction compares pooled mean widths with Kim. CMSM uses a per-model coverage oracle, so its coverage holds by construction.

Setting	Ours		Kim et al.		CMSM oracle		Khan et al.		Reduction vs. Kim (%)
	cov.	width	cov.	width	cov.	width	cov.	width	
Discrete SCMs	1.00	0.11181	1.00	0.13832	1.00	0.04810	0.14	0.00000	19.17
Continuous SCMs	1.00	0.10627	1.00	0.13071	1.00	0.04841	0.04	0.00000	18.70
IST	1.00	0.03726	1.00	0.03777	1.00	0.07558	0.04	0.00000	1.34

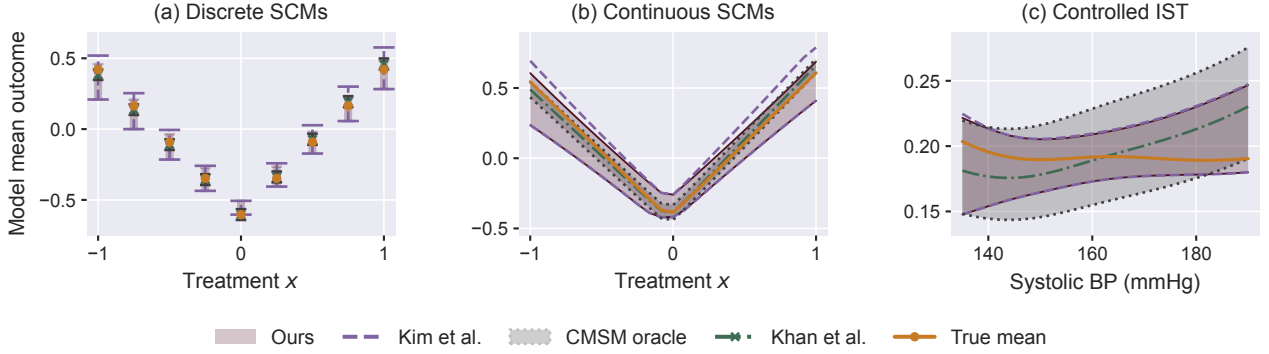


Figure 4: Example bounds and true means for discrete, smooth V-shaped continuous, and controlled IST models. Wine and gray shading denote ours and oracle CMSM; dashed and dotted boundaries denote Kim and CMSM. Discrete intervals use bars. Panels have separate outcome scales; example selection is specified in App. C.

of $b|x| + cU$ to the attained range $[-1, 1]$. We perturb the response and assignment parameters. Our mean width is 19.17% smaller than Kim’s, with full coverage for both. The gain comes from reference points near the interior response minimum, which Kim’s per-pair endpoints $\{x, x'\}$ omit.

Continuous SCMs. We draw 250 SCMs from each of two families, a V-shaped response and its smooth approximation. Both use a binary hidden confounder and the density $p(x | u) = (1 + \rho ux)/2$ on $[-1, 1]$. We perturb response scales, minimum locations, latent weights, and assignment strength. Across 12500 queries, our pooled mean width is 18.70% smaller than Kim’s; both methods achieve full coverage.

Controlled IST. Within three age tertiles of the International Stroke Trial (Carolei et al., 1997), consciousness defines a hidden categorical confounder. Smoothed binary death risks define a new SCM with known, overlapping assignment on 25 blood-pressure values per group. Target effects are computed from this model. Across 75 targets, our pooled width reduction is 1.34%, with full coverage for ours and Kim. Age-group reductions are 0.47%, 1.32%, and 1.76%. At the first, middle, and last query of each age group, a linear program over the ETT matrix, using only the observed diagonal, STR, STS, and the outcome range, reproduces our endpoints to within 10^{-8} .

Fig. 4 shows visible tightening over Kim in the synthetic examples and nearly coincident intervals in IST. Between oracle CMSM and ours, neither is uniformly better: CMSM is narrower in all synthetic families, while ours is narrower in all three IST groups. CMSM’s full coverage, however, relies on a target-based oracle Λ^* , calibrated using the known interventional target. Khan assumes unconfoundedness and, as expected, collapses to the observational regression.

6 CONCLUSIONS

We studied partial identification of CAPOs and CATEs for continuous treatments under unobserved confounding. Lipschitz response functions and Lipschitz propensity scores given the latent confounder imply the STR and STS conditions of Kim et al. (2018), with a selection cost that depends only on the observed treatment density. Under STR and STS with general costs, bounding through every diagonal reference point gives the sharp identified set of the CAPO at every treatment, inside or outside the observational support, whereas the two-point combined bound of Kim et al. (2018) is not sharp in general. Capping the difference of the upper and lower bounds at the response cost gives the sharp identified set of the CATE, without any monotonicity assumption.

References

- Baitairian, J.-B., Sebastien, B., Jreich, R., Katsahian, S., and Guilloux, A. (2025). Sharp bounds for continuous-valued treatment effects with unobserved confounders. *Biometrical Journal*, 67(5):e70084.
- Balazadeh Meresht, V., Syrgkanis, V., and Krishnan, R. G. (2022). Partial identification of treatment effects with implicit generative models. In *Advances in Neural Information Processing Systems*, volume 35, pages 22816–22829.
- Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1172–1176.
- Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973.
- Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. (2022). On pearl’s hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl*, pages 507–556.
- Bareinboim, E. and Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113:7345–7352.
- Bonvini, M., Kennedy, E. H., Ventura, V., and Wasserman, L. (2022). Sensitivity analysis for marginal structural models. *arXiv preprint arXiv:2210.04681*.
- Carolei, A. et al. (1997). The international stroke trial (ist): a randomized trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. *The Lancet*, 349:1569–1581.
- Chen, B., Kumor, D., and Bareinboim, E. (2017). Identification and model testing in linear structural equation models using auxiliary variables. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 757–766. PMLR.
- Dalal, A. and Tchetgen Tchetgen, E. J. (2025). Partial identification of causal effects for endogenous continuous treatments. In *arXiv preprint arXiv:2508.13946*.
- Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pages 1097–1104. Omnipress.
- Frangakis, C. and Rubin, D. (2002). Principal stratification in causal inference. *Biometrics*, 1(58):21–29.
- Frauen, D., Melnychuk, V., and Feuerriegel, S. (2023). Sharp bounds for generalized causal sensitivity analysis. In *Advances in Neural Information Processing Systems*, volume 36, pages 40556–40586.
- Gunsilius, F. (2020). A path-sampling method to partially identify causal effects in instrumental variable models. *arXiv preprint arXiv:1910.09502*.
- Hausman, J. A. and Newey, W. K. (2016). Individual heterogeneity and average welfare. *Econometrica*, 84(3):1225–1248.
- Huang, Y. and Valtorta, M. (2006). Pearl’s calculus of intervention is complete. In Dechter, R. and Richardson, T., editors, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 217–224. AUAI Press, Corvallis, OR.
- Jesson, A., Douglas, A., Manshausen, P., Solal, M., Meinshausen, N., Stier, P., Gal, Y., and Shalit, U. (2022). Scalable sensitivity and uncertainty analyses for causal-effect estimates of continuous-valued interventions. In *Advances in Neural Information Processing Systems*, volume 35, pages 13892–13907.
- Kallus, N. and Zhou, A. (2018). Confounding-robust policy improvement. In *Advances in neural information processing systems*, pages 9269–9279.
- Khan, S., Saveski, M., and Ugander, J. (2024). Off-policy evaluation beyond overlap: Sharp partial identification under smoothness. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*. PMLR.
- Kilbertus, N., Kusner, M. J., and Silva, R. (2020). A class of algorithms for general instrumental variable models. In *Advances in Neural Information Processing Systems*, volume 33.
- Kim, W., Kwon, K., Kwon, S., and Lee, S. (2018). The identification power of smoothness assumptions in models with counterfactual outcomes. *Quantitative Economics*, 9(2):617–642.
- Kocaoglu, M., Snyder, C., Dimakis, A. G., and Vishwanath, S. (2018). Causalgan: Learning causal implicit generative models with adversarial training. In *International Conference on Learning Representations*.
- Li, L., Munos, R., and Szepesvari, C. (2015). Toward minimax off-policy value estimation. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*.
- Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.
- Manski, C. F. and Pepper, J. V. (1998). Monotone instrumental variables with an application to the returns to schooling.

-
- Marmarelis, M. G., Haddad, E., Jesson, A., Jahan-shad, N., Galstyan, A., and Ver Steeg, G. (2023). Partial identification of dose responses with hidden confounders. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 1368–1379. PMLR.
- McShane, E. J. (1934). Extension of range of functions. *Bulletin of the American Mathematical Society*, 40(12):837–842.
- Munos, R., Stepleton, T., Harutyunyan, A., and Bellemare, M. (2016). Safe and efficient off-policy reinforcement learning. In *Advances in Neural Information Processing Systems*.
- Namkoong, H., Keramati, R., Yadlowsky, S., and Brunskill, E. (2020). Off-policy policy evaluation for sequential decisions under unobserved confounding. *Advances in Neural Information Processing Systems*, 33:18819–18831.
- Nasr-Esfahany, A., Alizadeh, M., and Shah, D. (2023). Counterfactual identifiability of bijective causal models. In *International Conference on Machine Learning*, pages 25733–25754. PMLR.
- Padh, K., Zeitler, J., Watson, D., Kusner, M., Silva, R., and Kilbertus, N. (2023). Stochastic causal programming for bounding treatment effects. In *Conference on Causal Learning and Reasoning*, pages 142–176. PMLR.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, NY. 2nd edition, 2009.
- Rosenbaum, P. and Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.
- Sandercock, P., Niewada, M., and Członkowska, A. (2011). International stroke trial database (version 2). Edinburgh DataShare, University of Edinburgh, Department of Clinical Neurosciences. <https://doi.org/10.7488/ds/104>.
- Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1219–1226.
- Tan, J., Blanchet, J., and Syrgkanis, V. (2024). Consistency of neural causal partial identification. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37.
- Thomas, P. and Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning*, pages 2139–2148.
- Wang, Y. and Blei, D. M. (2019). The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596.
- Whitney, H. (1934). Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89.
- Wright, S. (1934). The method of path coefficients. *The annals of mathematical statistics*, 5(3):161–215.
- Xia, K., Lee, K.-Z., Bengio, Y., and Bareinboim, E. (2021). The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 34:10823–10836.
- Xia, K., Pan, Y., and Bareinboim, E. (2022). Neural causal models for counterfactual identification and estimation. In *Proceedings of the 11th Eleventh International Conference on Learning Representations*.
- Zhang, J., Chen, J., and Bareinboim, E. (2026). Causal gaussian processes for robust treatment effect evaluation with unobserved confounding. *Syracuse Causal Science Laboratory, Technical Report (R-21)*.
- Zhang, J., Tian, J., and Bareinboim, E. (2022). Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, pages 26548–26558. PMLR.

Sharp Bounds on Continuous Treatment Effects under Unobserved Confounding:

Supplementary Materials

A RELATED WORK

This section expands the discussion of the Introduction, placing our bound among the main lines of work on identification, partial identification, and sensitivity analysis, and stating precisely how it relates to each.

Identification and adjustment. When the target effect is point-identified, for example under the back-door or ignorability condition (Pearl, 2000; Rosenbaum and Rubin, 1983), it can be recovered from observational data by covariate adjustment (Rosenbaum and Rubin, 1983; Bang and Robins, 2005) or, in the sequential and off-policy settings, by importance-weighted and doubly robust estimators (Dudík et al., 2011; Li et al., 2015; Munos et al., 2016; Thomas and Brunskill, 2016). Whether a given effect is identified at all is itself settled by the do-calculus, which is complete for interventional distributions (Huang and Valtorta, 2006; Shpitser and Pearl, 2006; Bareinboim et al., 2022). Our setting begins where these tools end: unobserved confounding leaves the CAPO non-identified, so no adjustment functional exists, and the effect can only be bounded.

Partial identification for discrete treatments. The classical response to non-identification is to bound the effect over all models compatible with the data (Manski, 1990; Balke and Pearl, 1997; Frangakis and Rubin, 2002). Manski (1990) gives the worst-case interval that assumes nothing about the unobserved response, which our Cor. 15 recovers as the no-smoothness corner of the family. Balke and Pearl (1997) tighten it under an instrument by linear programming over a discretized latent domain, and Zhang et al. (2022) extend such programs to counterfactual queries and to Monte-Carlo bounds for discrete outcomes; Manski and Pepper (1998) instead impose a monotone instrumental-variable structure. These bounds are sharp, but they rest on enumerating a finite latent domain, and that enumeration does not survive the move to continuous treatments and outcomes, which is the regime we target.

Sensitivity analysis and bounded confounding. A complementary line fixes a quantitative limit on how far confounding can distort the treatment mechanism and reports the resulting interval. For binary and discrete treatments, the marginal sensitivity model and its relatives bound the ratio between the observed and the latent-conditional propensity (Kallus and Zhou, 2018; Namkoong et al., 2020). For continuous treatments, Jesson et al. (2022) introduce the continuous marginal sensitivity model (CMSM), which we adopt as our main baseline, and bound both the CAPO $\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{z}]$ and its average. Frauen et al. (2023) derive sharp bounds, pointwise in the covariates, for a generalized sensitivity model that contains the CMSM, and Baitairian et al. (2025) give sharp closed-form CMSM bounds with a doubly robust estimator. Marmarelis et al. (2023) instead bound how fast the log density ratio may change with the treatment (the δ -MSM), Dalal and Tchetgen Tchetgen (2025) let the sensitivity parameter vary with the exposure level, and Bonvini et al. (2022) develop propensity-based, outcome-based, and subset-confounding sensitivity models for marginal structural models. Several of these methods bound conditional, pointwise effects; what separates them from ours is the form of the restriction. They constrain the treatment density relative to its observed value and are stated where that density is positive, so their intervals are built from the observed outcomes at the query and do not extend to treatments outside the support. Our restriction is instead on the smoothness of the mechanisms.

Identification under parametric structure. When the mechanisms are constrained, the effect can sometimes be point-identified despite confounding. Linear structural equation models identify effects through path coefficients (Wright, 1934) and auxiliary variables (Chen et al., 2017), and the deconfounder exploits many causes sharing a latent factor (Wang and Blei, 2019). We make no such parametric commitment. Our inputs are smoothness costs, and Zhang et al. (2026) show that smooth-mechanism SCMs are dense among the confounded models compatible with the data, matching the observational and every interventional law to arbitrary accuracy. Smoothness as such therefore rules out no observational law; the substantive, and untestable beyond Cor. 17,

assumption is the size of the constants, which we treat as sensitivity parameters.

Optimization-based and generative bounds. A further line computes bounds numerically, by optimizing over a class of structural models consistent with the data. For continuous treatments with an instrument, Kilbertus et al. (2020) optimize over response functions spanned by a chosen basis, Padh et al. (2023) extend this to instruments and leaky mediators by stochastic constrained optimization, and Gunsilius (2020) solves the resulting infinite-dimensional programs by path sampling. Balazadeh Meresht et al. (2022) fit implicit generative models to general causal graphs and bound average treatment derivatives by optimizing over them, and neural or generative SCMs are used in the same way elsewhere (Kocaoglu et al., 2018; Xia et al., 2021, 2022; Nasr-Esfahany et al., 2023). Tan et al. (2024) analyze when such neural estimates are consistent, and make explicit that an informative bound over continuous domains is impossible without positing a smoothness assumption, since two models can match the data yet disagree arbitrarily on a rarely observed treatment. The bounds in this line are valid relative to the chosen function class and are the output of a nonconvex optimization; ours is an explicit bound under stated Lipschitz constants.

Smoothness assumptions on counterfactual means. Closest to our work is a line in econometrics that bounds counterfactual means by limiting their rate of change. Hausman and Newey (2016) bound each individual’s income effect to partially identify average welfare under unrestricted heterogeneity. Following them, Kim et al. (2018) study $g(x, x') = \mathbb{E}[Y_x \mid X = x']$ for a scalar treatment, assume that g is Lipschitz in the intervened treatment (smooth treatment response, STR: $|g(x, x') - g(\tilde{x}, x')| \leq b|x - \tilde{x}|$) or in the received treatment (smooth treatment selection, STS: $|g(x, x'_0) - g(x, x'_1)| \leq a|x'_0 - x'_1|$), and derive closed-form bounds on $\mathbb{E}[Y_x]$ under each assumption, under both, and in combination with monotonicity; their bounds under monotone treatment response and selection (Props. 4.1–4.2) already route through intermediate treatment levels, with a monotone rather than Lipschitz selection step. Our ETT $g_z(\mathbf{x}, \mathbf{x}')$ is their g with covariates $\mathbf{z}$, our primary assumptions (STR) and (STS) are theirs with general costs, and our STR and STS bounds correspond to their propositions:

This paper	Kim et al. (2018)
Lem. 3 (intervened treatment)	Prop. 3.1 (STR)
Lem. 4 (received treatment)	Prop. D.1 (STS)
Cor. 7 with $\mathbb{I} = \{\mathbf{x}, \mathbf{x}'\}$ at each $\mathbf{x}'$	Prop. D.3 (STR and STS)
Thm. 10 (all reference points)	–

Our contribution is therefore not a new bounding principle: the STR and STS bounds and their combination go back to Kim et al. (2018), and the combined bound is a standard Lipschitz-extension bound (McShane, 1934). Relative to their work, five things are new. *First, sharpness.* Prop. D.3 of Kim et al. (2018) is stated as sharp, but its proof constructs a data-generating process with a constant regression, for which intermediate reference points never help. For a nonconstant regression the two-point bound is not sharp (Prop. 9), whereas our bound is the sharp identified set (Thm. 10); we also give the exact condition under which the full reference set is strictly tighter (Cor. 14) and the testable implication of the model (Cor. 17). *Second, the selection condition.* Kim et al. (2018) assume STS directly and motivate it informally; we derive it (Prop. 2) from a Lipschitz assignment density given the latent confounder, via a total-variation bound on the latent posteriors (Lem. 1), with the identified, density-adaptive cost κ_z^* , which needs no overlap floor. *Third, the scope.* We allow covariates, with effects conditional on them, and vector-valued treatments, and we use the bounded outcome range inside the bound, whereas their main results concern a scalar treatment without covariates and with unbounded outcomes. *Fourth, extrapolation.* Their STR bound is valid at treatments outside the support, but their combined bound then reduces to it, because the STS term needs the regression at the query; our bound still uses both steps there and remains the sharp identified set (Thm. 10). *Fifth, the CATE.* Under STR alone their CATE bound is the data-free interval $\pm b|t - t'|$ (Prop. 3.2), and their informative CATE bounds add monotonicity; under STR and STS, Thm. 11 gives the sharp identified set of the CATE without it.

Smoothness and off-policy evaluation beyond overlap. Khan et al. (2024) bound the off-policy value under a Lipschitz smoothness assumption through closed-form linear programs, targeting the unconfounded setting with limited or no overlap between the logging and target policies. Our setting differs in that confounding renders the effect non-identified even under full overlap. In the unconfounded limit our STR bound reduces to their bound (Cor. 12), so their result is the no-confounding case of ours; the STS bound, absent from their analysis, is what accounts for confounding breaking the equality of the observed regression and the true effect.

B PROOFS

This appendix collects the proofs of the theoretical results in the main text. App. B.1 proves the structural sufficient conditions (Lem. 1 and Prop. 2) that connect (A2) and (A3) to the STR and STS conditions; App. B.2 proves the STR and STS bounds (Lems. 3 and 4) and their unconfounded special case (Cor. 12); App. B.3 proves the combined bound (Lem. 5 and Thm. 6) together with its finite-reference corollary (Cor. 7) and the strictness and Manski corollaries (Cors. 14 and 15); and App. B.4 proves sharpness (Thms. 10 and 11, Cor. 17) and the comparison with the bound of Kim et al. (2018) (Props. 8 and 9). All validity proofs follow the same playbook: the ETT function $g_z(\mathbf{x}, \mathbf{x}')$ is bounded pointwise by chaining consistency on the diagonal with the STR and STS inequalities, and the pointwise bound is then averaged over the received treatment $\mathbf{X}' \sim P(\mathbf{X} \mid \mathbf{z})$ through the tower property $\tau(\mathbf{x}, \mathbf{z}) = \mathbb{E}[g_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}]$ of Eq. 3. The sharpness proofs run in the reverse direction: they exhibit an admissible ETT function that attains the bound and realize it by an SCM that reproduces the observational law.

Throughout, statements about a fixed covariate value $\mathbf{z}$ hold for P -almost every $\mathbf{z}$, with regular versions of the conditional laws, and all integrated functions of $(\mathbf{z}, \mathbf{x}, \mathbf{x}')$ are jointly measurable. Apart from Lem. 1 and Prop. 2, the proofs use only (A1), (STR), (STS), and consistency, with arbitrary costs; by Prop. 2 and the convention of Sec. 2.1, this covers (A1)–(A3), with S_z replaced by S_z^+ , which carries all of $P(\mathbf{X} \mid \mathbf{z})$ and whose points have neighbourhoods of positive mass (see the proof of Prop. 2).

B.1 Structural sufficient conditions

Lemma 1. *Under (A3), for every $\mathbf{z} \in \mathbb{Z}$ and every pair $\mathbf{x}'_0, \mathbf{x}'_1 \in \mathbb{X}$ with $p(\mathbf{x}'_0 \mid \mathbf{z}) > 0$ and $p(\mathbf{x}'_1 \mid \mathbf{z}) > 0$,*

$$\text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \leq \frac{L_X \|\mathbf{x}'_0 - \mathbf{x}'_1\|}{\max\{p(\mathbf{x}'_0 \mid \mathbf{z}), p(\mathbf{x}'_1 \mid \mathbf{z})\}}, \quad (5)$$

where $\text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \triangleq \text{TV}(P(\mathbf{U} \mid \mathbf{x}'_0, \mathbf{z}), P(\mathbf{U} \mid \mathbf{x}'_1, \mathbf{z}))$ is the total variation between latent posteriors.

Proof. Fix $\mathbf{z}$ and write

$$q \triangleq P(\mathbf{U} \mid \mathbf{z}), \quad p_{\mathbf{u}}(\mathbf{x}) \triangleq p(\mathbf{x} \mid \mathbf{z}, \mathbf{u}), \quad p(\mathbf{x}) \triangleq p(\mathbf{x} \mid \mathbf{z}) = \int p_{\mathbf{u}}(\mathbf{x}) dq(\mathbf{u}), \quad \Delta \triangleq \|\mathbf{x}'_0 - \mathbf{x}'_1\|.$$

Each $p_{\mathbf{u}}$ is L_X -Lipschitz by (A3). Since p is an average of these functions, it is L_X -Lipschitz as well:

$$|p(\mathbf{x}'_0) - p(\mathbf{x}'_1)| \leq \int |p_{\mathbf{u}}(\mathbf{x}'_0) - p_{\mathbf{u}}(\mathbf{x}'_1)| dq(\mathbf{u}) \leq L_X \Delta.$$

By Bayes' rule, whenever $p(\mathbf{x}') > 0$,

$$dP(\mathbf{u} \mid \mathbf{x}', \mathbf{z}) = \frac{p_{\mathbf{u}}(\mathbf{x}')}{p(\mathbf{x}')} dq(\mathbf{u}).$$

Hence the total variation between the two latent posteriors is

$$2 \text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) = \int \left| \frac{p_{\mathbf{u}}(\mathbf{x}'_0)}{p(\mathbf{x}'_0)} - \frac{p_{\mathbf{u}}(\mathbf{x}'_1)}{p(\mathbf{x}'_1)} \right| dq(\mathbf{u}).$$

Adding and subtracting $p_{\mathbf{u}}(\mathbf{x}'_1)/p(\mathbf{x}'_0)$ inside the absolute value and applying the triangle inequality, we obtain

$$2 \text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \leq \frac{1}{p(\mathbf{x}'_0)} \int |p_{\mathbf{u}}(\mathbf{x}'_0) - p_{\mathbf{u}}(\mathbf{x}'_1)| dq(\mathbf{u}) + \left| \frac{1}{p(\mathbf{x}'_0)} - \frac{1}{p(\mathbf{x}'_1)} \right| \int p_{\mathbf{u}}(\mathbf{x}'_1) dq(\mathbf{u}).$$

By (A3), the first term is at most $L_X \Delta / p(\mathbf{x}'_0)$. Since $\int p_{\mathbf{u}}(\mathbf{x}'_1) dq(\mathbf{u}) = p(\mathbf{x}'_1)$, the second term equals

$$\left| \frac{1}{p(\mathbf{x}'_0)} - \frac{1}{p(\mathbf{x}'_1)} \right| p(\mathbf{x}'_1) = \frac{|p(\mathbf{x}'_1) - p(\mathbf{x}'_0)|}{p(\mathbf{x}'_0)} \leq \frac{L_X \Delta}{p(\mathbf{x}'_0)},$$

where the inequality uses the Lipschitz property of p . Combining the two terms gives

$$2 \text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}) \leq \frac{L_X \Delta}{p(\mathbf{x}'_0)} + \frac{L_X \Delta}{p(\mathbf{x}'_0)} = \frac{2L_X \Delta}{p(\mathbf{x}'_0)}.$$

Exchanging the roles of $\mathbf{x}'_0$ and $\mathbf{x}'_1$ gives the same bound with $p(\mathbf{x}'_1)$ in the denominator. Taking the better of the two bounds gives Eq. 5. This completes the proof. $\square$

Proposition 2. *Under (A1), (A2), and (A3), for every $\mathbf{z} \in \mathbb{Z}$ the ETT function satisfies the following two conditions with the costs $\rho_T \triangleq L_Y \|\cdot\|$ on $\mathbb{X}$ and $\rho_S \triangleq \kappa_{\mathbf{z}}^*$ on $S_{\mathbf{z}}^+ \triangleq \{\mathbf{x}' \in S_{\mathbf{z}} : p(\mathbf{x}' | \mathbf{z}) > 0\}$.*

(STR) (Smooth treatment response.) *For all $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{X}$ and $\mathbf{x}' \in S_{\mathbf{z}}^+$,*

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - g_{\mathbf{z}}(\tilde{\mathbf{x}}, \mathbf{x}')| \leq \rho_T(\mathbf{x}, \tilde{\mathbf{x}}). \quad (7)$$

(STS) (Smooth treatment selection.) *For all $\mathbf{x} \in \mathbb{X}$ and $\mathbf{x}'_0, \mathbf{x}'_1 \in S_{\mathbf{z}}^+$,*

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_0) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_1)| \leq \rho_S(\mathbf{x}'_0, \mathbf{x}'_1). \quad (8)$$

Proof. We first prove (STR). Fix $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{X}$ and $\mathbf{x}' \in S_{\mathbf{z}}^+$. By Eq. 4, the two ETT values integrate different response functions against the same latent posterior:

$$g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - g_{\mathbf{z}}(\tilde{\mathbf{x}}, \mathbf{x}') = \int (h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) - h_{\mathbf{u}}(\tilde{\mathbf{x}}, \mathbf{z})) dP(\mathbf{u} | \mathbf{x}', \mathbf{z}).$$

By (A2), $|h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) - h_{\mathbf{u}}(\tilde{\mathbf{x}}, \mathbf{z})| \leq L_Y \|\mathbf{x} - \tilde{\mathbf{x}}\|$ for every $\mathbf{u}$. Since the bound holds for every stratum, it holds for the average:

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - g_{\mathbf{z}}(\tilde{\mathbf{x}}, \mathbf{x}')| \leq L_Y \|\mathbf{x} - \tilde{\mathbf{x}}\| = \rho_T(\mathbf{x}, \tilde{\mathbf{x}}).$$

This is (STR).

We next prove (STS). Fix $\mathbf{x} \in \mathbb{X}$ and $\mathbf{x}'_0, \mathbf{x}'_1 \in S_{\mathbf{z}}^+$. By Eq. 4, the two ETT values share the response functions and differ only through the signed measure

$$d\Delta(\mathbf{u}) \triangleq dP(\mathbf{u} | \mathbf{x}'_0, \mathbf{z}) - dP(\mathbf{u} | \mathbf{x}'_1, \mathbf{z}),$$

which has zero total mass. Let c be the midpoint of $\mathbb{Y}$. By (A1), each $h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) \in \mathbb{Y}$, so $|h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) - c| \leq \frac{1}{2} \text{diam}(\mathbb{Y})$. Since Δ has zero total mass, subtracting c from the integrand does not change the integral, and therefore

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_0) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}'_1)| = \left| \int (h_{\mathbf{u}}(\mathbf{x}, \mathbf{z}) - c) d\Delta(\mathbf{u}) \right| \leq \text{diam}(\mathbb{Y}) \text{TV}_{\mathbf{u}}(\mathbf{x}'_0, \mathbf{x}'_1; \mathbf{z}).$$

Since $p(\mathbf{x}'_0 | \mathbf{z}) > 0$ and $p(\mathbf{x}'_1 | \mathbf{z}) > 0$ on $S_{\mathbf{z}}^+$, Lem. 1 bounds the right side by $\kappa_{\mathbf{z}}(\mathbf{x}'_0, \mathbf{x}'_1)$; this is Eq. 6. To pass to the chain closure, let $\mathbf{a} = \mathbf{a}_0, \dots, \mathbf{a}_k = \mathbf{b}$ be a chain in $S_{\mathbf{z}}^+$. Applying Eq. 6 to each link and the triangle inequality, we obtain

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{a}) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{b})| \leq \sum_i |g_{\mathbf{z}}(\mathbf{x}, \mathbf{a}_i) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{a}_{i+1})| \leq \sum_i \kappa_{\mathbf{z}}(\mathbf{a}_i, \mathbf{a}_{i+1}).$$

Taking the infimum over chains gives $|g_{\mathbf{z}}(\mathbf{x}, \mathbf{a}) - g_{\mathbf{z}}(\mathbf{x}, \mathbf{b})| \leq \kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{b})$, which is (STS) with $\kappa_{\mathbf{z}}^*$.

We next verify that the cost is regular. The cost $\kappa_{\mathbf{z}}^*$ is symmetric and vanishes on the diagonal. Concatenating a chain from $\mathbf{a}$ to $\mathbf{b}$ with a chain from $\mathbf{b}$ to $\mathbf{c}$ gives a chain from $\mathbf{a}$ to $\mathbf{c}$, so

$$\kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{c}) \leq \kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{b}) + \kappa_{\mathbf{z}}^*(\mathbf{b}, \mathbf{c}).$$

Hence $\kappa_{\mathbf{z}}^*$ is a pseudometric. It is continuous on $S_{\mathbf{z}}^+$. Indeed, for $\mathbf{a} \in S_{\mathbf{z}}^+$ and $\mathbf{b} \rightarrow \mathbf{a}$, the one-link chain gives

$$\kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{b}) \leq \kappa_{\mathbf{z}}(\mathbf{a}, \mathbf{b}) \leq \frac{L_X \text{diam}(\mathbb{Y}) \|\mathbf{a} - \mathbf{b}\|}{p(\mathbf{a} | \mathbf{z})} \rightarrow 0,$$

and the triangle inequality gives

$$|\kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{b}) - \kappa_{\mathbf{z}}^*(\mathbf{a}', \mathbf{b}')| \leq \kappa_{\mathbf{z}}^*(\mathbf{a}, \mathbf{a}') + \kappa_{\mathbf{z}}^*(\mathbf{b}, \mathbf{b}').$$

Finally, we show that $S_{\mathbf{z}}^+$ has full measure. Since $P(\mathbf{X} | \mathbf{z})$ has density $p(\cdot | \mathbf{z})$,

$$P(p(\mathbf{X} | \mathbf{z}) = 0 | \mathbf{z}) = \int_{\{p=0\}} p(\mathbf{x} | \mathbf{z}) d\mathbf{x} = 0,$$

so $S_{\mathbf{z}}^+$ carries all of $P(\mathbf{X} | \mathbf{z})$. The proofs of Thms. 6 and 10 use the support only through this property and through the fact that each of its points has neighbourhoods of positive mass, which $S_{\mathbf{z}}^+ \subseteq S_{\mathbf{z}}$ inherits. Hence both theorems hold with $S_{\mathbf{z}}^+$ in place of $S_{\mathbf{z}}$, and no integral changes. This completes the proof. $\square$

B.2 STR and STS bounds

Lemma 3. Under (A1)–(A3), for every $(\mathbf{x}, \mathbf{z}) \in \mathbb{X} \times \mathbb{Z}$, the following inequalities hold:

$$\underline{w}(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{z}] \leq \overline{w}(\mathbf{x}, \mathbf{z}), \quad (9)$$

where, for a draw $\mathbf{X}' \sim P(\mathbf{X} \mid \mathbf{z})$ of the received treatment, the bounds are

$$\begin{aligned} \underline{w}(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{X}', \mathbf{z}) - \rho_T(\mathbf{x}, \mathbf{X}')) \vee \underline{y} \mid \mathbf{z}], \\ \overline{w}(\mathbf{x}, \mathbf{z}) &\triangleq \mathbb{E}[(\mu(\mathbf{X}', \mathbf{z}) + \rho_T(\mathbf{x}, \mathbf{X}')) \wedge \overline{y} \mid \mathbf{z}]. \end{aligned} \quad (10)$$

Proof. By the tower property over the received treatment $\mathbf{X}' \sim P(\cdot \mid \mathbf{z})$,

$$\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{z}] = \mathbb{E}[g_{\mathbf{z}}(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}].$$

It therefore suffices to bound $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}')$ pointwise for $\mathbf{x}' \in S_{\mathbf{z}}$. By consistency,

$$g_{\mathbf{z}}(\mathbf{x}', \mathbf{x}') = \mu(\mathbf{x}', \mathbf{z}), \quad \mathbf{x}' \in S_{\mathbf{z}}.$$

Applying (STR) at the received treatment $\mathbf{x}'$, between the intervened treatments $\mathbf{x}$ and $\mathbf{x}'$, we obtain

$$|g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - \mu(\mathbf{x}', \mathbf{z})| = |g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') - g_{\mathbf{z}}(\mathbf{x}', \mathbf{x}')| \leq \rho_T(\mathbf{x}, \mathbf{x}').$$

Moreover, $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') \in [\underline{y}, \overline{y}]$ by (A1), being an average of values of f_Y . Combining the two constraints, for every $\mathbf{x}' \in S_{\mathbf{z}}$,

$$(\mu(\mathbf{x}', \mathbf{z}) - \rho_T(\mathbf{x}, \mathbf{x}')) \vee \underline{y} \leq g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') \leq (\mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{x}, \mathbf{x}')) \wedge \overline{y}.$$

Substituting $\mathbf{x}' = \mathbf{X}'$ and taking expectations given $\mathbf{z}$, monotonicity of the expectation yields

$$\underline{w}(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[g_{\mathbf{z}}(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] = \mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{z}] \leq \overline{w}(\mathbf{x}, \mathbf{z}),$$

where the two sides are the definitions Eq. 10. This proves Eq. 9. $\square$

Corollary 12. Under (A1)–(A3), suppose additionally there is no unobserved confounding, $Y_{\mathbf{x}} \perp \mathbf{X} \mid \mathbf{Z}$. Then $\mu(\mathbf{x}, \mathbf{z}) = \tau(\mathbf{x}, \mathbf{z})$ on $S_{\mathbf{z}}$ and the ETT is constant in the received treatment, $\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', \mathbf{z}] = \tau(\mathbf{x}, \mathbf{z})$ for every $\mathbf{x}' \in S_{\mathbf{z}}$. Every observed treatment therefore constrains the same constant $\tau(\mathbf{x}, \mathbf{z})$, so the average over $P(\mathbf{X} \mid \mathbf{z})$ that defines the bounds $\underline{w}(\mathbf{x}, \mathbf{z}), \overline{w}(\mathbf{x}, \mathbf{z})$ can be replaced by the intersection of the per-treatment intervals, sharpening it to the Lipschitz bound of Khan et al. (2024)

$$\begin{aligned} \sup_{\mathbf{x}' \in S_{\mathbf{z}}} [\mu(\mathbf{x}', \mathbf{z}) - \rho_T(\mathbf{x}, \mathbf{x}')] &\leq \tau(\mathbf{x}, \mathbf{z}) \\ &\leq \inf_{\mathbf{x}' \in S_{\mathbf{z}}} [\mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{x}, \mathbf{x}')]. \end{aligned} \quad (22)$$

Proof. Under $Y_{\mathbf{x}} \perp \mathbf{X} \mid \mathbf{Z}$, conditioning on the received treatment does not change the mean of the potential outcome, so for every $\mathbf{x}' \in S_{\mathbf{z}}$,

$$g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}') = \mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', \mathbf{z}] = \mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{z}] = \tau(\mathbf{x}, \mathbf{z}),$$

independent of $\mathbf{x}'$. Taking $\mathbf{x}' = \mathbf{x} \in S_{\mathbf{z}}$ and using consistency $g_{\mathbf{z}}(\mathbf{x}, \mathbf{x}) = \mu(\mathbf{x}, \mathbf{z})$, we obtain $\mu(\mathbf{x}, \mathbf{z}) = \tau(\mathbf{x}, \mathbf{z})$.

The pointwise bound in the proof of Lem. 3 now constrains this constant: for every observed $\mathbf{x}' \in S_{\mathbf{z}}$,

$$\tau(\mathbf{x}, \mathbf{z}) \in [\mu(\mathbf{x}', \mathbf{z}) - \rho_T(\mathbf{x}, \mathbf{x}'), \mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{x}, \mathbf{x}')] \cap \mathbb{Y}.$$

Since every $\mathbf{x}'$ constrains the same constant $\tau(\mathbf{x}, \mathbf{z})$, the average over $P(\mathbf{X} \mid \mathbf{z})$ may be replaced by the intersection over $\mathbf{x}' \in S_{\mathbf{z}}$. Taking the supremum of the lower ends and the infimum of the upper ends gives Eq. 22.

It remains to show that each side is attained at a single observed treatment. Chaining the upper bound through an intermediate observed treatment $\mathbf{m} \in S_{\mathbf{z}}$, first from $\mathbf{x}'$ to $\mathbf{m}$ and then from $\mathbf{m}$ to $\mathbf{x}$, gives the upper value $\mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{m}, \mathbf{x}') + \rho_T(\mathbf{x}, \mathbf{m})$. By the triangle inequality for ρ_T ,

$$\mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{m}, \mathbf{x}') + \rho_T(\mathbf{x}, \mathbf{m}) \geq \mu(\mathbf{x}', \mathbf{z}) + \rho_T(\mathbf{x}, \mathbf{x}'),$$

so passing through $\mathbf{m}$ only relaxes the bound, and the nearest effective observed treatment determines it (the no-interaction property). The lower side is symmetric. This completes the proof. $\square$

Lemma 4. Under (A1)–(A3), for every $z \in \mathbb{Z}$ and $\mathbf{x} \in S_z$, the following inequalities hold:

$$\underline{w}_X(\mathbf{x}, z) \leq \mathbb{E}[Y_{\mathbf{x}} \mid z] \leq \overline{w}_X(\mathbf{x}, z), \quad (11)$$

where, for a draw $\mathbf{X}' \sim P(\mathbf{X} \mid z)$ of the received treatment, the bounds are

$$\begin{aligned} \underline{w}_X(\mathbf{x}, z) &\triangleq \mathbb{E}[(\mu(\mathbf{x}, z) - \rho_S(\mathbf{X}', \mathbf{x})) \vee \underline{y} \mid z], \\ \overline{w}_X(\mathbf{x}, z) &\triangleq \mathbb{E}[(\mu(\mathbf{x}, z) + \rho_S(\mathbf{X}', \mathbf{x})) \wedge \overline{y} \mid z]. \end{aligned} \quad (12)$$

Proof. The proof follows that of Lem. 3, with (STS) in place of (STR). By the tower property,

$$\mathbb{E}[Y_{\mathbf{x}} \mid z] = \mathbb{E}[g_z(\mathbf{x}, \mathbf{X}') \mid z],$$

while $\mu(\mathbf{x}, z)$ is constant in $\mathbf{X}'$. Since $\mathbf{x} \in S_z$, consistency gives

$$g_z(\mathbf{x}, \mathbf{x}) = \mu(\mathbf{x}, z).$$

Applying (STS) at the intervened treatment $\mathbf{x}$, between the received treatments $\mathbf{x}'$ and $\mathbf{x}$, we obtain

$$|g_z(\mathbf{x}, \mathbf{x}') - \mu(\mathbf{x}, z)| = |g_z(\mathbf{x}, \mathbf{x}') - g_z(\mathbf{x}, \mathbf{x})| \leq \rho_S(\mathbf{x}', \mathbf{x}).$$

As before, $g_z(\mathbf{x}, \mathbf{x}') \in [\underline{y}, \overline{y}]$. Hence, for every $\mathbf{x}' \in S_z$,

$$(\mu(\mathbf{x}, z) - \rho_S(\mathbf{x}', \mathbf{x})) \vee \underline{y} \leq g_z(\mathbf{x}, \mathbf{x}') \leq (\mu(\mathbf{x}, z) + \rho_S(\mathbf{x}', \mathbf{x})) \wedge \overline{y}.$$

Substituting $\mathbf{x}' = \mathbf{X}'$ and using monotonicity of the expectation gives Eq. 11, the two sides being Eq. 12. This completes the proof. $\square$

B.3 Combined bounds

Lemma 13. Let ρ_T and ρ_S be pseudometrics on $\mathbb{X}$ and S_z , and $d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{a}, \mathbf{b})) \triangleq \rho_T(\mathbf{x}, \mathbf{a}) + \rho_S(\mathbf{x}', \mathbf{b})$. A function $g : \mathbb{X} \times S_z \rightarrow \mathbb{R}$ satisfies the inequalities of (STR) and (STS) with these costs if and only if $|g(\mathbf{p}) - g(\mathbf{q})| \leq d_z(\mathbf{p}, \mathbf{q})$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{X} \times S_z$.

Proof. We first show that a d_z -Lipschitz function satisfies the two conditions. Let g be d_z -Lipschitz. Taking $\mathbf{p} = (\mathbf{x}, \mathbf{x}')$ and $\mathbf{q} = (\tilde{\mathbf{x}}, \mathbf{x}')$, and using $\rho_S(\mathbf{x}', \mathbf{x}') = 0$,

$$|g(\mathbf{x}, \mathbf{x}') - g(\tilde{\mathbf{x}}, \mathbf{x}')| \leq \rho_T(\mathbf{x}, \tilde{\mathbf{x}}) + \rho_S(\mathbf{x}', \mathbf{x}') = \rho_T(\mathbf{x}, \tilde{\mathbf{x}}),$$

which is the inequality of (STR). Taking $\mathbf{p} = (\mathbf{x}, \mathbf{x}'_0)$ and $\mathbf{q} = (\mathbf{x}, \mathbf{x}'_1)$, and using $\rho_T(\mathbf{x}, \mathbf{x}) = 0$,

$$|g(\mathbf{x}, \mathbf{x}'_0) - g(\mathbf{x}, \mathbf{x}'_1)| \leq \rho_T(\mathbf{x}, \mathbf{x}) + \rho_S(\mathbf{x}'_0, \mathbf{x}'_1) = \rho_S(\mathbf{x}'_0, \mathbf{x}'_1),$$

which is the inequality of (STS).

We next prove the converse. Suppose g satisfies the inequalities of (STR) and (STS), and let $\mathbf{p} = (\mathbf{x}, \mathbf{x}')$ and $\mathbf{q} = (\mathbf{a}, \mathbf{b})$. Passing through the intermediate point $(\mathbf{a}, \mathbf{x}')$ and applying (STR) to the first step and (STS) to the second,

$$|g(\mathbf{x}, \mathbf{x}') - g(\mathbf{a}, \mathbf{b})| \leq |g(\mathbf{x}, \mathbf{x}') - g(\mathbf{a}, \mathbf{x}')| + |g(\mathbf{a}, \mathbf{x}') - g(\mathbf{a}, \mathbf{b})| \leq \rho_T(\mathbf{x}, \mathbf{a}) + \rho_S(\mathbf{x}', \mathbf{b}).$$

The right side is $d_z(\mathbf{p}, \mathbf{q})$, so g is d_z -Lipschitz. This completes the proof. $\square$

Lemma 5. Under (A1)–(A3), for every $z \in \mathbb{Z}$ the ETT function is Lipschitz for the product cost: $|g_z(\mathbf{p}) - g_z(\mathbf{q})| \leq d_z(\mathbf{p}, \mathbf{q})$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{X} \times S_z$. In particular, for every $\mathbf{x} \in \mathbb{X}$ and $\mathbf{x}', \mathbf{s} \in S_z$,

$$|\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', z] - \mu(\mathbf{s}, z)| \leq \rho(\mathbf{s}, \mathbf{x}') \quad (13)$$

$$\text{with } \rho(\mathbf{s}, \mathbf{x}') \triangleq \rho_T(\mathbf{x}, \mathbf{s}) + \rho_S(\mathbf{x}', \mathbf{s}). \quad (14)$$

Proof. By Prop. 2 and the convention of Sec. 2.1, the ETT function satisfies (STR) and (STS) with the costs of Prop. 2. By Lem. 13, it is therefore d_z -Lipschitz:

$$|g_z(\mathbf{p}) - g_z(\mathbf{q})| \leq d_z(\mathbf{p}, \mathbf{q}), \quad \mathbf{p}, \mathbf{q} \in \mathbb{X} \times S_z.$$

Now take $\mathbf{p} = (\mathbf{x}, \mathbf{x}')$ and $\mathbf{q} = (\mathbf{s}, \mathbf{s})$ with $\mathbf{s} \in S_z$. By consistency, $g_z(\mathbf{s}, \mathbf{s}) = \mu(\mathbf{s}, \mathbf{z})$, and by the definition of the product cost,

$$d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{s}, \mathbf{s})) = \rho_T(\mathbf{x}, \mathbf{s}) + \rho_S(\mathbf{x}', \mathbf{s}) = \rho(\mathbf{s}, \mathbf{x}').$$

Substituting both into the Lipschitz inequality gives

$$|\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', \mathbf{z}] - \mu(\mathbf{s}, \mathbf{z})| = |g_z(\mathbf{x}, \mathbf{x}') - g_z(\mathbf{s}, \mathbf{s})| \leq \rho(\mathbf{s}, \mathbf{x}'),$$

which is Eq. 13. This completes the proof. $\square$

Optimality of the L-shaped path. Lem. 5 reaches the reference point along a two-step, L-shaped path: one step in the intervened treatment and one in the received treatment (Fig. 3, blue). Chaining the two conditions along any other path that moves one coordinate at a time, such as the zig-zag of Fig. 3 (grey), cannot give a smaller cost. To see this, let

$$(\mathbf{x}, \mathbf{x}') = (\mathbf{a}_0, \mathbf{b}_0), (\mathbf{a}_1, \mathbf{b}_1), \dots, (\mathbf{a}_k, \mathbf{b}_k) = (\mathbf{s}, \mathbf{s})$$

be such a path in $\mathbb{X} \times S_z$, where each step changes either the intervened treatment, at cost $\rho_T(\mathbf{a}_i, \mathbf{a}_{i+1})$ by (STR), or the received treatment, at cost $\rho_S(\mathbf{b}_i, \mathbf{b}_{i+1})$ by (STS). Summing the steps bounds $|g_z(\mathbf{x}, \mathbf{x}') - \mu(\mathbf{s}, \mathbf{z})|$ by

$$\sum_{i \in \mathcal{T}} \rho_T(\mathbf{a}_i, \mathbf{a}_{i+1}) + \sum_{i \in \mathcal{S}} \rho_S(\mathbf{b}_i, \mathbf{b}_{i+1}),$$

where $\mathcal{T}$ and $\mathcal{S}$ index the STR and the STS steps. The STR steps form a chain from $\mathbf{x}$ to $\mathbf{s}$ in the first argument, and the STS steps a chain from $\mathbf{x}'$ to $\mathbf{s}$ in the second, because each argument is unchanged by the steps of the other kind. By the triangle inequality for each cost,

$$\sum_{i \in \mathcal{T}} \rho_T(\mathbf{a}_i, \mathbf{a}_{i+1}) \geq \rho_T(\mathbf{x}, \mathbf{s}), \quad \sum_{i \in \mathcal{S}} \rho_S(\mathbf{b}_i, \mathbf{b}_{i+1}) \geq \rho_S(\mathbf{x}', \mathbf{s}),$$

and the sum of the two right sides is the cost $\rho(\mathbf{s}, \mathbf{x}')$ of the L-shaped path. The two orders of the L, through $(\mathbf{s}, \mathbf{x}')$ or through $(\mathbf{x}, \mathbf{s})$, cost the same. A path through several diagonal points does no better either: bounding $g_z(\mathbf{x}, \mathbf{x}')$ by $\mu(\mathbf{s}_1, \mathbf{z})$ and then $\mu(\mathbf{s}_1, \mathbf{z})$ by $\mu(\mathbf{s}_2, \mathbf{z})$ costs

$$d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{s}_1, \mathbf{s}_1)) + d_z((\mathbf{s}_1, \mathbf{s}_1), (\mathbf{s}_2, \mathbf{s}_2)) \geq d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{s}_2, \mathbf{s}_2)) = \rho(\mathbf{s}_2, \mathbf{x}'),$$

and the bounds of Eq. 15 already optimize over the endpoint. So the L-shaped path to the best reference point is the cheapest way to combine the two conditions, and Thm. 10 shows more: no argument at all can improve the resulting bound, because it is attained. The geometry is standard Lipschitz geometry; what is not standard is which bound is sharp, and that intermediate reference points are needed to reach it.

Theorem 6. *Under (A1)–(A3), for every $\mathbf{z} \in \mathbb{Z}$ the ETT function satisfies $L_z \leq g_z \leq U_z$ on $\mathbb{X} \times S_z$. Consequently, for every $\mathbf{x} \in \mathbb{X}$, including $\mathbf{x} \notin S_z$, and a draw $\mathbf{X}' \sim P(\mathbf{X} \mid \mathbf{z})$ of the received treatment,*

$$\mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] \leq \tau(\mathbf{x}, \mathbf{z}) \leq \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}]. \quad (16)$$

Proof. Let g_z be the ETT function of an SCM in $\mathcal{M}(\rho_T, \rho_S)$ that reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$. Its diagonal is $\mu(\cdot, \mathbf{z})$ by consistency, and it takes values in $\mathbb{Y}$ by (A1). Fix $(\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times S_z$ and a reference point $\mathbf{s} \in S_z$. By Lem. 13, g_z is d_z -Lipschitz, so consistency gives

$$g_z(\mathbf{x}, \mathbf{x}') \leq g_z(\mathbf{s}, \mathbf{s}) + d_z((\mathbf{x}, \mathbf{x}'), (\mathbf{s}, \mathbf{s})) = \mu(\mathbf{s}, \mathbf{z}) + \rho(\mathbf{s}, \mathbf{x}').$$

Since this holds for every $\mathbf{s} \in S_z$, taking the infimum over $\mathbf{s}$ and using $g_z \leq \bar{y}$ gives

$$g_z(\mathbf{x}, \mathbf{x}') \leq \left(\inf_{\mathbf{s} \in S_z} [\mu(\mathbf{s}, \mathbf{z}) + \rho(\mathbf{s}, \mathbf{x}')] \right) \wedge \bar{y} = U_z(\mathbf{x}, \mathbf{x}').$$

The lower bound is symmetric: $g_z(\mathbf{x}, \mathbf{x}') \geq \mu(\mathbf{s}, z) - \rho(\mathbf{s}, \mathbf{x}')$ for every $\mathbf{s} \in S_z$, and $g_z \geq \underline{y}$, so

$$g_z(\mathbf{x}, \mathbf{x}') \geq \left(\sup_{\mathbf{s} \in S_z} [\mu(\mathbf{s}, z) - \rho(\mathbf{s}, \mathbf{x}')] \right) \vee \underline{y} = L_z(\mathbf{x}, \mathbf{x}').$$

This proves $L_z \leq g_z \leq U_z$ on $\mathbb{X} \times S_z$.

We next average over the received treatment. The bounds U_z and L_z are bounded, so the expectations below exist. Substituting $\mathbf{x}' = \mathbf{X}' \sim P(\mathbf{X} \mid z)$ in the pointwise bounds and using $\tau(\mathbf{x}, z) = \mathbb{E}[g_z(\mathbf{x}, \mathbf{X}') \mid z]$, monotonicity of the expectation gives

$$\mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid z] \leq \mathbb{E}[g_z(\mathbf{x}, \mathbf{X}') \mid z] = \tau(\mathbf{x}, z) \leq \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid z],$$

which is Eq. 16. This completes the proof. $\square$

Corollary 7. *Under (A1)–(A3), let $\mathbb{I} \subseteq S_z$ be a nonempty reference set, and let $U_{z,\mathbb{I}}$ and $L_{z,\mathbb{I}}$ be the bounds of Eq. 15 with the infimum and supremum restricted to $\mathbb{I}$. Then $L_{z,\mathbb{I}} \leq L_z \leq U_z \leq U_{z,\mathbb{I}}$, so that*

$$\mathbb{E}[L_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z] \leq \tau(\mathbf{x}, z) \leq \mathbb{E}[U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z] \quad (17)$$

for every $\mathbf{x} \in \mathbb{X}$. The bounds tighten as $\mathbb{I}$ grows. For finite $\mathbb{I}$, they cost $O(n|\mathbb{I}|)$ per query when $P(\mathbf{X} \mid z)$ is an empirical law on n points.

Proof. Since $\mathbb{I} \subseteq S_z$, the infimum over $\mathbb{I}$ is at least the infimum over S_z : for every $(\mathbf{x}, \mathbf{x}')$,

$$\inf_{\mathbf{s} \in S_z} [\mu(\mathbf{s}, z) + \rho(\mathbf{s}, \mathbf{x}')] \leq \inf_{\mathbf{s} \in \mathbb{I}} [\mu(\mathbf{s}, z) + \rho(\mathbf{s}, \mathbf{x}').$$

Taking the minimum with $\bar{y}$ on both sides gives $U_z \leq U_{z,\mathbb{I}}$, and symmetrically $L_{z,\mathbb{I}} \leq L_z$. Validity of Eq. 17 then follows from Thm. 6 and monotonicity of the expectation:

$$\mathbb{E}[L_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z] \leq \mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid z] \leq \tau(\mathbf{x}, z) \leq \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid z] \leq \mathbb{E}[U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z].$$

It also follows directly from Lem. 5 at the points of $\mathbb{I}$, without Eq. 19. Enlarging $\mathbb{I}$ enlarges the index sets of the infimum and supremum, so the same comparison shows that the bounds tighten as $\mathbb{I}$ grows. Finally, when $P(\mathbf{X} \mid z)$ is an empirical law on n points, each of the n received treatments requires a minimum and a maximum over the $|\mathbb{I}|$ reference points, which gives the operation count $O(n|\mathbb{I}|)$. This completes the proof. $\square$

Corollary 14. *Fix $\mathbf{x} \in \mathbb{X}$ and a reference set $\mathbb{I} \subseteq S_z$. Then $\mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid z] < \mathbb{E}[U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z]$ if and only if $U_z(\mathbf{x}, \mathbf{X}') < U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}')$ with positive probability, that is, if and only if some reference point $\mathbf{s} \in S_z$ improves on the minimum over $\mathbb{I}$ (and on $\bar{y}$) for a set of received treatments of positive probability. The same holds for the lower bounds.*

Proof. By Cor. 7, the gap

$$D(\mathbf{x}') \triangleq U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{x}') - U_z(\mathbf{x}, \mathbf{x}') \geq 0$$

is nonnegative, and it is bounded. The expectation of a nonnegative function is positive if and only if the function is positive with positive probability, so

$$\mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid z] < \mathbb{E}[U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{X}') \mid z] \iff P(D(\mathbf{X}') > 0 \mid z) > 0.$$

By the definition Eq. 15, $D(\mathbf{x}') > 0$, that is, $U_z(\mathbf{x}, \mathbf{x}') < U_{z,\mathbb{I}}(\mathbf{x}, \mathbf{x}')$, holds exactly when some $\mathbf{s} \in S_z$ has

$$\mu(\mathbf{s}, z) + \rho(\mathbf{s}, \mathbf{x}') < \bar{y} \wedge \inf_{\mathbf{s}'' \in \mathbb{I}} [\mu(\mathbf{s}'', z) + \rho(\mathbf{s}'', \mathbf{x}')].$$

The lower bounds are symmetric. This completes the proof. $\square$

Corollary 15. *Let $\mathbb{X}$ be finite with $\mathbb{Y} = [\underline{y}, \bar{y}]$, and write $\pi(\mathbf{x}, z) \triangleq P(\mathbf{X} = \mathbf{x} \mid z)$. If $\rho_T = \rho_S = c \mathbb{1}\{\cdot \neq \cdot\}$ with $c \geq \text{diam}(\mathbb{Y})$ (no smoothness), then for every $\mathbf{x} \in \mathbb{X}$ the sharp set $\Theta(\mathbf{x}, z)$ of Thm. 10 (Eq. 18) is the conditional Manski interval (Manski, 1990)*

$$\begin{aligned} \Theta(\mathbf{x}, z) = & [\pi(\mathbf{x}, z) \mu(\mathbf{x}, z) + (1 - \pi(\mathbf{x}, z)) \underline{y}, \\ & \pi(\mathbf{x}, z) \mu(\mathbf{x}, z) + (1 - \pi(\mathbf{x}, z)) \bar{y}], \end{aligned} \quad (23)$$

with $\pi\mu$ read as 0 when $\pi(\mathbf{x}, z) = 0$. For $\mathbf{x} \in S_z$ and any ρ_T , the singleton reference set $\mathbb{I} = \{\mathbf{x}\}$ of Cor. 7 with $\rho_S = c \mathbb{1}\{\cdot \neq \cdot\}$ returns the same interval.

Proof. The costs $c\mathbb{1}\{\cdot \neq \cdot\}$ are pseudometrics, and they are continuous on the finite set $\mathbb{X}$. Moreover, for all $s \neq s'$,

$$|\mu(s, z) - \mu(s', z)| \leq \text{diam}(\mathbb{Y}) \leq 2c = \rho_T(s, s') + \rho_S(s, s').$$

So Eq. 19 holds, and Thm. 10 applies.

We next compute the bounds. Consider the terms $\mu(s, z) + \rho_T(x, s) + \rho_S(x', s)$ of Eq. 15. If $x' = x \in S_z$, the reference point $s = x$ gives the term $\mu(x, z)$, and Thm. 10(a) gives

$$U_z(x, x) = \mu(x, z).$$

If $x' \neq x$, every s differs from x or from x' , so at least one of the two costs equals c , and every term satisfies

$$\mu(s, z) + \rho_T(x, s) + \rho_S(x', s) \geq \underline{y} + c \geq \bar{y}.$$

Hence $U_z(x, x') = \bar{y}$ for $x' \neq x$. Symmetrically,

$$L_z(x, x) = \mu(x, z), \quad L_z(x, x') = \underline{y} \quad (x' \neq x).$$

Finally, we average over $X' \sim P(X | z)$. The atom $X' = x$ has probability $\pi(x, z)$ (zero if $x \notin S_z$), and every other received treatment takes the trivial values. Therefore

$$\mathbb{E}[U_z(x, X') | z] = \pi(x, z) \mu(x, z) + (1 - \pi(x, z)) \bar{y}, \quad \mathbb{E}[L_z(x, X') | z] = \pi(x, z) \mu(x, z) + (1 - \pi(x, z)) \underline{y},$$

and Thm. 10(b) gives Eq. 23.

For the singleton $\mathbb{I} = \{x\}$, the only term is

$$\mu(x, z) + \rho_T(x, x) + \rho_S(x', x) = \mu(x, z) + c\mathbb{1}\{x' \neq x\},$$

whatever ρ_T is. It equals $\mu(x, z)$ at $x' = x$ and is at least $\underline{y} + c \geq \bar{y}$ otherwise, so the same computation applies and returns the same interval. This completes the proof. $\square$

B.4 Sharpness analysis

Call a function $g : \mathbb{X} \times S_z \rightarrow \mathbb{Y}$ *admissible* if it satisfies the inequalities of (STR) and (STS) and has diagonal $\mu(\cdot, z)$. By Lem. 13, such a diagonal satisfies Eq. 19 and is continuous, so we take the continuous version of $\mu(\cdot, z)$. By consistency and (A1), the ETT function of every SCM in $\mathcal{M}(\rho_T, \rho_S)$ that reproduces $P(X, Z, Y)$ is admissible. The following lemma establishes the converse.

Lemma 16 (Realization). *Let $\{g_z\}_{z \in \mathbb{Z}}$ be a family of admissible functions, jointly measurable in (z, x, x') . Then there exists an SCM in $\mathcal{M}(\rho_T, \rho_S)$ that reproduces $P(X, Z, Y)$ and whose ETT function is g_z at every covariate value z .*

Proof. Let $(\tilde{Z}, \tilde{X}) \sim P(Z, X)$ and $\varepsilon \sim \text{Unif}(0, 1)$ be independent, and let $Q_{x', z}$ be the quantile function of Y given $(X, Z) = (x', z)$, chosen jointly measurable. Take the latent confounder $U \triangleq (\tilde{Z}, \tilde{X})$, the outcome noise $\varepsilon_Y \triangleq \varepsilon$, a degenerate ε_X , and set $Z \leftarrow \tilde{Z}$, $X \leftarrow \tilde{X}$, and

$$Y \leftarrow f_Y(X, Z, U, \varepsilon_Y), \quad f_Y(x, z, (\tilde{z}, \tilde{x}), \varepsilon) \triangleq \begin{cases} Q_{\tilde{x}, \tilde{z}}(\varepsilon), & x = \tilde{x}, \\ g_{\tilde{z}}(x, \tilde{x}), & x \neq \tilde{x}. \end{cases}$$

We first verify that this SCM reproduces the observational law. Observationally, $X = \tilde{X}$, so the first branch of f_Y applies, and

$$(Z, X) = (\tilde{Z}, \tilde{X}) \sim P(Z, X), \quad Y = Q_{X, Z}(\varepsilon).$$

Since $\varepsilon \sim \text{Unif}(0, 1)$ is independent of $(\tilde{Z}, \tilde{X})$, $Q_{X, Z}(\varepsilon)$ has the true conditional law of Y given (X, Z) . Hence $P(X, Z, Y)$ is reproduced. Moreover, all values lie in $\mathbb{Y}$: the conditional quantiles of Y lie in $\mathbb{Y}$ by (A1), and $g_{\tilde{z}}$ takes values in $\mathbb{Y}$ by admissibility.

It remains to verify that the ETT function of this SCM is g_z . Under $\text{do}(\mathbf{x})$, the potential outcome is

$$Y_{\mathbf{x}} = f_Y(\mathbf{x}, \tilde{\mathbf{Z}}, \mathbf{U}, \varepsilon),$$

and conditioning on the received treatment $\mathbf{x}'$ and the covariate value $\mathbf{z}$ amounts to conditioning on $(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}}) = (\mathbf{z}, \mathbf{x}')$. For $\mathbf{x} \neq \mathbf{x}'$, the second branch applies, and

$$\mathbb{E}[Y_{\mathbf{x}} \mid \mathbf{x}', \mathbf{z}] = g_z(\mathbf{x}, \mathbf{x}').$$

For $\mathbf{x} = \mathbf{x}'$, the first branch applies, and since ε is independent of $(\tilde{\mathbf{Z}}, \tilde{\mathbf{X}})$,

$$\mathbb{E}[Y_{\mathbf{x}'} \mid \mathbf{x}', \mathbf{z}] = \mathbb{E}[Q_{\mathbf{x}', \mathbf{z}}(\varepsilon)] = \mu(\mathbf{x}', \mathbf{z}) = g_z(\mathbf{x}', \mathbf{x}'),$$

where the last equality uses that g_z has diagonal $\mu(\cdot, \mathbf{z})$. So the ETT function of this SCM is g_z at every covariate value. Since each g_z is admissible, it satisfies (STR) and (STS), and the SCM lies in $\mathcal{M}(\rho_T, \rho_S)$. This completes the proof. $\square$

Theorem 10. *Under (A1), (STR), and (STS), fix $\mathbf{z} \in \mathbb{Z}$ and suppose $\mu(\cdot, \mathbf{z})$ satisfies Eq. 19. Then:*

- (a) (Attainment) U_z and L_z take values in $\mathbb{Y}$, satisfy the inequalities of (STR) and (STS), and equal $\mu(\mathbf{s}, \mathbf{z})$ at every diagonal point $(\mathbf{s}, \mathbf{s})$;
- (b) (Sharp set) for every $\mathbf{x} \in \mathbb{X}$, including $\mathbf{x} \notin S_z$,

$$\Theta(\mathbf{x}, \mathbf{z}) = \left[\mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}], \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] \right]. \quad (20)$$

If Eq. 19 fails on a set of covariate values of positive probability, no SCM in $\mathcal{M}(\rho_T, \rho_S)$ reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$, and the model is refuted.

Proof. We first prove (a). For each $\mathbf{s} \in S_z$, consider the map

$$\varphi_{\mathbf{s}}(\mathbf{p}) \triangleq \mu(\mathbf{s}, \mathbf{z}) + d_z(\mathbf{p}, (\mathbf{s}, \mathbf{s})), \quad \mathbf{p} \in \mathbb{X} \times S_z.$$

By the triangle inequality for d_z , each $\varphi_{\mathbf{s}}$ is d_z -Lipschitz:

$$|\varphi_{\mathbf{s}}(\mathbf{p}) - \varphi_{\mathbf{s}}(\mathbf{q})| = |d_z(\mathbf{p}, (\mathbf{s}, \mathbf{s})) - d_z(\mathbf{q}, (\mathbf{s}, \mathbf{s}))| \leq d_z(\mathbf{p}, \mathbf{q}).$$

An infimum of d_z -Lipschitz functions is d_z -Lipschitz whenever it is finite, and here it is bounded below:

$$\inf_{\mathbf{s} \in S_z} \varphi_{\mathbf{s}}(\mathbf{p}) \geq \inf_{\mathbf{s} \in S_z} \mu(\mathbf{s}, \mathbf{z}) \geq \underline{y}.$$

Taking the minimum with the constant $\bar{y}$ preserves the property. So $U_z = (\inf_{\mathbf{s}} \varphi_{\mathbf{s}}) \wedge \bar{y}$ is d_z -Lipschitz and takes values in $\mathbb{Y}$, and by Lem. 13 it satisfies the inequalities of (STR) and (STS).

It remains to compute U_z on the diagonal. Fix $\mathbf{s} \in S_z$. The reference point $\mathbf{s}$ itself gives

$$U_z(\mathbf{s}, \mathbf{s}) \leq \mu(\mathbf{s}, \mathbf{z}) + \rho_T(\mathbf{s}, \mathbf{s}) + \rho_S(\mathbf{s}, \mathbf{s}) = \mu(\mathbf{s}, \mathbf{z}).$$

Conversely, by Eq. 19, for every $\mathbf{s}' \in S_z$,

$$\mu(\mathbf{s}', \mathbf{z}) + \rho_T(\mathbf{s}, \mathbf{s}') + \rho_S(\mathbf{s}, \mathbf{s}') \geq \mu(\mathbf{s}, \mathbf{z}),$$

and $\bar{y} \geq \mu(\mathbf{s}, \mathbf{z})$. Taking the infimum over $\mathbf{s}'$ and the minimum with $\bar{y}$ gives $U_z(\mathbf{s}, \mathbf{s}) \geq \mu(\mathbf{s}, \mathbf{z})$. Hence $U_z(\mathbf{s}, \mathbf{s}) = \mu(\mathbf{s}, \mathbf{z})$. The argument for L_z is symmetric. This proves (a).

We next prove (b). By (a), U_z and L_z are admissible. The admissibility constraints are convex in g , so for every $\lambda \in [0, 1]$,

$$g_{\lambda} \triangleq \lambda U_z + (1 - \lambda) L_z$$

is admissible. The integrals $\mathbb{E}[g_{\lambda}(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}]$ exist, because $g_{\lambda}(\mathbf{x}, \cdot)$ is ρ_S -Lipschitz, hence continuous and bounded. By linearity,

$$\mathbb{E}[g_{\lambda}(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] = \lambda \mathbb{E}[U_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}] + (1 - \lambda) \mathbb{E}[L_z(\mathbf{x}, \mathbf{X}') \mid \mathbf{z}],$$

which sweeps the interval of Eq. 20 as λ runs over $[0, 1]$. By Thm. 6, $\Theta(\mathbf{x}, \mathbf{z})$ lies inside that interval. Conversely, fix $\lambda \in [0, 1]$ and take the family g_λ built at every covariate value from the same λ . Since (STR) and (STS) hold at every covariate value, $\mu(\cdot, \mathbf{z}')$ satisfies Eq. 19 for almost every $\mathbf{z}'$, so by (a) this family is admissible at almost every covariate value. By Lem. 16, it is the ETT function of an SCM $\mathcal{M} \in \mathcal{M}(\rho_T, \rho_S)$ that reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$, and by Eq. 3,

$$\mathbb{E}_{\mathcal{M}}[Y_{\mathbf{x}} | \mathbf{z}] = \mathbb{E}[g_\lambda(\mathbf{x}, \mathbf{X}') | \mathbf{z}].$$

Hence every point of the interval lies in $\Theta(\mathbf{x}, \mathbf{z})$, which proves (b).

Finally, we prove the refutation claim. If an SCM in $\mathcal{M}(\rho_T, \rho_S)$ reproduces the observational law, then for almost every $\mathbf{z}$ its ETT function has diagonal $\mu(\cdot, \mathbf{z})$, which then satisfies Eq. 19 by Lem. 13. Therefore, if Eq. 19 fails on a set of covariate values of positive probability, no such SCM exists. This completes the proof. $\square$

Corollary 17. *Under (A1), the class $\mathcal{M}(\rho_T, \rho_S)$ contains an SCM that reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$ if and only if Eq. 19 holds for P -almost every $\mathbf{z}$. With Euclidean costs this reads $|\mu(\mathbf{s}, \mathbf{z}) - \mu(\mathbf{s}', \mathbf{z})| \leq (L_Y + C)\|\mathbf{s} - \mathbf{s}'\|$ on $S_{\mathbf{z}}$: the Lipschitz constant of the observed regression on the support is an identified lower bound on $L_Y + C$. In particular, under (A1)–(A3), Eq. 19 with the costs of Prop. 2 is a testable implication.*

Proof. We first prove necessity. If an SCM in $\mathcal{M}(\rho_T, \rho_S)$ reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$, then Eq. 19 holds for P -almost every $\mathbf{z}$ by the refutation part of Thm. 10.

We next prove sufficiency. Suppose Eq. 19 holds for P -almost every $\mathbf{z}$. By Thm. 10(a), $U_{\mathbf{z}}$ is admissible at almost every $\mathbf{z}$, and Lem. 16, applied to the family $\{U_{\mathbf{z}}\}$, gives an SCM in $\mathcal{M}(\rho_T, \rho_S)$ that reproduces $P(\mathbf{X}, \mathbf{Z}, Y)$.

Finally, with Euclidean costs $\rho_T = L_Y \|\cdot\|$ and $\rho_S = C \|\cdot\|$,

$$\rho_T(\mathbf{s}, \mathbf{s}') + \rho_S(\mathbf{s}, \mathbf{s}') = (L_Y + C)\|\mathbf{s} - \mathbf{s}'\|,$$

so Eq. 19 takes the stated form. This completes the proof. $\square$

Proposition 8. *Let $\mathbb{X} \subset \mathbb{R}$, without covariates, and $\rho_T = b|\cdot|$, $\rho_S = a|\cdot|$. For any query $x \in S$, the combined STR+STS upper bound of Kim et al. (2018, Prop. D.3) gives the following*

$$u_2(x) \triangleq \mathbb{E}[\min\{\mu(x) + a|X' - x|, \mu(X') + b|X' - x|\}],$$

which is the bound of Cor. 7 with the two-point reference set $\{x, x'\}$ chosen separately at each received treatment x' (without the intersection with $\mathbb{Y}$).

Proof. With $\mathbb{I} = \{x, x'\}$ at the received treatment x' , Eq. 15 has two terms. The reference point $s = x$ contributes

$$\mu(x) + b \cdot 0 + a|x' - x| = \mu(x) + a|x' - x|,$$

and the reference point $s = x'$ contributes

$$\mu(x') + b|x - x'| + a \cdot 0 = \mu(x') + b|x' - x|.$$

The minimum of the two is the integrand of $u_2(x)$, so averaging over the received treatment gives the bound of Cor. 7 with this reference set. Validity follows from Lem. 13 and consistency at these two points, pointwise in x' . This completes the proof. $\square$

Proposition 9. *The bound u_2 of Prop. 8 is not sharp in general: for some costs, observational distribution, and query $x \in S$, $u_2(x)$ strictly exceeds $\mathbb{E}[Y_x]$ in every SCM satisfying the STR and STS conditions with these costs and reproducing the observational distribution.*

Proof. Take $a = b = 1$, the treatment uniform on $S = \{0, 1, 2\}$, $Y = \mu(X)$ with $\mu = (0, -1, 0)$, and the query $x = 2$. Write $g(x, x') \triangleq \mathbb{E}[Y_x | X = x']$ for $x, x' \in S$. The data fix the diagonal,

$$g(x', x') = \mu(x'), \quad x' \in S.$$

The two-point bound uses the columns $s \in \{2, x'\}$ of Table 1, which give the row values $(2, 0, 0)$ at $x' = 0, 1, 2$. Hence

$$u_2(2) = \frac{1}{3}(2 + 0 + 0) = \frac{2}{3}.$$

We first show that every model has $\mathbb{E}[Y_{x=2}] \leq 1/3$. Take any model whose ETT satisfies STR and STS with $a = b = 1$ and whose diagonal is μ . At $x' = 2$, the diagonal gives $g(2, 2) = \mu(2) = 0$. At $x' = 1$, STR gives

$$g(2, 1) \leq g(1, 1) + 1 = 0.$$

At $x' = 0$, STR and then STS at $x = 1$ give

$$g(2, 0) \leq g(1, 0) + 1 \leq g(1, 1) + 2 = 1.$$

Therefore

$$\mathbb{E}[Y_{x=2}] = \frac{1}{3}(g(2, 0) + g(2, 1) + g(2, 2)) \leq \frac{1}{3}(1 + 0 + 0) = \frac{1}{3}.$$

This uses only the two Lipschitz conditions and the diagonal, not any bound on the outcomes, so it holds over the whole class of Kim et al. (2018).

We next show that the value $1/3$ is attained. Let g be the function with rows indexed by x and columns by x' ,

$$(g(x, x'))_{x, x' \in S} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix};$$

this is the upper bound $\inf_s [\mu(s) + |x - s| + |x' - s|]$. Adjacent entries of every row and every column differ by at most 1, so g satisfies STR and STS with $a = b = 1$, and its diagonal is μ . The model $X \sim \text{Unif}(S)$, $Y_x \triangleq g(x, X)$ has ETT g and observational outcome $Y = g(X, X) = \mu(X)$, so it reproduces the observational law, and

$$\mathbb{E}[Y_{x=2}] = \frac{1}{3}(1 + 0 + 0) = \frac{1}{3}.$$

Combining the two steps, the sharp upper bound is $1/3 < 2/3 = u_2(2)$: the two-point bound is loose, in the setting of Kim et al. (2018) as well as in ours. This completes the proof. $\square$

Theorem 11. *Let $\mathbf{x}_1, \mathbf{x}_0 \in \mathbb{X}$, including treatments outside S_z , let $r \triangleq \rho_T(\mathbf{x}_1, \mathbf{x}_0)$, and let $\mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$ denote the interval*

$$\begin{aligned} & \left[-\mathbb{E}[\min\{U_z(\mathbf{x}_0, \mathbf{X}') - L_z(\mathbf{x}_1, \mathbf{X}'), r\} \mid \mathbf{z}], \right. \\ & \left. \mathbb{E}[\min\{U_z(\mathbf{x}_1, \mathbf{X}') - L_z(\mathbf{x}_0, \mathbf{X}'), r\} \mid \mathbf{z}] \right]. \end{aligned} \quad (21)$$

(a) (Validity) Under (A1)–(A3), $\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z}) \in \mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$.

(b) (Sharpness) Under (A1), (STR), and (STS), if $\mu(\cdot, \mathbf{z})$ satisfies Eq. 19, then $\mathcal{C}(\mathbf{x}_1, \mathbf{x}_0, \mathbf{z})$ is exactly the identified set of the CATE over the SCMs in $\mathcal{M}(\rho_T, \rho_S)$ that reproduce $P(\mathbf{X}, \mathbf{Z}, Y)$.

Proof. Write $U_j \triangleq U_z(\mathbf{x}_j, \cdot)$ and $L_j \triangleq L_z(\mathbf{x}_j, \cdot)$ on S_z for $j \in \{0, 1\}$, and $d \triangleq d_z$. By Thm. 6 applied to the admissible function U_z of Thm. 10(a), $L_j \leq U_j$.

We first prove (a). Let g be admissible and $\mathbf{x}' \in S_z$. By Thm. 6,

$$g(\mathbf{x}_1, \mathbf{x}') - g(\mathbf{x}_0, \mathbf{x}') \leq U_1(\mathbf{x}') - L_0(\mathbf{x}'),$$

and by (STR),

$$g(\mathbf{x}_1, \mathbf{x}') - g(\mathbf{x}_0, \mathbf{x}') \leq \rho_T(\mathbf{x}_1, \mathbf{x}_0) = r.$$

Hence $g(\mathbf{x}_1, \mathbf{x}') - g(\mathbf{x}_0, \mathbf{x}') \leq \min\{U_1(\mathbf{x}') - L_0(\mathbf{x}'), r\}$. Integrating the smaller of the two over $P(\mathbf{X} \mid \mathbf{z})$ and using Eq. 3 at both treatment levels gives

$$\tau(\mathbf{x}_1, \mathbf{z}) - \tau(\mathbf{x}_0, \mathbf{z}) \leq \mathbb{E}[\min\{U_1(\mathbf{X}') - L_0(\mathbf{X}'), r\} \mid \mathbf{z}],$$

the upper endpoint of Eq. 21. Exchanging $\mathbf{x}_1$ and $\mathbf{x}_0$ gives the lower one.

We next prove (b), starting with the upper endpoint. Define on S_z

$$a \triangleq U_1, \quad b \triangleq \max\{L_0, U_1 - r\}, \quad \text{so that} \quad a - b = \min\{U_1 - L_0, r\}$$

pointwise. It remains to exhibit an admissible g with $g(\mathbf{x}_1, \cdot) = a$ and $g(\mathbf{x}_0, \cdot) = b$. Let

$$A \triangleq (\{\mathbf{x}_1, \mathbf{x}_0\} \times S_{\mathbf{z}}) \cup D, \quad D \triangleq \{(\mathbf{s}, \mathbf{s}) : \mathbf{s} \in S_{\mathbf{z}}\},$$

and define g on A by a on $\{\mathbf{x}_1\} \times S_{\mathbf{z}}$, by b on $\{\mathbf{x}_0\} \times S_{\mathbf{z}}$, and by $\mu(\cdot, \mathbf{z})$ on the diagonal D . The triangle inequality for ρ_T gives

$$\rho_T(\mathbf{x}_1, \mathbf{s}) \leq r + \rho_T(\mathbf{x}_0, \mathbf{s}),$$

and the clipping at $\bar{y}$ and $\underline{y}$ preserves the resulting inequalities

$$U_1 \leq U_0 + r, \quad L_0 \leq L_1 + r. \quad (24)$$

We check that g is d -Lipschitz on A , pair type by pair type.

- *Diagonal–diagonal*: this is Eq. 19.
- *a–diagonal*: the requirement is

$$|a(\mathbf{x}') - \mu(\mathbf{s}, \mathbf{z})| \leq \rho_T(\mathbf{x}_1, \mathbf{s}) + \rho_S(\mathbf{x}', \mathbf{s}) \quad \text{for all } \mathbf{s}, \mathbf{x}' \in S_{\mathbf{z}},$$

which holds whenever $L_1 \leq a \leq U_1$, by the definitions Eq. 15; here $a = U_1$.

- *b–diagonal*: likewise, this requires $L_0 \leq b \leq U_0$. The lower inequality holds by the definition of b , and the upper one because $L_0 \leq U_0$ and $U_1 - r \leq U_0$ by Eq. 24.
- *a–a and b–b*: U_1 , L_0 , and $U_1 - r$ are ρ_S -Lipschitz by Thm. 10(a), and a maximum of ρ_S -Lipschitz functions is ρ_S -Lipschitz.
- *a(s) – b(s')* $\leq r + \rho_S(\mathbf{s}, \mathbf{s}')$: by the definition of b and the ρ_S -Lipschitz property of U_1 ,

$$b(\mathbf{s}') \geq U_1(\mathbf{s}') - r \geq U_1(\mathbf{s}) - \rho_S(\mathbf{s}, \mathbf{s}') - r.$$

- *b(s') – a(s)* $\leq r + \rho_S(\mathbf{s}, \mathbf{s}')$: if $b(\mathbf{s}') = U_1(\mathbf{s}') - r$, this follows from the ρ_S -Lipschitz property of U_1 . If $b(\mathbf{s}') = L_0(\mathbf{s}')$, then by Eq. 24,

$$L_0(\mathbf{s}') \leq L_1(\mathbf{s}') + r \leq U_1(\mathbf{s}') + r \leq U_1(\mathbf{s}) + \rho_S(\mathbf{s}, \mathbf{s}') + r.$$

The values lie in $\mathbb{Y}$, since $\underline{y} \leq L_0 \leq b \leq U_0 \leq \bar{y}$. Points of A that coincide receive consistent values. If $\mathbf{x}_1 \in S_{\mathbf{z}}$, then $a(\mathbf{x}_1) = U_1(\mathbf{x}_1) = \mu(\mathbf{x}_1, \mathbf{z})$ by Thm. 10(a). If $\mathbf{x}_0 \in S_{\mathbf{z}}$, then $L_0(\mathbf{x}_0) = \mu(\mathbf{x}_0, \mathbf{z})$, and the reference point $\mathbf{s} = \mathbf{x}_0$ in Eq. 15 gives

$$U_1(\mathbf{x}_0) - r \leq \mu(\mathbf{x}_0, \mathbf{z}) + \rho_T(\mathbf{x}_1, \mathbf{x}_0) + \rho_S(\mathbf{x}_0, \mathbf{x}_0) - r = \mu(\mathbf{x}_0, \mathbf{z}),$$

so $b(\mathbf{x}_0) = \mu(\mathbf{x}_0, \mathbf{z})$.

We now extend g from A to $\mathbb{X} \times S_{\mathbf{z}}$ by the McShane formula (McShane, 1934)

$$\hat{g}(\mathbf{q}) \triangleq \min\left\{\bar{y}, \inf_{\mathbf{p} \in A} [g(\mathbf{p}) + d(\mathbf{q}, \mathbf{p})]\right\}, \quad \mathbf{q} \in \mathbb{X} \times S_{\mathbf{z}}.$$

It is d -Lipschitz, it agrees with g on A because g is d -Lipschitz on A and bounded by $\bar{y}$, and it takes values in $\mathbb{Y}$ because $g \geq \underline{y}$ on A . By Lem. 13, $\hat{g}$ satisfies the inequalities of (STR) and (STS), and it has diagonal $\mu(\cdot, \mathbf{z})$ because $D \subseteq A$, so it is admissible. As in the proof of Thm. 10(b), $U_{\mathbf{z}'}$ is admissible at almost every covariate value $\mathbf{z}'$. By Lem. 16, with $\hat{g}$ at covariate value $\mathbf{z}$ and $U_{\mathbf{z}'}$ at every other covariate value, there is an SCM in $\mathcal{M}(\rho_T, \rho_S)$ that reproduces the observational law, and by Eq. 3 its CATE is

$$\mathbb{E}[a(\mathbf{X}') - b(\mathbf{X}') \mid \mathbf{z}] = \mathbb{E}[\min\{U_1(\mathbf{X}') - L_0(\mathbf{X}'), r\} \mid \mathbf{z}],$$

the upper endpoint.

Finally, exchanging the roles of $\mathbf{x}_1$ and $\mathbf{x}_0$ attains the lower endpoint. The admissible functions form a convex set and the CATE is linear in the ETT function, so the identified set is the whole interval. This completes the proof. $\square$

C EXPERIMENTAL SETUPS

This appendix specifies the three settings of Sec. 5: the V-shaped discrete SCMs (Sec. C.1), the continuous V-shaped and smooth V-shaped families (Sec. C.2), and the controlled IST-derived binary model (Sec. C.3). It then reports two additional experiments: smooth polynomial, logistic, phase, and curved-linear separation examples (Sec. C.4), and a controlled reference-set ablation on the counterexample of Prop. 9 (Sec. C.5). Conditional on each model, regressions and targets are finite latent sums; continuous bound integration uses the stated numerical quadrature. All experiments run on CPU on a single laptop (Apple M1 Pro, 16 GB memory); no GPUs are used.

Ours and Kim receive identical regressions, valid smoothness costs, outcome ranges, and integration weights. Kim uses the per-pair endpoints; ours adds the supported reference grid before clipping and averaging. Synthetic SCMs use their exact structural constants, with STR cost $L_Y|x-s|$ and STS cost $C|x'-s|$, where $C = L_X \text{diam}(\mathcal{Y})/p_{\min}$. Controlled IST instead uses exact finite-grid STR/STS constants of its constructed treatment-on-treated means. CMSM receives the same conditional outcome distribution.

CMSM uses the smallest single $\Lambda^* \geq 1$ whose interval contains every evaluated target within one SCM or IST age group. This target-based calibration gives full coverage by construction and is not an observable calibration rule. Let $w_{jq} = P(U = u_j | X = x_q)$ and y_{jq} be the conditional outcome atoms; for the IST-derived model these are Bernoulli atoms with probabilities $1 - \mu(x_q)$ and $\mu(x_q)$. For a target $t_q = \tau(x_q)$ define

$$A_q = \sum_j w_{jq}(y_{jq} - t_q)_+, \quad B_q = \sum_j w_{jq}(t_q - y_{jq})_+.$$

The minimum per-query tilt satisfies $\log \Lambda_q = \frac{1}{2}|\log A_q - \log B_q|$. We evaluate these sums in log space and use one $\Lambda^* = \max_q \Lambda_q$ per model, rather than choosing a separate reported parameter per query. Zero-mass atoms are removed; a constant conditional outcome equal to its target needs $\Lambda_q = 1$. A nonconstant law cannot reach an extreme atom with a finite tilt, and such an input fails explicitly. The exporter records the threshold separately from the 10^{-10} slack added to $\log_{10} \Lambda^*$ for numerical evaluation (no slack when $\Lambda^* = 1$). It checks coverage, failure just below a nontrivial threshold at its binding query, and that the threshold does not exceed the sufficient budget from both directions of $p(x | u)/p(x)$. Unit checks compare this inversion with forward numerical search and the Bernoulli formula. Coverage calibration uses the known target; it does not establish membership of the true assignment law in the calibrated CMSM.

We use a treatment-axis adaptation of the Lipschitz-envelope construction of Khan et al. (2024). Its anchors are the exact observational means on the sorted union of the observation and supported query grids, and its constant is the maximum absolute adjacent secant slope on that union. Every query has an exact anchor, so the envelope is the singleton $\{\mu(x_q)\}$ independently of the constant. The implementation verifies this identity before removing floating-point roundoff. Coverage is evaluated against $\tau(x_q)$ and may fail because these models are confounded; we do not assert full coverage for this baseline. This is an assumption-mismatch comparison, not a counterexample to Khan’s unconfoundedness-based guarantee. The resulting zero-width intervals cover 14.11% of discrete targets and 4% in each continuous family and IST group.

Figure 4 uses saved system 0 from the discrete and smooth V-shaped continuous families and the middle IST age tertile, without selecting systems for improvement. Discrete treatments use interval bars and target markers; the remaining panels connect evaluated grid points. The table pools both continuous families even though the figure displays only the smooth family. Run `plot_section5.py` to reproduce the figure from saved intervals without rerunning experiments.

C.1 Discrete SCMs with a V-shaped response

The discrete models use a response with a pronounced interior minimum, so that global references can tighten Kim’s per-pair endpoints. Let $U \in \{-1, +1\}$ with $P(U = -1) = p_-$, and let X have an odd, equally spaced support of size K on $[-1, 1]$. Define

$$P(X = x_j | U = u) = \frac{1 + \rho u x_j}{K}, \quad Y_x = \frac{2(b|x| + cu + c)}{b + 2c} - 1.$$

The canonical coefficients are $K = 7$, $b = 1$, $c = 1/2$, $\rho = 0.1$, and $p_- = 1/2$. Assignment is generated by an independent uniform variable through the conditional inverse CDF. Both latent types have positive probability,

Table 3: Stronger discrete variant with a V-shaped response. Widths are divided by the attained outcome range, $\text{diam}(\mathcal{Y}) = 2$. Ours and Kim use identical exact structural constants; all random draws are retained.

Setting	SCMs	Queries	Ours width	Kim width	Reduction (%)	Coverage (both)
Canonical	1	7	0.1119	0.1391	19.57	1.00/1.00
Random	500	3544	0.1118	0.1383	19.17	1.00/1.00

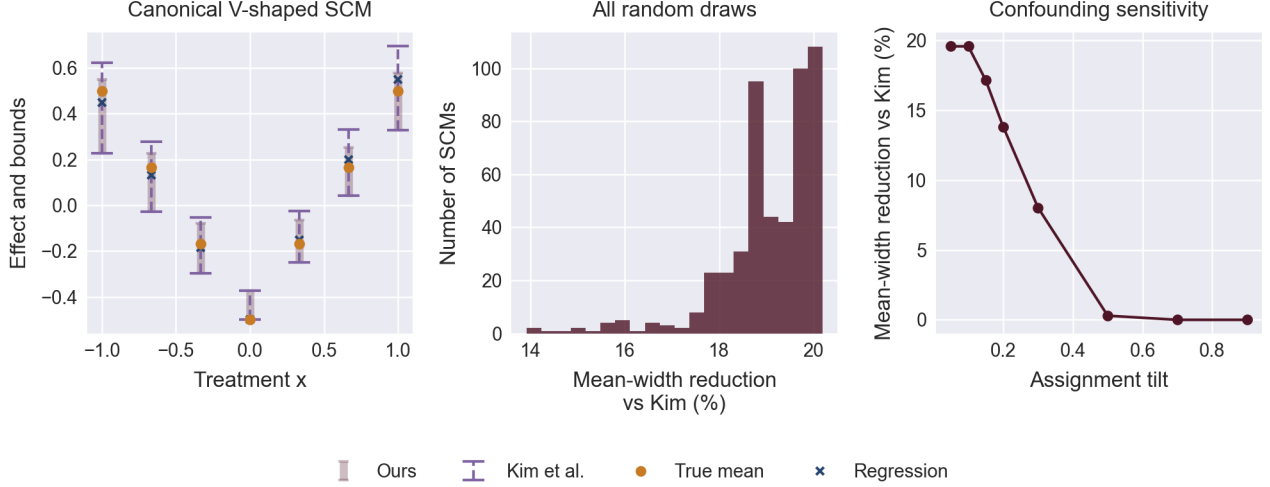


Figure 5: Discrete V-shaped SCMs with matched structural constants and exact outcome-range clipping. Left: the canonical V-shaped response, its confounded observational regression, and both intervals. Middle: mean-width reductions for all 500 random systems. Right: the reduction across the assignment-tilt sweep, including settings where the methods tie.

assignment has full overlap, and the outcome attains both -1 and $+1$ at supported treatments. Thus the supplied clipping range is exact, rather than enlarged to create an apparent improvement.

Both methods receive the maximum finite-support difference quotients of the response and assignment probabilities. Independently, these have the analytic values

$$L_Y^* = \frac{2b}{b+2c}, \quad L_X^* = \frac{\rho}{K}, \quad p_{\min} = \frac{1-\rho}{K}, \quad C = \frac{2\rho}{1-\rho}.$$

We use these structural constants rather than substituting smaller direct ETT constants. At the canonical coefficients, $\tau(x) = |x| - 1/2$ and $\mu(x) = |x| - 1/2 + 0.05x$, so confounding is nonzero, with maximum gap 2.5% of the outcome range. The global reference $s = 0$ supplies the response minimum that Kim’s per-pair endpoint references may omit. Adding this one reference to the endpoints reduces the normalized mean width from 0.13908 to 0.11187, a 19.57% reduction, and gives the same interval as using every supported reference. The improvement is strict at six of the seven queries. An independent finite LP agrees at every query; this verifies sharpness in the induced STR/STS class, not sharpness under the stronger structural assumptions.

We draw 500 systems with K uniform on $\{5, 7, 9\}$, $b \sim \text{Uniform}(0.9, 1.1)$, $c \sim \text{Uniform}(0.45, 0.55)$, $\rho \sim \text{Uniform}(0.08, 0.12)$, and $p_- \sim \text{Uniform}(0.35, 0.65)$, independently. This prior was designed after a feasibility check, so we do not interpret it as representative. No draw is rejected based on separation, coverage, or confounding. The pooled normalized widths are 0.11181 for ours and 0.13832 for Kim: a 19.17% reduction over 3544 queries, with per-system reductions ranging from 13.91% to 20.20% and full coverage for both methods (Table 3).

A separate canonical tilt sweep uses $\rho \in \{0.05, 0.1, 0.15, 0.2, 0.3, 0.5, 0.7, 0.9\}$, recalibrating both methods identically at every setting. At $\rho = 0.2$, the confounding gap is 5% of the outcome range and the mean-width reduction remains 13.82%. The improvement falls to zero at $\rho = 0.7$ and 0.9 ; coverage remains complete throughout (Fig. 5).

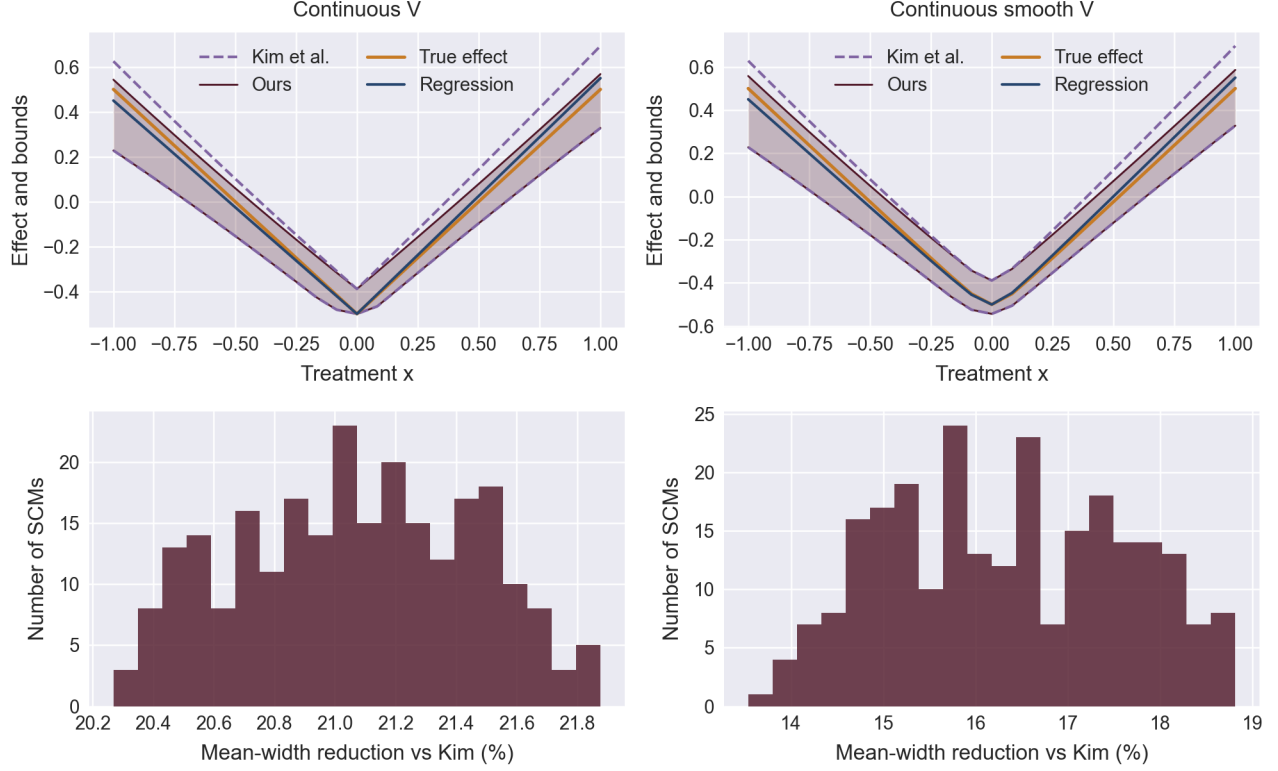


Figure 6: Continuous V-shaped and smooth V-shaped response families. Top: canonical bounds at identical exact structural constants. Bottom: mean-width reductions against Kim across all 250 random draws per family, including every draw without rejection.

The script `lpi_discrete_separation.py` exports every random system, query interval, tilt setting, and nested reference ablation, together with the analytic-constant checks and canonical LP endpoints.

C.2 Continuous SCMs with V-shaped responses

The continuous models extend the interior-minimum construction while retaining a strictly positive treatment density on the entire interval. Let $U \in \{-1, +1\}$, $P(U = -1) = p_-$, and define

$$p(x | u) = \frac{1 + \rho ux}{2}, \quad x \in [-1, 1], \quad Y = \frac{2\{b g_\epsilon(X) + cU + c\}}{b + 2c} - 1,$$

where, writing $D = 1 + |m|$ and $N_\epsilon = \sqrt{D^2 + \epsilon^2} - \epsilon$,

$$g_\epsilon(x) = \frac{\sqrt{(x - m)^2 + \epsilon^2} - \epsilon}{N_\epsilon}.$$

An independent uniform noise variable passed through the inverse conditional CDF defines the treatment equation. The response has its minimum at m and maximum at the farther support endpoint, so g_ϵ attains $[0, 1]$ and the outcome attains $[-1, 1]$. With $b, c > 0$, $0 < \rho < 1$, and $0 < p_- < 1$, the exact structural constants are

$$L_X = \rho/2, \quad p_{\min} = (1 - \rho)/2, \quad L_Y = \frac{2bD}{(b + 2c)N_\epsilon\sqrt{D^2 + \epsilon^2}}, \quad C = \frac{2\rho}{1 - \rho}.$$

The case $\epsilon = 0$ is a Lipschitz V-shaped response; $\epsilon > 0$ gives a differentiable smooth approximation. Both methods use these same structural constants, with no substitution of smaller direct ETT constants.

The canonical systems use $b = 1$, $c = 0.5$, $\rho = 0.1$, $p_- = 0.5$, $m = 0$, with $\epsilon = 0$ or 0.05 . Independently draw $b \sim U[0.9, 1.1]$, $c \sim U[0.45, 0.55]$, $\rho \sim U[0.08, 0.12]$, $p_- \sim U[0.35, 0.65]$, and $m \sim U[-0.1, 0.1]$. Set $\epsilon = 0$ for the

Table 4: Controlled IST-derived binary SCM: mean widths and coverage at identical oracle STR/STS constants. Each age group has 25 treatment queries. This model has a new, explicitly constructed target.

Age	Ours width	Kim width	Reduction (%)	Ours cov.	Kim cov.
16–67	0.01972	0.01981	0.47	1.00	1.00
68–77	0.04844	0.04909	1.32	1.00	1.00
78–99	0.04362	0.04440	1.76	1.00	1.00

V-shaped family and draw $\epsilon \sim U[0.025, 0.075]$ for the smooth family. Seed 0 generates 250 models per family, with every draw retained. This prior was selected after exploratory feasibility checks. Each model has 25 equally spaced queries on $[-1, 1]$. We use 257 trapezoidal integration nodes and a 257-point reference grid augmented by m and the exact pairwise endpoint references. Regression and interventional means are exact sums over the two latent types.

Canonical mean-width reductions are 21.27% and 16.27%; random-population reductions are 21.04% and 16.34%, with full coverage for ours and Kim (Fig. 6). Per-system reductions span 20.27–21.88% and 13.54–18.81%; there is no acceptance threshold based on the improvement. All random V-shaped queries and 93.824% of smooth-family queries improve strictly. Each model is checked for density normalization, overlap, attained outcome range, structural constants, coverage, and containment within Kim. Removing all additional references reproduces Kim at every query of all 502 canonical and random systems. Canonical ablations add the minimum alone, then nested grids of 3, 9, 33, 129, 257, 513, 1025 references augmented by the minimum, verifying interval nesting. A separate assignment-strength sweep retains $\rho \in \{0.05, 0.1, 0.15, 0.2, 0.3, 0.5, 0.7, 0.9\}$, including weaker separation at stronger confounding.

For each family, refinement checks use the canonical system and the first and last random draws. We refine integration and reference grids jointly and separately through 257, 513, 1025 points. Relative to the finest joint grid, the largest endpoint change is 1.674×10^{-5} of the outcome range, below the declared 10^{-4} tolerance. Run `lpi_continuous_separation.py --emit-paper-assets --assert` to export parameters, intervals, summaries, reference ablations, assignment sweeps, refinement records, and the figure. The consolidated exporter then adds oracle CMSM and Khan to these same saved models.

C.3 Controlled IST

The International Stroke Trial (Carolei et al., 1997), whose anonymised patient-level database is publicly released on Edinburgh DataShare under the repository’s end-user licence (Sandercock et al., 2011), is used to construct a model with a known target, not to estimate clinical effects. Blood pressure is not randomized in IST, so its raw regression is not a causal dose–response. We take systolic blood pressure at randomization (RSBP) as the treatment, death within six months (FDEAD) as the binary outcome, the age tertile z (roughly 6,400 patients each) as the covariate, and consciousness at randomization $u \in \{\text{F}, \text{D}, \text{U}\}$ (fully alert 14,921 / drowsy 4,254 / unconscious 260) as the hidden confounder. Consciousness strongly affects the outcome: the six-month death rates across its three levels are 0.150, 0.461, and 0.776. Within each tertile, the empirical category frequencies define $P(U = u \mid z)$. There are 25 equally spaced treatment values across the central 70% of that tertile’s observed systolic blood-pressure range, standardized to $t_j \in [-1, 1]$. For each consciousness category, Nadaraya–Watson regression of the binary outcomes supplies probabilities $r_z(t_j, u)$. All categories within a tertile use one bandwidth, twice the full-tertile Silverman bandwidth in mmHg. These fitted probabilities define the new model; they are not assumed to identify clinical causal effects.

With $v(\text{F}) = -1$ and $v(\text{D}) = v(\text{U}) = 1$, set

$$P(X = t_j \mid U = u, z) = \frac{1 + \eta v(u) t_j}{25}, \quad \eta = 0.3.$$

The symmetric grid makes this a normalized positive assignment law. An independent uniform variable transformed by the conditional inverse CDF defines X . A second, independent $E \sim \text{Uniform}(0, 1)$ defines the binary potential outcome $Y_{t_j} = \mathbf{1}\{E \leq r_z(t_j, U)\}$, with target $\tau_z(t_j) = \sum_u P(u \mid z) r_z(t_j, u)$. The full treatment-on-

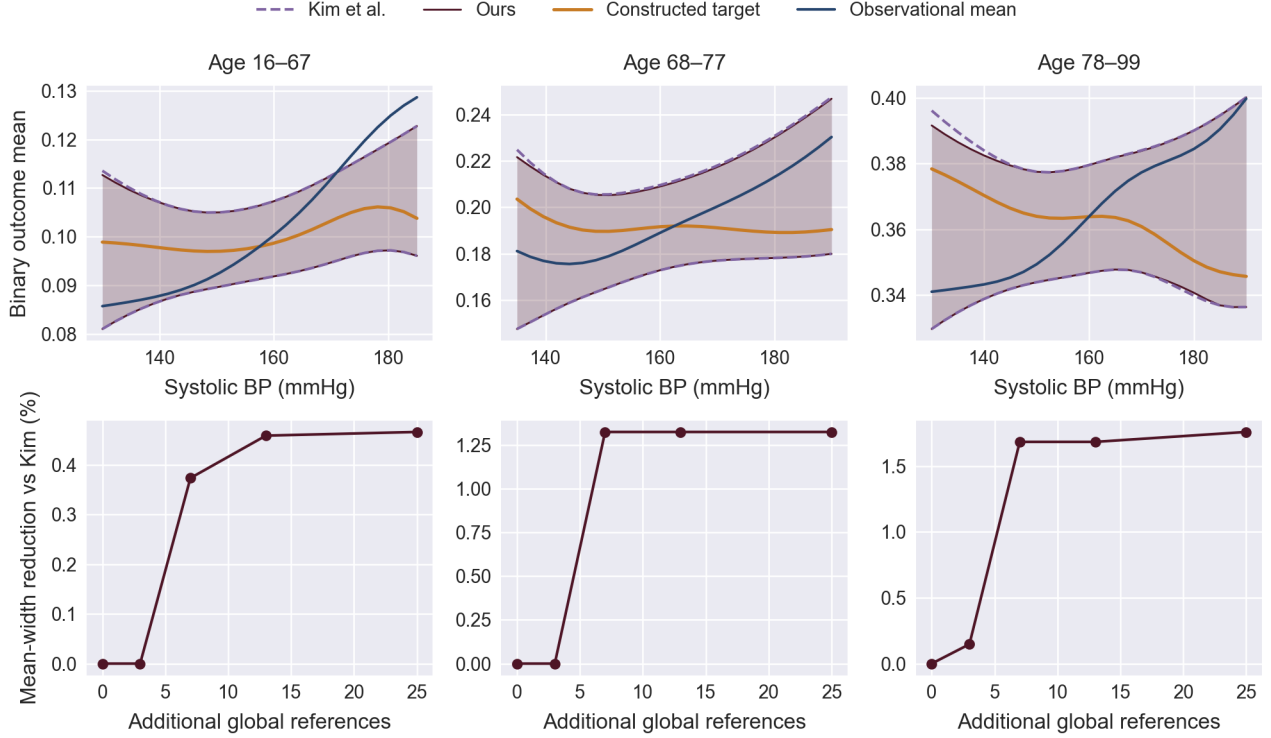


Figure 7: Controlled IST-derived binary SCM, with identical oracle STR/STS constants for both methods. Top: bounds, constructed interventional targets, and observational means in each age group. Bottom: the reduction in mean width as global references are added to the per-pair endpoints.

treated matrix is known by Bayes’ rule:

$$g_z(t_i, t_j) = \sum_u r_z(t_i, u) \frac{P(u | z)P(t_j | u, z)}{\sum_v P(v | z)P(t_j | v, z)}, \quad \mu_z(t_j) = g_z(t_j, t_j).$$

Every target, regression, and integration weight is an exact finite sum conditional on the fitted model. No simulated outcomes or estimated effect targets enter the evaluation.

For each age group, compute the maximum difference quotient of g_z over its first coordinate to obtain the STR coefficient b_z , and over its second coordinate to obtain the STS coefficient a_z . Both inequalities are checked over every matrix entry. These are oracle constants of the constructed mean-response model, not claims of pathwise smoothness for the binary outcome or of known propensity constants in the original data. Both methods receive exactly the same a_z, b_z , observational diagonal, treatment weights, and outcome range $[0, 1]$. Our full reference set includes every supported treatment; Kim uses the two endpoint references at each pair (t_i, t_j) . Constants are passed through a direct-cost interface without inventing overlap or density parameters.

Table 4 gives mean-width reductions of 0.47%, 1.32%, and 1.76%. Both methods cover all 75 constructed means, and our intervals are contained in Kim’s at every query. The constants and headline configuration were informed by a feasibility check, not preregistered. A nested reference ablation adds 0, 3, 7, 13, 25 global points to the per-pair endpoints, reproducing Kim exactly at zero and tightening monotonically thereafter (Fig. 7). An independent sparse linear program over the 25^2 ETT values, constrained only by the observed diagonal, STR, STS, and range, agrees with our full-reference endpoints to within 10^{-8} at the first, middle, and last query of every age group.

We retain every setting in sensitivity sweeps over $\eta \in \{0, 0.1, 0.3, 0.5, 0.7\}$; independent multipliers $\{1, 1.25, 2, 4\}$ of the oracle STR and STS constants; bandwidth multipliers $\{1, 2, 3\}$ of the full-tertile Silverman bandwidth; and grid sizes $\{13, 25, 49\}$. The bandwidth and grid sweeps rebuild the model and its oracle constants. Both methods retain full coverage in all settings, including settings where the width improvement vanishes. At

Table 5: Designed smooth continuous SCMs. Widths are normalized by each system’s outcome range. Random rows retain all 125 draws per family; reduction compares pooled mean widths. Canonical settings and neighborhoods were chosen after exploratory checks.

Setting	Family	Ours	Kim	Ours cov.	Kim cov.	Reduction (%)
Canonical	Polynomial	0.3480	0.3639	1.00	1.00	4.37
Canonical	Logistic	0.3243	0.3282	1.00	1.00	1.18
Canonical	Phase	0.2136	0.2279	1.00	1.00	6.29
Canonical	Curved linear	0.1316	0.1362	1.00	1.00	3.36
Random	Polynomial	0.3471	0.3616	1.00	1.00	4.01
Random	Logistic	0.3235	0.3271	1.00	1.00	1.08
Random	Phase	0.2172	0.2310	1.00	1.00	5.97
Random	Curved linear	0.1244	0.1276	1.00	1.00	2.51

$\eta = 0$, assignment is independent of the confounder, the STS coefficient is zero, and both bounds recover the model target exactly; relative width reductions are left undefined when Kim’s width is zero. The script `lpi_ist_controlled_variants.py` exports all settings, per-query intervals, fitted risks, latent proportions, calibration constants, and LP comparisons, with a checksum of the input data for reproducibility.

C.4 Additional continuous SCMs

We reuse four smooth, non-V-shaped response families with a bounded-propensity assignment mechanism to demonstrate separation from Kim beyond the V-shaped examples. These canonical settings and their random neighborhoods were chosen after exploratory checks; they are designed demonstrations, not a representative prior or a claim of universal gains. Let U have 201 equally spaced atoms on $[-2, 2]$, with probabilities proportional to $\exp(-u^2/2)$. On $[-1, 1]$, set

$$\begin{aligned}
p(x | u) &= \frac{1 + \rho v(u)x}{2}, \\
\text{Polynomial: } Y_x &= bx^2 + cU^2, & (b, c, \rho) &= (1, 1, 0.25), \\
\text{Logistic: } Y_x &= \left(1 + e^{-(bx - cU)}\right)^{-1}, & (b, c, \rho) &= (4, 1, 0.25), \\
\text{Phase: } Y_x &= \sin^2(wx) + \cos^2(wU) - 1, & (w, \rho) &= (1.3, 0.15), \\
\text{Curved linear: } Y_x &= bx + qx^2 + cU, & (b, q, c, \rho) &= (0.1, 1, 4, 0.10).
\end{aligned}$$

Here $v(u) = u^2/2 - 1$ for polynomial and phase, and $v(u) = u/2$ for logistic and curved linear. The curved-linear response includes a quadratic term and is not a strictly linear model. Applying the conditional inverse CDF to an independent uniform variable defines X . Every conditional density is normalized and positive. Both methods receive the same exact structural constants $L_X = \rho/2$, $p_{\min} = (1 - \rho)/2$, and $L_Y = 2b, b/4, w$, or $|b| + 2q$ for polynomial, logistic, phase, or curved linear, respectively, together with the exact outcome range over $x \in [-1, 1]$ and the finite latent support. Thus the STS coefficient is $C = L_X \text{diam}(\mathbb{Y})/p_{\min}$. Regressions and interventional means are exact latent sums. Both bounds use the same 257-point trapezoidal integration grid and 25 equally spaced queries; ours additionally uses all 257 reference points, augmented by the per-pair endpoints.

With seed 0, processing polynomial, phase, logistic, then curved linear, we draw 125 systems per family. Each response coefficient is multiplied independently by `Uniform(0.9, 1.1)`: (b, c) for polynomial and logistic, w for phase, and (b, q, c) for curved linear. Independently, ρ is drawn from $[0.20, 0.30]$ for polynomial and logistic, $[0.10, 0.20]$ for phase, or $[0.05, 0.15]$ for curved linear. All 500 draws are retained, without filtering by coverage, confounding, or separation. Table 5 reports mean interval widths divided by each system’s outcome range, with reductions computed from pooled mean widths. In the displayed order (polynomial, logistic, phase, curved linear), the canonical reductions are 4.37%, 1.18%, 6.29%, and 3.36%; the random-family reductions are 4.01%, 1.08%, 5.97%, and 2.51%. Both methods cover every query. Figure 8 shows the bounds for the four fixed canonical examples.

Checks verify conditional-density normalization, structural and ETT smoothness costs, coverage, interval or-

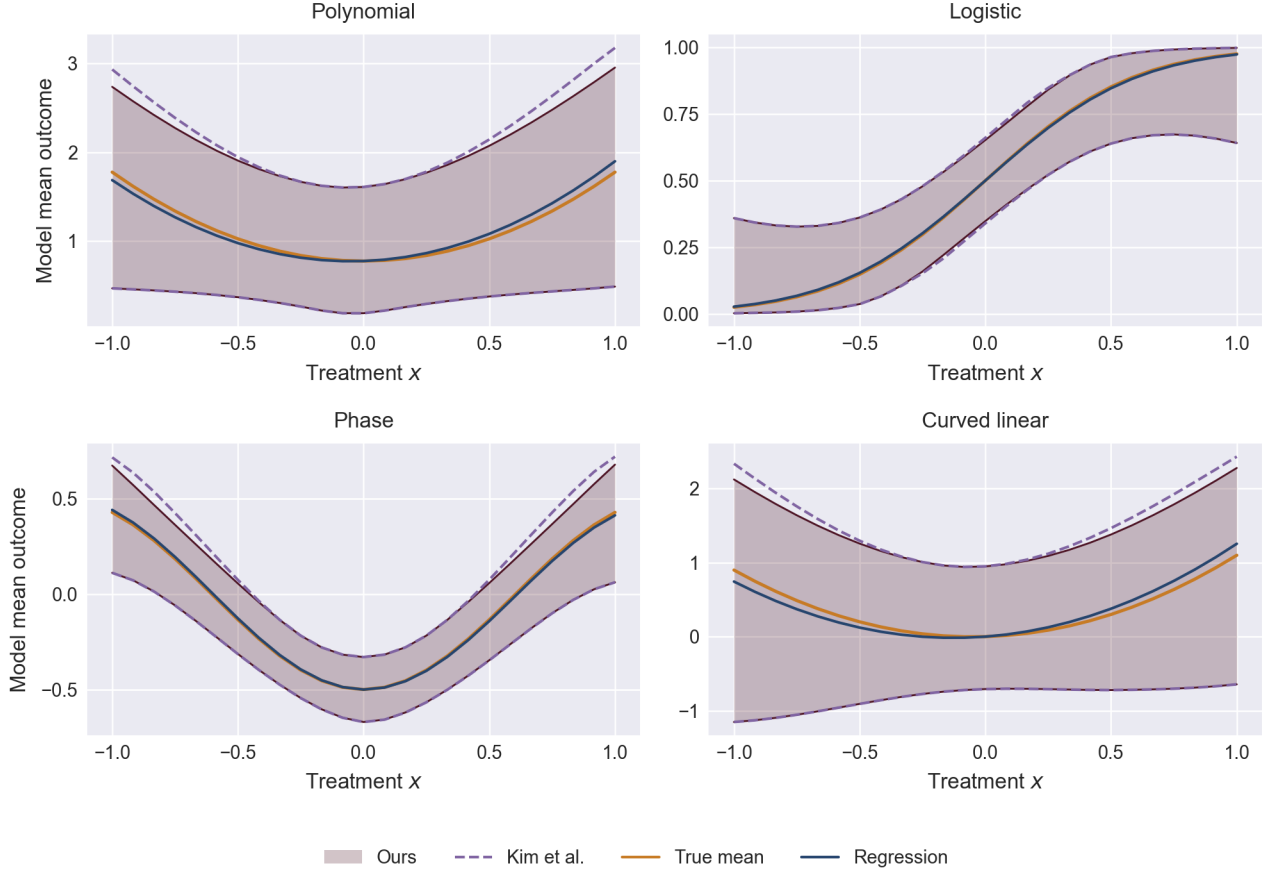


Figure 8: Designed smooth continuous SCMs with identical structural constants and outcome ranges for both methods. Each panel shows our full-reference bounds (dark red), Kim’s bounds (purple dashed), the true mean (orange), and the observational regression (blue). Top: polynomial and logistic; bottom: phase and curved linear. These are fixed canonical examples chosen after exploratory checks; Table 5 also reports all random draws in the stated neighborhoods.

dering, and containment within Kim. Endpoint-only references reproduce Kim; adding nested reference grids tightens monotonically. For each canonical example and the first and last random draw in each family, integration and reference grids are refined jointly and separately through 257, 513, and 1025 points, with endpoint changes required to stay below 10^{-4} of the outcome range. Run `lpi_additional_continuous.py --emit-paper-assets` to reproduce the parameters, intervals, aggregate results, reference ablations, refinement records, table, and figure; `--plot-only` redraws the saved results without rerunning the experiment.

C.5 Controlled reference-set ablation

To isolate the role of additional reference points, we evaluate the counterexample of Prop. 9 with the observational law and both costs held fixed: X is uniform on $\{0, 1, 2\}$, $\mu = (0, -1, 0)$, and $a = b = 1$. The supplied outcome range is $[-2, 2]$, shared by both methods; clipping to this range leaves the two-point envelopes unchanged. This experiment concerns the STR+STS class directly, rather than the structural-overlap constants of the synthetic SCMs above. We use the attaining SCM from App. B, with U uniform, $X = U$, and $Y_x = g(x, U)$, whose ETT rows are $(0, 0, 1)$, $(0, -1, 0)$, and $(1, 0, 0)$. Its true interventional means are $(1/3, -1/3, 1/3)$, and its STR and STS constants are both exactly one.

At each pair (x, x') , our reference set is $\{x, x'\} \cup R$, with R increasing through $\emptyset$, $\{0\}$, $\{0, 1\}$, and $\{0, 1, 2\}$. With $R = \emptyset$, our implementation reproduces Kim exactly. At $x = 2$, adding $s = 0$ does not help; adding the interior reference $s = 1$ lowers the upper endpoint from $2/3$ to $1/3$ while leaving the lower endpoint at -1 . Width

Table 6: Controlled reference ablation: fixed observational law, $a = b = 1$, and known outcome range $[-2, 2]$. Endpoints are $\{x, x'\}$ per stratum. Bounds and widths are at $x = 2$ in outcome units; coverage is over all three queries. Adding $s = 1$ reduces width by 20% without losing coverage.

References / method	Lower	Upper	Width	Coverage
Kim et al.	-1.000	0.667	1.667	1.00
Ours: endpoints	-1.000	0.667	1.667	1.00
Ours: endpoints + $\{0\}$	-1.000	0.667	1.667	1.00
Ours: endpoints + $\{0, 1\}$	-1.000	0.333	1.333	1.00
Ours: all references	-1.000	0.333	1.333	1.00

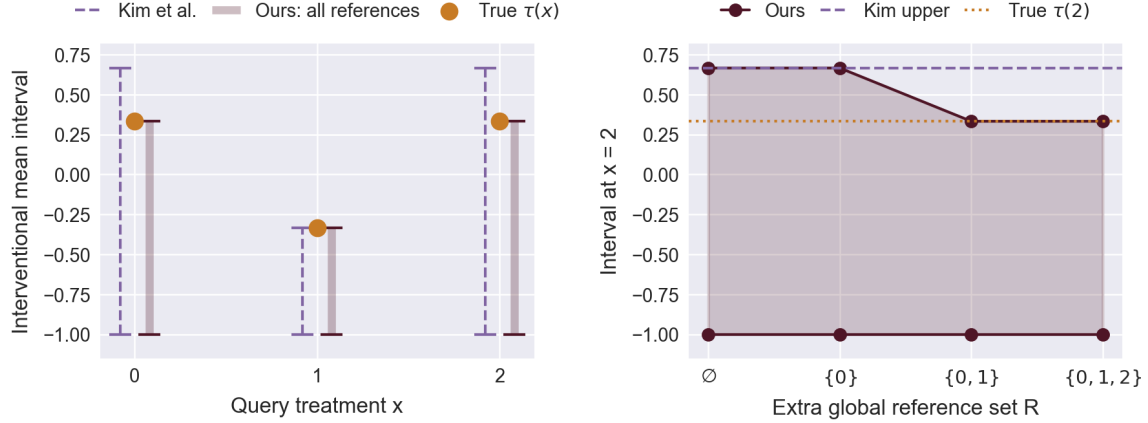


Figure 9: Controlled reference ablation with identical data and constants. Left: Kim’s intervals and our full-reference intervals at all three queries, with the true means marked. Right: the interval at $x = 2$ as the extra global reference set R grows. The interior reference $s = 1$ lowers the upper endpoint to the attained sharp value $1/3$; adding the remaining endpoint does not tighten it further.

therefore falls from $5/3$ to $4/3$, a 20% reduction, and all three true means remain covered at every stage (Table 6, Fig. 9). The final intervals agree with an independent linear program over all nine ETT values subject to the observed diagonal, STR, STS, and the outcome range. This is a controlled demonstration of strict improvement, not a claim of strict improvement on every system. A narrower supplied outcome range adds information and can remove the gap.