A Minimax Lower Bound for Interior Dose-Response Estimation

CausalSmith*

August 25, 2026

Abstract

This paper studies lower bounds for estimating an interior continuous-treatment dose-response partial mean. The target is

$$\theta_P(t_0) = \int \mu_P(t_0, x) p_{X,P}(x)\, dx,$$

interpretable as a causal dose-response mean under consistency, no unmeasured confounding, and local positivity. Over the same Hölder dose class $\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$, under iid sampling, bounded outcomes, an interior evaluation dose, the stated anisotropic Hölder restrictions, and a strict-slack baseline condition, we prove the same-class minimax lower bound

$$R_n\{\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0\} \geq c\, n^{-2\alpha/(2\alpha+1)}$$

for all sufficiently large n. The lower-bound exponent is the one-dimensional treatment-regression exponent and holds for every fixed $\beta > 0$, although the constant and slack-baseline feasibility may depend on β. We compare this lower floor with the published higher-order influence-function benchmark

$$\rho_n = n^{-2\alpha/(2\alpha+1)} \vee n^{-2/(1+d/(4s)+1/\alpha)}.$$

When $d \leq 4s$, the lower-bound exponent equals the exponent of this published comparator, whose upper theorem concerns a distinct localized-regularity class. When $4s < d$, the benchmark is governed by the covariate-smoothness term and has a strictly smaller exponent than the same-class lower-floor exponent. The paper establishes a same-class lower bound and an exact algebraic comparison with the external HOIF benchmark; a matching same-class upper analysis would complete the minimax characterization.

1 Introduction

Continuous treatments are central in many econometric and causal-inference problems: doses, prices, policy intensities, exposure levels, and other interventions are often better represented by a continuum than by a binary indicator. In the potential-outcome framework of Rubin [1974] and Robins [1986], the object of interest is commonly a dose-specific mean $E\{Y(t)\}$. Under consistency, conditional ignorability, and positivity, this causal object is identified by

*Machine-generated by the CausalSmith research pipeline. The displayed formal statements and proofs are Lean-verified under the paper's theorem-local citation interface; where a result from the literature is a formal dependency, it enters as a published input rather than a Lean-checked step, and cited work otherwise supplies framing and positioning.

an observed-data partial mean, averaging the conditional regression of Y on (A, X) at dose t over the marginal distribution of covariates. This paper studies the difficulty of estimating that fixed-dose partial mean at an interior dose.

The formal target is Definition 1,

$$\theta_P(t_0) = \int_{[0,1]^d} \mu_P(t_0, x) p_{X,P}(x)\, dx,$$

where $A \in [0,1]$, $X \in [0,1]^d$, and $\mu_P(a,x) = E_P[Y \mid A = a, X = x]$. The model class in Definition 2 imposes iid sampling, bounded outcomes, an interior evaluation dose, local positivity, Hölder smoothness of the outcome regression and treatment density in the treatment coordinate, and Hölder smoothness of the relevant covariate functions. The causal assumptions supply the interpretation of the same observed-data functional as a dose-response mean; the lower bound itself is a statement about the observed-data experiment and the minimax risk in Definition 3.

The continuous-treatment literature provides both the identification background and the estimator benchmarks for this problem. Early formulations include the generalized propensity-score and propensity-function approaches of Imbens [2000], Hirano and Imbens [2004], and Imai and van Dyk [2004], with comparisons developed by Zhao et al. [2013]. Semiparametric efficiency and influence-function methods provide the language for regular estimation and orthogonalization [Bickel et al., 1993, van der Vaart, 1998, van der Laan and Robins, 2003, Chernozhukov et al., 2018]. For nonparametric dose-response curves, Kennedy et al. [2017], Lee [2018], and Colangelo and Lee [2020] are direct antecedents. The closest upper-side benchmark for this paper is Bonvini and Kennedy [2022], which builds on higher-order influence-function methods [Robins et al., 2008, 2015] and reports the rate recorded here as Definition 4.

The main result is a same-class lower bound. Under the hypotheses of Theorem 1, including positive smoothness α, β, s, the stated interior and boundedness constants, and the strict-slack baseline condition in Assumption 12, there is a constant $c > 0$ such that, for all sufficiently large n,

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

This theorem establishes the lower half of the minimax analysis for the original Hölder dose class and fixes the treatment-regression obstruction at exponent $2\alpha/(2\alpha+1)$.

The comparison sequence is

$$\rho_n = n^{-2\alpha/(2\alpha+1)} \vee n^{-2/(1+d/(4s)+1/\alpha)}.$$

In the smooth-covariate regime $d \le 4s$, Proposition 1 and Theorem 2 show that the lower-bound exponent equals the exponent of ρ_n, giving an exact exponent comparison with the published benchmark. In the deficient-covariate regime $4s < d$, Theorem 3 shows that ρ_n is governed by the covariate-smoothness expression and has a strictly smaller exponent than the same-class lower floor. This ordering isolates the same-class upper behavior as the remaining question in that regime.

The lower-bound construction can be viewed as embedding the classical one-dimensional pointwise nonparametric regression obstruction inside the continuous-treatment partial-mean experiment. The embedding must preserve the observed-data restrictions of Definition 2: bounded outcomes, local positivity around t_0, smooth covariate density, treatment-density smoothness, and the causal dose-response interpretation. The strict-slack baseline condition in Assumption 12 is used to keep the least favorable alternatives inside the class while perturbing only

the treatment regression. This is why the resulting lower exponent is insensitive to the fixed treatment-density smoothness index β at the exponent level, while constants and feasibility may still depend on β.

The scope is deliberately narrower than neighboring continuous-exposure problems. Recent work studies modified treatment policies, weak-overlap or alternative identification regimes, nonparametric inference for dose-response curves, and incremental continuous-exposure interventions [Hejazi et al., 2022, Doss et al., 2022, Hudson et al., 2023, Shi et al., 2024, Zhang et al., 2024, Schindl et al., 2024, Bonvini et al., 2024]. Those estimands or positivity structures can change the role of the treatment density and the smoothing built into the target. The present paper concerns the fixed interior partial mean in Definition 1 under the same-class Hölder restrictions in Definition 2.

The precise proof scope is summarized in the verification appendix. The mathematical contribution is the same-class lower bound, the two algebraic comparisons with the published HOIF benchmark, and the explicit separation between those lower-bound statements and upper results proved for other localized-regularity classes.

2 Setup and observed-data class

We begin with the observed-data experiment. One observation is the observed unit $O = (Y, A, X)$, where the outcome Y is real-valued, the treatment dose A lies in $[0,1]$, and the covariate vector X lies in $[0,1]^d$. The single-observation law is denoted by P; its covariate marginal is P_X, with density $p_{X,P}$ on $[0,1]^d$. For $a \in [0,1]$ and $x \in [0,1]^d$, let the conditional mean and treatment density be

$$\mu_P(a,x) = E_P[Y \mid A = a, X = x], \qquad \pi_P(a \mid x) = \text{density of } A \mid X = x.$$

The outcome regression μ_P is the nuisance function whose value at the evaluation dose enters the target, while the conditional treatment density π_P governs the local amount of information available near that dose.

The sample consists of $O_1, \ldots, O_n$, generated from the product law P^n under the following sampling condition.

Assumption 1. [`ass:iid-sampling`] (Iid sampling) *The observations $O_1, \ldots, O_n$ are iid draws from P.*

Assumption 1 is the standard iid sampling condition for the nonparametric experiment and fixes P^n as the sampling law used in the risk calculations [Tsybakov, 2009].

Fix positive smoothness orders α, β, and s, a common envelope and Hölder radius M, a local positivity constant c_0, an interior evaluation dose t_0, and an interior half-width ε_0. Throughout, $\mathcal{H}^\gamma(M; \mathcal{D})$ denotes the Hölder ball of order γ and radius M on the domain $\mathcal{D}$. The analysis is local in the treatment coordinate, over the $[t_0 - \varepsilon_0, t_0 + \varepsilon_0]$.

We impose the basic observed-data restrictions before introducing the causal interpretation.

Assumption 2. [`ass:bounded-outcome`] (Bounded outcome) *Under P,*

$$|Y| \le M$$

almost surely.

Assumption 2 is the standard uniform bounded outcome condition; it keeps the loss and testing reductions within a common envelope over the model class [Bonvini and Kennedy, 2022].

Assumption 3. [`ass:interior-dose`] (Interior dose) *The exposure neighborhood around t_0 is contained in the interior of the exposure support:*

$$[t_0 - \varepsilon_0, t_0 + \varepsilon_0] \subset (0, 1).$$

Assumption 3 is the standard interior evaluation point condition, excluding boundary effects in the continuous-treatment coordinate [Colangelo and Lee, 2020].

Assumption 4. [`ass:local-positivity`] (Local positivity) *For every $a \in [t_0 - \varepsilon_0, t_0 + \varepsilon_0]$ and every $x \in [0, 1]^d$,*

$$\pi_P(a \mid x) \geq c_0.$$

Assumption 4 is the standard local positivity condition; it requires enough treatment density in the neighborhood of t_0 for the interior dose-response target to be statistically meaningful [Bonvini and Kennedy, 2022].

The smoothness restrictions separate regularity in the treatment coordinate from regularity in the covariate coordinate. This separation is useful because the lower bound varies with α, β, s, and d through distinct approximation and testing constraints.

Assumption 5. [`ass:mu-treatment-holder`] (Treatment-regression dose smoothness) *For every $x \in [0, 1]^d$, the map*

$$a \longmapsto \mu_P(a, x)$$

belongs to the Hölder ball $\mathcal{H}^\alpha(M; [t_0 - \varepsilon_0, t_0 + \varepsilon_0])$.

Assumption 5 is the standard Hölder smoothness in the treatment coordinate for the regression function near the target dose [Tsybakov, 2009].

Assumption 6. [`ass:pi-treatment-holder`] (Treatment-density dose smoothness) *For every $x \in [0, 1]^d$, the map*

$$a \longmapsto \pi_P(a \mid x)$$

belongs to the Hölder ball $\mathcal{H}^\beta(M; [t_0 - \varepsilon_0, t_0 + \varepsilon_0])$.

Assumption 6 is the standard Hölder smoothness condition for the conditional treatment density in the treatment coordinate [Bonvini and Kennedy, 2022].

Assumption 7. [`ass:mu-covariate-holder`] (Treatment-regression covariate smoothness) *The map*

$$x \longmapsto \mu_P(t_0, x)$$

belongs to the Hölder ball $\mathcal{H}^s(M; [0, 1]^d)$.

Assumption 7 is the standard Hölder smoothness condition in the covariate coordinate for the target-dose regression surface [Tsybakov, 2009].

Assumption 8. [`ass:pi-covariate-holder`] (Treatment-density covariate smoothness) *The map*

$$x \longmapsto \pi_P(t_0 \mid x)$$

belongs to the Hölder ball $\mathcal{H}^s(M; [0, 1]^d)$.

Assumption 8 imposes the standard covariate smoothness condition on the marginal covariate density [Bonvini and Kennedy, 2022].

Assumption 9. [ass:px-holder] (Covariate-density smoothness) *The marginal covariate density $p_{X,P}$ belongs to the Hölder ball*

$$\mathcal{H}^s(M;[0,1]^d)$$

and satisfies

$$0 \le p_{X,P}(x) \le M$$

for every $x \in [0,1]^d$.

Assumption 9 is the standard Hölder-smooth covariate density condition, together with a uniform upper bound on the marginal density [Bonvini and Kennedy, 2022].

The statistical target is the partial mean of the target-dose regression surface over the covariate distribution.

Definition 1. [def:theta-functional] (Dose-response partial mean $\theta_P(t_0)$) *The dose-response partial mean at the evaluation dose t_0 is*

$$\theta_P(t_0) = \int_{[0,1]^d} \mu_P(t_0, x)\, p_{X,P}(x)\, dx.$$

Definition 1 is an observed-data functional: it depends only on the conditional mean of Y given (A, X) and on the marginal law of X. The next subsection overlays the causal interpretation under which this same functional is the dose-response mean.

The causal overlay introduces potential outcomes $Y(a)$, indexed by $a \in [0,1]$, on the same probability space as the observed data. The following two assumptions are the identification conditions used to interpret Definition 1 causally.

Assumption 10. [ass:consistency] (Consistency) *Under P,*

$$Y = Y(A)$$

almost surely.

Assumption 10 is the standard consistency condition, linking the observed outcome to the potential outcome at the realized dose [Kennedy et al., 2017].

Assumption 11. [ass:no-unmeasured-confounding] (No unmeasured confounding) *For every $a \in [0,1]$, the potential outcome $Y(a)$ is independent of A conditional on X under P.*

Assumption 11 is the standard conditional ignorability condition for continuous treatments [Kennedy et al., 2017]. Together with local positivity in Assumption 4, it identifies the causal dose-response mean at the interior dose t_0 with the observed-data partial mean in Definition 1.

Combining these restrictions gives the observed-data class used for the lower-bound analysis. The minimax risk and the published benchmark rate referenced in the definition are introduced formally in Definition 3 and Definition 4.

Definition 2. [def:holder-dose-class] (Hölder Dose Class $\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$) *With the minimax risk $R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)$ and the published HOIF benchmark ρ_n as in Definitions 3 and 4, define*

$$\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$$

to be the collection of observed-data laws P satisfying all of the following conditions:

- ***(Sampling.)*** *The observations satisfy Assumption 1.*
- ***(Causal identification.)*** *Consistency and no unmeasured confounding hold as in Assumptions 10 and 11.*
- ***(Outcome boundedness.)*** *The bounded-outcome condition holds as in Assumption 2.*
- ***(Interior dose.)*** *The evaluation dose t_0 and local radius ε_0 satisfy Assumption 3.*
- ***(Local positivity.)*** *The treatment density obeys the local positivity condition with constant c_0 as in Assumption 4.*
- ***(Treatment-direction smoothness.)*** *The maps $a \mapsto \mu_P(a, x)$ and $a \mapsto \pi_P(a \mid x)$ satisfy the treatment-direction Hölder conditions in Assumptions 5 and 6.*
- ***(Covariate-direction smoothness and density range.)*** *The maps $x \mapsto \mu_P(t_0, x)$ and $x \mapsto \pi_P(t_0 \mid x)$ satisfy the covariate-direction Hölder conditions in Assumptions 7 and 8, and $p_{X,P}$ satisfies the Hölder, nonnegativity, and upper-envelope condition in Assumption 9.*
- ***(Semantic ties to the observed-data law.)*** *The regression, covariate-density, and treatment-density symbols are tied to P: $\mu_P(A, X)$ equals the conditional mean of Y given (A, X) P-almost surely; $p_{X,P}$ is the density of the X-marginal on $[0,1]^d$; and π_P is nonnegative on $[0,1] \times [0,1]^d$ with the (A, X)-law density $(a, x) \mapsto \pi_P(a \mid x) p_{X,P}(x)$ relative to Lebesgue measure on $[0,1] \times [0,1]^d$.*

Definition 2 is a statistical model for observed laws equipped with the stated causal interpretation. The lower-bound alternatives are constructed as observed-data laws in this class; the potential-outcome overlay supplies the causal reading of the same target while preserving the statistical experiment.

For the least favorable submodels used later, we also require a baseline law at a positive distance from the model boundary.

Assumption 12. [`ass:baseline-submodel-slack`] (Baseline submodel slack) *There exist a covariate density p_0 on $[0,1]^d$, a conditional treatment density q_0 on $[0,1]$, a constant $\eta_0 > 0$, and a constant B_0 with $0 < B_0 < M$, such that $p_0(x) \geq 0$ for all $x \in [0,1]^d$, $q_0(a) \geq 0$ for all $a \in \mathbb{R}$,*

$$\int_{[0,1]^d} p_0(x)\, dx = 1 \qquad \text{and} \qquad \int_0^1 q_0(a)\, da = 1,$$

$$\|p_0\|_{\mathcal{H}^s([0,1]^d)} \leq M - \eta_0,$$

q_0 has Hölder norm at most $M - \eta_0$ of order β on $[t_0 - \varepsilon_0, t_0 + \varepsilon_0]$,

$$q_0(a) \geq c_0 + \eta_0 \qquad \text{for all } a \in [t_0 - \varepsilon_0, t_0 + \varepsilon_0],$$

and

$$p_0(x) \leq M - \eta_0 \qquad \text{for all } x \in [0,1]^d.$$

Assumption 12 is specific to this analysis. It provides an interior baseline density p_0, an interior baseline treatment density q_0, and a positive margin η_0, so that the perturbations used in the lower-bound construction can remain inside Definition 2.

3 Main results

The target in Definition 1 is evaluated under squared loss over the observed-data class in Definition 2. We first fix the risk criterion. The estimator is allowed to be any measurable function of the iid sample, with range restricted to the outcome envelope; this normalization is immaterial for the rate comparisons below because the target itself lies in the same bounded range.

Definition 3. [`def:minimax-risk`] (Minimax risk $R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)$) *The minimax risk over the model class $\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$ at the evaluation dose t_0 is*

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) = \inf_{\substack{\hat{\theta}_n \text{ measurable} \\ \hat{\theta}_n(O_1,\ldots,O_n)\in[-M,M] \text{ for every sample}}} \sup_{P\in\mathcal{P}_{\alpha,\beta,s}(M,c_0,\varepsilon_0,t_0)} E_{P^{\otimes n}}\Big[\Big\{\hat{\theta}_n(O_1,\ldots,O_n) - \theta_P(t_0)$$

where $P^{\otimes n}$ is the product law of the iid sample $O_1, \ldots, O_n$.

The benchmark used for comparison is the rate reported for higher-order influence-function methods in the continuous-exposure literature. We use $x \vee y$ to denote $\max\{x, y\}$.

Definition 4. [`def:published-hoif-rate`] (Published HOIF benchmark ρ_n) *The published higher-order influence-function benchmark rate is*

$$\rho_n = n^{-2\alpha/(2\alpha+1)} \vee n^{-2/(1+d/(4s)+1/\alpha)}.$$

Definition 4 records the comparison sequence from the published localized-regularity upper result. The first term is the one-dimensional nonparametric regression scale in the treatment coordinate, while the second is the higher-order influence-function benchmark involving the covariate dimension and covariate smoothness [Robins et al., 2008, 2015, Kennedy et al., 2017, Bonvini and Kennedy, 2022]. The lower bound below is the paper's same-class result.

Theorem 1. [`thm:sharp-pointwise-lower-bound`] (Sharp Pointwise Lower Bound) *For every covariate dimension d and every real constants $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0$, suppose that*

- ***(Positive smoothness.)*** *$\alpha > 0$, $\beta > 0$, and $s > 0$.*
- ***(Regime constants.)*** *$M > 0$, $c_0 > 0$, $t_0 \in (0, 1)$, $\varepsilon_0 \in (0, 1/2)$, and the local window determined by (t_0, ε_0) is contained in $(0, 1)$.*
- ***(Strict-slack baseline.)*** *Assumption 12 holds for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$.*

Then there exists a constant $c > 0$ such that, for all sufficiently large sample sizes n,

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,,$$

where $\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$ is the model class of Definition 2 and R_n is the minimax risk of Definition 3.

Theorem 1 gives a lower floor for the original Hölder dose class, uniformly over every positive value of the treatment-density smoothness parameter β. Its construction varies the treatment regression while keeping the density components within the slack allowed by Assumption 12. The result fixes the same-class treatment-regression obstruction and supplies the lower half of the minimax analysis.

The next statements reduce the comparison with ρ_n to algebraic regimes. When the covariate smoothness is high enough relative to dimension, the maximum in Definition 4 is attained by the treatment-regression term.

Proposition 1. [prop:oracle-regime-reduction] (Oracle Regime Reduction) *Let d be the covariate dimension, and let $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0$ be constants. Suppose that*

- ***(Well-formed constants.)*** *$\alpha > 0$, $\beta > 0$, $s > 0$, $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, and the window $[t_0 - \varepsilon_0, t_0 + \varepsilon_0]$ is contained in $(0,1)$.*
- ***(Oracle regime.)*** *$d \le 4s$.*
- ***(Baseline slack.)*** *Assumption 12 holds for $d, \beta, s, M, c_0, \varepsilon_0, t_0$.*

Then there exists a constant $c > 0$ such that, for all sufficiently large sample sizes n,

$$\rho_n = n^{-2\alpha/(2\alpha+1)}$$

for the published HOIF benchmark rate ρ_n of Definition 4, and

$$c\,\rho_n \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,,$$

where the model class is Definition 2 and the minimax risk is Definition 3.

Proposition 1 gives the exact comparison statement for $d \le 4s$: the certified lower-bound exponent agrees with the exponent of the published benchmark sequence. Construction of an estimator over the same Hölder class is the corresponding upper-analysis question.

Theorem 2. [thm:sharp-minimax-smooth-covariate] (Smooth Covariate Minimax Floor) *Let $d \in \mathbb{N}$ and let $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0 \in \mathbb{R}$. Suppose that:*

- ***(Smoothness.)*** *$\alpha > 0$, $\beta > 0$, and $s > 0$.*
- ***(Regime constants.)*** *$M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, and the window $[t_0 - \varepsilon_0, t_0 + \varepsilon_0]$ is contained in $(0,1)$.*
- ***(Smooth-covariate regime.)*** *$d \le 4s$.*
- ***(Strict-slack baseline.)*** *Assumption 12 holds for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$.*

Then there exists a constant $c > 0$ such that, for all sufficiently large n,

$$c\,n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,,$$

where R_n is the minimax risk of Definition 3 over the same-class model of Definition 2. Moreover,

$$n^{-2\alpha/(2\alpha+1)} = \rho_n,$$

with ρ_n as in Definition 4.

Theorem 2 packages the same conclusion in the smooth-covariate regime: it establishes a lower bound for Definition 2 and an exact exponent comparison with ρ_n.

When $4s < d$, the published benchmark is governed by the covariate-smoothness term. The lower bound remains on the treatment-regression scale, and the benchmark exponent is strictly smaller.

Theorem 3. [thm:frontier-bracket-deficient] (Deficient Frontier Bracket) *Fix a covariate dimension $d \in \mathbb{N}$ and constants $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0 \in \mathbb{R}$. Assume:*

- ***(Regime constants.)*** $\alpha > 0$, $\beta > 0$, $s > 0$, $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, *and the window* $[t_0 - \varepsilon_0, t_0 + \varepsilon_0]$ *lies in* $(0,1)$.
- ***(Deficient covariate smoothness.)*** $4s < d$.
- ***(Strict-slack baseline.)*** *Assumption 12 holds for* $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$.

Then there exists a constant $c > 0$ *such that, for all sufficiently large sample sizes* n,

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,,$$

where the minimax risk is as in Definition 3 and the class is Definition 2. Moreover, for the published HOIF benchmark of Definition 4,

$$\rho_n = n^{-\,2/(1+d/(4s)+1/\alpha)}$$

and its exponent is strictly smaller than the certified lower-bound exponent:

$$\frac{2}{1 + d/(4s) + 1/\alpha} < \frac{2\alpha}{2\alpha + 1}.$$

Thus, in the low-covariate-smoothness regime, Theorem 3 determines the strict ordering between the same-class lower floor and the exponent in the published benchmark sequence. The same-class upper frontier remains open among the benchmark rate, an intermediate rate, and a rate depending more directly on β.

For orientation, we record the external higher-order influence-function result used as a literature comparator. The cited bound from Bonvini and Kennedy [2022] is stated over an explicitly defined, more regular subclass of the Hölder dose class in Definition 2 and supplies the corresponding upper-side benchmark.

Cited result 1. [`lem:published-upper-bound-cited`] (Published HOIF Comparator) *(imported result; machine-checked against the cited source, not proved here.) For covariate dimension* d *and parameters* $\alpha, \beta, s, \gamma_1, \gamma_2$, *consider any Bonvini–Kennedy HOIF estimator specification* E *consisting of an observed-data law* P, *evaluation dose* t_0, *nuisance estimators* $\hat{\mu}_n, \hat{\pi}_n, \hat{p}_n$, *kernel* K, *bandwidths* h_n, *projection dimensions* k_n, *and true and estimated projection kernels* Π_n *and* $\widehat{\Pi}_n$. *Suppose the following conditions of Bonvini and Kennedy [2022, Theorem 3.1 and §3.4] hold for these same inputs:* P *is a probability law;* μ_P, $p_{X,P}$, *and* π_P *are respectively the outcome regression, covariate density, and conditional treatment density of* P; *there exist* $0 < \ell \le u$ *such that, for every* n, a, x, $\ell \le \pi_P(a \mid x) \le u$ *and* $\ell \le \hat{\pi}_n(a \mid x) \le u$; *there exists* $B > 0$ *such that* $|Y| \le B$ *and* $|A| \le B$ P*-almost surely and* $|\hat{\mu}_n(a,x)| \le B$ *for every* n, a, x; *and, for every* n *and* x, *the maps* $a \mapsto \mu_P(a,x)$ *and* $a \mapsto \hat{\mu}_n(a,x)$ *lie in the one-dimensional Hölder ball of order* α *and radius* 1 *on* $\mathbb{R}$, *while* $a \mapsto \pi_P(a \mid x)$ *and* $a \mapsto \hat{\pi}_n(a \mid x)$ *lie in the corresponding ball of order* β. *Writing* $K_{h_n,t_0}(a) = h_n^{-1} K((a - t_0)/h_n)$, *assume also that there exist* $K_{\max} > 0$ *and* $0 < g_\ell \le g_u$ *such that* $|K(u)| \le K_{\max}$, $\operatorname{supp}(K) \subseteq [-1,1]$, $\int K(u)\,du = 1$, $\int u^j K(u)\,du = 0$ *for* $1 \le j \le \lceil \alpha \rceil - 1$, $h_n > 0$, $\int K_{h_n,t_0}(a)\,da = 1$, *and, for every* n, x, $g_\ell \le \int K_{h_n,t_0}(a)\pi_P(a \mid x) p_{X,P}(x)\,da \le g_u$; *that there exists* $C_\Pi > 0$ *with* $\Pi_n(x,x) \le C_\Pi k_n$ *and* $\widehat{\Pi}_n(x,x) \le C_\Pi k_n$ *for every* n, x; *and that there exist* $0 < r_\ell \le r_u$ *such that, for every* n, x,

$$r_\ell \le \frac{\int K_{h_n,t_0}(a)\pi_P(a \mid x) p_{X,P}(x)\,da}{\int K_{h_n,t_0}(a)\hat{p}_n(a,x)\,da} \le r_u.$$

Assume further that $d > 0$, $\alpha > 0$, $\beta > 0$, $s > 0$, $\alpha \le \beta$, $\gamma_1 = \gamma_2 = s$, and that there exists $B_X > 0$ such that, for every a and n, the maps $x \mapsto \mu_P(a, x)$, $x \mapsto \hat{\mu}_n(a, x)$, $x \mapsto \pi_P(a \mid x)$, and $x \mapsto \hat{\pi}_n(a \mid x)$ lie in the s-Hölder ball of radius B_X on $[0,1]^d$. Let $v_n(x) = \hat{\mu}_n(t_0, x) - \mu_P(t_0, x)$ and $q_n(x) = \hat{\pi}_n(t_0 \mid x)^{-1} - \pi_P(t_0 \mid x)^{-1}$. The rate specialization requires

$$\|v_n - \Pi_n v_n\|_{L^2([0,1]^d)} \|q_n - \Pi_n q_n\|_{L^2([0,1]^d)} \lesssim k_n^{-(\gamma_1+\gamma_2)/d}$$

eventually, $\|v_n\|_{L^2([0,1]^d)} \asymp \|q_n\|_{L^2([0,1]^d)}$ eventually, and

$$\left(\|v_n\|_{L^2([0,1]^d)} \|q_n\|_{L^2([0,1]^d)} \|\hat{p}_n(t_0, \cdot) - \pi_P(t_0 \mid \cdot) p_{X,P}(\cdot)\|_{\infty,[0,1]^d}\right)^2 \lesssim \rho_n$$

eventually. The tuning satisfies, when $d \le 4s$, $h_n \asymp n^{-1/(2\alpha+1)}$ and $k_n \asymp n h_n$ eventually, and, when $4s < d$, $h_n \asymp n^{-4s/(\alpha(4s+d)+4s)}$ and $k_n \asymp (n h_n)^{2d/(d+4s)}$ eventually. Then the specified Bonvini–Kennedy HOIF estimator $\hat{\theta}_n$, with its nuisance-training outputs held fixed, attains the published benchmark conditional mean-squared-error rate of Definition 4: there exists $C > 0$ such that, for all sufficiently large n,

$$\int (\hat{\theta}_n(O_1, \ldots, O_n) - \theta_P(t_0))^2 \, dP^n \le C\rho_n, \qquad \rho_n = \max\left\{n^{-2\alpha/(2\alpha+1)}, n^{-2/(1+d/(4s)+1/\alpha)}\right\}.$$

The notation inside the cited display belongs to the imported comparator. In the present paper, it motivates the sequence ρ_n and provides an upper-side benchmark from a strictly smaller, more regular class; Definition 3 reserves $R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)$ for the same-class risk.

The final two remarks summarize the established rate comparison and identify the remaining same-class upper frontier.

Remark 1. [`def:beta-frontier-handle`] (Beta frontier handle $\mathfrak{B}_{\alpha,\beta,s}$) *The beta comparison handle $\mathfrak{B}_{\alpha,\beta,s}$ denotes the following frontier statement. For every $\beta > 0$, there is a constant $c > 0$ such that, for all sufficiently large n,*

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

For every $n \ge 1$, when $d \le 4s$, the published benchmark satisfies

$$\rho_n = n^{-2\alpha/(2\alpha+1)}.$$

For every $n \ge 1$, when $0 < s$ and $4s < d$, the published benchmark satisfies

$$\rho_n = n^{-2/(1+d/(4s)+1/\alpha)}$$

and its exponent obeys

$$\frac{2}{1 + d/(4s) + 1/\alpha} < \frac{2\alpha}{2\alpha + 1}.$$

The upper-frontier alternatives are recorded as the following disjunction: either there is a constant $C > 0$ such that, for all sufficiently large n,

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \le C\rho_n,$$

or there are κ and $C > 0$ with

$$\frac{2}{1 + d/(4s) + 1/\alpha} < \kappa < \frac{2\alpha}{2\alpha + 1}$$

such that, for all sufficiently large n,

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \leq Cn^{-\kappa},$$

or there are $\beta' > 0$, $\beta' \neq \beta$, exponents κ, κ' with $\kappa' \neq \kappa$, and positive constants c, C, c', C' such that, for all sufficiently large n,

$$cn^{-\kappa} \leq R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \leq Cn^{-\kappa}$$

and

$$c'n^{-\kappa'} \leq R_n\big(\mathcal{P}_{\alpha,\beta',s}(M, c_0, \varepsilon_0, t_0), t_0\big) \leq C'n^{-\kappa'}.$$

Remark 1 is an interpretive shorthand for the preceding lower-bound and comparison statements. The word "frontier" denotes this exponent-level comparison and highlights the benchmark-aligned smooth-covariate regime.

4 Discussion, limitations, and extensions

The lower-bound analysis isolates one obstruction that is already present in the standard interior partial mean problem. The target in Definition 1 is an observed-data partial mean at a fixed interior dose, and the risk in Definition 3 is taken over the same Hölder dose class of Definition 2. Within that class, Theorem 1 gives a treatment-regression lower floor. The comparison statements in Theorem 2 and Theorem 3 relate this floor to the published benchmark sequence ρ_n from Definition 4, while Cited result 1 supplies the external Bonvini–Kennedy comparator on its localized-regularity class.

For each fixed $\beta > 0$, under the corresponding slack-baseline condition in Assumption 12, the exponent of the same-class lower bound is invariant to the treatment-density smoothness index. The constant in the lower bound, and the feasibility of the slack baseline itself, may depend on β and on the other fixed model constants. Thus the conclusion is exponent-level insensitivity to β, with constants calibrated separately for each fixed smoothness value.

This distinction matters most when $4s < d$. In that regime, Theorem 3 shows that the published benchmark sequence has a smaller exponent than the same-class lower-bound exponent. The unresolved same-class question is whether the upper behavior in low covariate smoothness follows the published benchmark, an intermediate exponent, or a rate that depends more directly on the smoothness of the treatment density.

The scope is tied to the interior positivity regime. Assumption 4 imposes a fixed lower bound on the treatment density in a neighborhood of t_0, and Assumption 3 selects an interior evaluation window. These restrictions match the usual interior continuous-exposure formulation. Weak-overlap, design-adaptive, and trimmed targets lead to a different statistical experiment in which the amount of information near t_0 enters the rate calculation directly. Recent work on continuous exposures and overlap-sensitive causal estimands emphasizes these complementary regimes [Hejazi et al., 2022, Doss et al., 2022, Hudson et al., 2023, Shi et al., 2024, Zhang et al., 2024, Schindl et al., 2024, Bonvini et al., 2024].

Neighboring estimands raise a similar qualification. The partial mean studied here evaluates the regression surface at a single interior dose and averages over the marginal law of X. Other continuous-exposure targets average over dose neighborhoods, consider stochastic interventions, or target modified exposure policies. Their definitions can smooth the treatment coordinate or alter the role of the conditional treatment density. The present lower bound is consequently a

benchmark for the fixed-dose interior partial mean, while those neighboring functionals invite estimand-specific minimax analyses.

Finally, the comparison with higher-order influence-function methods is a comparison across stated model classes. The imported result in Cited result 1 supplies an upper-side benchmark under additional localized regularity, while the lower bound here is proved over the Hölder class in Definition 2. Two routes can bridge the statements: a same-class upper bound under Definition 2, or an inclusion argument covering the present Hölder class by the localized-regularity class with the needed constants.

Appendices

A auxiliary lemmas and proofs

This appendix collects the auxiliary mathematical statements used to support the lower-bound and rate-comparison results in Theorem 1–Theorem 3. The statements are organized according to their role: first the two-point testing ingredients, then the algebra for the benchmark sequence ρ_n, and finally the combined lower-bound comparison used to package the main conclusions. Throughout this appendix, $\mathrm{KL}(Q_0, Q_1)$ denotes the Kullback–Leibler divergence from Q_0 to Q_1.

The testing reduction begins with a bounded binary outcome channel. It is convenient because the least favorable alternatives only need to separate conditional means while keeping outcomes inside the common envelope. The following elementary inequality controls the information distance between two such channels by the squared separation of their means.

Lemma 1. [lem:bernoulli-mean-channel-kl] (Bernoulli mean-channel KL bound) *Fix $B > 0$. Let Q_u and Q_v be distributions on $\{-B, B\}$ with means u and v, respectively, so that*

$$Q_u(B) = \frac{1 + u/B}{2}, \qquad Q_v(B) = \frac{1 + v/B}{2}.$$

If $|u| \le B/2$ and $|v| \le B/2$, then

$$\mathrm{KL}(Q_u, Q_v) \le \frac{2(u - v)^2}{B^2}.$$

Proof of Lemma 1. Throughout, $\mathrm{Bern}(r)$ denotes the two-point law on $\{0, 1\}$ with

$$\mathrm{Bern}(r)(\{1\}) = r, \qquad \mathrm{Bern}(r)(\{0\}) = 1 - r.$$

Set

$$p = \frac{1 + u/B}{2}, \qquad q = \frac{1 + v/B}{2},$$

so that, by the definition of the channels in the statement,

$$Q_u(\{B\}) = p, \quad Q_u(\{-B\}) = 1 - p, \qquad Q_v(\{B\}) = q, \quad Q_v(\{-B\}) = 1 - q.$$

Step 1 (the two success probabilities lie in the central band). Since $B > 0$,

$$p - \frac{1}{2} = \frac{u}{2B}, \qquad q - \frac{1}{2} = \frac{v}{2B},$$

and the hypotheses $|u| \le B/2$ and $|v| \le B/2$ give $|u/(2B)| \le 1/4$ and $|v/(2B)| \le 1/4$. Hence

$$\frac{1}{4} \le p \le \frac{3}{4}, \qquad \frac{1}{4} \le q \le \frac{3}{4}.$$

Step 2 (transport to the $\{0,1\}$ parametrization). Let

$$T(x) = 2Bx - B.$$

Because $B > 0$, the map T is an affine bijection of the real line whose inverse is again affine, hence measurable, and $T(0) = -B$, $T(1) = B$. Consequently the image of $\mathrm{Bern}(p)$ under T puts mass p at B and mass $1-p$ at $-B$, which is exactly Q_u; the same computation with q in place of p gives Q_v as the image of $\mathrm{Bern}(q)$:

$$Q_u = \mathrm{Bern}(p) \circ T^{-1}, \qquad Q_v = \mathrm{Bern}(q) \circ T^{-1}.$$

Applying one and the same measurable bijection to both arguments leaves the Kullback–Leibler divergence unchanged, since the Radon–Nikodym derivative of the two image laws is the transport of the original one along T. Therefore

$$\mathrm{KL}(Q_u, Q_v) = \mathrm{KL}\big(\mathrm{Bern}(p), \mathrm{Bern}(q)\big).$$

Step 3 (quadratic band bound for the Bernoulli divergence). For $p, q \in (0,1)$ the divergence of two-point laws is the explicit sum

$$\mathrm{KL}\big(\mathrm{Bern}(p), \mathrm{Bern}(q)\big) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q},$$

and on the central band this is dominated by a quadratic:

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \le 4(p-q)^2 \qquad \text{whenever } p, q \in \left[\tfrac{1}{4}, \tfrac{3}{4}\right].$$

Indeed, writing $D(t) = t \log t + (1-t)\log(1-t)$ for the negative Bernoulli entropy, the left-hand side is the Bregman remainder $D(p) - D(q) - D'(q)(p-q)$, so the assertion is that the function

$$H_q(t) = 4(t-q)^2 + D'(q)(t-q) + D(q) - D(t)$$

is nonnegative at $t = p$; and it is, because $H_q(q) = 0$, $H_q'(q) = 0$, and H_q is convex on $[1/4, 3/4]$, its second derivative there being

$$H_q''(t) = 8 - \frac{1}{t} - \frac{1}{1-t} \;\ge\; 8 - 4 - 4 = 0,$$

so that q minimizes H_q over that interval.

Step 4 (back to the mean parametrization). Step 1 supplies the band hypothesis needed in Step 3, and

$$p - q = \frac{u - v}{2B},$$

so

$$4(p-q)^2 = 4\left(\frac{u-v}{2B}\right)^2 = \frac{(u-v)^2}{B^2} \le \frac{2(u-v)^2}{B^2},$$

the last inequality because $(u-v)^2/B^2 \ge 0$. Chaining Steps 2–4,

$$\mathrm{KL}(Q_u, Q_v) \le \frac{2(u-v)^2}{B^2},$$

which is the claim. □

Lemma 1 supplies the local information calculation for the outcome part of the two-point experiment. The boundedness condition $|u|, |v| \le B/2$ keeps both Bernoulli probabilities uniformly away from zero and one, so the quadratic control of the divergence is uniform over the perturbations used in the construction.

The next ingredient is the standard two-point minimax testing bound. It translates a separation in the parameter into a lower bound on mean-squared error whenever the two induced sample laws remain close in information distance; this is the usual Le Cam device for nonparametric lower bounds [Tsybakov, 2009].

Lemma 2. [lem:le-cam-two-point-mse-source] (Le Cam Two-Point Bound) *Let Q_0 and Q_1 be two laws for the n-sample experiment, and let θ_0 and θ_1 be the corresponding values of a real parameter. Set*

$$\Delta = |\theta_1 - \theta_0|.$$

If

$$\mathrm{KL}(Q_0, Q_1) \le K < \infty,$$

then there exists a constant $c_K > 0$, depending only on K, such that every estimator T satisfies

$$\max_{i=0,1} E_{Q_i}\big[(T - \theta_i)^2\big] \ge c_K \Delta^2.$$

Proof of Lemma 2. Throughout the proof, $\mathrm{KL}(Q_0, Q_1)$ denotes the extended-real Kullback–Leibler divergence from Q_0 to Q_1; when this divergence is finite, its real value is

$$\int \log\left(\frac{\mathrm{d}Q_0}{\mathrm{d}Q_1}\right) \mathrm{d}Q_0.$$

Write

$$\mathrm{TV}(Q_0, Q_1) = \sup_{S \text{ measurable}} \big|Q_0(S) - Q_1(S)\big|$$

for total variation. Let $K_+ = \max\{K, 0\}$ and define, before any laws or estimator are chosen,

$$c_K = \frac{\exp(-K_+)}{32} > 0.$$

Fix probability laws Q_0, Q_1 on a common measurable sample space satisfying the finite-budget hypothesis, and fix a measurable estimator T such that the two squared losses $(T - \theta_0)^2$ under Q_0 and $(T - \theta_1)^2$ under Q_1 are integrable. Put

$$r = \frac{|\theta_1 - \theta_0|}{2} = \frac{\Delta}{2} \ge 0.$$

Step 1 (estimation to testing). Since $2r = |\theta_0 - \theta_1|$, the two measurable error sets

$$S_0 = \{|T - \theta_0| \ge r\}, \qquad S_1 = \{|T - \theta_1| \ge r\}$$

cover the sample space in the following sense: if S_0 fails, then $|T - \theta_0| < r$, and the triangle inequality gives

$$|T - \theta_1| \ge |\theta_0 - \theta_1| - |T - \theta_0| > 2r - r = r.$$

Thus $S_0^c \subseteq S_1$. The testing characterization of total variation therefore gives

$$Q_0(S_0) + Q_1(S_1) \ge Q_0(S_0) + Q_1(S_0^c) \ge 1 - \mathrm{TV}(Q_0, Q_1).$$

Since a sum of two real numbers is at most twice their maximum,

$$\frac{1-\mathrm{TV}(Q_0,Q_1)}{2} \le \max\Big\{Q_0(|T-\theta_0|\ge r), Q_1(|T-\theta_1|\ge r)\Big\}.$$

Step 2 (a testing floor for a finite budget). The budget hypothesis is an inequality in the extended nonnegative reals against the finite value associated with the real number K. Hence the divergence is finite, so $Q_0 \ll Q_1$, and its real value satisfies

$$\mathrm{KL}(Q_0,Q_1) \le K_+.$$

The Bretagnolle–Huber affinity bound applies under these two side conditions and yields

$$\frac{1}{2}\exp\{-\mathrm{KL}(Q_0,Q_1)\} \le 1-\mathrm{TV}(Q_0,Q_1).$$

Because $\mathrm{KL}(Q_0,Q_1) \le K_+$, monotonicity of the exponential gives $\exp(-K_+) \le \exp\{-\mathrm{KL}(Q_0,Q_1)\}$. Combining this inequality with the preceding display and Step 1 gives

$$\frac{\exp(-K_+)}{4} \le \frac{1-\mathrm{TV}(Q_0,Q_1)}{2} \le \max\Big\{Q_0(|T-\theta_0|\ge r), Q_1(|T-\theta_1|\ge r)\Big\}.$$

Step 3 (testing back to squared error). For $i=0,1$, the event $|T-\theta_i|\ge r$ is the event $(T-\theta_i)^2 \ge r^2$. The squared loss is nonnegative and integrable under the corresponding law, so the integral form of Markov's inequality gives

$$r^2\, Q_i(|T-\theta_i|\ge r) \le E_{Q_i}\big\{(T-\theta_i)^2\big\}, \qquad i=0,1.$$

Writing

$$p_i = Q_i(|T-\theta_i|\ge r), \qquad m_i = E_{Q_i}\big\{(T-\theta_i)^2\big\},$$

and distinguishing whether $p_0 \le p_1$ or $p_1 \le p_0$, these two inequalities imply

$$r^2 \max\Big\{Q_0(|T-\theta_0|\ge r), Q_1(|T-\theta_1|\ge r)\Big\} \le \max_{i=0,1} E_{Q_i}\big\{(T-\theta_i)^2\big\}.$$

Step 4 (conclusion). Multiplying the lower bound from Step 2 by $r^2 \ge 0$ and using Step 3,

$$r^2\,\frac{\exp(-K_+)}{4} \le \max_{i=0,1} E_{Q_i}\big\{(T-\theta_i)^2\big\}.$$

Since $r^2 = (\theta_1-\theta_0)^2/4 = \Delta^2/4$,

$$c_K(\theta_1-\theta_0)^2 = \frac{\exp(-K_+)}{32}\,\Delta^2 \le r^2\,\frac{\exp(-K_+)}{4}.$$

Therefore

$$\max_{i=0,1} E_{Q_i}\big\{(T-\theta_i)^2\big\} \ge \frac{\exp(-K_+)}{32}\,\Delta^2 = c_K\Delta^2.$$

The constant c_K was chosen from K alone, before the laws and the estimator, and the argument applies to every measurable estimator with the two stated integrability properties. This is the claimed two-point bound. □

Lemma 2 is used only at the level stated here. In the lower-bound construction, the alternatives are chosen so that their target values differ by Δ of the desired order, while the product-law divergence is bounded by a fixed constant K. The resulting mean-squared error lower bound is therefore proportional to Δ^2.

We next record the algebraic reductions for the comparison sequence ρ_n from Definition 4. These statements identify exactly which exponent appears in the published benchmark under the two covariate-smoothness regimes.

Lemma 3. [`lem:rho-oracle-regime-algebra`] (Smooth-Covariate Rate Collapse) *Let ρ_n be the published HOIF benchmark of Definition 4, evaluated at (n, α, s, d). Suppose that*

- ***(Positive treatment smoothness.)*** $\alpha > 0$.
- ***(Positive covariate smoothness.)*** $s > 0$.
- ***(Smooth-covariate regime.)*** $d \le 4s$.
- ***(Sample size.)*** $n \ge 1$.

Then

$$\rho_n = n^{-2\alpha/(2\alpha+1)}.$$

Proof of Lemma 3. Throughout, ρ_n is the maximum of the two power terms in Definition 4, so the claim is that the first term dominates.

Case $n = 1$. Every real power of 1 equals 1, so both terms coincide,

$$1^{-2\alpha/(2\alpha+1)} = 1 = 1^{-2/(1+d/(4s)+1/\alpha)},$$

and hence $\rho_1 = \max\{1, 1\} = 1 = 1^{-2\alpha/(2\alpha+1)}$, which is the asserted identity.

Case $n > 1$. Since $s > 0$ we have $4s > 0$, so the smooth-covariate hypothesis $d \le 4s$ can be divided through:

$$\frac{d}{4s} \le 1.$$

Adding $1 + 1/\alpha$ to both sides,

$$1 + \frac{d}{4s} + \frac{1}{\alpha} \le 2 + \frac{1}{\alpha}.$$

Both sides are strictly positive, since $d/(4s) \ge 0$ and $1/\alpha > 0$ by $\alpha > 0$. Dividing the constant $2 > 0$ by the two sides therefore reverses the inequality:

$$\frac{2}{2 + 1/\alpha} \le \frac{2}{1 + d/(4s) + 1/\alpha}.$$

Multiplying numerator and denominator of the left-hand side by $\alpha > 0$ gives the elementary identity

$$\frac{2}{2 + 1/\alpha} = \frac{2\alpha}{2\alpha + 1},$$

so that

$$\frac{2\alpha}{2\alpha + 1} \le \frac{2}{1 + d/(4s) + 1/\alpha}.$$

Since $n > 1$, the map $x \mapsto n^x$ is strictly increasing; negating the two exponents reverses the last inequality and then applying $x \mapsto n^x$ preserves it, so

$$n^{-2/(1+d/(4s)+1/\alpha)} \leq n^{-2\alpha/(2\alpha+1)}.$$

Thus the maximum in Definition 4 is attained by the treatment-regression term, that is,

$$\rho_n = n^{-2\alpha/(2\alpha+1)}.$$

□

Lemma 3 states that, when $d \leq 4s$, the maximum defining ρ_n is attained by the treatment-regression term. This is the algebra behind the smooth-covariate comparison in Proposition 1 and Theorem 2.

Lemma 4. [`lem:rho-deficient-regime-algebra`] (Deficient Regime Algebra) *For any sample size n, treatment smoothness α, covariate smoothness s, and covariate dimension d, suppose that*

- ***(Positive treatment smoothness.)*** $\alpha > 0$;
- ***(Positive covariate smoothness.)*** $s > 0$;
- ***(Deficient-covariate regime.)*** $4s < d$;
- ***(Nonzero sample size.)*** $n \geq 1$.

Then the published HOIF benchmark ρ_n from Definition 4 satisfies

$$\rho_n = n^{-2/(1+d/(4s)+1/\alpha)},$$

and the deficient-regime exponent is strictly smaller than the oracle exponent:

$$\frac{2}{1+d/(4s)+1/\alpha} < \frac{2\alpha}{2\alpha+1}.$$

Proof of Lemma 4. Step 1 (strict comparison of the two exponents). Since $s > 0$ we have $4s > 0$, so the deficient-regime hypothesis $4s < d$ can be divided through:

$$1 < \frac{d}{4s}.$$

Adding $1 + 1/\alpha$ to both sides,

$$2 + \frac{1}{\alpha} < 1 + \frac{d}{4s} + \frac{1}{\alpha}.$$

Both sides are strictly positive, because $1/\alpha > 0$ by $\alpha > 0$. Dividing the constant $2 > 0$ by the two sides therefore reverses the strict inequality:

$$\frac{2}{1+d/(4s)+1/\alpha} < \frac{2}{2+1/\alpha}.$$

Multiplying numerator and denominator by $\alpha > 0$ gives the elementary identity

$$\frac{2}{2+1/\alpha} = \frac{2\alpha}{2\alpha+1},$$

and hence the strict exponent comparison asserted in the statement,

$$\frac{2}{1+d/(4s)+1/\alpha} < \frac{2\alpha}{2\alpha+1}.$$

Step 2 (identification of the maximum in ρ_n). If $n = 1$, every real power of 1 equals 1, so both terms of Definition 4 coincide,

$$1^{-2\alpha/(2\alpha+1)} = 1 = 1^{-2/(1+d/(4s)+1/\alpha)},$$

and $\rho_1 = \max\{1, 1\} = 1 = 1^{-2/(1+d/(4s)+1/\alpha)}$, as claimed. If instead $n > 1$, the map $x \mapsto n^x$ is strictly increasing; negating the two exponents reverses the strict inequality of Step 1, and applying $x \mapsto n^x$ then preserves the reversed inequality, so

$$n^{-2\alpha/(2\alpha+1)} \leq n^{-2/(1+d/(4s)+1/\alpha)}.$$

In either case the maximum in Definition 4 is attained by the covariate-smoothness term, that is,

$$\rho_n = n^{-2/(1+d/(4s)+1/\alpha)}.$$

Together with the strict comparison of Step 1, this is the desired conclusion. □

Lemma 4 gives the complementary algebra for $4s < d$. In that regime the published benchmark is governed by the covariate-smoothness expression, and its exponent is smaller than the exponent appearing in the same-class lower floor. The lemma thereby supplies the deficient-regime exponent comparison.

The lower-bound input for the main results is the oracle floor obtained by perturbing the treatment regression while keeping the density components within the slack baseline introduced in Assumption 12. The construction is the usual local nonparametric testing argument in the treatment coordinate [Stone, 1982, Tsybakov, 2009, Goldenshluger and Lepski, 2020], specialized to the interior partial-mean functional.

Lemma 5. [`lem:oracle-dose-regression-lower-all-beta`] (All-Beta Oracle Floor) *Fix a covariate dimension $d \in \mathbb{N}$ and constants $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0 \in \mathbb{R}$. Suppose that:*

- ***(Positive smoothness.)*** *$\alpha > 0$, $\beta > 0$, and $s > 0$.*
- ***(Regime constants.)*** *$M > 0$, $c_0 > 0$, $t_0 \in (0, 1)$, $\varepsilon_0 \in (0, 1/2)$, and the window of radius ε_0 around t_0 lies in the interior dose region.*
- ***(Slack baseline.)*** *The strict-slack baseline condition of Assumption 12 holds for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$.*

Then there exists a constant $c_{\mathrm{or}} > 0$ such that, for all sufficiently large $n \in \mathbb{N}$,

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \geq c_{\mathrm{or}}\, n^{-2\alpha/(2\alpha+1)}.$$

Here $\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$ is the model class of Definition 2, and R_n is the minimax risk of Definition 3.

Proof of Lemma 5. Step 1 (baseline, outcome scale, and the reference bump). By the strict-slack baseline condition of Assumption 12, fix a covariate density p_0 on $[0,1]^d$, a conditional treatment density q_0 on $[0,1]$, and a slack $\eta_0 > 0$ with the properties listed there. Taking the order-0 bound in the Hölder condition on q_0 and combining it with $q_0 \geq 0$ gives, in particular,

$$0 \leq q_0(a) \leq M - \eta_0 \qquad \text{for every } a \in [t_0 - \varepsilon_0, t_0 + \varepsilon_0],$$

so $M - \eta_0 \geq 0$. Put

$$B = \frac{M}{2} > 0.$$

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a fixed smooth reference bump with

$$0 \leq \psi \leq 1, \qquad \psi(z) = 1 \ \text{ for } |z| \leq \tfrac{1}{2}, \qquad \psi(z) = 0 \ \text{ for } |z| \geq 1;$$

in particular $\psi(0) = 1$, $|\psi| \leq 1$, and ψ has compact support, so all of its derivatives are bounded.

Step 2 (the amplitude gate). There is an amplitude λ with

$$0 < \lambda \leq \frac{M}{4} = \frac{B}{2}$$

such that, for every sign $\zeta \in \{-1, 1\}$ and every bandwidth $h \in (0, 1]$, the scaled signed bump lies in the treatment-direction Hölder ball required by Definition 2:

$$\Big(a \mapsto \zeta\, \lambda\, h^{\alpha}\, \psi\big(\tfrac{a - t_0}{h}\big)\Big) \in \mathcal{H}^{\alpha}\big(M; [t_0 - \varepsilon_0, t_0 + \varepsilon_0]\big).$$

Indeed, write $k = \lceil \alpha \rceil - 1$ for the largest integer strictly below α, let $C_j = \sup_z |\psi^{(j)}(z)| < \infty$ for $j \leq k$, let C_H be an $(\alpha - k)$-Hölder constant of $\psi^{(k)}$ on $\mathbb{R}$, and set

$$\lambda = \min\left\{ \frac{M}{4},\ \frac{M}{1 + \max\{C_H,\ C_0 + \cdots + C_k,\ 1\}} \right\} > 0.$$

For $j \leq k$ the j-th derivative of the scaled bump at a is $\zeta \lambda h^{\alpha - j} \psi^{(j)}((a - t_0)/h)$, and $h^{\alpha - j} \leq 1$ because $0 < h \leq 1$ and $j \leq k \leq \alpha$; hence its modulus is at most $\lambda C_j \leq M$. For the top order, writing $u_x = (x - t_0)/h$ and using $|u_x - u_y| = |x - y|/h$,

$$\Big|\zeta \lambda h^{\alpha - k} \big\{\psi^{(k)}(u_x) - \psi^{(k)}(u_y)\big\}\Big| \leq \lambda h^{\alpha - k} C_H\, \frac{|x - y|^{\alpha - k}}{h^{\alpha - k}} = \lambda C_H\, |x - y|^{\alpha - k} \leq M\, |x - y|^{\alpha - k},$$

which is exactly the top-order Hölder requirement of radius M.

Step 3 (the divergence budget and the constant). Set

$$K = \frac{16 \lambda^2 (M - \eta_0)}{B^2},$$

let $c_K > 0$ be the constant supplied by Lemma 2 for this budget, and define

$$c_{\mathrm{or}} = 4 c_K \lambda^2 > 0.$$

Step 4 (bandwidth). For $n \geq 1$ put

$$h = h_n = n^{-1/(2\alpha + 1)}.$$

Since $\alpha > 0$ we have $2\alpha + 1 > 0$, so $h_n \to 0$; hence for all sufficiently large n both $0 < h \le 1$ and $h \le \varepsilon_0$ hold. Moreover, for every $n \ge 1$,

$$n\, h^{2\alpha+1} = n \cdot n^{-(2\alpha+1)/(2\alpha+1)} = n \cdot n^{-1} = 1.$$

Fix such an n from here on.

Step 5 (the two least favourable laws). For $\zeta \in \{-1, 1\}$ define the perturbed treatment regression

$$\mu_\zeta(a, x) = \zeta\, \lambda\, h^\alpha\, \psi\left(\frac{a - t_0}{h}\right), \qquad a \in [0, 1],\ x \in [0, 1]^d,$$

and let P_ζ be the observed-data law generated as follows: draw X on $[0, 1]^d$ with density p_0; independently draw A on $[0, 1]$ with density q_0; and, given $(A, X) = (a, x)$, draw Y from the two-point channel $Q_{\mu_\zeta(a,x)}$ on $\{-B, B\}$ of Lemma 1. The potential-outcome overlay is

$$Y(a) = \begin{cases} Y, & A = a, \\ \mu_\zeta(a, X), & A \ne a. \end{cases}$$

By construction the law-side nuisances of P_ζ are

$$\mu_{P_\zeta}(a, x) = \mu_\zeta(a, x), \qquad \pi_{P_\zeta}(a \mid x) = q_0(a), \qquad p_{X,P_\zeta} = p_0.$$

Write $P_0 = P_{-1}$ and $P_1 = P_{+1}$. Because $|\zeta| = 1$, $0 \le \psi \le 1$, $0 < h \le 1$ and $\alpha > 0$ (so $h^\alpha \le 1$),

$$|\mu_\zeta(a, x)| = \lambda h^\alpha\, \psi\left(\frac{a - t_0}{h}\right) \le \lambda \le \frac{B}{2},$$

so in particular $|\mu_\zeta| \le B$ and each channel $Q_{\mu_\zeta(a,x)}$ is a genuine probability measure on $\{-B, B\}$.

Step 6 (both laws lie in the model class). We check the conditions listed in Definition 2 for P_ζ, $\zeta \in \{-1, 1\}$.

- *Sampling.* $p_0 \ge 0$ with $\int_{[0,1]^d} p_0 = 1$, $q_0 \ge 0$ with $\int_0^1 q_0 = 1$, and each outcome channel is a probability measure by the last display of Step 5; so P_ζ is a probability law under which $A \in [0, 1]$ and $X \in [0, 1]^d$ almost surely, and the sample is n independent draws from it.
- *Bounded outcome.* $Y \in \{-B, B\}$, so $|Y| = B = M/2 \le M$ almost surely.
- *Interior dose.* $[t_0 - \varepsilon_0, t_0 + \varepsilon_0] \subset (0, 1)$ is a hypothesis of the present lemma.
- *Local positivity.* $\pi_{P_\zeta}(a \mid x) = q_0(a) \ge c_0 + \eta_0 \ge c_0$ for every a in the window and every $x \in [0, 1]^d$, by Assumption 12 and $\eta_0 > 0$.
- *Treatment-direction smoothness of the regression.* This is exactly the conclusion of Step 2, since $\mu_{P_\zeta}(\cdot, x)$ does not depend on x.
- *Treatment-direction smoothness of the density.* q_0 has Hölder norm at most $M - \eta_0$ of order β on the window, and $M - \eta_0 \le M$, so $q_0 \in \mathcal{H}^\beta(M; [t_0 - \varepsilon_0, t_0 + \varepsilon_0])$.
- *Covariate-direction smoothness of the regression.* Since $\psi(0) = 1$,

  $$\mu_{P_\zeta}(t_0, x) = \zeta \lambda h^\alpha \qquad \text{for every } x \in [0, 1]^d,$$

  a constant function of x of modulus at most $B \le M$ by Step 5; a constant of modulus at most M lies in $\mathcal{H}^s(M; [0, 1]^d)$, all its derivatives vanishing.

- *Covariate-direction smoothness of the density.* Likewise $\pi_{P_\zeta}(t_0 \mid x) = q_0(t_0)$ is constant in x, of modulus at most $M - \eta_0 \le M$, hence in $\mathcal{H}^s(M; [0,1]^d)$.
- *Covariate density.* $p_0 \in \mathcal{H}^s(M - \eta_0; [0,1]^d) \subseteq \mathcal{H}^s(M; [0,1]^d)$, and $0 \le p_0 \le M - \eta_0 \le M$ on the cube.
- *Semantic ties.* The channel Q_u has mean u, so $E_{P_\zeta}[Y \mid A = a, X = x] = \mu_\zeta(a,x) = \mu_{P_\zeta}(a,x)$; the X-marginal of P_ζ is p_0 times Lebesgue measure on $[0,1]^d$; and the (A,X)-law has Lebesgue density $(a,x) \mapsto q_0(a)p_0(x) = \pi_{P_\zeta}(a \mid x)p_{X,P_\zeta}(x)$ on $[0,1] \times [0,1]^d$, which is exactly the required factorization.
- *Causal overlay.* Consistency $Y = Y(A)$ holds by the definition of $Y(\cdot)$ in Step 5. For ignorability, fix a; because A has a Lebesgue density the event $\{A = a\}$ is null, so $Y(a) = \mu_\zeta(a, X)$ almost surely, a deterministic function of X alone; hence $Y(a)$ is conditionally independent of A given X.

Therefore $P_0, P_1 \in \mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$.

Step 7 (one-observation divergence). The two laws share the same (A,X)-law, with density $q_0(a)p_0(x)$, and differ only in the conditional outcome channel; since the observed pair (A,X) is a function of the observation and each fibre channel is dominated by its counterpart, the chain rule for divergence gives

$$\mathrm{KL}(P_0, P_1) = \int_{[0,1]^d} \int_0^1 \mathrm{KL}\big(Q_{\mu_{-1}(a,x)}, Q_{\mu_1(a,x)}\big)\, q_0(a)\, p_0(x)\, \mathrm{d}a\, \mathrm{d}x.$$

Fibrewise, $|\mu_{-1}(a,x)| \le B/2$ and $|\mu_1(a,x)| \le B/2$ by Step 5, so Lemma 1 applies with $u = \mu_{-1}(a,x)$ and $v = \mu_1(a,x)$; as

$$\mu_{-1}(a,x) - \mu_1(a,x) = -2\lambda h^\alpha \psi\left(\frac{a - t_0}{h}\right),$$

it yields

$$\mathrm{KL}\big(Q_{\mu_{-1}(a,x)}, Q_{\mu_1(a,x)}\big) \le \frac{2\big(\mu_{-1}(a,x) - \mu_1(a,x)\big)^2}{B^2} = \frac{8\lambda^2 h^{2\alpha}}{B^2}\, \psi\left(\frac{a - t_0}{h}\right)^2.$$

The bump concentrates the remaining integral: $\psi((a - t_0)/h) = 0$ whenever $|a - t_0| \ge h$, so the integrand vanishes outside $(t_0 - h, t_0 + h)$; on that interval, which is contained in the window because $h \le \varepsilon_0$, we have $\psi^2 \le 1$ and $q_0 \le M - \eta_0$ by Step 1; and the interval has length $2h$. Hence

$$\int_0^1 \psi\left(\frac{a - t_0}{h}\right)^2 q_0(a)\, \mathrm{d}a \le 2(M - \eta_0)\, h.$$

Using $\int_{[0,1]^d} p_0 = 1$, the two previous displays combine to the one-observation estimate

$$\mathrm{KL}(P_0, P_1) \le \frac{8\lambda^2 h^{2\alpha}}{B^2} \cdot 2(M - \eta_0)h = \frac{16\lambda^2(M - \eta_0)}{B^2}\, h^{2\alpha+1}.$$

Step 8 (tensorization and the budget). Each fibre channel puts mass at least $1/4$ on each of the two atoms $\pm B$, because $|\mu_\zeta| \le B/2$; consequently P_0 and P_1 are mutually absolutely

continuous with integrable log-likelihood ratio, and the divergence between the n-fold product laws tensorizes:

$$\mathrm{KL}\big(P_0^{\otimes n}, P_1^{\otimes n}\big) \le n\,\mathrm{KL}(P_0, P_1) \le n\,\frac{16\lambda^2(M-\eta_0)}{B^2}\,h^{2\alpha+1} = \frac{16\lambda^2(M-\eta_0)}{B^2} = K,$$

the last equality by the calibration $n\,h^{2\alpha+1} = 1$ of Step 4.

Step 9 (functional separation). Since $\psi(0) = 1$ and $\int_{[0,1]^d} p_0 = 1$,

$$\theta_{P_\zeta}(t_0) = \int_{[0,1]^d} \mu_\zeta(t_0, x)\, p_0(x)\,\mathrm{d}x = \zeta\lambda h^\alpha \psi(0) \int_{[0,1]^d} p_0(x)\,\mathrm{d}x = \zeta\lambda h^\alpha,$$

so the two target values are separated by

$$\theta_{P_1}(t_0) - \theta_{P_0}(t_0) = 2\lambda h^\alpha.$$

Step 10 (two-point reduction inside the minimax risk). Let $\hat\theta_n$ be any estimator admissible in Definition 3, that is, measurable with values in $[-M, M]$. For every P in the model class, $|\mu_P(t_0, \cdot)| \le M$ and $0 \le p_{X,P} \le M$, and $[0,1]^d$ has unit Lebesgue measure, so

$$|\theta_P(t_0)| = \left|\int_{[0,1]^d} \mu_P(t_0, x) p_{X,P}(x)\,\mathrm{d}x\right| \le M^2, \qquad \text{hence} \qquad \big|\hat\theta_n - \theta_P(t_0)\big| \le M + M^2.$$

Thus under every n-fold product law the squared loss is bounded by $(M+M^2)^2$, so it is integrable and the worst-case risk over the class is a finite supremum. Apply Lemma 2 with the n-sample laws $Q_0 = P_0^{\otimes n}$, $Q_1 = P_1^{\otimes n}$, parameter values $\theta_i = \theta_{P_i}(t_0)$, budget K (available by Step 8) and estimator $T = \hat\theta_n$:

$$c_K\big(\theta_{P_1}(t_0) - \theta_{P_0}(t_0)\big)^2 \le \max_{i=0,1} \int \big(\hat\theta_n - \theta_{P_i}(t_0)\big)^2\,\mathrm{d}P_i^{\otimes n}.$$

Both P_0 and P_1 belong to the class by Step 6, so the right-hand side is at most the supremum over $P \in \mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0)$ of the risk of $\hat\theta_n$. As $\hat\theta_n$ was arbitrary, taking the infimum over admissible estimators preserves the bound, and with the separation of Step 9,

$$c_K\big(2\lambda h^\alpha\big)^2 \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

Step 11 (conclusion). Finally, by the choice of h in Step 4,

$$c_K\big(2\lambda h^\alpha\big)^2 = 4c_K\lambda^2 h^{2\alpha} = c_{\mathrm{or}}\left(n^{-1/(2\alpha+1)}\right)^{2\alpha} = c_{\mathrm{or}}\,n^{-2\alpha/(2\alpha+1)}.$$

Combining the last two displays,

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \ge c_{\mathrm{or}}\,n^{-2\alpha/(2\alpha+1)}$$

for all sufficiently large n, with $c_{\mathrm{or}} = 4c_K\lambda^2 > 0$ depending only on the fixed model constants and the slack baseline, not on n. □

Lemma 5 is the same-class lower-bound engine. Its conclusion is uniform in the sense relevant for the paper's exponent comparison: for each fixed admissible $\beta > 0$, the lower exponent is $2\alpha/(2\alpha+1)$. The constant c_{or} may depend on the fixed model parameters and on the slack available in Assumption 12; the stated uniformity concerns each fixed β.

B verification note

This appendix records the verification scope for the formal statements used in the paper. The observed-data experiment, causal overlay, Hölder dose class, target functional, minimax risk, published benchmark sequence, auxiliary testing statements, algebraic rate comparisons, and lower-bound consequences are represented by the numbered assumptions, definitions, lemmas, propositions, theorems, and remarks displayed in the paper. The corresponding theorem statements and their derivations are machine-checked in Lean 4.

The machine-checked part covers the consequences of the stated assumptions and auxiliary lemmas as they are formulated here. In particular, Assumption 1–Assumption 12 define the statistical and causal restrictions used by the displayed results; Definition 1, Definition 2, Definition 3, and Definition 4 fix the target, model class, risk, and comparison sequence; and the auxiliary appendix lemmas supply the testing and algebraic inputs used to derive Theorem 1, Theorem 2, and Theorem 3 together with Proposition 1. The verification establishes the same-class lower-bound and rate-comparison statements, with the upper frontier identified as a separate research question.

Several ingredients enter as mathematical assumptions or cited comparators. The causal identification primitives in Assumption 10 and Assumption 11 are imposed conditions on the observed-data law with its potential-outcome interpretation. The formal argument uses the classical Le Cam minimax reduction at the level of the auxiliary testing lemma stated in the appendix and verifies its consequences for the present model and rates. The Bonvini–Kennedy result in Cited result 1, imported from Bonvini and Kennedy [2022], motivates the benchmark sequence ρ_n on its localized-regularity class; Definition 2 specifies the distinct same-class lower-bound model studied here.

The Lean-oriented labels are therefore confined to cross-references and this reproducibility note. In the main text they serve only to identify the precise assumptions and statements being invoked. The econometric content of the paper is the lower bound over Definition 2, the algebraic comparison with Definition 4, and the explicit separation between those same-class lower results and the external higher-order influence-function comparator.

C Proofs of the main results

Proof of Theorem 1. 1. *Instantiate the oracle floor.* The theorem hypotheses provide the inputs required by Lemma 5 for the same covariate dimension d and the same constants $\alpha, \beta, s, M, c_0, \varepsilon_0, t_0$. The positive-smoothness assumptions give $\alpha > 0$, $\beta > 0$, and $s > 0$. The regime-constants assumptions give $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, and the interior-window condition recorded in Assumption 3. The strict-slack baseline assumption is Assumption 12 for the tuple $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$, matching the slack-baseline input of the lemma.

2. *Apply the oracle floor.* Applying Lemma 5 with these inputs yields a constant $c_{\mathrm{or}} > 0$ such that, for all sufficiently large $n \in \mathbb{N}$,

$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \geq c_{\mathrm{or}}\, n^{-2\alpha/(2\alpha+1)}.$$

Here the class is the model class of Definition 2, and R_n is the minimax risk of Definition 3. The displayed tail bound is the conclusion of the cited oracle floor for the fixed treatment-density smoothness parameter $\beta > 0$.

3. *Conclusion.* Set $c = c_{\mathrm{or}}$. Then $c > 0$, and the inequality from Step 2 is equivalently

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)$$

for all sufficiently large sample sizes n. This is the asserted positive constant and eventual lower bound over the model class of Definition 2 for the minimax risk of Definition 3.

□

Proof of Proposition 1. 1. The positive-smoothness, regime-constant, and baseline-slack hypotheses assumed here – $\alpha > 0$, $\beta > 0$, $s > 0$, $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, $[t_0 - \varepsilon_0, t_0 + \varepsilon_0] \subset (0,1)$, and Assumption 12 for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$ – are the hypotheses required by Theorem 1. Applying that result gives a constant $c > 0$ such that, for all sufficiently large sample sizes n,

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

Fix this c for the rest of the proof.

2. Intersect the eventual lower-bound event supplied by Theorem 1 with the eventual sample-size event $n \ge 1$. Equivalently, after increasing the sufficiently-large threshold if necessary, every sample size under consideration satisfies both $n \ge 1$ and the lower-bound inequality supplied by Theorem 1.

3. For any such n, the hypotheses $\alpha > 0$, $s > 0$, $d \le 4s$, and $n \ge 1$ are the hypotheses of Lemma 3. Hence the published benchmark rate satisfies

$$\rho_n = n^{-2\alpha/(2\alpha+1)}.$$

4. For the same n, the lower-bound inequality supplied by Theorem 1 gives

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

The identity supplied by Lemma 3 rewrites $c\rho_n$ as $c\, n^{-2\alpha/(2\alpha+1)}$, and therefore

$$c\, \rho_n \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,.$$

This holds on the same eventual set of sample sizes. Together with $c > 0$ and the displayed identity for ρ_n, it is the assertion of Proposition 1.

□

Proof of Theorem 2. 1. The smoothness, regime-constant and strict-slack hypotheses assumed here — $\alpha > 0$, $\beta > 0$, $s > 0$, $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, $[t_0 - \varepsilon_0, t_0 + \varepsilon_0] \subset (0,1)$, and Assumption 12 for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$ — are exactly the hypotheses of Theorem 1; the smooth-covariate condition $d \le 4s$ is not needed for this step and is used only in Step 3. That theorem therefore supplies a constant $c > 0$ and a threshold n_1 with

$$c\, n^{-2\alpha/(2\alpha+1)} \le R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \qquad \text{for every } n \ge n_1.$$

Fix this c; it is the constant asserted in the statement.

2. Set $n_0 = \max\{n_1, 1\}$. For every $n \geq n_0$ both the displayed lower bound of Step 1 and the elementary condition $n \geq 1$ hold; this is the intersection of two eventual conditions, and the conjunction asserted by the theorem is claimed only for such n.

3. Fix $n \geq n_0$. The hypotheses $\alpha > 0$, $s > 0$, $d \leq 4s$ and $n \geq 1$ are exactly those of Lemma 3, so
$$\rho_n = n^{-2\alpha/(2\alpha+1)}, \qquad \text{equivalently} \qquad n^{-2\alpha/(2\alpha+1)} = \rho_n,$$
which is the second assertion of the theorem, in the orientation in which it is stated. Holding this identity together with the lower bound of Step 1 on the same range $n \geq n_0$ gives, for the same constant $c > 0$, the asserted eventual conjunction of the lower bound and the exponent identity.

□

Proof of Theorem 3. 1. The regime-constant and strict-slack hypotheses assumed here — $\alpha > 0$, $\beta > 0$, $s > 0$, $M > 0$, $c_0 > 0$, $t_0 \in (0,1)$, $\varepsilon_0 \in (0, 1/2)$, $[t_0 - \varepsilon_0, t_0 + \varepsilon_0] \subset (0,1)$, and Assumption 12 for $(d, \beta, s, M, c_0, \varepsilon_0, t_0)$ — are exactly the hypotheses of Lemma 5; the deficient-covariate condition $4s < d$ is not used for this step and enters only in Step 3. That lemma therefore supplies a constant $c > 0$ and a threshold n_1 with
$$R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0) \geq c\, n^{-2\alpha/(2\alpha+1)} \qquad \text{for every } n \geq n_1.$$
Fix this c; it is the constant asserted in the statement. Note that the lower floor is established for the given fixed $\beta > 0$ and its exponent does not involve β.

2. Set $n_0 = \max\{n_1, 1\}$. For every $n \geq n_0$ both the display of Step 1 and the elementary condition $n \geq 1$ hold; this is the intersection of two eventual conditions, and the conjunction asserted by the theorem is claimed only for such n.

3. Fix $n \geq n_0$. The hypotheses $\alpha > 0$, $s > 0$, $4s < d$ and $n \geq 1$ are exactly those of Lemma 4, which gives both the deficient-regime formula
$$\rho_n = n^{-\,2/(1+d/(4s)+1/\alpha)}$$
and the strict comparison of exponents
$$\frac{2}{1 + d/(4s) + 1/\alpha} < \frac{2\alpha}{2\alpha + 1}.$$

4. On the common range $n \geq n_0$, the lower bound of Step 1, rewritten in the orientation of the statement as
$$c\, n^{-2\alpha/(2\alpha+1)} \leq R_n(\mathcal{P}_{\alpha,\beta,s}(M, c_0, \varepsilon_0, t_0), t_0)\,,$$
holds together with the two conclusions of Step 3. That conjunction, with the constant $c > 0$ fixed in Step 1, is exactly the assertion of the theorem: the certified lower floor sits at the treatment-regression exponent $2\alpha/(2\alpha + 1)$, while the published benchmark sequence is the covariate-smoothness power, whose exponent is strictly smaller.

□

References

Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York, 1993.

Matteo Bonvini and Edward H. Kennedy. Fast convergence rates for dose-response estimation, 2022. Version 2 revised 12 April 2026.

Matteo Bonvini, Edward H. Kennedy, Oliver Dukes, and Sivaraman Balakrishnan. Doubly-robust inference and optimality in structure-agnostic models with smoothness, 2024.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097.

Kyle Colangelo and Ying-Ying Lee. Double debiased machine learning nonparametric inference with continuous treatments, 2020.

Charles R. Doss, Guangwei Weng, Lan Wang, Ira Moscovice, and Tongtan Chantarat. A nonparametric doubly robust test for a continuous treatment effect, 2022.

Alexander Goldenshluger and Oleg Lepski. Minimax estimation of norms of a probability density: I. lower bounds, 2020.

Nima S. Hejazi, David Benkeser, Iván Díaz, and Mark J. van der Laan. Efficient estimation of modified treatment policy effects based on the generalized propensity score, 2022.

Keisuke Hirano and Guido W. Imbens. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li Meng, editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pages 73–84. Wiley, Chichester, 2004.

Aaron Hudson, Elvin H. Geng, Thomas A. Odeny, Elizabeth A. Bukusi, Maya L. Petersen, and Mark J. van der Laan. An approach to nonparametric inference on the causal dose response function, 2023.

Kosuke Imai and David A van Dyk. Causal inference with general treatment regimes. *Journal of the American Statistical Association*, 99(467):854–866, 2004. doi: 10.1198/016214504000001187.

Guido W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000. doi: 10.1093/biomet/87.3.706.

Edward H. Kennedy, Zongming Ma, Matthew D. McHugh, and Dylan S. Small. Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1229–1245, 2017. doi: 10.1111/rssb.12212.

Ying-Ying Lee. Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models, 2018.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9–12):1393–1512, 1986. doi: 10.1016/0270-0255(86)90088-6.

James Robins, Lingling Li, Eric Tchetgen Tchetgen, and Aad van der Vaart. Higher order influence functions and minimax estimation of nonlinear functionals. *IMS Collections*, 2: 335–421, 2008. doi: 10.1214/193940307000000527.

James Robins, Lingling Li, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad van der Vaart. Higher order estimating equations for high-dimensional models, 2015.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Kyle Schindl, Shuying Shen, and Edward H. Kennedy. Incremental effects for continuous exposures, 2024. Version 3 revised 27 January 2026.

Junming Shi, Wenxin Zhang, Alan E. Hubbard, and Mark J. van der Laan. Hal-based plug-in estimation with pointwise asymptotic normality of the causal dose-response curve, 2024. Version 4 revised 27 August 2025.

Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982. doi: 10.1214/aos/1176345969.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. doi: 10.1007/b13794.

Mark J. van der Laan and James M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York, 2003.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. doi: 10.1017/CBO9780511802256.

Yikun Zhang, Yen-Chi Chen, and Alexander Giessing. Nonparametric inference on dose-response curves without the positivity condition, 2024.

Shandong Zhao, David A. van Dyk, and Kosuke Imai. Causal inference in observational studies with non-binary treatments, 2013.