A Lower-Bound Calibration for Joint Margin–Overlap Decay in Offline Policy Learning

CausalSmith*

August 25, 2026

Abstract

This paper studies offline policy learning for deterministic treatment rules when overlap can deteriorate near either propensity boundary in the same region where the treatment contrast is small. The target is observed-law welfare regret, defined directly from the conditional treatment contrast. The law class imposes bounded outcomes, positivity, a margin condition for small treatment contrasts, a zero-effect convention, and a joint overlap-decay restriction that ties weak treatment-arm information to the small-contrast region. Under the margin-window normalization and the auxiliary calibrations used by the two-point construction, the paper proves an observed-law minimax lower bound

$$M_n(\alpha,\gamma) \geq c\, n^{-r_\star(\alpha,\gamma)}, \qquad r_\star(\alpha,\gamma) = \frac{1+\alpha}{2+\alpha+\beta_{\alpha,\gamma}}, \qquad \beta_{\alpha,\gamma} = \begin{cases} 0, & \gamma = 0, \\ \alpha\gamma/(\alpha+1), & \gamma > 0. \end{cases}$$

The exponent is obtained by balancing margin mass, local contrast size, and the probability of observing the informative treatment arm. The paper also gives a conditional analysis of a specified clipped cross-fitted AIPW empirical welfare rule with supplied nuisance estimates satisfying explicit rate, boundedness, cross-fitting, and localized empirical-process conditions. In the nonbinding nuisance-and-clipping regime, this conditional upper exponent matches the lower-bound exponent up to logarithmic factors; in the binding regime, the analysis gives the rule's nuisance-limited exponent.

1 Introduction

Policy learning in observational and logged data asks how well a data-dependent rule can approximate the treatment assignment that maximizes welfare. The classical treatment-choice literature formulates this as a statistical decision problem under sampling uncertainty [Manski, 2004, 2009, Stoye, 2009], and empirical welfare maximization makes the connection to classification over structured policy classes explicit [Kitagawa and Tetenov, 2018]. Modern causal policy-learning procedures often use inverse-propensity and doubly robust scores, combined with sample splitting and flexible first-stage estimation, to estimate welfare criteria under overlap and nuisance-regularity conditions [Athey and Wager, 2021, Chernozhukov et al., 2018, 2022].

*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.

This paper focuses on how the regret lower-bound calibration changes when overlap deteriorates in the same region where the treatment contrast is small. The experiment is an offline observed-law policy-learning problem. One observation is $O = (X, A, Y)$, with propensity score $e_P(x) = P(A = 1 \mid X = x)$, treatment-arm regressions $\mu_a(x) = E_P[Y \mid A = a, X = x]$, and contrast $\tau_P(x) = \mu_1(x) - \mu_0(x)$. A deterministic policy π is evaluated by the contrast-weighted welfare and regret in Definition 1. At the observed-law level this defines the welfare problem directly; the usual potential-outcome identification conditions additionally supply its causal interpretation.

The distinctive feature of the law class is a joint overlap-decay condition. Let $p_P(x) = \min\{e_P(x), 1 - e_P(x)\}$. The class $\mathcal{P}_{\alpha,\gamma}$ in Definition 4 combines bounded outcomes, positivity, a margin condition for small nonzero contrasts, a zero-effect convention, the overlap-decay restriction, and the strict-overlap endpoint when $\gamma = 0$. The condition permits weak overlap broadly while controlling the probability that weak overlap and small welfare-relevant contrasts occur together. This is the region that matters for regret, because Theorem 1 writes regret as a weighted classification error with weight $|\tau_P(X)|$.

The main lower-bound result is an observed-law minimax lower bound over $\mathcal{P}_{\alpha,\gamma}$. Under the margin-window normalization and the witness-law calibration used in the construction, Theorem 3 shows that

$$M_n(\alpha, \gamma) \geq c\, n^{-r_\star(\alpha,\gamma)}, \qquad r_\star(\alpha, \gamma) = \frac{1 + \alpha}{2 + \alpha + \beta_{\alpha,\gamma}}.$$

Here $\beta_{\alpha,\gamma} = 0$ under strict overlap and $\beta_{\alpha,\gamma} = \alpha\gamma/(\alpha + 1)$ when $\gamma > 0$. The denominator has three components: 2 from testing local Bernoulli means separated by the contrast scale, α from the mass of the margin block, and $\beta_{\alpha,\gamma}$ from the probability of observing the informative treatment arm under overlap decay.

The lower-bound proof uses a two-point Le Cam construction. The two laws agree outside a small active block and differ only in the sign of the local treatment contrast on that block. The active block has mass of order h^α, the contrast has size h, and the informative arm is sampled with probability of order $h^{\beta_{\alpha,\gamma}}$. The overlap-envelope calculation in Proposition 1 identifies $\beta_{\alpha,\gamma}$ as the largest admissible weak-arm exponent within the tight-window algebraic calibration used in the construction. The chi-square divergence is bounded above at scale $h^{2+\alpha+\beta_{\alpha,\gamma}}$, while the regret separation is bounded below at scale $h^{1+\alpha}$. Taking $h = n^{-1/(2+\alpha+\beta_{\alpha,\gamma})}$ gives the displayed exponent.

The paper separates the converse calculation from feasible achievability. The lower bound ranges over all measurable policy estimators taking values in the policy class and characterizes the observed-law converse. A complementary analysis studies a computable clipped AIPW rule under explicit side conditions, because weak overlap affects both testing difficulty and the stability of inverse-propensity and doubly robust scores.

The specified procedure is a clipped cross-fitted AIPW empirical welfare maximizer with supplied nuisance estimates. The score is defined in Definition 9, the measurable empirical rule in Definition 10, and the conditional risk domain in Definition 11. Theorem 4 gives a uniform upper bound for this particular rule over the stated side-condition domain: laws in $\mathcal{P}_{\alpha,\gamma}$ with the oracle policy in class, supplied cross-fitted nuisances satisfying the displayed rate and boundedness conditions, fixed balanced folds, and primitive localized empirical-process and offset-envelope conditions. These inputs define the theorem's conditional domain and isolate uniform nuisance construction as a separate research question.

For $\gamma > 0$, the feasible exponent is obtained by balancing four terms: the clipped empirical-process envelope, the clipped product nuisance remainder, the localized weak-overlap drift, and the outcome-regression error outside the localization window. In the notation of Definition 7, this gives

$$r_{\text{up}} = \min\{r_\star(\alpha, \gamma), g_{\text{joint}}(\alpha, \gamma, a, c)\}.$$

When $g_{\text{joint}} \geq r_\star(\alpha, \gamma)$, the conditional upper bound matches the observed-law lower-bound exponent up to logarithmic factors. When $g_{\text{joint}} < r_\star(\alpha, \gamma)$, the clipped-AIPW analysis yields the rule's slower nuisance-limited exponent.

Definition 12 delineates the strict-gap branch and records the two relevant routes for a sharp feasible minimax characterization: a stronger feasible construction under comparable side conditions or a lower bound that incorporates nuisance learning into the experiment.

The analysis is related to four literatures. First, it uses the treatment-choice and empirical-welfare framework developed by Manski [2004, 2009], Stoye [2009] and Kitagawa and Tetenov [2018]. Second, the feasible rule follows the semiparametric and doubly robust tradition of Robins et al. [1994], Hahn [1998], Hirano et al. [2003], Newey [1994], Bickel et al. [1993], van der Laan and Rose [2011] and the cross-fitting approach of Chernozhukov et al. [2018, 2022]. Third, the exponent calculation parallels margin and low-noise excess-risk theory [Audibert and Tsybakov, 2007, Massart and Nédélec, 2006, Tsybakov, 2009] and the policy-regret margin analysis of Luedtke and Chambaz [2017]. Fourth, the motivation is close to work on limited and weak overlap, trimming, weighting instability, and clipping-based procedures [Li et al., 2016, D'Amour et al., 2017, Ben-Michael and Keele, 2022, Hill and Chaudhuri, 2024, Susmann et al., 2025, Zhao et al., 2023, Liu et al., 2026]. The point of departure is the coupling between weak overlap and the small-contrast region that drives welfare regret.

The verification appendix records the formal scope of the paper. The rest of the paper proceeds as follows. The next section defines the observed-law policy problem, the regret target, and the law class. The following section gives the minimax lower bound. A later section analyzes the clipped cross-fitted AIPW rule under the stated side conditions. The final section discusses the strict-gap branch and the feasible-tightness question.

2 Related Literature

This paper is closest to four strands of work: econometric treatment choice, semiparametric and doubly robust estimation, margin-based excess-risk theory, and empirical learning under weak or limited overlap. The observed-law policy-learning experiment studied here keeps the target at the level of welfare regret for deterministic treatment rules, while allowing the information available for learning the sign of the conditional treatment contrast to deteriorate in the same region where the contrast is small.

The treatment-choice literature frames policy learning as a welfare maximization problem under sampling uncertainty. Manski [2004, 2009] and Stoye [2009] formulate finite-sample and minimax perspectives on treatment choice, while Kitagawa and Tetenov [2018] develop empirical welfare maximization over policy classes and connect regret to classification-type complexity. The modern policy-learning literature extends these ideas using doubly robust and orthogonal scores, sample splitting, and high-dimensional nuisance estimation [Athey and Wager, 2021, Chernozhukov et al., 2018, 2022]. The present paper places joint margin-overlap decay directly in the lower-bound calibration and quantifies its interaction with margin complexity.

The semiparametric literature provides the estimation architecture behind the feasible rule considered later. Efficient influence-function calculations and doubly robust estimating equations originate in the general semiparametric theory of Bickel et al. [1993] and Newey [1994], and in the causal-inference developments of Robins et al. [1994], Hahn [1998], Hirano et al. [2003], and van der Laan and Rose [2011]. Cross-fitting and Neyman-orthogonal score construction make these ideas compatible with flexible first-stage estimation [Chernozhukov et al., 2018, 2022]. Our conditional upper bound applies this orthogonal-score logic to a clipped AIPW empirical welfare rule under explicit nuisance-rate and empirical-process side conditions.

A second point of contact is individualized treatment-rule estimation. Outcome-weighted learning and related classification reductions appear in Qian and Murphy [2011], Zhao et al. [2012], and Zhang et al. [2012]; off-policy and contextual-bandit work studies related inverse-propensity and doubly robust objectives [Dudik et al., 2011, Swaminathan and Joachims, 2015, Kallus, 2017]. More recent contributions include R-learning and related robust policy learning under flexible nuisance estimation [Nie and Wager, 2017, Athey and Wager, 2021], instrumental-variable welfare analysis via marginal treatment effects [Sasaki and Ura, 2020], constrained empirical welfare maximization [Sun, 2021], and time-series policy choice by empirical welfare maximization [Kitagawa et al., 2022]. These papers emphasize feasible algorithms and statistical guarantees under their respective overlap and nuisance assumptions. The present contribution instead separates an observed-law minimax lower bound from a conditional clipped-AIPW achievability statement, so the feasible analysis should be read as one procedure-specific upper bound under stated side conditions.

The exponent calculations are also related to margin-based excess-risk theory. The role played here by small conditional treatment contrasts is analogous to the low-noise or margin conditions in binary classification [Audibert and Tsybakov, 2007, Massart and Nédélec, 2006, Tsybakov, 2009]. Under strict overlap, margin structure yields the familiar improvement from a square-root empirical-process rate toward faster regret rates. With joint overlap decay, however, observations in the small-contrast region can have less informative treatment-arm sampling. This changes the testing difficulty in the local alternatives used for the lower bound and makes the exponent depend jointly on the margin and overlap-decay parameters.

Limited-overlap and weak-overlap problems have been studied extensively in causal inference and off-policy evaluation. Diagnostics, trimming, weighting instability, and design sensitivity are discussed by Li et al. [2016], D'Amour et al. [2017], Ben-Michael and Keele [2022], Hill and Chaudhuri [2024], and Susmann et al. [2025]. Recent work also develops weak-overlap guarantees and clipping-based procedures for treatment-effect or policy-value estimation [Zhao et al., 2023, Liu et al., 2026]; related adaptive-data and online policy-learning guarantees under dependent sampling, diminishing exploration, or self-normalized inequalities include Zhan et al. [2021] and Girard et al. [2025]. Related doubly robust off-policy and dynamic-regime guarantees appear in Chen et al. [2020], Sakaguchi [2024]. This paper couples overlap decay to the margin region for welfare regret, concentrating the difficulty where small treatment contrasts and weak treatment-arm information coincide.

Table 1 summarizes the closest comparisons at the level of target, overlap condition, and conclusion. The table is intentionally limited to high-level rate structure; constants, logarithmic factors, and side-condition domains differ across papers.

Finally, the lower-bound argument follows the classical testing tradition of Le Cam [1986] and its use in nonparametric and classification lower bounds [Tsybakov, 2009]. The least favorable alternatives are calibrated so that regret separation and product divergence balance after

Table 1: Closest comparisons in the related literature

Reference	Target	Overlap treatment	Relation to this paper
Kitagawa and Tetenov [2018]	Welfare regret for empirical welfare maximization	Standard overlap conditions for observational welfare learning	Establishes the empirical-welfare framework; the present paper changes the local testing calibration under joint margin-overlap decay.
Luedtke and Chambaz [2017]	Optimal treatment-rule regret under margin structure	Strict-overlap setting	Provides the strict-overlap benchmark in which the rate is driven by the margin exponent; here the lower-bound exponent also includes an overlap-decay component.
Chernozhukov et al. [2018, 2022]	Orthogonal and cross-fitted semiparametric estimation	Overlap enters through score regularity and nuisance control	Supplies the orthogonal-score methodology used by the feasible clipped-AIPW rule; the present paper adds the joint overlap-decay minimax lower bound.
Liu et al. [2026]	Clipping and weak-overlap upper bounds for policy learning or related causal targets	Allows weak-overlap behavior controlled through clipping and nuisance conditions	Closest to the feasible upper-bound side: this paper adds a separate observed-law lower-bound exponent and keeps the clipped-AIPW achievability statement conditional.

accounting for the mass of the margin region and the information loss from joint overlap decay. The paper thereby gives a lower-bound exponent for the law class and a separate conditional upper exponent for one clipped cross-fitted AIPW rule, including its nuisance-limited regime under the stated side conditions.

3 Setup and Assumptions

We work with an observed law P for one offline observation $O = (X, A, Y)$ on

$$\Omega = \mathcal{X} \times \{0, 1\} \times [-1, 1].$$

Here X is a pretreatment covariate, $A \in \{0, 1\}$ is the treatment indicator, and Y is the observed outcome. The sample size is n, and $O_1, \ldots, O_n$ denote the offline observations. We write P_X for the marginal law of X. The observed-law propensity score is $e_P(x) = P(A = 1 \mid X = x)$, and the treatment-arm outcome regressions are $\mu_a(x) = E_P[Y \mid A = a, X = x]$, $a \in \{0, 1\}$. The treatment contrast is $\tau_P(x) = \mu_1(x) - \mu_0(x)$. The overlap score used throughout the joint overlap condition is $p_P(x)= \min\{e_P(x), 1-e_P(x)\}$, the distance from the propensity score to the nearest endpoint. Potential-outcome notation $Y(a)$ is used for the boundedness assumption below; the regret and minimax statements themselves are stated in terms of the observed law P, e_P, μ_a, and τ_P.

Assumption 1. [`ass:iid`] (Independent sampling) *The observed data units*

$$O_1, \dots, O_n$$

are independent and identically distributed draws from P.

This is the standard i.i.d. sampling condition for offline policy learning, fixing the experiment in which welfare regret is evaluated [Athey and Wager, 2021].

Assumption 2. [`ass:bounded-outcome`] (Bounded potential outcomes) *Under the observed law P, the observed outcome is bounded almost surely,*

$$Y \in [-1, 1] \quad P\text{-a.s.},$$

and, for every covariate value x,

$$\mu_0(x) \in [-1, 1] \quad \text{and} \quad \mu_1(x) \in [-1, 1].$$

This standard bounded-outcome condition controls the scale of welfare, regret, and inverse-propensity scores; it also implies bounded observed-law conditional means in the treatment arms [Athey and Wager, 2021].

Assumption 3. [`ass:positivity`] (Positivity) *The propensity score satisfies*

$$0 < e_P(X) < 1$$

P_X-almost surely.

This is the standard positivity condition, requiring both treatment arms to occur with positive probability at almost every covariate value relevant under P_X [Imbens and Rubin, 2015].

A deterministic policy is a measurable map $\pi : \mathcal{X} \to \{0, 1\}$. The welfare criterion uses the contrast-weighted value of a policy, so policies are compared only through how they assign treatment on covariate regions with positive or negative τ_P.

Definition 1. [`def:welfare-regret`] (Welfare $V_P(\pi)$ and Regret $R_P(\pi)$) *For a policy π, its welfare under P is*

$$V_P(\pi) = \mathbb{E}_P[\pi(X)\tau_P(X)] \, .$$

The oracle policy is

$$\pi^\star_P(x) = \mathbf{1}\{\tau_P(x) \geq 0\}.$$

The regret of π under P is

$$R_P(\pi) = V_P(\pi^\star_P) - V_P(\pi).$$

Thus $V_P(\pi)$ is normalized by dropping the treatment-invariant baseline component, $\pi^\star_P$ is the threshold rule induced by the sign of the treatment contrast, and $R_P(\pi)$ is the welfare loss from deviating from that rule.

Assumption 4. [ass:policy-class] (Finite-VC policy class) *The policy class* Π *is a pointwise measurable class of measurable deterministic policies* $\pi : \mathcal{X} \to \{0, 1\}$ *with finite VC dimension* $d_\Pi < \infty$*. There exists a countable subclass*

$$\Pi_0 = \{\pi_j : j \geq 1\} \subset \Pi$$

such that every $\pi \in \Pi$ *is the pointwise limit of a sequence of policies from* Π_0*.*

This is the standard pointwise measurable finite-VC policy-class condition, used to make empirical welfare maximization over Π statistically and measurably well behaved [Kitagawa and Tetenov, 2018].

Assumption 5. [ass:optimal-in-class] (Oracle policy inclusion) *For every law* P *in the law class, the oracle policy* $\pi_P^\star$ *belongs to* Π.

This is the standard optimum-in-class condition for welfare learning over restricted policy classes, ensuring that the oracle benchmark is attainable within Π when an upper-bound argument requires such an in-class comparator [Kitagawa and Tetenov, 2018]. It is kept separate from the observed-law class defined below: the minimax lower-bound class is an observed-law class, whereas oracle inclusion is imposed only where the relevant statement explicitly requires comparison to an element of Π.

The next assumptions restrict the distribution of small treatment contrasts. The margin condition is the analogue of the low-noise condition in classification: it limits how much covariate mass can lie near the decision boundary $\tau_P(x) = 0$.

Assumption 6. [ass:margin] (Margin condition) *The parameters satisfy* $0 \leq \alpha$, $0 < C_m$*, and* $0 < u_0$*. For every* u *with* $0 < u \leq u_0$*,*

$$P\{0 < |\tau_P(X)| \leq u\} \leq C_m u^\alpha.$$

This is the standard Tsybakov decision margin condition, controlling the probability of covariate values at which the welfare consequences of the treatment decision are small [Tsybakov, 2004].

Assumption 7. [ass:zero-effect] (Zero-effect agreement) *Either*

$$P_X\{\tau_P(X) = 0\} = 0,$$

or every policy $\pi \in \Pi$ *agrees with the oracle policy* $\pi_P^\star$ P_X*-almost everywhere on the set* $\{\tau_P(X) = 0\}$*.*

This convention removes welfare-irrelevant disagreements on covariate points where both treatment choices have the same value for the regret target.

Assumption 8. [ass:margin-window] (Margin window) *The margin window satisfies*

$$0 < u_0 < 2.$$

This margin-window normalization fixes the range on which the small-contrast condition is imposed. Since the bounded-outcome condition restricts treatment-arm means to $[-1, 1]$, treatment contrasts lie in $[-2, 2]$, and $u_0 < 2$ keeps the window inside that scale.

Definition 2. [`def:disagreement`] (Disagreement Set D_π) *For a policy π, its disagreement set is*

$$D_\pi = \{x \in \mathcal{X} : \pi(x) \neq \pi_P^\star(x)\}.$$

The disagreement set records exactly where a policy makes a different treatment assignment from the oracle threshold rule. Regret can therefore be read as a weighted classification error, with weights given by the absolute treatment contrast.

Theorem 1. [`thm:welfare-identity`] (Welfare regret identity) *Under Assumption 2, for every deterministic policy $\pi \in \Pi$,*

$$R_P(\pi) = E_P[|\tau_P(X)|\,\mathbf{1}\{\pi(X) \neq \pi_P^\star(X)\}] = \int_{\mathcal{X}} |\tau_P(x)|\,\mathbf{1}\{\pi(x) \neq \pi_P^\star(x)\}\,dP_X(x).$$

The identity makes the welfare-regret target explicit: mistakes on high-contrast covariate regions are costly, while mistakes near $\tau_P(x) = 0$ contribute little to regret.

Theorem 2. [`thm:margin-localization`] (Margin Localization) *Suppose that the following conditions hold:*

- ***Margin condition.*** *Assumption 6 holds.*
- ***Zero-effect convention.*** *Assumption 7 holds.*
- ***Bounded outcomes.*** *Assumption 2 holds.*
- ***Disagreement sets.*** *D_π is defined as in Definition 2.*

Then there exists a constant

$$C = C(C_m, u_0, \alpha)$$

such that, for every $\pi \in \Pi$,

$$P_X(D_\pi) \leq C\, R_P(\pi)^{\alpha/(1+\alpha)}.$$

The localization result converts small regret into a small disagreement region under the margin condition, a step that parallels the standard excess-risk localization used in margin-based classification analysis [Audibert and Tsybakov, 2007, Massart and Nédélec, 2006, Tsybakov, 2009].

The distinctive restriction in this paper governs the interaction between weak overlap and small contrasts. For $\gamma > 0$, the law class controls the probability that weak overlap and small welfare-relevant contrasts occur together.

Assumption 9. [`ass:overlap-decay`] (Overlap decay) *For all u and v with $0 < u \leq u_0$ and $v > 0$ satisfying*

$$v \leq c_o u^\gamma,$$

the joint small-contrast and weak-overlap probability obeys

$$P\{p_P(X) \leq v,\ 0 < |\tau_P(X)| \leq u\} \leq C_o u^\alpha v^{1/\gamma},$$

with the convention that $v^{1/\gamma} = 1$ when $\gamma = 0$.

This condition is specific to this analysis: it permits overlap to deteriorate near either propensity boundary, but only at a rate tied to the small-contrast region that drives welfare regret. The lower-bound witness itself uses one boundary.

Assumption 10. [`ass:strict-overlap-endpoint`] (Strict-overlap endpoint) *When $\gamma = 0$, there exists a constant $\underline{p} \in (0, 1/2]$ such that*

$$p_P(X) \geq \underline{p}$$

holds P_X-almost surely.

This is the standard strict-overlap, or uniform-positivity, endpoint condition for the $\gamma = 0$ case [Imbens and Rubin, 2015].

Definition 3. [`def:exponents`] (Exponents $\beta_{\alpha,\gamma}$, $D_{\alpha,\gamma}$, and $r_\star(\alpha, \gamma)$) *The overlap calibration exponent, denominator exponent, and minimax exponent are*

$$\beta_{\alpha,\gamma} = \begin{cases} 0, & \gamma = 0, \\ \dfrac{\alpha\gamma}{\alpha + 1}, & \gamma > 0, \end{cases} \qquad D_{\alpha,\gamma} = 2 + \alpha + \beta_{\alpha,\gamma}, \qquad r_\star(\alpha, \gamma) = \frac{1 + \alpha}{D_{\alpha,\gamma}}.$$

The term $\beta_{\alpha,\gamma}$ is zero under strict overlap and, for $\alpha > 0$ and $\gamma > 0$, positive when weak overlap is allowed to concentrate near small contrasts. The denominator $D_{\alpha,\gamma}$ is the exponent balance that will enter the two-point lower-bound construction, and $r_\star(\alpha, \gamma)$ is the corresponding lower-bound regret exponent.

Definition 4. [`def:law-class`] (Law class $\mathcal{P}_{\alpha,\gamma}$) *Fix the policy class Π and the constants $C_m, u_0, C_o, c_o, \underline{p}$. The class $\mathcal{P}_{\alpha,\gamma}$ consists of all well-formed observed laws P on $\Omega = \mathcal{X} \times \{0, 1\} \times [-1, 1]$ for which the following properties hold: $Y \in [-1, 1]$ P-a.s. and $\mu_0(x), \mu_1(x) \in [-1, 1]$ for every x; $0 < e_P(x) < 1$ P_X-a.s.; $0 \leq \alpha$, $0 < C_m$, $0 < u_0$, and $P_X\{0 < |\tau_P(X)| \leq u\} \leq C_m u^\alpha$ for every $0 < u \leq u_0$; either $P_X\{\tau_P(X) = 0\} = 0$, or every $\pi \in \Pi$ agrees with $\pi_P^\star$ on $\{\tau_P = 0\}$ up to P_X-null sets; for every $0 < u \leq u_0$ and $0 < v \leq c_o u^\gamma$,*

$$P_X\{p_P(X) \leq v,\ 0 < |\tau_P(X)| \leq u\} \leq C_o u^\alpha \begin{cases} 1, & \gamma = 0, \\ v^{1/\gamma}, & \gamma \neq 0; \end{cases}$$

and, when $\gamma = 0$, $0 < \underline{p} \leq 1/2$ and $p_P(x) \geq \underline{p}$ P_X-a.s.

Equivalently, $\mathcal{P}_{\alpha,\gamma}$ is the observed-law class satisfying bounded outcomes, positivity, the margin and zero-effect restrictions, the joint overlap-decay condition, and the strict-overlap endpoint convention when $\gamma = 0$. The sampling condition in Assumption 1, the finite-VC structure in Assumption 4, the margin-window normalization in Assumption 8, and the oracle-in-class condition in Assumption 5 are separate structural conditions invoked by the statements that need them.

4 Minimax Lower Bound

This section gives the observed-law minimax lower bound over the class $\mathcal{P}_{\alpha,\gamma}$ in Definition 4. The argument uses a two-point testing construction in the sense of Le Cam [1986]: two laws agree away from a small active block, differ only in the sign of a local treatment contrast on that block, and are calibrated so that their product distributions remain statistically close. The regret separation is then read through the welfare identity in Theorem 1. The result characterizes the observed-law converse exponent for this law class.

Definition 5. [`def:minimax-regret`] (Minimax regret $M_n(\alpha,\gamma)$) *The minimax regret over the law class $\mathcal{P}_{\alpha,\gamma}$ is*

$$M_n(\alpha,\gamma) = \inf_{\widehat{\pi}_n} \sup_{P\in\mathcal{P}_{\alpha,\gamma}} \mathbb{E}_P[R_P(\widehat{\pi}_n)]\,,$$

where the infimum ranges over all measurable data-dependent estimators $\widehat{\pi}_n$ taking values in Π.

The risk in Definition 5 is evaluated under the observed law P, and the supremum is taken only over laws satisfying the boundedness, positivity, margin, zero-effect, overlap-decay, and endpoint restrictions collected in Definition 4. The estimator is required to output an element of the policy class Π, while regret remains measured relative to the oracle rule in Definition 1.

The lower bound is driven by the balance between three quantities: the block mass allowed by the margin condition, the local contrast size, and the treatment-arm sampling probability on the weak-overlap cell. The construction below fixes the contrast scale h_n, makes the active block have mass of order h_n^{α}, and assigns the weak treatment arm probability of order $h_n^{\beta_{\alpha,\gamma}}$ when overlap decay permits it.

Definition 6. [`def:two-point-witness`] (Two-point witness laws $P_{n,\sigma}$) *Let $\mathcal{X}=[0,1]$ and let P_X be Lebesgue measure restricted to* $[0,1]$. *Set*

$$h_n = n^{-1/(2+\alpha+\beta_{\alpha,\gamma})}, \qquad q_n = \begin{cases} 1/4, & \beta_{\alpha,\gamma}=0,\\ h_n^{\beta_{\alpha,\gamma}}, & \beta_{\alpha,\gamma}>0, \end{cases}$$

and

$$B_n = \{x\in\mathbb{R} : 0\le x\le c_B h_n^{\alpha}\}.$$

Set $\tau_0=(u_0+2)/2$. For each $\sigma\in\{1,-1\}$, the law $P_{n,\sigma}$ has the same covariate law P_X, the same logging propensity

$$e(x) = \begin{cases} q_n, & x\in B_n,\\ 1/2, & x\notin B_n, \end{cases}$$

and the same off-block outcome law supported on $\{-1,1\}$, with

$$E[Y\mid X=x, A=1] = \tau_0/2, \qquad E[Y\mid X=x, A=0] = -\tau_0/2, \qquad x\notin B_n.$$

Thus the common off-block optimal label is 1, *and the common off-block contrast is τ_0. On B_n, set $Y=0$ almost surely when $A=0$. On the active treated cell $\{X\in B_n, A=1\}$, let Y be supported on $\{-1,1\}$ with*

$$P(Y=1\mid X\in B_n, A=1) = \frac{1+\sigma h_n}{2}, \qquad P(Y=-1\mid X\in B_n, A=1) = \frac{1-\sigma h_n}{2}.$$

Consequently,

$$E[Y\mid X\in B_n, A=1] = \sigma h_n, \qquad \tau_{P_{n,\sigma}}(x) = \sigma h_n \quad \text{for } x\in B_n.$$

The two laws agree everywhere except on the treated active cell. The constants satisfy

$$8c_B < \log 5.$$

The construction in Definition 6 is localized in the relevant sense: the two laws differ only in the treated observations on B_n, and the probability of entering that informative cell is $P_X(B_n)q_n$. Thus weak overlap reduces information exactly where the sign of the contrast is hardest to learn.

The exponent $\beta_{\alpha,\gamma}$ in Definition 3 is the largest weak-arm exponent within the tight-window algebraic calibration used for a block of margin mass h^α. The next proposition records this calibration.

Proposition 1. [`prop:overlap-envelope`] (Overlap envelope calibration) *For $\alpha \geq 0$, $\gamma > 0$, $h \in (0,1)$, and $\beta \geq 0$, consider the tight window*

$$v = h^\beta, \qquad u = h^{\beta/\gamma}.$$

Then

$$u^\alpha v^{1/\gamma} = h^{(\alpha+1)\beta/\gamma},$$

and

$$u^\alpha v^{1/\gamma} \geq h^\alpha \quad \Longleftrightarrow \quad \beta \leq \beta_{\alpha,\gamma} = \frac{\alpha\gamma}{\alpha+1},$$

with equality in the displayed power comparison whenever $\beta = \beta_{\alpha,\gamma}$. Moreover $\beta_{\alpha,\gamma} \geq 0$, and for every $\beta' \geq 0$, the corresponding tight window $v = h^{\beta'}$, $u = h^{\beta'/\gamma}$ satisfies

$$u^\alpha v^{1/\gamma} \geq h^\alpha \quad \Longleftrightarrow \quad \beta' \leq \beta_{\alpha,\gamma}.$$

Thus, within this algebraic tight-window envelope, $\beta_{\alpha,\gamma}$ is the largest admissible weak-arm exponent.

Proposition 1 is the source of the denominator $2 + \alpha + \beta_{\alpha,\gamma}$. The term 2 comes from distinguishing Bernoulli means separated by h_n, the term α from the block mass allowed by the margin condition, and the term $\beta_{\alpha,\gamma}$ from the probability of observing the informative treatment arm.

The witness laws must also satisfy the restrictions defining $\mathcal{P}_{\alpha,\gamma}$. The next statement verifies membership and records the two oracle rules induced by the sign change on B_n.

Lemma 1. [`lem:witness-membership`] (Witness-law membership) *Assume $\alpha \geq 0$, $\gamma \geq 0$, the margin-window normalization $0 < u_0 < 2$ in Assumption 8, $C_m, C_o, c_o, c_B, \underline{p} > 0$, $c_B \leq C_m$, $c_B \leq C_o$, and $\underline{p} \leq 1/4$. Also assume the witness constant satisfies the overlap-decay smallness conditions*

$$0 < \gamma,\ 0 < \alpha \quad \Longrightarrow \quad c_B \leq C_o c_o^{-\alpha/\gamma}, \qquad 0 < \gamma,\ \alpha = 0 \quad \Longrightarrow \quad c_B \leq C_o 4^{-1/\gamma}.$$

Then, for each witness sign $\sigma \in \{+,-\}$, for all sufficiently large n, the witness law $P_{n,\sigma}$ of Definition 6 satisfies the observed-law membership conditions defining the class $\mathcal{P}_{\alpha,\gamma}$ in Definition 4. Hence, for all sufficiently large n,

$$P_{n,+}, P_{n,-} \in \mathcal{P}_{\alpha,\gamma}.$$

Moreover, the two corresponding witness-optimal policies are

$$x \mapsto 1 \qquad \textit{and} \qquad x \mapsto \mathbf{1}\{x \notin B_n\},$$

where $B_n = \{x : 0 \leq x \leq c_B h_n^\alpha\}$. If these two policies belong to Π, then they are available as the two policy actions in the associated two-point lower-bound reduction. The construction uses the off-block contrast $\tau_0 = (u_0 + 2)/2$, so $\tau_0 \in (u_0, 2)$ under the normalization $0 < u_0 < 2$.

The proof is deferred to Section E.

Lemma 1 has two roles. First, it places both local alternatives inside the same observed-law class used in the minimax risk. Second, it identifies the two witness-optimal rules. When those two rules lie in Π, they are the two admissible policy actions in the associated two-action testing reduction; the regret lower bound itself is still stated for the minimax risk in Definition 5, with regret evaluated relative to the oracle rule.

It remains to check that the two laws are hard to distinguish at sample size n. Because the laws differ only on a set with covariate mass m_n, treatment probability q_n, and conditional mean separation of order h_n, the one-observation divergence has order $m_n q_n h_n^2$.

Lemma 2. [`lem:two-point-divergence`] (Two-point divergence) *For the witness laws in Definition 6, under Assumption 8, assume in addition that*

$$0 < c_B, \qquad \alpha \geq 0, \qquad \gamma \geq 0, \qquad 8c_B < \log 5.$$

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

$$P_{n,+}^{\otimes n} \ll P_{n,-}^{\otimes n}, \qquad \int \left(\frac{dP_{n,+}^{\otimes n}}{dP_{n,-}^{\otimes n}} - 1 \right)^2 dP_{n,-}^{\otimes n} < \infty.$$

Moreover the one-observation chi-square divergence satisfies

$$\chi^2(P_{n,+} \,\|\, P_{n,-}) \leq C\, h_n^{2+\alpha+\beta_{\alpha,\gamma}}, \qquad h_n = n^{-1/(2+\alpha+\beta_{\alpha,\gamma})}.$$

For the same constant C, the product divergence is bounded:

$$\chi^2(P_{n,+}^{\otimes n} \,\|\, P_{n,-}^{\otimes n}) \leq C.$$

The proof is deferred to Section E.

The bandwidth h_n in Definition 6 is chosen so that $nh_n^{2+\alpha+\beta_{\alpha,\gamma}}$ is bounded. Hence no estimator can reliably determine which sign generated the active block merely from the observed sample.

The same active block creates the welfare separation. Under one law the oracle treats B_n, and under the other law the oracle does not. Any single policy must disagree with at least one of these two assignments on a nontrivial part of the block.

Lemma 3. [`lem:regret-separation`] (Regret separation) *For the witness laws in Definition 6, suppose Assumption 8 holds, $\alpha, \gamma \geq 0$, $c_B, C_m, C_o, c_o, \underline{p} > 0$, $c_B \leq C_m$, $c_B \leq C_o$, $\underline{p} \leq 1/4$, every $\pi \in \Pi$ is measurable, and the block-size constant satisfies*

$$\gamma > 0,\ \alpha > 0 \implies c_B \leq C_o c_o^{-\alpha/\gamma}, \qquad \gamma > 0,\ \alpha = 0 \implies c_B \leq C_o 4^{-1/\gamma}.$$

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

$$\forall \pi \in \Pi, \qquad \max_{\sigma \in \{+,-\}} R_{P_{n,\sigma}}(\pi) \geq c\, h_n^{1+\alpha}.$$

For such n, the two optimal policies disagree on the active block. Hence every $\pi \in \Pi$ misclassifies the active block under at least one of the two laws, and the sum of the two regrets is bounded below by the contrast scale times a covariate subblock of mass $(\min\{c_B, 1\})h_n^{\alpha}$. Passing from the sum to the maximum gives the stated separation of order $h_n^{1+\alpha}$.

The proof is deferred to Section E.

Lemma 3 converts the testing ambiguity into welfare regret. The separation is h_n per misclassified point on a set of mass h_n^{α}, giving $h_n^{1+\alpha}$.

The testing step uses the standard chi-square form of Le Cam's two-point method [Le Cam, 1986, Tsybakov, 2009].

Lemma 4. [`lem:le-cam-two-point-chisq`] (Chi-square testing bound) *Let P_+ and P_- be probability laws on a common measurable space, and assume the two product laws satisfy the regularity conditions*

$$P_+^{\otimes n} \ll P_-^{\otimes n}, \qquad \int \left(\frac{dP_+^{\otimes n}}{dP_-^{\otimes n}} - 1\right)^2 dP_-^{\otimes n} < \infty.$$

If, in addition,

$$\chi^2(P_+^{\otimes n} \,\|\, P_-^{\otimes n}) \le C_\chi,$$

then every measurable test ψ taking values in $\{+,-\}$ satisfies

$$P_+^{\otimes n}(\psi \ne +) + P_-^{\otimes n}(\psi \ne -) \ge c(C_\chi) > 0,$$

where the floor $c(C_\chi)$ depends only on the budget C_χ, and not on the pair of laws or on the test. In particular, for $c_0 \ge 0$ and $n \ge 1$, this conclusion holds whenever the one-observation pair obeys the same two regularity conditions,

$$P_+ \ll P_-, \qquad \int \left(\frac{dP_+}{dP_-} - 1\right)^2 dP_- < \infty,$$

and the one-observation chi-square divergence is bounded by c_0/n: the two displayed product conditions are then inherited from the one-observation pair, and

$$1 + \chi^2(P_+^{\otimes n} \,\|\, P_-^{\otimes n}) = \big(1 + \chi^2(P_+ \,\|\, P_-)\big)^n \le \left(1 + \frac{c_0}{n}\right)^n \le e^{c_0}.$$

The proof is deferred to Section E.

Combining Lemma 2, Lemma 3, and Lemma 4 yields the minimax lower bound. Since $h_n^{1+\alpha} = n^{-(1+\alpha)/(2+\alpha+\beta_{\alpha,\gamma})}$, the lower-bound exponent is $r_\star(\alpha,\gamma)$.

Theorem 3. [`thm:minimax-lower`] (Minimax lower bound) *Under Assumption 8, suppose that $\alpha, \gamma \ge 0$, $C_m, C_o, c_o, c_B, \underline{p} > 0$, $c_B \le C_m$, $c_B \le C_o$, $0 < \underline{p} \le 1/4$, and $8c_B < \log 5$. In addition, when $\gamma > 0$ and $\alpha > 0$, assume $c_B \le C_o c_o^{-\alpha/\gamma}$, and when $\gamma > 0$ and $\alpha = 0$, assume $c_B \le C_o 4^{-1/\gamma}$. If the policy class Π is nonempty and every $\pi \in \Pi$ is measurable, then for the minimax regret $M_n(\alpha,\gamma)$ of Definition 5 over the law class $\mathcal{P}_{\alpha,\gamma}$ of Definition 4, there is a constant $c > 0$, independent of n, such that for all sufficiently large n,*

$$M_n(\alpha,\gamma) = \inf_{\widehat{\pi}} \sup_{P \in \mathcal{P}_{\alpha,\gamma}} \mathbb{E}_P R_P(\widehat{\pi}) \ge c\, n^{-r_\star(\alpha,\gamma)}.$$

Theorem 3 is an observed-law lower bound stated entirely in terms of the law class in Definition 4, the regret functional in Definition 1, and the minimax risk in Definition 5. Theorem 4 develops the complementary procedure-specific achievability analysis.

Writing the exponent of Definition 3 explicitly, $r_\star(\alpha,\gamma) = (1+\alpha)/(2+\alpha+\beta_{\alpha,\gamma})$. Under strict overlap, $\beta_{\alpha,\gamma} = 0$, and the lower-bound exponent reduces to the usual margin-driven benchmark. When $\alpha > 0$ and $\gamma > 0$, the additional positive term $\beta_{\alpha,\gamma}$ reflects the loss of treatment-arm information in the small-contrast region; at the boundary $\alpha = 0$, it is zero. The section that follows considers a clipped cross-fitted AIPW empirical welfare rule and gives a separate conditional upper bound under explicit nuisance-rate and empirical-process side conditions.

5 Conditional Analysis of a Clipped Cross-Fitted AIPW Rule

We now turn from the observed-law lower bound to one specified empirical rule. The rule uses a clipped augmented inverse-propensity-weighted score for the treatment contrast and maximizes the resulting empirical welfare criterion over the same pointwise measurable finite-VC policy class used above. This section establishes a conditional rate for the clipped cross-fitted AIPW ERM with supplied nuisance estimates under the nuisance-rate, cross-fitting, boundedness, and localized empirical-process side conditions stated below. These explicit inputs define the procedure-specific achievability domain.

The AIPW score follows the semiparametric and doubly robust construction of Robins et al. [1994], Hahn [1998], Hirano et al. [2003], van der Laan and Rose [2011], while cross-fitting is used to separate first-stage nuisance estimation from foldwise policy evaluation as in double/debiased machine learning [Chernozhukov et al., 2018, 2022]. Clipping is needed because the overlap-decay class permits small propensity-boundary distance in the margin region.

Assumption 11. [`ass:vc-localized-envelope`] (Localized VC envelope) *For the pointwise measurable finite-VC policy class Π, there exist constants $C > 0$ and $p \geq 0$ such that, for every sample size $m \geq 1$, every $B \geq 0$, every localized radius $r \geq 0$, and every policy-compatible policy-indexed increment class g_π satisfying $|g_\pi(O)| \leq B$ for all $\pi \in \Pi$ and all O, and*

$$\int g_\pi(O)^2 \, dP(O) \leq B^2 P_X(D_\pi), \qquad \pi \in \Pi,$$

the expected fixed-radius localized centered supremum obeys

$$\mathbb{E}\left[\sup_{\pi\in\Pi:\, R_P(\pi)\leq r} \left| m^{-1}\sum_{i=1}^{m} g_\pi(O_i) - \int g_\pi(O)\, dP(O) \right|\right] \leq CBm^{-1/2} r^{\alpha/(2+2\alpha)} (\log m)^p.$$

This is a primitive high-level localized empirical-process condition for the policy-compatible increment classes used below. It is motivated by finite-VC localization results [Bartlett, 2005]; the displayed uniform bound enters the analysis as an explicit assumption beyond finite VC dimension.

Assumption 12. [`ass:nuisance-rate`] (Cross-fitted nuisance rates) *The cross-fitted nuisance estimators satisfy, for each $a \in \{0,1\}$,*

$$\|\widehat{\mu}_a - \mu_a\|_{L_2(P)} \leq r_{\mu,n}, \qquad \|\widehat{e} - e\|_{L_2(P)} \leq r_{e,n},$$

and their product rate satisfies

$$r_{\mu,n} r_{e,n} = O(n^{-1/2}).$$

This is the standard cross-fitted $L_2(P)$ nuisance-rate condition for orthogonal score analysis; the product requirement is the usual second-order remainder control in double machine learning [Chernozhukov et al., 2018].

Assumption 13. [`ass:bounded-crossfit-nuisances`] (Bounded Cross-Fit Nuisances) *The cross-fitted outcome-regression estimators $\widehat{\mu}_0$ and $\widehat{\mu}_1$ take values in $[-1, 1]$.*

This is the standard bounded nuisance truncation condition; it matches the bounded-outcome scale and keeps the foldwise clipped scores uniformly controlled [Chernozhukov et al., 2018].

Assumption 14. [`ass:polynomial-nuisance-exponents`] (Polynomial Nuisance Rates) *There exist exponents $a \geq 0$ and $c \geq 1/2$, and constants C_μ and C_{prod}, such that, for all sufficiently large n,*

$$r_{\mu,n} \leq C_\mu n^{-a}, \qquad r_{\mu,n} r_{e,n} \leq C_{\text{prod}} n^{-c}.$$

This is the standard polynomial first-stage nuisance-rate condition; the exponents a and c make explicit which nuisance terms can limit the final regret exponent [Chernozhukov et al., 2018].

Assumption 15. [`ass:fixed-crossfit-fold-count`] (Fixed Cross-Fit Fold Count) *The number K of cross-fitting folds is a fixed finite integer independent of n, and the deterministic partition $I_1, \ldots, I_K$ of the observation indices is balanced.*

This is the fixed-K cross-fitting convention used by the conditional upper bound. The balanced deterministic folds are a side condition for the foldwise pooling step in the present empirical-process argument; the general use of cross-fitting follows the double/debiased machine-learning approach [Chernozhukov et al., 2018].

Assumption 16. [`ass:vc-localized-offset-envelope`] (Localized Offset Envelope) *For the pointwise measurable finite-VC policy class Π, there exist constants $C > 0$ and $p \geq 0$ such that, for every integer $n \geq 1$, every $B \geq 0$, and every policy-compatible increment class g_π (that is, $g_\pi(O)$ factors through O and the decision $\pi(X)$), the centered empirical process*

$$z_\pi = \frac{1}{n} \sum_{i=1}^{n} g_\pi(O_i) - \int g_\pi(O)\, dP(O) = (P_n - P) g_\pi$$

built from an i.i.d. sample from P, with envelope $|g_\pi(O)| \leq B$ for all $\pi \in \Pi$ and O, and second moment bounded by

$$\int g_\pi(O)^2\, dP(O) \leq B^2 P_X(D_\pi), \qquad \pi \in \Pi,$$

satisfies

$$\mathbb{E} \sup_{\pi \in \Pi} \left\{ 2|z_\pi| - \frac{R_P(\pi)}{4} \right\}_+ \leq C \left(\frac{B^2}{n} \right)^{A_\alpha} (\log n)^p, \qquad A_\alpha = \frac{1+\alpha}{2+\alpha}.$$

This is a second primitive high-level condition, now for an offset empirical process. It converts stochastic fluctuation into an expected regret contribution at the margin-dependent exponent; the exact statement enters as an explicit assumption beyond finite VC dimension [Bartlett, 2005].

The tuning schedule balances four terms: the clipped empirical-process envelope, the product nuisance remainder after clipping, the localized weak-overlap bias, and the outcome-regression error on the complement of the localization window. The next definition records this balance for a fixed nuisance regime.

Definition 7. [`def:feasible-rate`] (Feasible upper exponent $r_{\mathrm{up}} = r_{\mathrm{feas}}$) *Fix one nuisance regime* $(a, c, C_\mu, C_{\mathrm{prod}})$*, and set*

$$A_\alpha = \frac{1+\alpha}{2+\alpha}.$$

If $\gamma > 0$*, fix constants*

$$\bar{u} \in (0, u_0], \qquad q_0 \in \big(0, \min\{1/2, c_o \bar{u}^\gamma\}\big].$$

For $0 \le s \le 1/2$ *and* $0 \le t \le s/\gamma$*, define*

$$\phi(s,t) = \min\left\{A_\alpha(1-2s),\, c - s,\, a + \frac{s}{2\gamma} + \frac{\alpha t}{2},\, 2a - t\right\}.$$

Let $(s_{\mathrm{feas}}, t_{\mathrm{feas}})$ *be any maximizer of* ϕ *over this compact feasible set, and set*

$$g_{\mathrm{joint}}(\alpha, \gamma, a, c) = \phi(s_{\mathrm{feas}}, t_{\mathrm{feas}}), \qquad q_n = q_0 n^{-s_{\mathrm{feas}}}, \qquad u_n = \bar{u} n^{-t_{\mathrm{feas}}}.$$

Then $q_n \le c_o u_n^\gamma$ *for all sufficiently large* n*, and the solved conditional feasible upper exponent is*

$$r_{\mathrm{up}} = r_{\mathrm{feas}} = \min\left\{r_\star(\alpha,\gamma),\, g_{\mathrm{joint}}(\alpha,\gamma,a,c)\right\}.$$

If $\gamma = 0$*, take a fixed clipping sequence* $q_n = q_0$ *with* $q_0 \in (0, \underline{p}/2]$*, and set*

$$r_{\mathrm{up}} = r_{\mathrm{feas}} = \min\{A_\alpha, c\}.$$

For $\gamma > 0$, the constraint $t \le s/\gamma$ is the exponent form of the admissibility condition $q_n \le c_o u_n^\gamma$ in the overlap-decay region. The exponent r_{up} in Definition 7 is therefore the conditional exponent delivered by this balance, truncated by the lower-bound exponent $r_\star(\alpha,\gamma)$ from Definition 3.

Definition 8. [`def:clipped-propensity`] (Clipped propensity $e_q(x;\eta)$) *For a nuisance triple* $\eta = (\mu_0, \mu_1, e)$ *and a clipping level* $q \in (0, 1/2]$*, the clipped propensity score is*

$$e_q(x;\eta) = \min\{1-q, \max\{q, e(x)\}\}.$$

The nuisance triple $\eta = (\mu_0, \mu_1, e)$ collects the two outcome regressions and the propensity score used in the plug-in score. Clipping replaces the propensity component by a value in $[q, 1-q]$, controlling inverse-propensity weights while preserving the target welfare functional in Definition 1.

Definition 9. [`def:clipped-aipw-score`] (Clipped AIPW score $\Gamma_q(O;\eta)$) *For* $O = (X, A, Y)$*, nuisance triple* $\eta = (\mu_0, \mu_1, e)$*, and clipping level* $q \in (0, 1/2]$*, the clipped augmented inverse-propensity-weighted score for the treatment contrast is*

$$\Gamma_q(O;\eta) = \mu_1(X) - \mu_0(X) + \frac{A}{e_q(X;\eta)}\{Y - \mu_1(X)\} - \frac{1-A}{1-e_q(X;\eta)}\{Y - \mu_0(X)\}.$$

When the nuisance functions are equal to the true observed-law regressions and propensity and no clipping binds, Γ_q is the usual AIPW score for the contrast $\tau_P(X)$. Under overlap decay, clipping introduces a deterministic drift that requires explicit control beyond the strict-overlap simplification.

Definition 10. [`def:feasible-erm`] (Feasible clipped-AIPW ERM $\widehat{\pi}_n$) *Fix the countable pointwise dense subclass*

$$\Pi_0 = \{\pi_j : j \geq 1\}$$

of Π*, and fix a deterministic balanced* K*-fold partition*

$$I_1, \ldots, I_K$$

of $\{1, \ldots, n\}$*, with each fold having size* $\lfloor n/K \rfloor$ *or* $\lceil n/K \rceil$*. Let* $k(i)$ *denote the evaluation fold containing observation* i*. Given foldwise cross-fitted nuisance estimators* $\widehat{\eta}^{(-k)}$ *and a clipping level* $q \in (0, 1/2]$*, define the empirical welfare criterion*

$$\widehat{V}_{n,q}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \pi(X_i)\, \Gamma_q\Big(O_i; \widehat{\eta}^{(-k(i))}\Big).$$

Let j_n *be the smallest index* $j \geq 1$ *such that*

$$\widehat{V}_{n,q}(\pi_j) \geq \sup_{\ell \geq 1} \widehat{V}_{n,q}(\pi_\ell) - \frac{1}{n},$$

and set

$$\widehat{\pi}_n = \pi_{j_n}.$$

This measurable Π*-valued* $1/n$*-empirical risk maximizer over* Π_0 *is the feasible clipped-AIPW ERM. The clipping level* q *is the tuning parameter chosen by the bias-variance balance.*

Definition 10 fixes the enumeration, the balanced fold partition, and the foldwise supplied nuisance estimates. The superscript $(-k)$ indicates that the nuisance estimate used on fold I_k is trained away from that fold; its construction is otherwise abstracted into the side conditions above.

Definition 11. [`def:upper-risk`] (Conditional upper risk $U_n(\alpha, \gamma, a, c; \widehat{\eta})$) *Fix a nuisance regime* $(a, c, C_\mu, C_{\mathrm{prod}})$*, the associated deterministic schedules* (q_n, u_n)*, and the supplied policy class, enumeration, nuisance estimators, and fold assignment. The conditional upper risk is*

$$U_n(\alpha, \gamma, a, c; \widehat{\eta}) = \sup E_P[R_P(\widehat{\pi}_n)]\,,$$

where the supremum is over all laws $P \in \mathcal{P}_{\alpha,\gamma}$ *for which the oracle policy belongs to* Π*, the supplied cross-fitted nuisance estimators* $\widehat{\eta}^{(-k)}$ *satisfy the stated* $L_2(P)$ *nuisance-rate bounds, boundedness conditions, and polynomial nuisance-rate bounds with exponents* a, c *and constants* C_μ, C_{prod}*, the supplied policy class* Π *satisfies the stated finite-VC, localized-envelope, and localized-offset-envelope conditions, the supplied cross-fitting scheme has fixed fold count* K*, and the supplied enumeration is a countable pointwise-dense skeleton* $\Pi_0 = \{\pi_j : j \geq 1\}$ *of* Π*.*

The risk in Definition 11 is evaluated over the stated side-condition domain: laws in $\mathcal{P}_{\alpha,\gamma}$, oracle-in-class policy problems, supplied cross-fitted nuisances with the required rates and boundedness, policy classes satisfying the localized empirical-process conditions, and fixed balanced cross-fitting schemes. The definition precisely describes the conditional risk controlled by this clipped-AIPW ERM.

The basic inequality is the usual consequence of near-maximizing the empirical welfare criterion over the dense countable subclass.

Lemma 5. [`lem:feasible-erm-basic-inequality`] (Feasible ERM Inequality) *Under Assumption 4, Definition 10 defines a measurable Π-valued estimator $\widehat{\pi}_n$ satisfying*

$$\widehat{V}_{n,q}(\widehat{\pi}_n) \geq \sup_{\pi \in \Pi} \widehat{V}_{n,q}(\pi) - \frac{1}{n}.$$

Hence, for every comparator $\pi^b \in \Pi$,

$$P_n\Big[(\widehat{\pi}_n - \pi^b)\Gamma_q\Big] \geq -\frac{1}{n}.$$

Under Assumption 5, this conclusion applies with $\pi^b = \pi^\star_P$.

The proof is deferred to Section E.

The comparator $\pi^\star_P$ is available only when Assumption 5 is imposed. This is why the upper-bound analysis carries oracle inclusion as an explicit side condition even though the lower bound in Theorem 3 is stated as an observed-law result over $\mathcal{P}_{\alpha,\gamma}$.

Lemma 6. [`lem:crude-clipped-score-envelope`] (Clipped-Score Envelope) *Under the bounded-outcome condition of Assumption 2, for every $q \in (0, 1/2]$ and every cross-fitted nuisance triple $\widehat{\eta}^{(-k)} = (\widehat{\mu}_0^{(-k)}, \widehat{\mu}_1^{(-k)}, \widehat{e}^{(-k)})$ whose outcome-regression components satisfy $\widehat{\mu}_0^{(-k)}(x), \widehat{\mu}_1^{(-k)}(x) \in [-1, 1]$ for every x, there exists a constant $C > 0$ such that the clipped AIPW score of Definition 9 satisfies*

$$\left|\Gamma_q\Big(O; \widehat{\eta}^{(-k)}\Big)\right| \leq \frac{C}{q} \quad \textit{almost surely.}$$

Consequently, with the same constant,

$$\Gamma_q\Big(O; \widehat{\eta}^{(-k)}\Big)^2 \leq \frac{C}{q^2} \quad \textit{almost surely.}$$

The proof is deferred to Section E.

Lemma 6 identifies the stochastic price of clipping at level q. Smaller q reduces clipping bias but increases the empirical-process envelope, producing the term $(nq^2)^{-A_\alpha}$ in the master bound below.

Lemma 7. [`lem:localized-clipped-drift-bound`] (Localized Clipped Drift Bound) *Under Assumption 12, Assumption 14, Assumption 9, Assumption 7, Assumption 2, and Assumption 10, and Lemma 13, let P be a well-formed observed law, let every $\pi \in \Pi$ be measurable, and fix a clipping level q with $0 < q \leq 1/2$. With the cross-fitted nuisance estimators fixed on the evaluation fold, write $\widehat{\eta}^{(-k)} = (\widehat{\mu}_0^{(-k)}, \widehat{\mu}_1^{(-k)}, \widehat{e}^{(-k)})$, and suppose $r_{\mu,n}, r_{e,n} \geq 0$,*

$$\int (\widehat{\mu}_0^{(-k)}(x) - \mu_0(x))^2 \, dP_X(x) \leq r_{\mu,n}^2, \quad \int (\widehat{\mu}_1^{(-k)}(x) - \mu_1(x))^2 \, dP_X(x) \leq r_{\mu,n}^2,$$

$$\int (\widehat{e}^{(-k)}(x) - e_P(x))^2 \, dP_X(x) \leq r_{e,n}^2,$$

and that these three nuisance-error functions belong to $L_2(P_X)$. Define

$$b_q(x) = \mathbb{E}_P\Big[\Gamma_q(O; \widehat{\eta}^{(-k)}) \mid X = x\Big] - \tau_P(x).$$

If $\gamma > 0$, set $C_0 = 4\{1 + \sqrt{\max\{C_o, 1\}}\}$. Then, for every $\pi \in \Pi$ and every localization window u satisfying $0 < u \le u_0$, writing $r = R_P(\pi)$,

$$|P[(\pi - \pi_P^\star)b_q]| \le C_0 \left\{ \frac{r_{\mu,n} r_{e,n}}{q} + r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)} + r_{\mu,n} \left(\frac{r}{u}\right)^{1/2} \right\}$$

whenever $q \le c_o u^\gamma$.

If $\gamma = 0$ and $q \le \underline{p}/2$, set $C_1 = \max\{1, 2/q\}$. Then, for every $\pi \in \Pi$,

$$|P[(\pi - \pi_P^\star)b_q]| \le C_1 r_{\mu,n} r_{e,n}.$$

The proof is deferred to Section E.

The drift b_q is the conditional mean bias of the clipped and estimated score relative to the true contrast. In the weak-overlap case, Lemma 7 separates the drift into a clipped product remainder, a small-contrast weak-overlap term, and a localization-complement term; under strict overlap, only the product nuisance remainder remains.

Lemma 8. [`lem:crude-localized-master-bound`] (Crude Localized Master Bound) *Let (q_n) and (u_n) be deterministic clipping and contrast-window schedules, and let $\widehat{\pi}_n$ be the cross-fitted clipped-AIPW $1/n$-ERM of Definition 10, computed with clip q_n over the enumeration $\Pi_0 = \{\pi_j : j \ge 1\}$.*

Assume:

- ***(Policy class, skeleton, and oracle.)*** *The policy class satisfies Assumption 4; welfare and regret are as in Definition 1. The enumeration used by the ERM is the pointwise-dense skeleton of Π: every $\pi_j \in \Pi$, and every $\pi \in \Pi$ is the pointwise limit of a subsequence from Π_0. For each law P to which the bound is applied, the oracle satisfies Assumption 5.*
- ***(Sampling.)*** *For each law P to which the bound is applied, the observations are i.i.d. as in Assumption 1.*
- ***(Nuisance rates.)*** *The polynomial nuisance exponent condition in Assumption 14 holds. For each law P to which the bound is applied, each foldwise nuisance triple satisfies Assumption 12. In addition, $r_{\mu,n} \ge 0$ and $r_{e,n} \ge 0$ for all sufficiently large n.*
- ***(Cross-fitting and boundedness.)*** *The number of folds is fixed and the fold assignment is balanced as in Assumption 15; for each law P to which the bound is applied, the foldwise cross-fitted outcome nuisances satisfy Assumption 13.*
- ***(Clipping schedule.)*** *For all sufficiently large n, $0 < q_n \le 1/2$. If $\gamma = 0$, then there is a fixed $q_0 > 0$ such that $q_n = q_0$ for all sufficiently large n.*
- ***(Measurability and L^2 regularity.)*** *For every n and fold k, the nuisance functions $\widehat{\mu}_{0,n}^{(-k)}$, $\widehat{\mu}_{1,n}^{(-k)}$, and $\widehat{e}_n^{(-k)}$ are measurable. For all sufficiently large n, every $P \in \mathcal{P}_{\alpha,\gamma}$, and every fold k, the errors $\widehat{\mu}_{0,n}^{(-k)} - \mu_0$, $\widehat{\mu}_{1,n}^{(-k)} - \mu_1$, and $\widehat{e}_n^{(-k)} - e_P$ belong to $L^2(P_X)$.*
- ***(Localized empirical processes.)*** *The uniform class-level localized envelope and offset envelope conditions in Assumption 11 and Assumption 16 hold.*

Table 2: Terms entering the conditional feasible exponent when $\gamma > 0$

Master-bound term	Source	Exponent component
$(nq_n^2)^{-A_\alpha}$	clipped empirical-process envelope	$A_\alpha(1-2s)$
$r_{\mu,n}r_{e,n}/q_n$	clipped product nuisance remainder	$c-s$
$r_{\mu,n}u_n^{\alpha/2}q_n^{1/(2\gamma)}$	localized weak-overlap drift	$a+s/(2\gamma)+\alpha t/2$
$r_{\mu,n}^2/u_n$	localization-complement outcome error	$2a-t$

Then there exist constants C, p, with $0 < C$ and $0 \le p$, such that for all sufficiently large n and every observed law P, if $P \in \mathcal{P}_{\alpha,\gamma}$ and the law-specific oracle, i.i.d. sampling, foldwise nuisance-rate, and bounded cross-fit nuisance conditions above hold, then the following two implications hold.

If $0 < \gamma$, $0 < u_n \le u_0$, and

$$q_n \le c_o u_n^\gamma,$$

then

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \le C\left\{n^{-r_\star} + (nq_n^2)^{-A_\alpha} + \frac{r_{\mu,n}r_{e,n}}{q_n} + r_{\mu,n}u_n^{\alpha/2}q_n^{1/(2\gamma)} + \frac{r_{\mu,n}^2}{u_n}\right\}(\log n)^p.$$

If $\gamma = 0$ and $q_n \le \underline{p}/2$, then

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \le C\left\{n^{-A_\alpha} + r_{\mu,n}r_{e,n}\right\}(\log n)^p.$$

The proof is deferred to Section E.

Lemma 8 is the organizing inequality for the upper-bound calculation. Its assumptions collect the policy, sampling, nuisance, cross-fitting, boundedness, empirical-process, and drift inputs needed by the clipped-AIPW rule.

For reference, Table 2 maps the $\gamma > 0$ terms in the master bound to the exponent components optimized in Definition 7, after substituting $q_n = q_0 n^{-s}$, $u_n = \bar{u}n^{-t}$, $r_{\mu,n} \lesssim n^{-a}$, and $r_{\mu,n}r_{e,n} \lesssim n^{-c}$.

The balance in Definition 7 chooses the clipping and localization exponents to maximize the minimum of the four entries in the last column. The next lemma records the resulting algebraic rate.

Lemma 9. [`lem:clip-balance-exponent`] (Clipping Balance Exponent) *Under the polynomial nuisance-rate exponents of Assumption 14, assume also that the nuisance constants satisfy $C_\mu \ge 0$ and $C_{\text{prod}} \ge 0$, that $r_{\mu,n} \ge 0$ for all sufficiently large n, and that $q_0 > 0$. If $\gamma > 0$, assume in addition that $\bar{u} > 0$ and that the deterministic schedules below are eventually admissible, $q_n \le c_o u_n^\gamma$. Then the deterministic clipping and localization schedules satisfy the following exponent balance.*

For $\gamma > 0$, let

$$A_\alpha = \frac{1+\alpha}{2+\alpha}, \qquad q_n = q_0 n^{-s_{\text{feas}}}, \qquad u_n = \bar{u}\, n^{-t_{\text{feas}}},$$

and

$$g_{\text{joint}} = \max_{\substack{0\le s\le 1/2\\ 0\le t\le s/\gamma}} \min\left\{A_\alpha(1-2s),\, c-s,\, a+\frac{s}{2\gamma}+\frac{\alpha t}{2},\, 2a-t\right\}.$$

There exist constants $C>0$ and $p\ge 0$ such that, for all sufficiently large n,

$$n^{-r_\star(\alpha,\gamma)} + (nq_n^2)^{-A_\alpha} + \frac{r_{\mu,n}r_{e,n}}{q_n} + r_{\mu,n}u_n^{\alpha/2}q_n^{1/(2\gamma)} + \frac{r_{\mu,n}^2}{u_n} \le Cn^{-r_{\text{feas}}}(\log n)^p,$$

where

$$r_{\text{feas}} = \min\{r_\star(\alpha,\gamma), g_{\text{joint}}\}.$$

Thus the optimized upper-bound exponent in the positive-overlap-decay regime is

$$r_{\text{up}} = r_{\text{feas}}.$$

For $\gamma=0$, with the fixed clipping schedule $q_n=q_0$, the same constants may be chosen so that, for all sufficiently large n,

$$n^{-A_\alpha} + r_{\mu,n}r_{e,n} \le Cn^{-r_{\text{feas}}}(\log n)^p, \qquad r_{\text{feas}} = \min\{A_\alpha, c\}.$$

The proof is deferred to Section E.

Lemma 9 shows exactly when this clipped-AIPW rule reaches the lower-bound exponent from Theorem 3: this happens in the nonbinding branch for $g_{\text{joint}}(\alpha,\gamma,a,c)$. In the binding branch, the result gives the conditional upper exponent for the specified rule.

We will refer to the side-condition domain of Definition 11 as the collection of laws, policy classes, nuisance estimates, and fold schemes satisfying the assumptions explicitly listed in that definition and in the theorem below. This terminology abbreviates exactly those displayed hypotheses.

Theorem 4. [`oeq:feasible-upper`] (Feasible Upper Exponent) *Fix parameters with $0\le\gamma$ and a nuisance regime $(a,c,C_\mu,C_{\text{prod}})$ satisfying Assumption 14, with $C_\mu\ge 0$ and $C_{\text{prod}}\ge 0$. Consider the law class $\mathcal{P}_{\alpha,\gamma}$ of Definition 4, the feasible cross-fitted clipped-AIPW $1/n$-ERM $\widehat{\pi}_n$ of Definition 10, formed using an enumeration of a pointwise-dense skeleton of Π, and the conditional upper risk $U_n(\alpha,\gamma,a,c;\widehat{\eta})$ of Definition 11. Let the feasible clipping and localization constants $q_0,\bar{u}$ lie in the input domain of Definition 7: if $\gamma>0$, then*

$$0<\bar{u}\le u_0, \qquad 0<q_0\le\min\{1/2, c_o\bar{u}^\gamma\},$$

and if $\gamma=0$, then

$$0<q_0\le \underline{p}/2.$$

Assume:

- ***(Policy class and skeleton.)*** *The policy class satisfies Assumption 4, and the enumeration used by $\widehat{\pi}_n$ is a pointwise-dense skeleton of Π.*
- ***(Cross-fitting.)*** *The number of folds satisfies Assumption 15.*
- ***(Nuisance boundedness.)*** *The cross-fitted outcome nuisance estimators satisfy Assumption 13.*

- ***(Nuisance measurability and integrability.)*** *For every n and fold k, $\widehat{\mu}_{0,n}^{(-k)}$, $\widehat{\mu}_{1,n}^{(-k)}$, and $\widehat{e}_n^{(-k)}$ are measurable. The rate sequences satisfy $r_{\mu,n} \geq 0$ and $r_{e,n} \geq 0$ for all sufficiently large n. For all sufficiently large n, every $P \in \mathcal{P}_{\alpha,\gamma}$, and every fold k, the errors $\widehat{\mu}_{0,n}^{(-k)} - \mu_{0,P}$, $\widehat{\mu}_{1,n}^{(-k)} - \mu_{1,P}$, and $\widehat{e}_n^{(-k)} - e_P$ belong to $L^2(P_X)$.*
- ***(Localized envelopes.)*** *The uniform localized empirical-process and offset envelope conditions in Assumption 11 and Assumption 16 hold.*

Then there exist constants C, p with $0 < C$ and $0 \leq p$ such that, for all sufficiently large n, the conditional upper risk of $\widehat{\pi}_n$, uniformly over the side-condition domain in Definition 11, satisfies

$$U_n(\alpha, \gamma, a, c; \widehat{\eta}) \leq Cn^{-r_{\mathrm{up}}}(\log n)^p,$$

where r_{up} is the feasible upper exponent supplied by Definition 7 for the above q_0 and $\bar{u}$.

Theorem 4 is the promised conditional achievability statement for the plain clipped cross-fitted AIPW empirical welfare rule. Its uniformity is over the domain specified in Definition 11 and the theorem's displayed hypotheses. Assumption 11–Assumption 16 are explicit inputs to this domain. The discussion section compares the strict-gap branch with the observed-law converse.

6 Discussion, Limitations, and Open Question

The preceding sections separate two complementary objects. Theorem 3 gives an observed-law minimax lower bound over the class $\mathcal{P}_{\alpha,\gamma}$ in Definition 4. Theorem 4 gives a conditional upper bound for the clipped cross-fitted AIPW empirical welfare rule over the side-condition domain in Definition 11. Thus the exponent $r_\star(\alpha, \gamma)$ is a converse benchmark for the law class, whereas r_{up} is the exponent delivered by one feasible procedure under explicit nuisance-rate, cross-fitting, boundedness, and localized empirical-process conditions.

This distinction affects the displayed exponent in regimes where the nuisance and clipping balance binds. In the nonbinding branch of Definition 7, the optimized nuisance exponent is at least the lower-bound exponent, so the conditional upper bound has the same polynomial exponent as the observed-law lower bound, up to logarithmic factors. In the strict-gap branch, the feasible exponent is smaller because one of the terms in the master bound of Lemma 8 controls the rate. The two conclusions are consistent: the lower bound ranges over all measurable policy estimators, while the upper bound analyzes a particular clipped AIPW ERM with estimated cross-fitted nuisance functions.

A statistical source represented in the bound is the weak-overlap mechanism emphasized in the limited-overlap literature [Li et al., 2016, D'Amour et al., 2017, Ben-Michael and Keele, 2022, Hill and Chaudhuri, 2024, Susmann et al., 2025]. The lower-bound alternatives make the informative treatment arm rare precisely on the small-contrast block that determines regret. The observed-law minimax lower bound shows that the law class has a testing difficulty encoded in $r_\star(\alpha, \gamma)$. A feasible clipped-score procedure must also estimate the nuisance functions and choose a clipping level that controls inverse-propensity variability. When nuisance learning is slow in the weak arm, the product remainder and localized drift terms can dominate the purely testing-driven benchmark. This connects to recent weak-overlap and clipping analyses of how overlap, nuisance estimation, and localization interact [Zhao et al., 2023, Liu et al., 2026].

Definition 12. [`oeq:feasible-tight`] (Feasible Tightness Question) *In the strict-gap branch of the conditional feasible achievability bound for the regime-indexed risk U_n, it is open whether a genuinely feasible estimator using estimated cross-fitted nuisance functions can attain the converse exponent $r_\star(\alpha, \gamma)$, or whether weak-arm nuisance learning imposes the slower conditional exponent derived by the feasible upper bound.*

Definition 12 is intentionally stated as an open question. It asks whether the gap in the strict-gap branch is an artifact of the plain clipped AIPW ERM and its analysis, or whether it reflects an intrinsic cost of learning the nuisance functions in the same weak-overlap region that drives the lower-bound construction. Resolving this would require either a feasible procedure whose risk reaches $r_\star(\alpha, \gamma)$ under comparable nuisance-side conditions, or a converse argument that incorporates nuisance learning into the lower-bound experiment.

Theorem 4 controls $U_n(\alpha, \gamma, a, c; \widehat{\eta})$ for the specified clipped cross-fitted rule and the stated side-condition domain. The broader feasible minimax problem also includes other estimators, adaptive clipping schemes, and procedures that exploit additional structure in the nuisance functions. Theorem 3 supplies the observed-law converse free of nuisance-estimation restrictions. The strict-gap branch therefore poses a genuine feasible minimax question beyond the completed observed-law lower-bound calculation.

Appendices

A Algebra, Testing, and Witness-Law Details

This appendix collects the algebraic details behind the lower-bound construction in Theorem 3. The purpose is only to connect the calibration, membership, divergence, separation, and testing ingredients used by Theorem 3. Throughout, the rate statement is a lower-bound statement over the observed-law class in Definition 4.

The exponent calculation begins with the overlap envelope in Proposition 1. For a local contrast scale h, an active block of mass proportional to h^α, and weak-arm probability proportional to h^β, the overlap-decay condition can be tested on the tight admissible window

$$v = h^\beta, \qquad u = h^{\beta/\gamma}.$$

On that window,

$$u^\alpha v^{1/\gamma} = h^{\alpha\beta/\gamma} h^{\beta/\gamma} = h^{(\alpha+1)\beta/\gamma}.$$

Since $0 < h < 1$, the envelope $u^\alpha v^{1/\gamma}$ is at least of order h^α precisely when

$$\frac{(\alpha+1)\beta}{\gamma} \leq \alpha, \qquad \text{or equivalently} \qquad \beta \leq \frac{\alpha\gamma}{\alpha+1}.$$

This is the calibration recorded in Proposition 1. Within the tight-window algebra used by the witness construction, it identifies the weakest admissible treatment-arm probability on the active block, and therefore explains why the denominator in Definition 3 contains the additional term $\beta_{\alpha,\gamma}$ when $\gamma > 0$.

The witness laws in Definition 6 are designed so that this calibration is binding. The covariate law is Lebesgue measure on $[0, 1]$, the active block has mass $P_X(B_n) = c_B h_n^\alpha$, and the logging propensity on that block is q_n, equal to $1/4$ in the strict-overlap endpoint and to $h_n^{\beta_{\alpha,\gamma}}$ in the weak-overlap case. Off the block, both laws have the same strictly positive contrast τ_0. On the block, they differ only by the sign of the treated-arm conditional mean, so that the two oracle rules are $x \mapsto 1$ and $x \mapsto \mathbf{1}\{x \notin B_n\}$, as stated in Lemma 1. The membership lemma is also where the availability of these two rules in the policy class is recorded for the two-action reduction.

The chi-square calculation in Lemma 2 has only one contributing cell. The laws coincide outside B_n, and they also coincide on the untreated part of B_n. Conditional on $X \in B_n$ and $A = 1$, the two outcome distributions are Bernoulli laws on $\{-1, 1\}$ with success probabilities $(1 + h_n)/2$ and $(1 - h_n)/2$. Their one-cell chi-square divergence is of order h_n^2. Multiplying by the probability of entering the cell gives

$$\chi^2(P_{n,+} \,\|\, P_{n,-}) \lesssim P_X(B_n) q_n h_n^2 \asymp h_n^{\alpha+\beta_{\alpha,\gamma}+2}.$$

With

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

the product divergence remains bounded because

$$n h_n^{2+\alpha+\beta_{\alpha,\gamma}} \asymp 1.$$

Equivalently, the product formula

$$1 + \chi^2(P_{n,+}^{\otimes n} \,\|\, P_{n,-}^{\otimes n}) = \big(1 + \chi^2(P_{n,+} \,\|\, P_{n,-})\big)^n$$

is bounded by a constant, as in Lemma 2.

The regret calculation in Lemma 3 uses the fact that the two oracle assignments disagree exactly on B_n. Any single deterministic policy must assign at least half of the block incorrectly for one of the two laws, up to the constant convention in the lemma. On that law, the welfare identity in Theorem 1 gives a pointwise regret weight h_n on the misclassified part of B_n. Thus the loss is bounded below by a constant multiple of

$$P_X(B_n) h_n \asymp h_n^{1+\alpha}.$$

Substituting the calibrated bandwidth gives

$$h_n^{1+\alpha} = n^{-(1+\alpha)/(2+\alpha+\beta_{\alpha,\gamma})} = n^{-r_\star(\alpha,\gamma)}.$$

The final step is the two-point testing reduction. Lemma 4 is the chi-square version of Le Cam's method [Le Cam, 1986, Tsybakov, 2009]: bounded product divergence prevents every test from identifying the sign with vanishing total error. Applying that testing bound to the two witness laws converts the regret separation into the minimax lower bound in Theorem 3. Its exponent characterizes the converse construction, while feasible policy-learning upper rates are developed through a separate achievability analysis.

B Empirical-Process and Cross-Fitting Details

This appendix collects the empirical-process inputs used by the conditional feasible upper bound in Theorem 4. The statements translate the localized offset-envelope side condition in Assumption 16 into the cross-fitted and self-bounding forms used by Lemma 8; the localized VC envelope condition in Assumption 11 enters the conditional domain directly. Fixed K, balanced folds, and the displayed empirical-process envelopes thereby define the conditional feasible-analysis domain.

The first auxiliary statement is the self-bounding step that converts an offset empirical-process control into an expected regret bound. It is the margin-localized analogue of the standard offset argument in excess-risk analysis for VC classes [Audibert and Tsybakov, 2007, Massart and Nédélec, 2006, Tsybakov, 2009].

Lemma 10. [`lem:localized-vc-self-bound`] (Localized Self Bound) *Let*

$$\rho_n = \left(\frac{B^2}{n}\right)^{A_\alpha} (\log n)^p, \qquad A_\alpha = \frac{1+\alpha}{2+\alpha}, \qquad B \geq 1, \qquad p \geq 0,$$

and let $C > 0$ be the constant of the cross-fit localized offset bound, so that the offset positive part of the centered Π-indexed process z_π is controlled at scale $C\rho_n$. Assume the sample size is large enough that

$$\frac{1}{n} \leq C\rho_n.$$

Then every Π-indexed estimator $\widetilde{\pi}$ satisfying

$$R_P(\widetilde{\pi}) \leq 2|z_{\widetilde{\pi}}| + \delta_n, \qquad \delta_n \geq 0,$$

also satisfies

$$\mathbb{E}_P R_P(\widetilde{\pi}) \leq \tfrac{8}{3}\,(C\rho_n + \delta_n).$$

If, in addition, $\delta_n \leq 1/n$, then—using $1/n \leq C\rho_n$—the δ_n term is absorbed into the leading rate and

$$\mathbb{E}_P R_P(\widetilde{\pi}) \leq \tfrac{8}{3}\,C\,\rho_n.$$

Proof of Lemma 10. Write

$$\rho = C\rho_n = C\left(\frac{B^2}{n}\right)^{A_\alpha} (\log n)^p$$

for the offset scale supplied to the lemma, so that the hypothesis on the centered Π-indexed process z_π reads

$$\mathbb{E}_P\,\Omega_{\text{off}} \leq \rho, \qquad \Omega_{\text{off}} = \sup_{\pi\in\Pi}\left\{2|z_\pi| - \frac{R_P(\pi)}{4}\right\}_+,$$

and the large-n hypothesis reads $1/n \leq \rho$. In particular $\rho > 0$.

Step 1: a samplewise bound. Fix a realization of the sample. Since $\widetilde{\pi} \in \Pi$, the value of the offset functional at $\pi = \widetilde{\pi}$ is one of the terms in the supremum defining Ω_{off}, so

$$\left\{2|z_{\widetilde{\pi}}| - \frac{R_P(\widetilde{\pi})}{4}\right\}_+ \leq \Omega_{\text{off}}.$$

Put $R = R_P(\widetilde{\pi})$, $z = z_{\widetilde{\pi}}$, and $\Delta = 2|z| - R/4$, so that $2|z| = \Delta + R/4$. Substituting this into the assumed selection inequality $R \leq 2|z| + \delta_n$ gives

$$R \leq \Delta + \frac{R}{4} + \delta_n, \qquad \text{that is,} \qquad \frac{3}{4}R \leq \Delta + \delta_n.$$

If $\Delta \geq 0$, then $\{\Delta\}_+ = \Delta$ and multiplying by $4/3$ gives $R \leq \frac{4}{3}\Delta + \frac{4}{3}\delta_n = \frac{4}{3}\{\Delta\}_+ + \frac{4}{3}\delta_n$. If $\Delta < 0$, then $\{\Delta\}_+ = 0$ and $\frac{3}{4}R \leq \Delta + \delta_n \leq \delta_n$, so $R \leq \frac{4}{3}\delta_n = \frac{4}{3}\{\Delta\}_+ + \frac{4}{3}\delta_n$. In both cases, using the first display of this step,

$$R_P(\widetilde{\pi}) \leq \frac{4}{3}\,\{\Delta\}_+ + \frac{4}{3}\,\delta_n \leq \frac{4}{3}\,\Omega_{\text{off}} + \frac{4}{3}\,\delta_n.$$

Step 2: integrate. Both sides are integrable by the stated regularity, so integrating the samplewise bound and using $\mathbb{E}_P\Omega_{\text{off}} \leq \rho$ gives

$$\mathbb{E}_P R_P(\widetilde{\pi}) \leq \frac{4}{3}\rho + \frac{4}{3}\delta_n.$$

Since $\rho + \delta_n \geq 0$, we may enlarge the factor $4/3$ to $8/3$, obtaining

$$\mathbb{E}_P R_P(\widetilde{\pi}) \leq \frac{8}{3}(\rho + \delta_n) = \frac{8}{3}\left(C\rho_n + \delta_n\right),$$

which is the asserted bound.

Step 3: absorbing the slack. If in addition $\delta_n \leq 1/n$, then the assumed large-n comparison $1/n \leq C\rho_n = \rho$ gives $\delta_n \leq \rho$, and the bound of Step 2 becomes

$$\mathbb{E}_P R_P(\widetilde{\pi}) \leq \frac{4}{3}\rho + \frac{4}{3}\delta_n \leq \frac{4}{3}\rho + \frac{4}{3}\rho = \frac{8}{3}\rho = \frac{8}{3}\,C\,\rho_n,$$

so the δ_n term is absorbed into the localized rate. □

Lemma 10 is the device that absorbs the stochastic term appearing after the ERM basic inequality. The $1/n$ slack in Definition 10 is negligible relative to the localized rate ρ_n, so the near-maximization error preserves the polynomial exponent.

The final empirical-process input is the offset version needed by the self-bound. It is the cross-fitted counterpart of the localized offset-envelope condition in Assumption 16.

Lemma 11. [`lem:crossfit-localized-offset-control`] (Cross-Fit Offset Control) *Let $G_{\text{cf}}(\pi)$ be the pooled cross-fit centered process formed from foldwise increments $g_{k,\pi}(O)$, indexed by $\pi \in \Pi$.*

Assume:

- ***(Data and folds.)*** *The observations are i.i.d. as in Assumption 1, the observed law is well formed, and the fixed balanced cross-fitting scheme of Assumption 15 is used.*
- ***(Policies and localization.)*** *The policy class satisfies Assumption 4, the margin and zero-effect conditions of Assumption 6 and Assumption 7 hold, and D_π is the disagreement set in Definition 2.*
- ***(Localized offset envelope.)*** *The uniform class-level localized offset-envelope condition of Assumption 16 holds for Π and exponent α.*
- ***(Outcome.)*** *The bounded-outcome condition of Assumption 2 holds.*
- ***(Foldwise increments.)*** *For every sample size $n \geq 1$, every $B \geq 0$, and every foldwise increment family $g_{k,\pi}$, each fold map is policy-compatible, each $g_{k,\pi}$ is measurable for*

$\pi \in \Pi$, the pooled offset supremum and the corresponding foldwise offset suprema are integrable, and

$$|g_{k,\pi}(O)| \le B, \qquad \int g_{k,\pi}(O)^2 \, dP(O) \le B^2 P_X(D_\pi)$$

for every fold k, every $\pi \in \Pi$, and every observation O.

Then there exist constants $C > 0$ and $p \ge 0$ such that

$$E_P \sup_{\pi \in \Pi} \left\{ 2|G_{\rm cf}(\pi)| - \frac{R_P(\pi)}{4} \right\}_+ \le C \left(\frac{B^2}{n} \right)^{(1+\alpha)/(2+\alpha)} (\log n)^p.$$

Proof of Lemma 11. Write $m_k = |I_k|$ for the size of fold k, $w_k = m_k/n$ for its weight, and

$$A_\alpha = \frac{1+\alpha}{2+\alpha}.$$

The pooled cross-fit centered process is

$$G_{\rm cf}(\pi) = \frac{1}{n} \sum_{i=1}^n \left\{ g_{k(i),\pi}(O_i) - \int g_{k(i),\pi}(O) \, dP(O) \right\}.$$

For every fold, use the totalized foldwise centered process

$$G_k(\pi) = \begin{cases} \dfrac{1}{m_k} \displaystyle\sum_{i \in I_k} \left\{ g_{k,\pi}(O_i) - \int g_{k,\pi}(O) \, dP(O) \right\}, & m_k > 0, \\ 0, & m_k = 0. \end{cases}$$

With this convention, empty folds are included in fold sums and have zero weight. Grouping the pooled sum by fold gives the exact decomposition

$$G_{\rm cf}(\pi) = \sum_{k=1}^K w_k G_k(\pi), \qquad \sum_{k=1}^K w_k = 1, \qquad w_k \ge 0,$$

the weight identity because $I_1, \ldots, I_K$ partitions $\{1, \ldots, n\}$.

Apply Assumption 16 in its uniform class-level form: it supplies constants $C_0 > 0$ and $p \ge 0$, not depending on the law, and we set

$$C = C_0 K,$$

which is positive because Assumption 15 gives $K \ge 1$. Also, Assumption 6 gives $\alpha \ge 0$, hence

$$A_\alpha = \frac{1+\alpha}{2+\alpha} \le 1.$$

Step 1: a pointwise foldwise bound on the offset supremum. Fix a sample and a policy $\pi \in \Pi$, and write $R = R_P(\pi)$. By the triangle inequality and $\sum_k w_k = 1$,

$$2\,|G_{\rm cf}(\pi)| - \frac{R}{4} \le \sum_{k=1}^K w_k \left\{ 2|G_k(\pi)| - \frac{R}{4} \right\} \le \sum_{k=1}^K w_k \left\{ 2|G_k(\pi)| - \frac{R}{4} \right\}_+,$$

the second step because $w_k \geq 0$ and $z \leq \{z\}_+$. The right-hand side is a nonnegative sum, so it also dominates 0, and therefore dominates the positive part on the left:

$$\left\{2\,|G_{\mathrm{cf}}(\pi)| - \frac{R_P(\pi)}{4}\right\}_+ \leq \sum_{k=1}^K w_k \left\{2|G_k(\pi)| - \frac{R_P(\pi)}{4}\right\}_+ .$$

The regret is bounded below by a fixed constant on all of Π: the setup identity $\tau_P = \mu_1 - \mu_0$ and Assumption 2 give $|\tau_P(x)| \leq 2$ for every x, every $\pi \in \Pi$ is measurable by Assumption 4, and the oracle $\pi_P^\star = \mathbf{1}\{\tau_P \geq 0\}$ is measurable because τ_P is; hence

$$|V_P(\pi)| = \left|\int \mathbf{1}\{\pi(x) = 1\}\tau_P(x)\,dP_X(x)\right| \leq 2 \qquad \text{for } \pi \in \Pi \text{ and for } \pi = \pi_P^\star,$$

so $R_P(\pi) = V_P(\pi_P^\star) - V_P(\pi) \geq -2 - 2 = -4$. Each foldwise offset term is therefore bounded above uniformly over $\pi \in \Pi$: the envelope hypothesis gives $|G_k(\pi)| \leq 2B$, with the totalized definition giving the same bound when $m_k = 0$, and $R_P(\pi) \geq -4$. Thus each foldwise supremum

$$\Omega_k = \sup_{\pi\in\Pi} \left\{2|G_k(\pi)| - \frac{R_P(\pi)}{4}\right\}_+$$

is finite, and taking the supremum over $\pi \in \Pi$ in the previous display yields the pointwise-in-sample bound

$$\sup_{\pi\in\Pi} \left\{2\,|G_{\mathrm{cf}}(\pi)| - \frac{R_P(\pi)}{4}\right\}_+ \leq \sum_{k=1}^K w_k\,\Omega_k.$$

Step 2: integrate and reduce each nonempty fold to an i.i.d. sample of its own size. The stated regularity hypotheses give integrability of the pooled offset supremum and of the foldwise offset suprema on their fold product spaces. The coordinate projection onto fold k is measure preserving, so the corresponding full-sample foldwise supremum is integrable. Integrating the last display over the product law and exchanging the finite sum with the integral gives

$$\mathbb{E}_P \sup_{\pi\in\Pi} \left\{2\,|G_{\mathrm{cf}}(\pi)| - \frac{R_P(\pi)}{4}\right\}_+ \leq \sum_{k=1}^K w_k\,\mathbb{E}_P\Omega_k.$$

If $m_k = 0$, then $w_k = 0$ and that fold contributes nothing. If $m_k > 0$, the same measure-preserving projection identifies $\mathbb{E}_P\Omega_k$ with the offset-supremum expectation for an i.i.d. sample of size m_k. For such a fold, $g_{k,\pi}$ is policy-compatible with envelope $|g_{k,\pi}(O)| \leq B$ and second moment

$$\int g_{k,\pi}(O)^2\,dP(O) \leq B^2 P_X(D_\pi),$$

so Assumption 16 applies at sample size m_k and gives

$$\mathbb{E}_P\Omega_k \leq C_0 \left(\frac{B^2}{m_k}\right)^{A_\alpha} (\log m_k)^p.$$

Step 3: compare each weighted fold rate with the target n-rate. Put $x = m_k/n \in (0, 1]$. Then $B^2/m_k = (B^2/n)/x$, so

$$w_k \left(\frac{B^2}{m_k}\right)^{A_\alpha} = x \cdot \frac{(B^2/n)^{A_\alpha}}{x^{A_\alpha}} = \left(\frac{B^2}{n}\right)^{A_\alpha} x^{1-A_\alpha} \leq \left(\frac{B^2}{n}\right)^{A_\alpha},$$

because $0 < x \le 1$ and $1 - A_\alpha \ge 0$. Moreover $m_k \le n$ and $p \ge 0$ give

$$(\log m_k)^p \le (\log n)^p.$$

Combining the last three displays, every fold contributes at most

$$w_k \, \mathbb{E}_P \Omega_k \le C_0 \left(\frac{B^2}{n} \right)^{A_\alpha} (\log n)^p.$$

Conclusion. Summing over the K folds,

$$\mathbb{E}_P \sup_{\pi \in \Pi} \left\{ 2|G_{\mathrm{cf}}(\pi)| - \frac{R_P(\pi)}{4} \right\}_+ \le C_0 K \left(\frac{B^2}{n} \right)^{A_\alpha} (\log n)^p = C \left(\frac{B^2}{n} \right)^{(1+\alpha)/(2+\alpha)} (\log n)^p,$$

which is the asserted bound. □

Lemma 11 supplies the offset positive-part bound required by Lemma 10. In the feasible clipped-AIPW analysis, the envelope B is later instantiated at the clipped-score scale from Lemma 6; the resulting term is the empirical-process component $(nq^2)^{-A_\alpha}$ in Lemma 8. Together, Lemma 10–Lemma 11 provide only the empirical-process part of the conditional upper bound; the drift and clipping bias are handled separately in Appendix C.

C Clipped-Score Drift and Bias Localization

This appendix records the deterministic bias calculations used by the clipped-score analysis in Theorem 4. The key point is that clipping a doubly robust score changes its conditional mean unless the clipped plug-in propensity agrees with the observed-law propensity, or the outcome-regression errors vanish at the same covariate value. Thus the usual orthogonality logic for AIPW scores [Robins et al., 1994, Hahn, 1998, Hirano et al., 2003, Chernozhukov et al., 2018] must be combined with an explicit localization argument near the clipped propensity region.

Lemma 12. [`lem:clip-bias`] (Clipped Score Drift) *Under Assumption 2, Assumption 3, and Assumption 13, and with the clipped propensity and clipped AIPW score of Definition 8 and Definition 9, let $\bar\eta = (\bar\mu_0, \bar\mu_1, \bar e)$ be any measurable plug-in nuisance triple. Define*

$$\bar e_q(x) = \min\{1-q, \max\{q, \bar e(x)\}\}, \qquad \Delta_a(x) = \bar\mu_a(x) - \mu_a(x), \qquad b_q(x) = (\bar e_q(x) - e_P(x)) \left\{ \frac{\Delta_1(x)}{\bar e_q(x)} + \frac{\Delta_0(x)}{1 - \bar e_q(x)} \right\}.$$

Then the clipped score reproduces the contrast up to the drift b_q in the tested (population) sense: for every bounded measurable φ,

$$\mathbb{E}_P[\varphi(X)\, \Gamma_q(O; \bar\eta)] - \mathbb{E}_P[\varphi(X)\, \tau_P(X)] = \mathbb{E}_P[\varphi(X)\, b_q(X)],$$

equivalently $\mathbb{E}_P[\Gamma_q(O; \bar\eta) \mid X = x] - \tau_P(x) = b_q(x)$ for P_X-almost every x. In particular, the drift vanishes whenever $\bar e_q(x) = e_P(x)$ or $\Delta_0(x) = \Delta_1(x) = 0$, since either makes a factor of $b_q(x)$ vanish; these conditions are sufficient but not necessary, as the bracketed term $\Delta_1(x)/\bar e_q(x) + \Delta_0(x)/(1 - \bar e_q(x))$ may itself vanish. Cancellation does not, however, follow merely from $p_P(x) > q$: there exist laws and covariate values with $p_P(x) > q$ at which the drift is nonzero.

Proof of Lemma 12. Fix a bounded measurable covariate test function φ, and write $\bar{e}_q(x) = e_q(x;\bar{\eta})$ for the clipped plug-in propensity of Definition 8, whose clipping level is fixed in $q \in (0,1/2]$. Since $q \le 1/2$, we have $q \le 1-q$, and the definition $\bar{e}_q(x) = \min\{1-q, \max\{q, \bar{e}(x)\}\}$ therefore gives the two-sided bound

$$q \le \bar{e}_q(x) \le 1-q \qquad \text{for every } x,$$

so both denominators are bounded away from zero: $\bar{e}_q(x) \ge q > 0$ and $1-\bar{e}_q(x) \ge q > 0$. Together with Assumption 2 (which gives $|Y| \le 1$ P-a.s. and $|\mu_a| \le 1$) and the fixed plug-in regularity for this drift calculation (measurability of $\bar{e}$, measurability of $\bar{\mu}_0, \bar{\mu}_1$, and uniform boundedness of $\bar{\mu}_0, \bar{\mu}_1$), every integrand appearing below is measurable and bounded, hence P- and P_X-integrable; this is the only role of the boundedness hypotheses.

Step 1: expand the score and pass to covariate integrals. Multiplying the score of Definition 9 by $\varphi(X)$ and splitting it into five terms,

$$\varphi(X)\Gamma_q(O;\bar{\eta}) = \varphi(X)\{\bar{\mu}_1(X)-\bar{\mu}_0(X)\}+\frac{AY\varphi(X)}{\bar{e}_q(X)}-\frac{A\bar{\mu}_1(X)\varphi(X)}{\bar{e}_q(X)}-\frac{(1-A)Y\varphi(X)}{1-\bar{e}_q(X)}+\frac{(1-A)\bar{\mu}_0(X)\varphi(X)}{1-\bar{e}_q(X)}.$$

Each of the five terms is integrable, so the integral of the sum splits into the sum of the integrals. The defining conditional-mean relations of the observed law, namely $e_P(x) = P(A=1 \mid X=x)$ and $\mu_a(x) = E_P[Y \mid A=a, X=x]$, state that for every bounded measurable ψ,

$$E_P[\psi(X)] = \int \psi\, dP_X, \qquad E_P[A\,\psi(X)] = \int e_P\psi\, dP_X, \qquad E_P[(1-A)\,\psi(X)] = \int (1-e_P)\psi\, dP_X,$$

$$E_P[A\,Y\,\psi(X)] = \int e_P\mu_1\psi\, dP_X, \qquad E_P[(1-A)\,Y\,\psi(X)] = \int (1-e_P)\mu_0\psi\, dP_X.$$

Applying these with $\psi = \varphi/\bar{e}_q$, $\psi = \bar{\mu}_1\varphi/\bar{e}_q$, $\psi = \varphi/(1-\bar{e}_q)$, and $\psi = \bar{\mu}_0\varphi/(1-\bar{e}_q)$ – all bounded and measurable by the previous paragraph – converts the five terms into P_X-integrals:

$$\begin{aligned}\int \varphi(X)\Gamma_q(O;\bar{\eta})\, dP = &\int \varphi(\bar{\mu}_1-\bar{\mu}_0)\, dP_X \\ &+ \int e_P\mu_1 \frac{\varphi}{\bar{e}_q}\, dP_X - \int e_P\bar{\mu}_1 \frac{\varphi}{\bar{e}_q}\, dP_X \\ &- \int (1-e_P)\mu_0 \frac{\varphi}{1-\bar{e}_q}\, dP_X + \int (1-e_P)\bar{\mu}_0 \frac{\varphi}{1-\bar{e}_q}\, dP_X.\end{aligned}$$

Step 2: collect the integrand pointwise. Fix x. Using $\bar{\mu}_a = \mu_a + \Delta_a$ and $\tau_P = \mu_1 - \mu_0$, the two treated terms combine as

$$\Delta_1(x) - \frac{e_P(x)\Delta_1(x)}{\bar{e}_q(x)} = \frac{\{\bar{e}_q(x)-e_P(x)\}\Delta_1(x)}{\bar{e}_q(x)},$$

and the two control terms combine as

$$-\Delta_0(x) + \frac{\{1-e_P(x)\}\Delta_0(x)}{1-\bar{e}_q(x)} = \frac{\{\bar{e}_q(x)-e_P(x)\}\Delta_0(x)}{1-\bar{e}_q(x)}.$$

Both rearrangements are legitimate because $\bar{e}_q(x)$ and $1-\bar{e}_q(x)$ are nonzero. Adding them to $\tau_P(x)$ gives the pointwise identity

$$\varphi(x)(\bar{\mu}_1-\bar{\mu}_0)(x)+\frac{e_P\mu_1\varphi}{\bar{e}_q}(x)-\frac{e_P\bar{\mu}_1\varphi}{\bar{e}_q}(x)-\frac{(1-e_P)\mu_0\varphi}{1-\bar{e}_q}(x)+\frac{(1-e_P)\bar{\mu}_0\varphi}{1-\bar{e}_q}(x) = \varphi(x)\{\tau_P(x)+b_q(x)\},$$

with b_q exactly the function displayed in the statement.

Step 3: integrate. Both $\varphi\tau_P$ and φb_q are bounded measurable, hence P_X-integrable, so the integral of the right-hand side splits and Step 1 becomes

$$\int \varphi(X)\Gamma_q(O;\bar{\eta})\,dP - \int \varphi(x)\tau_P(x)\,dP_X(x) = \int \varphi(x)\,\{\bar{e}_q(x) - e_P(x)\}\left\{\frac{\Delta_1(x)}{\bar{e}_q(x)} + \frac{\Delta_0(x)}{1-\bar{e}_q(x)}\right\} dP_X(x) = \int \varphi\, b_q\, dP_X$$

Since this holds for every bounded measurable φ, it is precisely the asserted conditional-mean drift identity $E_P[\Gamma_q(O;\bar{\eta}) \mid X = x] - \tau_P(x) = b_q(x)$ for P_X-almost every x.

Step 4: the cancellation mechanisms. The closed form

$$b_q(x) = \{\bar{e}_q(x) - e_P(x)\}\left\{\frac{\Delta_1(x)}{\bar{e}_q(x)} + \frac{\Delta_0(x)}{1-\bar{e}_q(x)}\right\}$$

is a product of two factors. If $\bar{e}_q(x) = e_P(x)$, the first factor is 0; if $\Delta_0(x) = \Delta_1(x) = 0$, the second factor is 0. In either case $b_q(x) = 0$.

Step 5: no cancellation from $p_P(x) > q$ alone. It remains to show that the stated assumptions do not permit replacing this algebra by a cancellation claim on the region $\{p_P > q\}$. Take the covariate space $\mathbb{R}$, and let P' be the law that puts covariate mass 1 at $x = 0$, assigns treatment by a fair coin, and returns $Y = 0$ always; explicitly, P' is the two-atom law

$$P' = \tfrac{1}{2}\,\delta_{(0,1,0)} + \tfrac{1}{2}\,\delta_{(0,0,0)}, \qquad P'_X = \delta_0, \qquad e_{P'} \equiv \tfrac{1}{2}, \qquad \mu_0 \equiv \mu_1 \equiv 0, \qquad \tau_{P'} \equiv 0.$$

This P' is a well-formed observed law with bounded outcomes and satisfies positivity, since $0 < 1/2 < 1$. Take $q' = 1/4$ and the plug-in triple $\bar{\mu}_0 \equiv 0$, $\bar{\mu}_1 \equiv 1$, $\bar{e} \equiv 0$, and evaluate at $x = 0$. The overlap score is

$$p_{P'}(0) = \min\{e_{P'}(0), 1 - e_{P'}(0)\} = \tfrac{1}{2} > \tfrac{1}{4} = q',$$

so $x = 0$ lies strictly inside the region where clipping is claimed not to matter. Yet the clipped plug-in propensity is $\bar{e}_{q'}(0) = \min\{3/4, \max\{1/4, 0\}\} = 1/4$, while $\Delta_1(0) = 1 - 0 = 1$ and $\Delta_0(0) = 0 - 0 = 0$, so the drift formula gives

$$b_{q'}(0) = \left(\tfrac{1}{4} - \tfrac{1}{2}\right)\left\{\frac{1}{1/4} + \frac{0}{3/4}\right\} = -\tfrac{1}{4}\cdot 4 = -1 \neq 0.$$

Hence $p_P(x) > q$ alone does not force the drift to vanish, and no additive clipped-region cancellation bound follows from the stated boundedness, positivity, and observed-law conditions alone. $\square$

Lemma 12 is the exact drift identity behind the bias term in Lemma 7. It separates the propensity discrepancy from the regression errors and shows that correct handling of clipping requires the full product identity. In particular, stronger cancellation requires assumptions beyond the stated boundedness and positivity conditions.

The remaining issue is to control where clipping can matter for regret. Under the joint overlap-decay condition, the relevant clipped region becomes small after intersection with the welfare-relevant disagreement and small-contrast regions. The next lemma gives that localization.

Lemma 13. [`lem:clipped-region-localization`] (Clipped Region Localization) *Suppose that the following conditions hold:*

- ***Positive weak-overlap exponent.*** *$\gamma > 0$.*

- ***Well-formed law.*** *The observed law P is well formed, with P_X a probability covariate marginal, $\tau_P = \mu_1 - \mu_0$, measurable nuisance functionals, and $p_P(x) = \min\{e_P(x), 1 - e_P(x)\}$.*

- ***Overlap decay.*** *Assumption 9 holds with constants $C_o, c_o, \alpha, \gamma, u_0$.*

- ***Zero-effect convention.*** *Assumption 7 holds for the policy class Π.*

- ***Bounded outcomes.*** *Assumption 2 holds.*

- ***Measurable policies.*** *Every $\pi \in \Pi$ is measurable.*

- ***Disagreement sets.*** *D_π is defined as in Definition 2.*

Then $\max\{C_o, 1\} > 0$, and every policy $\pi \in \Pi$ with regret $r = R_P(\pi)$ satisfies

$$P_X(D_\pi \cap \{x : p_P(x) \le q\}) \le \max\{C_o, 1\} u^\alpha q^{1/\gamma} + \frac{r}{u}$$

whenever $0 < u \le u_0$ and $0 < q \le c_o u^\gamma$.

Proof of Lemma 13. Since C_o is a real constant, $\max\{C_o, 1\} \ge 1 > 0$, which is the first assertion.

Fix a policy $\pi \in \Pi$, let $D = D_\pi$ and $r = R_P(\pi)$, and fix u, q with $0 < u \le u_0$, $q > 0$, and $q \le c_o u^\gamma$. Put

$$T = D \cap \{x : p_P(x) \le q\},$$

the set whose P_X-mass is to be bounded. We cover T by the three sets

$$Z = \{x : \tau_P(x) = 0,\ \pi(x) \ne \pi_P^\star(x)\},$$

$$S = \{x : p_P(x) \le q,\ 0 < |\tau_P(x)|,\ |\tau_P(x)| \le u\},$$

and

$$B = D \cap \{x : u < |\tau_P(x)|\}.$$

Indeed, let $x \in T$. If $\tau_P(x) = 0$, then, since $x \in D$ means $\pi(x) \ne \pi_P^\star(x)$, we get $x \in Z$. If $\tau_P(x) \ne 0$ and $|\tau_P(x)| \le u$, then, since $p_P(x) \le q$, we get $x \in S$. Otherwise $u < |\tau_P(x)|$, and since $x \in D$, we get $x \in B$. Thus

$$T \subseteq Z \cup S \cup B.$$

All three sets are measurable, because τ_P, p_P, and π are measurable.

The zero-contrast piece is null. Assumption 7 gives

$$P_X(Z) = 0.$$

In its first alternative $P_X\{\tau_P(X) = 0\} = 0$, so its subset Z is null by monotonicity. In its second alternative, applied to the present policy $\pi \in \Pi$, it says exactly that π agrees with $\pi_P^\star$ P_X-almost everywhere on $\{\tau_P = 0\}$, that is, $P_X\{x : \tau_P(x) = 0,\ \pi(x) \ne \pi_P^\star(x)\} = 0$.

The small-contrast weak-overlap piece. Assumption 9 applies with the present u and $v = q$, because $0 < u \le u_0$, $q > 0$, and $q \le c_o u^\gamma$. Since $\gamma > 0$, the convention $v^{1/\gamma} = 1$ is not in force and the assumption reads

$$P_X(S) = P_X\{p_P(X) \le q,\ 0 < |\tau_P(X)| \le u\} \le C_o u^\alpha q^{1/\gamma} \le \max\{C_o, 1\}\, u^\alpha q^{1/\gamma},$$

the last step because $C_o \le \max\{C_o, 1\}$ and $u^\alpha q^{1/\gamma} \ge 0$.

The large-contrast piece. Here we use the welfare regret identity of Theorem 1,

$$R_P(\pi) = \int_{\mathcal{X}} |\tau_P(x)| \, \mathbf{1}\{x \in D\} \, dP_X(x),$$

whose integrand is nonnegative and, on B, at least u (since $B \subseteq D$ and $|\tau_P| > u$ on B). Restricting the integral to B and then bounding the integrand below by the constant u gives

$$u \, P_X(B) \le \int_B |\tau_P(x)| \, \mathbf{1}\{x \in D\} \, dP_X(x) \le \int_{\mathcal{X}} |\tau_P(x)| \, \mathbf{1}\{x \in D\} \, dP_X(x) = R_P(\pi),$$

the middle inequality because the integrand is nonnegative. The integral is finite because Assumption 2 gives $|\tau_P| = |\mu_1 - \mu_0| \le 2$. Dividing by $u > 0$,

$$P_X(B) \le \frac{R_P(\pi)}{u} = \frac{r}{u}.$$

Conclusion. By monotonicity of P_X along $T \subseteq Z \cup S \cup B$, subadditivity, and $P_X(Z) = 0$,

$$P_X(T) \le P_X(Z) + P_X(S) + P_X(B) = P_X(S) + P_X(B) \le \max\{C_o, 1\} \, u^\alpha q^{1/\gamma} + \frac{r}{u},$$

which is the asserted bound. □

Lemma 13 is the measure bound used to localize the drift from Lemma 12. The first term is the overlap-decay contribution on the small-contrast window $\{0 < |\tau_P(X)| \le u\}$, while the second term is the regret-controlled contribution from disagreements outside that window. The admissibility condition $q \le c_o u^\gamma$ is the same one used in the feasible tuning schedule of Definition 7.

Together, Lemma 12 and Lemma 13 supply the deterministic bias inputs for Lemma 7. They bound the clipped-score drift by combining nuisance-rate control, overlap-decay localization, and the regret radius of the policy under consideration.

D Verification Note

This appendix records the scope of the formal layer underlying the statements in the paper. The object of analysis is the observed-law offline policy-learning experiment introduced in Assumption 1: one observation has law P on $\Omega = \mathcal{X} \times \{0, 1\} \times [-1, 1]$, and the regret target is the contrast-weighted welfare loss in Definition 1. The potential-outcome notation used in Assumption 2 fixes the outcome scale, but the lower-bound and upper-bound statements themselves are expressed through the observed law, the propensity score, the outcome regressions, the contrast, and the induced oracle policy.

The observed-law side of the analysis consists of the welfare identity in Theorem 1, the disagreement notation in Definition 2, the margin localization statement in Theorem 2, and the law class in Definition 4. The exponents used in the lower bound are those in Definition 3; in particular, the denominator $2 + \alpha + \beta_{\alpha,\gamma}$ and the lower-bound exponent $r_\star(\alpha, \gamma)$ are fixed there before the minimax statement is formed.

The lower-bound argument is represented by the two-point witness construction in Definition 6, the overlap-envelope calibration in Proposition 1, witness membership in Lemma 1, the product-divergence control in Lemma 2, the regret separation in Lemma 3, and the chi-square testing reduction in Lemma 4. These statements support the minimax lower bound in Theorem 3; the conditional upper analysis is represented by the separate clipped-AIPW chain below.

The feasible rule is represented separately. The clipping schedule and feasible exponent are defined in Definition 7; the clipped propensity and clipped AIPW score are defined in Definition 8 and Definition 9; and the cross-fitted empirical welfare rule is defined in Definition 10. The basic ERM inequality in Lemma 5, the clipped-score envelope in Lemma 6, the clipped-bias identity in Lemma 12, the drift localization in Lemma 7, and the master bound in Lemma 8 are the components used for the conditional upper bound in Theorem 4.

The upper-bound result has an explicit side-condition domain in addition to the observed-law class. This domain includes the nuisance-rate assumptions in Assumption 12 and Assumption 14, the bounded cross-fitted nuisance condition in Assumption 13, the fixed balanced fold condition in Assumption 15, and the localized empirical-process and offset-envelope conditions in Assumption 11 and Assumption 16. The process reductions in Lemma 10 and Lemma 11 are invoked within that conditional domain.

Accordingly, Theorem 4 is a procedure-specific conditional achievability statement for the clipped cross-fitted AIPW ERM under Definition 11. Definition 12 records the open strict-gap question: whether another genuinely feasible estimator can attain the converse exponent $r_\star(\alpha, \gamma)$, or whether the slower nuisance-limited exponent is intrinsic.

The theorem statements and proofs described above are machine-checked in Lean 4. The machine-checked scope covers the observed-law objects, exponent definitions, witness construction, lower-bound skeleton, clipped-score identities, ERM inequality, and conditional upper-bound dependencies. Identification primitives, nuisance-rate conditions, fixed-fold cross-fitting, localized VC envelope bounds, and offset-envelope controls enter as stated assumptions defining the conditional domain.

E Proofs of the main results

Proof of Theorem 1. Fix a well-formed observed law P satisfying Assumption 2, and fix $\pi \in \Pi$. By Assumption 4, π is measurable. Well-formedness gives that P_X is a probability measure, that τ_P is measurable, and that $\tau_P(x) = \mu_1(x) - \mu_0(x)$ for every x. The regression part of Assumption 2 gives $\mu_0(x), \mu_1(x) \in [-1, 1]$ for every x, and therefore

$$|\tau_P(x)| = |\mu_1(x) - \mu_0(x)| \le |\mu_1(x)| + |\mu_0(x)| \le 2.$$

The set $\{x : 0 \le \tau_P(x)\}$ is measurable because τ_P is measurable, and $\{x : \pi(x) = 1\}$ is measurable because π is measurable. Hence the functions

$$x \mapsto \begin{cases} \tau_P(x), & 0 \le \tau_P(x), \\ 0, & \text{otherwise}, \end{cases} \qquad x \mapsto \begin{cases} \tau_P(x), & \pi(x) = 1, \\ 0, & \text{otherwise}, \end{cases}$$

are measurable. Since P_X is a probability measure and $|\tau_P| \le 2$, both functions are integrable with respect to P_X, so the difference of the two welfare integrals may be written as one P_X-integral.

Writing Definition 1 in its covariate-marginal form,

$$V_P(\rho) = \int_{\mathcal{X}} \mathbf{1}\{\rho(x) = 1\} \tau_P(x)\, dP_X(x),$$

with $\pi_P^\star(x) = \mathbf{1}\{\tau_P(x) \geq 0\}$ and $R_P(\pi) = V_P(\pi_P^\star) - V_P(\pi)$, the preceding integrability gives

$$R_P(\pi) = \int_{\mathcal{X}} [\mathbf{1}\{0 \leq \tau_P(x)\}\tau_P(x) - \mathbf{1}\{\pi(x) = 1\}\tau_P(x)] \, dP_X(x).$$

Fix x. If $0 \leq \tau_P(x)$, then $\pi_P^\star(x) = 1$. When $\pi(x) = 1$, the two terms in the bracket are both $\tau_P(x)$, so the bracket and the disagreement indicator are both zero. When $\pi(x) \neq 1$, the bracket is $\tau_P(x) = |\tau_P(x)|$, and the disagreement indicator is one. If $\tau_P(x) < 0$, then $\pi_P^\star(x) = 0$. When $\pi(x) = 1$, the bracket is $-\tau_P(x) = |\tau_P(x)|$, and the disagreement indicator is one. When $\pi(x) \neq 1$, the two terms in the bracket are both zero, and the disagreement indicator is zero. Therefore, for every x,

$$\mathbf{1}\{0 \leq \tau_P(x)\}\tau_P(x) - \mathbf{1}\{\pi(x) = 1\}\tau_P(x) = |\tau_P(x)| \, \mathbf{1}\{\pi(x) \neq \pi_P^\star(x)\}.$$

Substituting this pointwise identity into the preceding integral yields

$$R_P(\pi) = \int_{\mathcal{X}} |\tau_P(x)| \, \mathbf{1}\{\pi(x) \neq \pi_P^\star(x)\} \, dP_X(x).$$

This is the $|\tau_P|$-weighted disagreement representation of regret with respect to the covariate marginal. □

Proof of Theorem 2. 1. *The constant.* Fix

$$C := |C_m| + 2 + u_0^{-\alpha}.$$

Assumption 6 gives $0 \leq \alpha$, $C_m > 0$, and $u_0 > 0$, so every summand is nonnegative and $C > 0$; moreover

$$C \geq C_m + 1 \qquad \text{and} \qquad C \geq u_0^{-\alpha}.$$

This constant depends only on C_m, u_0, and α. It remains to prove the displayed inequality for an arbitrary well-formed observed law P, policy class Π whose members are measurable, and every $\pi \in \Pi$.

2. *Two elementary facts.* Let $D_\pi = \{x : \pi(x) \neq \pi_P^\star(x)\}$ as in Definition 2. By Theorem 1,

$$R_P(\pi) = \int_{\mathcal{X}} |\tau_P(x)| \, \mathbf{1}_{D_\pi}(x) \, dP_X(x).$$

The integrand is nonnegative, so $R_P(\pi) \geq 0$. The standing well-formed-law conditions give $\tau_P = \mu_1 - \mu_0$, and Assumption 2 gives $|\mu_0| \leq 1$ and $|\mu_1| \leq 1$; hence $|\tau_P| \leq 2$, which supplies the integrability used in the identity. Also $P_X(D_\pi) \leq 1$, since well-formedness makes P_X a probability measure.

3. *The basic decomposition.* For every $u \in (0, u_0]$,

$$P_X(D_\pi) \leq C_m u^\alpha + \frac{R_P(\pi)}{u}.$$

Indeed, write

$$Z = \{x : \tau_P(x) = 0, \, x \in D_\pi\}, \qquad S_u = \{x : 0 < |\tau_P(x)| \leq u\}, \qquad B_u = D_\pi \cap \{x : u < |\tau_P(x)|\}.$$

Every $x \in D_\pi$ lies in one of these three sets, according to whether $\tau_P(x) = 0$, $0 < |\tau_P(x)| \leq u$, or $|\tau_P(x)| > u$; hence

$$D_\pi \subseteq Z \cup S_u \cup B_u.$$

Assumption 7 gives $P_X(Z) = 0$: in its first alternative $P_X\{\tau_P = 0\} = 0$, so its subset Z is null; in its second alternative, applied to $\pi \in \Pi$, the set Z of zero-contrast disagreements is null by hypothesis. Assumption 6, applied at the admissible window $u \leq u_0$, gives

$$P_X(S_u) \leq C_m u^\alpha.$$

Finally, on B_u the integrand in the regret identity is at least u, and the integrand is nonnegative, so restricting the integral to B_u gives

$$u\, P_X(B_u) \leq \int_{B_u} |\tau_P(x)|\, \mathbf{1}_{D_\pi}(x)\, dP_X(x) \leq R_P(\pi), \qquad \text{hence} \qquad P_X(B_u) \leq \frac{R_P(\pi)}{u}.$$

Monotonicity and subadditivity of P_X along the displayed inclusion now give the asserted decomposition. The required measurability comes from the well-formed-law measurability of τ_P together with the policy-measurability part of Assumption 4; boundedness and finiteness are supplied by the well-formed-law identity $\tau_P = \mu_1 - \mu_0$, Assumption 2, and the probability property of P_X.

4. *The case $\alpha = 0$.* Then $\alpha/(1+\alpha) = 0$ and $R_P(\pi)^0 = 1$, while $u_0^{-0} = 1$ gives $C = C_m + 3 \geq 1$. Hence

$$P_X(D_\pi) \leq 1 \leq C = C\, R_P(\pi)^{\alpha/(1+\alpha)}.$$

5. *The case $\alpha > 0$, $R_P(\pi) > 0$, small window.* Set

$$u = R_P(\pi)^{1/(1+\alpha)} > 0,$$

and suppose first $u \leq u_0$. Then the two exponent identities

$$u^\alpha = R_P(\pi)^{\alpha/(1+\alpha)}, \qquad \frac{R_P(\pi)}{u} = R_P(\pi)^{1-1/(1+\alpha)} = R_P(\pi)^{\alpha/(1+\alpha)}$$

hold, so Step 3 gives

$$P_X(D_\pi) \leq C_m u^\alpha + \frac{R_P(\pi)}{u} = (C_m + 1)\, R_P(\pi)^{\alpha/(1+\alpha)} \leq C\, R_P(\pi)^{\alpha/(1+\alpha)},$$

the last step by $C_m + 1 \leq C$ and $R_P(\pi)^{\alpha/(1+\alpha)} \geq 0$.

6. *The case $\alpha > 0$, $R_P(\pi) > 0$, large window.* Suppose instead $u > u_0$. Since $\alpha > 0$, raising to the power α is increasing on nonnegative arguments, so

$$u_0^\alpha \leq u^\alpha = R_P(\pi)^{\alpha/(1+\alpha)}.$$

Multiplying by $u_0^{-\alpha} > 0$ and using $u_0^{-\alpha} \leq C$,

$$1 = u_0^{-\alpha} u_0^\alpha \leq u_0^{-\alpha}\, R_P(\pi)^{\alpha/(1+\alpha)} \leq C\, R_P(\pi)^{\alpha/(1+\alpha)}.$$

Combining with $P_X(D_\pi) \leq 1$ from Step 2 gives the desired bound in this subcase.

7. *The case $\alpha > 0$, $R_P(\pi) = 0$.* Let $\varepsilon > 0$, and put

$$u = \min\left\{u_0, \left(\frac{\varepsilon}{C_m + 1}\right)^{1/\alpha}\right\}.$$

Then $0 < u \le u_0$, and monotonicity of positive powers gives

$$u^\alpha \le \frac{\varepsilon}{C_m + 1}.$$

Step 3 with $R_P(\pi) = 0$ therefore gives

$$P_X(D_\pi) \le C_m u^\alpha + \frac{0}{u} = C_m u^\alpha \le \frac{C_m}{C_m + 1}\,\varepsilon \le \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary and $P_X(D_\pi) \ge 0$, $P_X(D_\pi) = 0$. Also $\alpha/(1+\alpha) > 0$, so

$$R_P(\pi)^{\alpha/(1+\alpha)} = 0^{\alpha/(1+\alpha)} = 0,$$

and therefore

$$P_X(D_\pi) = 0 \le C\, R_P(\pi)^{\alpha/(1+\alpha)}.$$

□

Proof of Proposition 1. Throughout, the tight window is $v = h^b$ and $u = h^{b/\gamma}$ for a real exponent b, so that the envelope value in question is

$$u^\alpha v^{1/\gamma} = \left(h^{b/\gamma}\right)^\alpha \left(h^b\right)^{1/\gamma}.$$

For any real b, the power laws (legitimate since $h > 0$ and $\gamma > 0$) give

$$\left(h^{b/\gamma}\right)^\alpha \left(h^b\right)^{1/\gamma} = h^{(b/\gamma)\alpha}\, h^{b/\gamma} = h^{(b/\gamma)\alpha + b/\gamma} = h^{(\alpha+1)b/\gamma}.$$

Since $0 < h < 1$, the map $t \mapsto h^t$ is strictly order-reversing, so

$$h^{(\alpha+1)b/\gamma} \ge h^\alpha \quad\Longleftrightarrow\quad \frac{(\alpha+1)b}{\gamma} \le \alpha.$$

As $\alpha + 1 > 0$ and $\gamma > 0$, multiplying by $\gamma/(\alpha + 1) > 0$ turns the right-hand inequality into

$$b \le \frac{\alpha\gamma}{\alpha + 1} = \beta_{\alpha,\gamma},$$

with $\beta_{\alpha,\gamma}$ as in Definition 3. Applying the two displays with $b = \beta$ gives the asserted identity

$$u^\alpha v^{1/\gamma} = h^{(\alpha+1)\beta/\gamma}$$

and the asserted equivalence

$$u^\alpha v^{1/\gamma} \ge h^\alpha \quad\Longleftrightarrow\quad \beta \le \beta_{\alpha,\gamma}.$$

If $\beta = \beta_{\alpha,\gamma}$, the exponent evaluates exactly:

$$\frac{\alpha+1}{\gamma} \cdot \frac{\alpha\gamma}{\alpha+1} = \alpha, \qquad \text{so} \qquad u^{\alpha} v^{1/\gamma} = h^{\alpha},$$

which is the equality case in the power comparison. Finally, $\beta_{\alpha,\gamma} = \alpha\gamma/(\alpha+1) \geq 0$ because $\alpha \geq 0$, $\gamma > 0$, and $\alpha + 1 > 0$; and since b above was an arbitrary real, the same equivalence

$$\left(h^{\beta'/\gamma}\right)^{\alpha} \left(h^{\beta'}\right)^{1/\gamma} \geq h^{\alpha} \quad \Longleftrightarrow \quad \beta' \leq \beta_{\alpha,\gamma}$$

holds for every $\beta' \geq 0$. Thus the largest, and therefore least informative, weak-arm exponent allowed within this tight-window overlap-envelope calibration is precisely $\beta_{\alpha,\gamma}$. □

Proof of Lemma 1. Fix a witness sign $\sigma \in \{1, -1\}$. Write $h = h_n$, $q = q_n$, $\tau_0 = (u_0 + 2)/2$, and $B_n = [0, c_B h^{\alpha}]$. Since $0 < u_0 < 2$ by Assumption 8, we have $\tau_0 \in (u_0, 2)$. After discarding finitely many n, the construction of Definition 6 has $n \geq 1$, $0 < h \leq 1/2$, and $0 < q \leq 1/2$. By construction the contrast, propensity, and overlap score of $P_{n,\sigma}$ are

$$\tau_{P_{n,\sigma}}(x) = \begin{cases} \sigma h, & x \in B_n, \\ \tau_0, & x \notin B_n, \end{cases} \qquad e_{P_{n,\sigma}}(x) = \begin{cases} q, & x \in B_n, \\ 1/2, & x \notin B_n, \end{cases} \qquad p_{P_{n,\sigma}}(x) = \begin{cases} q, & x \in B_n, \\ 1/2, & x \notin B_n, \end{cases}$$

the last because $q \leq 1/2$ makes $\min\{q, 1-q\} = q$. The arm regressions are $\mu_0 = 0$ and $\mu_1 = \sigma h$ on B_n, and $\mu_0 = -\tau_0/2$, $\mu_1 = \tau_0/2$ off B_n. We check the defining conditions of Definition 4 in turn.

Well-formedness and bounded outcomes. The construction specifies the covariate law, propensity, outcome kernels, and nuisance functions explicitly; these fields are measurable, the contrast is the difference of the two regressions, and the tested propensity and regression identities agree with the kernels, so $P_{n,\sigma}$ is a well-formed observed law. Its outcome is supported on $\{-1, 1\}$ or on $\{0\}$, so $Y \in [-1, 1]$ almost surely; and the arm regressions take only the values 0, σh, and $\pm\tau_0/2$, which lie in $[-1, 1]$ because $h \leq 1/2$ and $\tau_0 < 2$.

Positivity. From the display above, $e_{P_{n,\sigma}}(x) \in \{q, 1/2\}$, and $0 < q \leq 1/2 < 1$, so $0 < e_{P_{n,\sigma}}(x) < 1$ for every x.

Zero-effect. The contrast takes only the values $\sigma h \neq 0$ and $\tau_0 > 0$, so $\{\tau_{P_{n,\sigma}} = 0\} = \varnothing$ and in particular $P_X\{\tau_{P_{n,\sigma}}(X) = 0\} = 0$.

Margin. Fix $0 < u \leq u_0$ and put $E_m = \{x : 0 < |\tau_{P_{n,\sigma}}(x)| \leq u\}$. Off B_n the contrast has magnitude $\tau_0 > u_0 \geq u$, so $E_m \subseteq B_n$, and since P_X is Lebesgue measure restricted to $[0, 1]$,

$$P_X(E_m) \leq P_X(B_n) \leq c_B h^{\alpha}.$$

If $h > u$, then on B_n the contrast has magnitude $h > u$, so $E_m = \varnothing$ and the bound is trivial. If $h \leq u$, then $h^{\alpha} \leq u^{\alpha}$ because $\alpha \geq 0$, and, using $c_B \leq C_m$,

$$P_X(E_m) \leq c_B h^{\alpha} \leq c_B u^{\alpha} \leq C_m u^{\alpha}.$$

Strict-overlap endpoint. Suppose $\gamma = 0$. Then $\beta_{\alpha,\gamma} = 0$, so Definition 6 gives $q = 1/4$, and the overlap score takes only the values $1/4$ and $1/2$. Since $0 < \underline{p} \leq 1/4$, we get $p_{P_{n,\sigma}}(x) \geq \underline{p}$ for every x.

Overlap decay. Fix u, v with $0 < u \leq u_0$, $v > 0$, and $v \leq c_o u^{\gamma}$, and let

$$E = \{x : \; p_{P_{n,\sigma}}(x) \leq v, \; 0 < |\tau_{P_{n,\sigma}}(x)| \leq u\}.$$

As in the margin check, off-block points have contrast magnitude $\tau_0 > u_0 \geq u$, so

$$E \subseteq B_n \qquad \text{and} \qquad P_X(E) \leq P_X(B_n) \leq c_B h^\alpha.$$

Moreover, on B_n the contrast magnitude is h. Hence if $h > u$, then $E = \varnothing$ and the required bound holds because its right-hand side is nonnegative. It remains to consider the case $h \leq u$, and we split first on γ.

If $\gamma = 0$, the required bound is $P_X(E) \leq C_o u^\alpha$ by the convention factor in Definition 4. From $h \leq u$, $\alpha \geq 0$, and $c_B \leq C_o$,

$$P_X(E) \leq c_B h^\alpha \leq c_B u^\alpha \leq C_o u^\alpha.$$

Now suppose $\gamma > 0$. On B_n the overlap score is q, so if $q > v$, then $E = \varnothing$ and the required bound again follows from nonnegativity of the right-hand side. Assume $q \leq v$.

If $\alpha = 0$, then $\beta_{\alpha,\gamma} = 0$, so $q = 1/4$, and $q \leq v$ reads $1/4 \leq v$. Raising to the nonnegative power $1/\gamma$ gives $4^{-1/\gamma} \leq v^{1/\gamma}$, so with the assumed $c_B \leq C_o 4^{-1/\gamma}$ and $u^\alpha = u^0 = 1$,

$$P_X(E) \leq c_B h^0 = c_B \leq C_o 4^{-1/\gamma} \leq C_o v^{1/\gamma} = C_o u^\alpha v^{1/\gamma}.$$

If $\alpha > 0$, then $q = h^{\beta_{\alpha,\gamma}}$ with $\beta_{\alpha,\gamma} = \alpha\gamma/(\alpha+1) > 0$, and the calibration identity is

$$\beta_{\alpha,\gamma} \cdot \frac{\alpha+1}{\gamma} = \alpha, \qquad \text{so} \qquad h^\alpha = \left(h^{\beta_{\alpha,\gamma}}\right)^{(\alpha+1)/\gamma} = \left(h^{\beta_{\alpha,\gamma}}\right)^{1/\gamma} \left(h^{\beta_{\alpha,\gamma}}\right)^{\alpha/\gamma}.$$

Using $h^{\beta_{\alpha,\gamma}} = q \leq v$ in both factors, and then $v \leq c_o u^\gamma$ in the second factor,

$$h^\alpha \leq v^{1/\gamma}\, v^{\alpha/\gamma} \leq v^{1/\gamma}\, (c_o u^\gamma)^{\alpha/\gamma} = v^{1/\gamma} c_o^{\alpha/\gamma} u^\alpha.$$

Multiplying by c_B and using the assumed $c_B \leq C_o c_o^{-\alpha/\gamma}$,

$$P_X(E) \leq c_B h^\alpha \leq C_o c_o^{-\alpha/\gamma}\, c_o^{\alpha/\gamma}\, u^\alpha v^{1/\gamma} = C_o u^\alpha v^{1/\gamma}.$$

Thus the overlap-decay condition holds in every case.

Conclusion. All the defining conditions of Definition 4 hold for the fixed sign σ. Applying the preceding argument for $\sigma = 1$ and $\sigma = -1$, and intersecting the two eventual ranges of n, gives $P_{n,+}, P_{n,-} \in \mathcal{P}_{\alpha,\gamma}$ for all sufficiently large n. For the oracle rules: under $P_{n,+}$ the contrast equals $h \geq 0$ on B_n and $\tau_0 > 0$ off B_n, so $\pi^\star_{P_{n,+}}(x) = \mathbf{1}\{\tau_{P_{n,+}}(x) \geq 0\} = 1$ for every x. Under $P_{n,-}$ the contrast equals $-h < 0$ on B_n and $\tau_0 > 0$ off B_n, so $\pi^\star_{P_{n,-}}(x) = 1$ exactly when $x \notin B_n$, that is $\pi^\star_{P_{n,-}}(x) = \mathbf{1}\{x \notin B_n\}$. If these two maps belong to Π, the displayed oracle identities identify the two policy actions used in the associated two-point reduction. Finally, the normalization $0 < u_0 < 2$ is exactly what permits the displayed choice $\tau_0 = (u_0+2)/2 \in (u_0, 2)$, which is both strictly above the margin window and compatible with the outcome bound $|Y| \leq 1$. □

Proof of Lemma 2. Take

$$C = \max\{8c_B, 5\} > 0,$$

and fix $n \geq 1$ in the eventual range where $0 < h_n \leq 1/2$ and $0 < q_n \leq 1/2$. Write $h = h_n$, $q = q_n$, $P_+ = P_{n,+}$, and $P_- = P_{n,-}$.

Step 1: the one-observation divergence. By Definition 6, the two laws have the same covariate law, the same propensity, and the same outcome law except on the treated active cell. Let

$$s_1 = \{X \in B_n,\ A = 1,\ Y = 1\}, \qquad s_2 = \{X \in B_n,\ A = 1,\ Y = -1\}, \qquad s_0 = (s_1 \cup s_2)^c.$$

These three measurable cells partition the full observation space. On s_0, the restrictions of P_+ and P_- agree: this includes the common off-cell part of the law and the residual part of the active treated cell outside the two charged outcome atoms, which has zero mass under both laws. On the two charged cells the outcome probabilities in Definition 6 give constant restriction ratios

$$c_0 = 1, \qquad c_1 = \frac{1+h}{1-h}, \qquad c_2 = \frac{1-h}{1+h},$$

so $P_+ \ll P_-$, and the squared likelihood-ratio deviation is P_--integrable.

Set

$$A_n = P_X(B_n)\, q_n.$$

Under P_-, the two charged cells have masses

$$P_-(s_1) = A_n \frac{1-h}{2}, \qquad P_-(s_2) = A_n \frac{1+h}{2}.$$

The finite-partition chi-square formula for these restriction ratios yields

$$\chi^2(P_+ \,\|\, P_-) = \sum_{i=0}^{2} (c_i - 1)^2 P_-(s_i) = \left(\frac{2h}{1-h}\right)^2 \frac{A_n(1-h)}{2} + \left(\frac{2h}{1+h}\right)^2 \frac{A_n(1+h)}{2} = \frac{4A_n h^2}{1-h^2}.$$

Since $0 \le h \le 1/2$, we have $1 - h^2 \ge 1/2$, and hence

$$\chi^2(P_+ \,\|\, P_-) \le 8 A_n h^2.$$

Because P_X is Lebesgue measure restricted to $[0,1]$ and $B_n = \{x : 0 \le x \le c_B h^\alpha\}$, its block mass obeys $P_X(B_n) \le c_B h^\alpha$. Thus

$$\chi^2(P_+ \,\|\, P_-) \le 8\, c_B h^\alpha\, q_n\, h^2.$$

The weak-arm level satisfies $q_n \le h^{\beta_{\alpha,\gamma}}$: if $\beta_{\alpha,\gamma} = 0$, then $q_n = 1/4 \le 1 = h^0$, while otherwise $q_n = h^{\beta_{\alpha,\gamma}}$, as specified in Definition 6. Therefore

$$\chi^2(P_+ \,\|\, P_-) \le 8 c_B\, h_n^{2+\alpha+\beta_{\alpha,\gamma}} \le C\, h_n^{2+\alpha+\beta_{\alpha,\gamma}},$$

which is the asserted one-observation bound.

Step 2: the product experiment. The one-observation absolute continuity and square-integrability just obtained tensorize, so $P_+^{\otimes n} \ll P_-^{\otimes n}$ and

$$\int \left(\frac{dP_+^{\otimes n}}{dP_-^{\otimes n}} - 1\right)^2 dP_-^{\otimes n} < \infty.$$

By Definition 3, $D_{\alpha,\gamma} = 2 + \alpha + \beta_{\alpha,\gamma} > 0$, and Definition 6 gives $h_n = n^{-1/D_{\alpha,\gamma}}$. Hence

$$h_n^{D_{\alpha,\gamma}} = n^{-1}, \qquad \text{and therefore} \qquad n\, \chi^2(P_+ \,\|\, P_-) \le 8 c_B.$$

The chi-square divergence tensorizes as

$$1 + \chi^2(P_+^{\otimes n} \,\|\, P_-^{\otimes n}) = \left(1 + \chi^2(P_+ \,\|\, P_-)\right)^n.$$

Using $1 + x \le e^x$ for $x = \chi^2(P_+\|P_-) \ge 0$ and the assumption $8c_B < \log 5$,

$$\big(1 + \chi^2(P_+ \,\|\, P_-)\big)^n \le \exp\big\{n\,\chi^2(P_+ \,\|\, P_-)\big\} \le \exp\{8c_B\} < 5.$$

Therefore $1 + \chi^2(P_+^{\otimes n}\|P_-^{\otimes n}) \le 5$, and so

$$\chi^2(P_+^{\otimes n} \,\|\, P_-^{\otimes n}) \le 4 \le 5 \le C.$$

□

Proof of Lemma 3. Take the separation constant

$$c_{\text{sep}} = \frac{\min\{c_B, 1\}}{4} > 0,$$

and work at all sufficiently large n for which Lemma 1 applies to both signs, so that $P_{n,+}, P_{n,-} \in \mathcal{P}_{\alpha,\gamma}$ and their oracle rules are

$$\pi^\star_{P_{n,+}}(x) = 1 \qquad \text{and} \qquad \pi^\star_{P_{n,-}}(x) = \mathbf{1}\{x \notin B_n\}.$$

Write $h = h_n$, $\tau_\pm = \tau_{P_{n,\pm}}$, and shrink the active block to

$$E = [0, \ell], \qquad \ell = \min\{c_B, 1\}\, h^\alpha.$$

Since $0 < h \le 1$ and $\alpha \ge 0$, we have $0 \le h^\alpha \le 1$ and hence $0 \le \ell \le 1$; as P_X is Lebesgue measure restricted to $[0, 1]$, the covariate mass of E is exactly

$$P_X(E) = \ell.$$

Moreover $\ell = \min\{c_B, 1\}h^\alpha \le c_B h^\alpha$, so

$$E \subseteq B_n = \{x : 0 \le x \le c_B h^\alpha\}.$$

Fix any policy $\pi \in \Pi$. Let D_+ and D_- be its disagreement sets under $P_{n,+}$ and $P_{n,-}$, and set

$$B_+ = D_+ \cap \{x : h/2 < |\tau_+(x)|\}, \qquad B_- = D_- \cap \{x : h/2 < |\tau_-(x)|\}.$$

Under each of the two witness laws, which are well formed with bounded outcomes, the welfare identity of Theorem 1 writes the regret as $\int |\tau|\mathbf{1}_D\, dP_X$ with a nonnegative integrand exceeding $h/2$ on the respective set $B_\pm$; restricting the integral to that set therefore gives

$$\frac{h}{2}\, P_X(B_+) \le R_{P_{n,+}}(\pi), \qquad \frac{h}{2}\, P_X(B_-) \le R_{P_{n,-}}(\pi).$$

On $E \subseteq B_n$, the contrasts are $\tau_+(x) = h$ and $\tau_-(x) = -h$, so both have magnitude $h > h/2$; and the two oracle rules take opposite values there,

$$\pi^\star_{P_{n,+}}(x) = 1, \qquad \pi^\star_{P_{n,-}}(x) = 0 \qquad (x \in E).$$

Since $\pi(x) \in \{0, 1\}$, it differs from at least one of these two values at every $x \in E$: if $\pi(x) = 1$ then $x \in D_-$, and if $\pi(x) = 0$ then $x \in D_+$. Combined with the contrast bound this gives

$$E \subseteq B_+ \cup B_-.$$

The two witness laws share the covariate marginal P_X, so by monotonicity and subadditivity,

$$\ell = P_X(E) \le P_X(B_+) + P_X(B_-).$$

Multiplying by $h/2 > 0$ and inserting the two regret bounds gives

$$\frac{h}{2}\,\ell \le \frac{h}{2} P_X(B_+) + \frac{h}{2} P_X(B_-) \le R_{P_{n,+}}(\pi) + R_{P_{n,-}}(\pi).$$

Since $h^{1+\alpha} = h \cdot h^{\alpha}$ and $\ell = \min\{c_B, 1\} h^{\alpha}$, the left-hand side is

$$\frac{h}{2}\,\ell = \frac{\min\{c_B,1\}}{2}\,h^{1+\alpha}, \qquad \text{so} \qquad \frac{\min\{c_B,1\}}{2} h^{1+\alpha} \le R_{P_{n,+}}(\pi) + R_{P_{n,-}}(\pi).$$

Finally, a sum of two reals is at most twice their maximum,

$$R_{P_{n,+}}(\pi) + R_{P_{n,-}}(\pi) \le 2\max\{R_{P_{n,+}}(\pi), R_{P_{n,-}}(\pi)\},$$

so dividing by 2,

$$\max_{\sigma\in\{+,-\}} R_{P_{n,\sigma}}(\pi) \ge \frac{\min\{c_B,1\}}{4}\,h_n^{1+\alpha} = c_{\text{sep}}\,h_n^{1+\alpha}.$$

Since $\pi \in \Pi$ was arbitrary, this holds for every policy in Π, for all such sufficiently large n. $\square$

Proof of Lemma 4. Write $P = P_+^{\otimes n}$ and $Q = P_-^{\otimes n}$, and let

$$A = \{\psi = +\}$$

be the (measurable) region on which the test chooses $+$, so that

$$P_+^{\otimes n}(\psi \ne +) = P(A^c), \qquad P_-^{\otimes n}(\psi \ne -) = Q(A).$$

Throughout we work under the two standing regularity conditions that make the chi-square divergence a genuine density functional: $P \ll Q$, and $(dP/dQ - 1)^2$ is Q-integrable. Both are displayed hypotheses of the lemma. Under them the chi-square divergence has the second-moment form

$$\int \left(\frac{dP}{dQ}\right)^2 dQ = 1 + \chi^2(P\,\|\,Q) \le C_\chi + 1,$$

obtained by expanding $(dP/dQ - 1)^2 = (dP/dQ)^2 - 2\,dP/dQ + 1$ and using $\int (dP/dQ)\,dQ = 1$. Note $C_\chi + 1 \ge 1 > 0$ since $\chi^2 \ge 0$.

Step 1: Cauchy–Schwarz mass transfer. Writing the P-mass of A as a Q-integral against the density and applying Cauchy–Schwarz,

$$P(A) = \int \mathbf{1}_A \frac{dP}{dQ}\,dQ \le \left(\int \mathbf{1}_A^2\,dQ\right)^{1/2} \left(\int \left(\frac{dP}{dQ}\right)^2 dQ\right)^{1/2} \le \sqrt{Q(A)}\,\sqrt{C_\chi + 1} = \sqrt{(C_\chi+1)\,Q(A)}.$$

Step 2: the testing floor. Let $e = P(A^c) + Q(A)$ be the combined error. Since $P(A^c) = 1 - P(A)$ and $Q(A) \ge 0$, Step 1 gives

$$1 - e = P(A) - Q(A) \le P(A) \le \sqrt{(C_\chi+1)\,Q(A)}.$$

Suppose, to get a contradiction, that

$$e < \frac{1}{4(C_\chi + 1)}.$$

Since $P(A^c) \geq 0$, we have $Q(A) \leq e$, hence $Q(A) < 1/(4(C_\chi + 1))$ and therefore

$$(C_\chi + 1)Q(A) < \tfrac{1}{4}, \qquad \text{so} \qquad \sqrt{(C_\chi + 1)Q(A)} < \tfrac{1}{2}.$$

Also $C_\chi + 1 \geq 1$ gives $1/(4(C_\chi + 1)) \leq 1/4$, so $e < 1/4$ and hence $1 - e > 1/2$. This contradicts the previous display. Therefore

$$P_+^{\otimes n}(\psi \neq +) + P_-^{\otimes n}(\psi \neq -) = P(A^c) + Q(A) \geq \frac{1}{4(C_\chi + 1)} = c(C_\chi) > 0,$$

a floor that depends only on C_χ and not on the particular pair of laws or on the test.

Step 3: the one-observation sufficient condition. Suppose now $c_0 \geq 0$, $n > 0$, and the single-observation divergence satisfies $\chi^2(P_+ \,\|\, P_-) \leq c_0/n$, the two standing regularity conditions above now being imposed on the single-observation pair (P_+, P_-). The product experiment obeys the tensorization identity

$$1 + \chi^2(P_+^{\otimes n} \,\|\, P_-^{\otimes n}) = \left(1 + \chi^2(P_+ \,\|\, P_-)\right)^n.$$

Using $1 + x \leq e^x$ with $x = \chi^2(P_+\|P_-) \geq 0$ and then $n\chi^2(P_+\|P_-) \leq c_0$,

$$\left(1 + \chi^2(P_+ \,\|\, P_-)\right)^n \leq \exp\left\{n\,\chi^2(P_+ \,\|\, P_-)\right\} \leq e^{c_0},$$

so the product divergence is bounded, $\chi^2(P_+^{\otimes n}\|P_-^{\otimes n}) \leq e^{c_0} - 1$. Applying Steps 1–2 to the product experiment with the budget $C_\chi = e^{c_0} - 1 \geq 0$ yields the uniform testing floor

$$P_+^{\otimes n}(\psi \neq +) + P_-^{\otimes n}(\psi \neq -) \geq \frac{1}{4\{(e^{c_0} - 1) + 1\}} = \frac{1}{4e^{c_0}} > 0.$$

□

Proof of Theorem 3. 1. *The three ingredients and the lower-bound constant.* By Lemma 3, there are constants $c_{\text{sep}} > 0$ and N_1 such that, for all $n \geq N_1$ and every $\pi \in \Pi$,

$$\max_{\sigma \in \{+,-\}} R_{P_{n,\sigma}}(\pi) \geq c_{\text{sep}}\, h_n^{1+\alpha}.$$

By Lemma 2, there are $C_\chi > 0$ and N_2 such that, for all $n \geq N_2$, the product laws satisfy $P_{n,+}^{\otimes n} \ll P_{n,-}^{\otimes n}$ with square-integrable density deviation and

$$\chi^2(P_{n,+}^{\otimes n} \,\|\, P_{n,-}^{\otimes n}) \leq C_\chi.$$

Applying Lemma 4 to this fixed budget gives a testing floor $c_{\text{test}} = c_{\text{test}}(C_\chi) > 0$. Equivalently, applying the binary test that chooses $+$ on a measurable event $\mathcal{B}$ and $-$ on $\mathcal{B}^c$, every measurable $\mathcal{B}$ in the n-sample space satisfies

$$P_{n,+}^{\otimes n}(\mathcal{B}^c) + P_{n,-}^{\otimes n}(\mathcal{B}) \geq c_{\text{test}}.$$

Set

$$\kappa_{\mathrm{lb}} = \frac{c_{\mathrm{sep}}\, c_{\mathrm{test}}}{2} > 0.$$

2. *Membership and the rate identity.* By Lemma 1, enlarging the eventual lower bound on n if necessary, both witness laws $P_{n,+}$ and $P_{n,-}$ belong to $\mathcal{P}_{\alpha,\gamma}$ for all n under consideration. The displays in Definitions 3 and 6 give

$$h_n = n^{-1/(2+\alpha+\beta_{\alpha,\gamma})}, \qquad r_\star(\alpha,\gamma) = \frac{1+\alpha}{2+\alpha+\beta_{\alpha,\gamma}},$$

so

$$h_n^{1+\alpha} = n^{-r_\star(\alpha,\gamma)}.$$

The policy class is nonempty, so the collection of measurable Π-valued estimators is nonempty: a constant estimator at any fixed policy in Π qualifies. For laws in $\mathcal{P}_{\alpha,\gamma}$, the well-formedness and bounded-outcome components of Definition 4 imply that the regret maps are nonnegative and bounded by 2, hence the displayed expectations below are finite.

3. *The threshold event.* Fix such an n, and let $\widehat{\pi}$ be any measurable Π-valued estimator. Write

$$R_+ = \mathbb{E}_{P_{n,+}^{\otimes n}} R_{P_{n,+}}(\widehat{\pi}), \qquad R_- = \mathbb{E}_{P_{n,-}^{\otimes n}} R_{P_{n,-}}(\widehat{\pi}),$$

put

$$s_n = c_{\mathrm{sep}} h_n^{1+\alpha} > 0, \qquad \mathcal{E}_n = \left\{\text{samples}: \ s_n \le R_{P_{n,-}}(\widehat{\pi})\right\}.$$

The event $\mathcal{E}_n$ is measurable because the map from the sample to $R_{P_{n,-}}(\widehat{\pi})$ is measurable, as required for admissible estimators in Definition 5.

4. *Two Markov-type threshold bounds.* For a probability measure Q, a nonnegative Q-integrable function f, and a level $s > 0$, the pointwise inequality $f \ge s\,\mathbf{1}\{f \ge s\}$ integrates to

$$s\, Q\{f \ge s\} \le \int f\, dQ.$$

Applying this to the nonnegative integrable regret maps under the two product laws gives

$$s_n\, P_{n,-}^{\otimes n}(\mathcal{E}_n) \le R_-, \qquad s_n\, P_{n,+}^{\otimes n}\left\{s_n \le R_{P_{n,+}}(\widehat{\pi})\right\} \le R_+.$$

On $\mathcal{E}_n^c$, the minus-regret is below s_n. The separation bound above applies pointwise to the realized policy $\widehat{\pi} \in \Pi$, so

$$s_n \le \max\left\{R_{P_{n,+}}(\widehat{\pi}), R_{P_{n,-}}(\widehat{\pi})\right\}.$$

Hence the plus-regret is at least s_n on $\mathcal{E}_n^c$, and

$$\mathcal{E}_n^c \subseteq \left\{s_n \le R_{P_{n,+}}(\widehat{\pi})\right\}.$$

By monotonicity of $P_{n,+}^{\otimes n}$ and the second threshold bound,

$$s_n\, P_{n,+}^{\otimes n}(\mathcal{E}_n^c) \le R_+.$$

5. *Combining with the testing floor.* Applying the testing floor to $\mathcal{E}_n$ and adding the two threshold bounds yields

$$s_n\, c_{\mathrm{test}} \le s_n \left\{P_{n,+}^{\otimes n}(\mathcal{E}_n^c) + P_{n,-}^{\otimes n}(\mathcal{E}_n)\right\} \le R_+ + R_-.$$

Since $R_+ + R_- \le 2\max\{R_+, R_-\}$,

$$\max\{R_+, R_-\} \ge \frac{s_n c_{\mathrm{test}}}{2} = \frac{c_{\mathrm{sep}} c_{\mathrm{test}}}{2}\, h_n^{1+\alpha} = \kappa_{\mathrm{lb}}\, h_n^{1+\alpha} = \kappa_{\mathrm{lb}}\, n^{-r_\star(\alpha,\gamma)}.$$

6. *Passing to the minimax risk.* Because $P_{n,+}, P_{n,-} \in \mathcal{P}_{\alpha,\gamma}$, the two expectations R_+ and R_- are values indexed by admissible laws in the supremum of Definition 5. Therefore

$$\max\{R_+, R_-\} \le \sup_{P\in\mathcal{P}_{\alpha,\gamma}} \mathbb{E}_{P^{\otimes n}} R_P(\widehat{\pi}).$$

Combining this with the preceding display, every admissible estimator satisfies

$$\sup_{P\in\mathcal{P}_{\alpha,\gamma}} \mathbb{E}_{P^{\otimes n}} R_P(\widehat{\pi}) \ge \kappa_{\mathrm{lb}}\, n^{-r_\star(\alpha,\gamma)},$$

where κ_{lb} depends only on the fixed theorem parameters. Taking the infimum over all measurable Π-valued estimators in Definition 5 gives

$$M_n(\alpha,\gamma) \ge \kappa_{\mathrm{lb}}\, n^{-r_\star(\alpha,\gamma)}$$

for all sufficiently large n. □

Proof of Theorem 4. 1. *The schedules.* The clipping and localization schedules are those of Definition 7: for $\gamma > 0$,

$$q_n = q_0 n^{-s_{\mathrm{feas}}}, \qquad u_n = \bar{u}\, n^{-t_{\mathrm{feas}}},$$

while for $\gamma = 0$ the clipping schedule is fixed, $q_n = q_0$ for every n. The input restrictions assumed in the theorem give $q_0 > 0$ in both regimes; when $\gamma > 0$ they also give $\bar{u} > 0$, $q_0 \le \min\{1/2, c_o \bar{u}^\gamma\}$, and the eventual admissibility recorded in Definition 7,

$$q_n \le c_o u_n^\gamma.$$

Thus $q_n > 0$ for all sufficiently large n.

2. *The empty law-class branch.* Suppose $\gamma = 0$ and the law class in Definition 4 is empty. Then the image set whose supremum defines Definition 11 is empty, since every law in that side-condition domain must first belong to $\mathcal{P}_{\alpha,\gamma}$. We use the real-supremum convention $\sup\emptyset = 0$. Hence $U_n(\alpha,\gamma,a,c;\widehat{\eta}) = 0$, and for $n \ge 2$ the candidate bound with $C = 1$ and $p = 0$ is nonnegative. This proves the asserted eventual bound in this branch; the remainder treats the complementary branch.

3. *The clip lies in the admissible clipping range.* If $\gamma = 0$, the complementary branch supplies some law $P \in \mathcal{P}_{\alpha,\gamma}$. Its strict-overlap endpoint condition in Definition 4 gives $\underline{p} \le 1/2$, and the input restriction $q_0 \le \underline{p}/2$ gives

$$q_n = q_0 \le \frac{1}{2}.$$

If $\gamma > 0$, the maximizer of Definition 7 lies in the feasible box, so $s_{\mathrm{feas}} \ge 0$. Hence $n^{-s_{\mathrm{feas}}} \le 1$ for $n \ge 1$, and therefore

$$q_n = q_0 n^{-s_{\mathrm{feas}}} \le q_0 \le \frac{1}{2}.$$

Likewise, when $\gamma > 0$, $t_{\mathrm{feas}} \ge 0$ gives $n^{-t_{\mathrm{feas}}} \le 1$ for $n \ge 1$, so with $\bar{u} \le u_0$,

$$0 < u_n = \bar{u}\, n^{-t_{\mathrm{feas}}} \le \bar{u} \le u_0.$$

4. *The two ingredients.* Lemma 8, applied with these deterministic schedules and the theorem hypotheses, supplies constants $C_{\mathrm{M}} > 0$ and $p_{\mathrm{M}} \geq 0$ such that, for all sufficiently large n, every law P satisfying the side-condition domain of Definition 11 obeys the corresponding master inequality. The schedule hypotheses needed there are the eventual bounds $0 < q_n \leq 1/2$, together with eventual constancy of q_n when $\gamma = 0$, established above. Lemma 9, applied to the same schedules, the polynomial nuisance-rate exponents of Assumption 14, the nonnegative nuisance constants, the eventual nonnegativity of $r_{\mu,n}$, and the positive-γ admissibility above, supplies constants $C_{\mathrm{B}} > 0$ and $p_{\mathrm{B}} \geq 0$ bounding the deterministic bracket of the master bound by

$$C_{\mathrm{B}} n^{-r_{\mathrm{feas}}} (\log n)^{p_{\mathrm{B}}}, \qquad r_{\mathrm{feas}} = r_{\mathrm{up}}$$

by Definition 7. Both constant pairs are chosen before n. Set

$$C = C_{\mathrm{M}} C_{\mathrm{B}} > 0, \qquad p = p_{\mathrm{M}} + p_{\mathrm{B}} \geq 0.$$

5. *The case $\gamma = 0$.* For every P in the domain of Definition 11 and all sufficiently large n, the strict-overlap branch of Lemma 8 applies because $q_n = q_0 \leq \underline{p}/2$, and gives

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C_{\mathrm{M}} \left\{ n^{-A_\alpha} + r_{\mu,n} r_{e,n} \right\} (\log n)^{p_{\mathrm{M}}}.$$

The $\gamma = 0$ branch of Lemma 9 bounds the brace by $C_{\mathrm{B}} n^{-r_{\mathrm{feas}}} (\log n)^{p_{\mathrm{B}}}$. Since $\log n > 0$ for all sufficiently large n, the two logarithmic factors multiply as $(\log n)^{p_{\mathrm{B}}} (\log n)^{p_{\mathrm{M}}} = (\log n)^p$, and therefore

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C\, n^{-r_{\mathrm{feas}}} (\log n)^p.$$

6. *The case $\gamma > 0$.* For all sufficiently large n we have $u_n > 0$, $u_n \leq u_0$, and $q_n \leq c_o u_n^{\gamma}$ by the schedule facts above. The positive-γ branch of Lemma 8 applies and yields, for each P in the domain,

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C_{\mathrm{M}} \left\{ n^{-r_\star(\alpha,\gamma)} + (n q_n^2)^{-A_\alpha} + \frac{r_{\mu,n} r_{e,n}}{q_n} + r_{\mu,n} u_n^{\alpha/2} q_n^{1/(2\gamma)} + \frac{r_{\mu,n}^2}{u_n} \right\} (\log n)^{p_{\mathrm{M}}}.$$

The positive-γ branch of Lemma 9 bounds the brace by $C_{\mathrm{B}} n^{-r_{\mathrm{feas}}} (\log n)^{p_{\mathrm{B}}}$, and the same logarithm identity gives

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C\, n^{-r_{\mathrm{feas}}} (\log n)^p.$$

7. *Passing to the supremum.* Fix a sufficiently large n, and set

$$B_n = C\, n^{-r_{\mathrm{feas}}} (\log n)^p.$$

The quantity B_n is nonnegative and uniform over the side-condition domain of Definition 11. The preceding case analysis shows that every value in the image set defining $U_n(\alpha, \gamma, a, c; \widehat{\eta})$ is at most B_n. If that image set is empty, the convention $\sup \emptyset = 0$ gives the same conclusion because $B_n \geq 0$; otherwise the pointwise bound gives the supremum bound. Therefore

$$U_n(\alpha, \gamma, a, c; \widehat{\eta}) \leq C\, n^{-r_{\mathrm{feas}}} (\log n)^p = C\, n^{-r_{\mathrm{up}}} (\log n)^p,$$

for all sufficiently large n, with the single pair (C, p) fixed above. This is the claimed bound.

□

Proof of Lemma 5. Work throughout in the positive-sample-size case $n \geq 1$. Fix a realized sample, the pointwise-dense skeleton $\Pi_0 = \{\pi_j : j \geq 1\}$ used in Definition 10, and measurable foldwise nuisance components $\widehat{\mu}_0^{(-k)}$, $\widehat{\mu}_1^{(-k)}$, and $\widehat{e}^{(-k)}$ for each fold k. Abbreviate the observed foldwise scores of Definition 9 and the empirical criterion of Definition 10 by

$$\Gamma_i = \Gamma_q\left(O_i; \widehat{\eta}^{(-k(i))}\right), \qquad \widehat{V}_{n,q}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\pi(X_i) = 1\}\,\Gamma_i.$$

The criterion $\widehat{V}_{n,q}(\pi)$ depends on π only through the finitely many values $\pi(X_1), \ldots, \pi(X_n)$.

Step 1: a $1/n$-near maximizer over the skeleton exists. Since each indicator is 0 or 1,

$$\widehat{V}_{n,q}(\pi_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\pi_j(X_i) = 1\}\,\Gamma_i \leq \frac{1}{n}\sum_{i=1}^{n} |\Gamma_i| \qquad \text{for every } j \geq 1.$$

Thus the set of skeleton scores $\{\widehat{V}_{n,q}(\pi_j) : j \geq 1\}$ is nonempty and bounded above. Write $S = \sup_{j\geq 1} \widehat{V}_{n,q}(\pi_j)$, which is finite. Because $n \geq 1$, $1/n > 0$, so $S - 1/n < S$. By the defining property of the supremum, there is an index j with $\widehat{V}_{n,q}(\pi_j) > S - 1/n$, and hence

$$\widehat{V}_{n,q}(\pi_{j'}) \leq \widehat{V}_{n,q}(\pi_j) + \frac{1}{n} \qquad \text{for every } j' \geq 1.$$

The set of such indices is a nonempty set of positive integers and has a least element. This least element is the selector index j_n in Definition 10; consequently

$$\widehat{V}_{n,q}(\pi_{j'}) \leq \widehat{V}_{n,q}(\widehat{\pi}_n) + \frac{1}{n} \qquad \text{for every } j' \geq 1.$$

Step 2: the selector is measurable and the selected policy lies in Π. Every enumerated policy π_j belongs to Π by the skeleton condition in Assumption 4 and Definition 10, so $\widehat{\pi}_n = \pi_{j_n} \in \Pi$. For measurability, the event $\{j_n = j\}$ is exactly the event that j is a $1/n$-near maximizer and every smaller index is outside the near-maximizer set:

$$\{j_n = j\} = N_j \cap (\text{complement of } N_1) \cap \ldots \cap (\text{complement of } N_{j-1}), \qquad N_k = \left\{\widehat{V}_{n,q}(\pi_{j'}) \leq \widehat{V}_{n,q}(\pi_k) + \tfrac{1}{n} \ \text{ for every}\right.$$

Each score $\widehat{V}_{n,q}(\pi_j)$ is a measurable function of the sample: π_j is measurable by Assumption 4, and the foldwise components $\widehat{\mu}_0^{(-k)}$, $\widehat{\mu}_1^{(-k)}$, and $\widehat{e}^{(-k)}$ are measurable by the regularity condition fixed above. Therefore each N_k is a countable intersection of measurable score comparisons, and $\{j_n = j\}$ is a finite intersection of such sets and their complements. Hence every fiber of j_n is measurable, so j_n is a measurable integer-valued selector. Since $j \mapsto R_P(\pi_j)$ is a function on a countable set, the regret map $s \mapsto R_P(\pi_{j_n(s)}) = R_P(\widehat{\pi}_n(s))$ is measurable.

Step 3: the skeleton inequality reaches every comparator in Π. Let $\pi^b \in \Pi$. By Assumption 4 and the pointwise-dense skeleton fixed in Definition 10, there is a sequence $(\pi_{j_t})_{t\geq 1}$ from Π_0 such that $\pi_{j_t}(x)$ is eventually equal to $\pi^b(x)$ at each covariate value x. Applying this eventual equality to the finite set $X_1, \ldots, X_n$, choose one index t such that

$$\pi_{j_t}(X_i) = \pi^b(X_i) \qquad \text{for all } i = 1, \ldots, n.$$

Since $\widehat{V}_{n,q}$ depends on the policy only through those n values,

$$\widehat{V}_{n,q}(\pi_{j_t}) = \widehat{V}_{n,q}(\pi^b).$$

Applying the near-maximality from Step 1 to the skeleton index j_t gives

$$\widehat{V}_{n,q}(\pi^b) = \widehat{V}_{n,q}(\pi_{j_t}) \le \widehat{V}_{n,q}(\widehat{\pi}_n) + \frac{1}{n}.$$

Thus every comparator $\pi^b \in \Pi$ has empirical score at most $\widehat{V}_{n,q}(\widehat{\pi}_n) + 1/n$, which is the Π-wide $1/n$-ERM inequality in pointwise-comparator form.

Step 4: the empirical-process form. For any comparator $\pi^b \in \Pi$, the definition of $\widehat{V}_{n,q}$ gives

$$\widehat{V}_{n,q}(\widehat{\pi}_n) - \widehat{V}_{n,q}(\pi^b) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\mathbf{1}\{\widehat{\pi}_n(X_i)=1\} - \mathbf{1}\{\pi^b(X_i)=1\}\Big\}\,\Gamma_i = P_n\Big[(\widehat{\pi}_n - \pi^b)\Gamma_q\Big].$$

The inequality from Step 3 is therefore

$$P_n\Big[(\widehat{\pi}_n - \pi^b)\Gamma_q\Big] \ge -\frac{1}{n}.$$

Finally, under Assumption 5 the oracle policy satisfies $\pi_P^\star \in \Pi$, so the comparator inequality applies with $\pi^b = \pi_P^\star$. ☐

Proof of Lemma 6. Fix $q \in (0, 1/2]$ and one foldwise nuisance triple, and write

$$e_q = e_q\Big(X; \widehat{\eta}^{(-k)}\Big) = \min\Big\{1-q, \max\{q, \widehat{e}^{(-k)}(X)\}\Big\}.$$

Since $q \le 1/2 \le 1-q$, the outer minimum and inner maximum give the two-sided clip

$$q \le e_q \le 1-q, \qquad \text{hence} \qquad q \le 1 - e_q,$$

so both denominators are at least $q > 0$ and

$$\frac{1}{e_q} \le \frac{1}{q}, \qquad \frac{1}{1-e_q} \le \frac{1}{q}.$$

The treatment indicators satisfy $|A| \le 1$ and $|1-A| \le 1$. By Assumption 2, $|Y| \le 1$ almost surely, and by the assumed foldwise range of the cross-fitted outcome regressions, $|\widehat{\mu}_0^{(-k)}(X)| \le 1$ and $|\widehat{\mu}_1^{(-k)}(X)| \le 1$. Therefore, by the triangle inequality,

$$\Big|\widehat{\mu}_1^{(-k)}(X) - \widehat{\mu}_0^{(-k)}(X)\Big| \le 2, \qquad \Big|Y - \widehat{\mu}_a^{(-k)}(X)\Big| \le 2, \quad a = 0, 1.$$

Multiplying the indicator, inverse-denominator, and residual bounds gives, for the two inverse-propensity terms,

$$\left|\frac{A}{e_q}\Big\{Y - \widehat{\mu}_1^{(-k)}(X)\Big\}\right| \le 1 \cdot \frac{1}{q} \cdot 2 = \frac{2}{q}, \qquad \left|\frac{1-A}{1-e_q}\Big\{Y - \widehat{\mu}_0^{(-k)}(X)\Big\}\right| \le \frac{2}{q}.$$

Applying the triangle inequality to the three summands of Definition 9,

$$\Big|\Gamma_q\Big(O; \widehat{\eta}^{(-k)}\Big)\Big| \le \Big|\widehat{\mu}_1^{(-k)}(X) - \widehat{\mu}_0^{(-k)}(X)\Big| + \left|\frac{A}{e_q}\Big\{Y - \widehat{\mu}_1^{(-k)}(X)\Big\}\right| + \left|\frac{1-A}{1-e_q}\Big\{Y - \widehat{\mu}_0^{(-k)}(X)\Big\}\right| \le 2 + \frac{2}{q} + \frac{2}{q}.$$

Since $q \le 1/2 < 1$, we have $2 \le 2/q$, so the constant summand is absorbed and

$$\Big|\Gamma_q\Big(O; \widehat{\eta}^{(-k)}\Big)\Big| \le \frac{6}{q} \quad \text{almost surely.}$$

As $q > 0$ gives $6/q \le 36/q$, the choice $C = 36$ delivers the asserted absolute envelope

$$\left|\Gamma_q\left(O; \widehat{\eta}^{(-k)}\right)\right| \le \frac{C}{q} \quad \text{almost surely.}$$

Squaring the sharper bound $|\Gamma_q| \le 6/q$, whose right-hand side is nonnegative, yields

$$\Gamma_q\left(O; \widehat{\eta}^{(-k)}\right)^2 \le \left(\frac{6}{q}\right)^2 = \frac{36}{q^2} = \frac{C}{q^2} \quad \text{almost surely,}$$

so the same constant $C = 36$ serves both conclusions. □

Proof of Lemma 9. Throughout, we use the elementary monotonicity rule for powers of n: if $n \ge 1$ and $r \le e$, then

$$n^{-e} \le n^{-r}.$$

Choice of the constants. The pair (C, p) is chosen once, before the case distinction on γ, and the same pair serves both branches. Set

$$K_{\mathrm{off}} = (q_0^2)^{-A_\alpha}, \qquad K_{\mathrm{prod}} = \frac{C_{\mathrm{prod}}}{q_0}, \qquad K_\mu = C_\mu\, |\bar{u}|^{\alpha/2} q_0^{1/(2\gamma)}, \qquad K_{\mathrm{sq}} = \frac{C_\mu^2}{|\bar{u}|},$$

and

$$C = 2 + K_{\mathrm{off}} + K_{\mathrm{prod}} + K_\mu + K_{\mathrm{sq}} + C_{\mathrm{prod}}, \qquad p = 0.$$

The four auxiliary constants are formed from $|\bar{u}|$ rather than $\bar{u}$, and are read under the convention that a quotient with vanishing denominator is 0 and that the exponent $1/(2\gamma)$ is 0 when $\gamma = 0$; with this convention all four are defined for every value of the parameters, including the fixed-overlap regime, where neither $\bar{u} > 0$ nor $\gamma > 0$ is available. They are nonnegative there: K_{off} is a real power of $q_0^2 \ge 0$; K_{prod} and K_{sq} are quotients of the nonnegative numbers C_{prod} and C_μ^2 by the nonnegative numbers q_0 and $|\bar{u}|$; and K_μ is a product of the nonnegative factors C_μ, $|\bar{u}|^{\alpha/2}$ and $q_0^{1/(2\gamma)}$. Hence $C \ge 2 > 0$ and $p \ge 0$. Neither branch below alters these constants: the fixed-overlap branch never uses the values of $K_{\mathrm{off}}, K_{\mathrm{prod}}, K_\mu, K_{\mathrm{sq}}$, which enter its estimate only as nonnegative slack inside C.

Case $\gamma > 0$. Here $\bar{u} > 0$, so $|\bar{u}| = \bar{u}$ and the two $\bar{u}$-dependent constants take their stated values $K_\mu = C_\mu \bar{u}^{\alpha/2} q_0^{1/(2\gamma)}$ and $K_{\mathrm{sq}} = C_\mu^2/\bar{u}$. Write $s = s_{\mathrm{feas}}$, $t = t_{\mathrm{feas}}$, so that the deterministic schedules of Definition 7 are

$$q_n = q_0 n^{-s}, \qquad u_n = \bar{u}\, n^{-t}.$$

By Definition 7,

$$r_{\mathrm{feas}} = \min\{r_\star(\alpha, \gamma), g_{\mathrm{joint}}\}, \qquad \text{so} \qquad r_{\mathrm{feas}} \le r_\star(\alpha, \gamma) \quad \text{and} \quad r_{\mathrm{feas}} \le g_{\mathrm{joint}}.$$

The pair (s, t) is selected in Definition 7 as a maximizer of ϕ over the feasible box; in particular it is a point of that box,

$$0 \le s \le \tfrac{1}{2}, \qquad 0 \le t \le \frac{s}{\gamma},$$

so the two schedules above are the ones the definition prescribes and ϕ is evaluated at an admissible pair. The box is nonempty and compact and ϕ is continuous on it, so ϕ attains its supremum there, and the selected pair realizes that value:

$$g_{\mathrm{joint}} = \phi(s, t) = \min\left\{A_\alpha(1 - 2s),\, c - s,\, a + \frac{s}{2\gamma} + \frac{\alpha t}{2},\, 2a - t\right\}.$$

Since a minimum is at most each of its entries, this yields the four exponent inequalities

$$g_{\text{joint}} \le A_\alpha(1-2s), \quad g_{\text{joint}} \le c-s, \quad g_{\text{joint}} \le a + \frac{s}{2\gamma} + \frac{\alpha t}{2}, \quad g_{\text{joint}} \le 2a - t,$$

and therefore, combining with $r_{\text{feas}} \le g_{\text{joint}}$,

$$r_{\text{feas}} \le A_\alpha(1-2s), \quad r_{\text{feas}} \le c-s, \quad r_{\text{feas}} \le a + \frac{s}{2\gamma} + \frac{\alpha t}{2}, \quad r_{\text{feas}} \le 2a - t.$$

Work on the eventual index set on which $n \ge 1$, the polynomial nuisance bounds of Assumption 14,

$$r_{\mu,n} \le C_\mu n^{-a}, \qquad r_{\mu,n} r_{e,n} \le C_{\text{prod}} n^{-c},$$

hold, and $r_{\mu,n} \ge 0$. Write $b_n = n^{-r_{\text{feas}}} \ge 0$ for the target rate, and bound the five master-bound terms one at a time.

For the converse term, $r_{\text{feas}} \le r_\star(\alpha,\gamma)$ gives directly

$$n^{-r_\star(\alpha,\gamma)} \le 1 \cdot b_n.$$

For the empirical-process term, $nq_n^2 = q_0^2 n^{1-2s}$, so raising to the power $-A_\alpha$ is an exact identity, and the exponent inequality then applies:

$$(nq_n^2)^{-A_\alpha} = (q_0^2)^{-A_\alpha}\, n^{-A_\alpha(1-2s)} \le K_{\text{off}}\, b_n.$$

For the clipped product-nuisance remainder, dividing the product bound by $q_n = q_0 n^{-s} > 0$,

$$\frac{r_{\mu,n} r_{e,n}}{q_n} \le \frac{C_{\text{prod}} n^{-c}}{q_0 n^{-s}} = \frac{C_{\text{prod}}}{q_0}\, n^{-(c-s)} \le K_{\text{prod}}\, b_n.$$

For the localized weak-overlap drift term, the factor $u_n^{\alpha/2} q_n^{1/(2\gamma)}$ is nonnegative, so it may be multiplied into the bound on $r_{\mu,n}$:

$$r_{\mu,n} u_n^{\alpha/2} q_n^{1/(2\gamma)} \le C_\mu n^{-a}\, (\bar{u} n^{-t})^{\alpha/2} (q_0 n^{-s})^{1/(2\gamma)} = C_\mu \bar{u}^{\alpha/2} q_0^{1/(2\gamma)}\, n^{-\{a + s/(2\gamma) + \alpha t/2\}} \le K_\mu\, b_n.$$

For the localization-complement outcome term, the squaring step uses both $0 \le r_{\mu,n}$ and $C_\mu \ge 0$, so that $r_{\mu,n} \le C_\mu n^{-a}$ implies $r_{\mu,n}^2 \le C_\mu^2 n^{-2a}$; dividing by $u_n = \bar{u} n^{-t} > 0$,

$$\frac{r_{\mu,n}^2}{u_n} \le \frac{C_\mu^2 n^{-2a}}{\bar{u} n^{-t}} = \frac{C_\mu^2}{\bar{u}}\, n^{-(2a-t)} \le K_{\text{sq}}\, b_n.$$

Adding the five displays gives the sum bound with coefficient $1 + K_{\text{off}} + K_{\text{prod}} + K_\mu + K_{\text{sq}}$; since $C_{\text{prod}} \ge 0$, that coefficient is at most C, and $b_n \ge 0$, so

$$n^{-r_\star(\alpha,\gamma)} + (nq_n^2)^{-A_\alpha} + \frac{r_{\mu,n} r_{e,n}}{q_n} + r_{\mu,n} u_n^{\alpha/2} q_n^{1/(2\gamma)} + \frac{r_{\mu,n}^2}{u_n} \le \big(1 + K_{\text{off}} + K_{\text{prod}} + K_\mu + K_{\text{sq}}\big) b_n \le C\, n^{-r_{\text{feas}}} (\log n)^0,$$

which is the asserted positive-γ bound with $r_{\text{feas}} = \min\{r_\star(\alpha,\gamma), g_{\text{joint}}\}$ and $p = 0$.

Case $\gamma = 0$. The constants are those fixed above; nothing is rechosen. Here Definition 7 gives the fixed clipping schedule $q_n = q_0$ and

$$r_{\text{feas}} = \min\{A_\alpha, c\}, \qquad \text{so} \qquad r_{\text{feas}} \le A_\alpha \quad \text{and} \quad r_{\text{feas}} \le c.$$

On the eventual index set where $n \geq 1$ and the polynomial product bound holds, the same monotonicity rule gives the two termwise bounds, again with $b_n = n^{-r_{\text{feas}}} \geq 0$,

$$n^{-A_\alpha} \leq 1 \cdot b_n, \qquad r_{\mu,n} r_{e,n} \leq C_{\text{prod}} n^{-c} \leq C_{\text{prod}}\, b_n.$$

Adding them gives the sum bound with coefficient $1 + C_{\text{prod}}$; since $K_{\text{off}}, K_{\text{prod}}, K_\mu, K_{\text{sq}} \geq 0$, that coefficient is at most C, whence

$$n^{-A_\alpha} + r_{\mu,n} r_{e,n} \leq \big(1 + C_{\text{prod}}\big) b_n \leq C\, n^{-r_{\text{feas}}} (\log n)^0 = C\, n^{-\min\{A_\alpha, c\}} (\log n)^0.$$

$\square$

Proof of Lemma 8. Constants. Lemma 11 supplies constants $C^0_{\text{off}} > 0$ and $p \geq 0$; put

$$C_{\text{off}} = \max\{C^0_{\text{off}}, 1\} \ \ (\geq 1), \qquad C_d = 4\left\{1 + \sqrt{\max\{C_o, 1\}}\right\},$$

where C_d is the constant appearing in the $\gamma > 0$ branch of Lemma 7. Choose a branchwise auxiliary constant q_{fix} as follows: if $\gamma = 0$, take $q_{\text{fix}} > 0$ with $q_n = q_{\text{fix}}$ eventually, as supplied by the fixed-clip hypothesis; otherwise set $q_{\text{fix}} = 1$. Set $C_q = \max\{1, 2/q_{\text{fix}}\}$, which in the strict-overlap branch is the constant appearing in the $\gamma = 0$ branch of Lemma 7. Set

$$K_{\text{pos}} = (36^2)^{A_\alpha}, \qquad K_{\text{zero}} = \left(\left(\frac{36}{q_{\text{fix}}}\right)^2\right)^{A_\alpha}, \qquad C = 100 + 20\left(C_{\text{off}} K_{\text{pos}} + C_{\text{off}} K_{\text{zero}} + C_d + C_d^2 + C_q\right),$$

so that $C > 0$. The logarithmic power in the conclusion is this same p.

Large-n restriction. Work on the eventual index set on which $n \geq 1$, $\log n \geq 1$, $0 < q_n \leq 1/2$, $r_{\mu,n} \geq 0$, $r_{e,n} \geq 0$, the foldwise nuisance errors lie in $L^2(P_X)$, and, when $\gamma = 0$, $q_n = q_{\text{fix}}$. Fix such an n and a law P satisfying the stated law-class, oracle, sampling, nuisance-rate, and bounded-nuisance conditions. The measurability, integrability, and boundedness-above properties of the selected regret and of the offset suprema follow from the bounded cross-fitted nuisances, the measurability of the foldwise nuisance estimators, the positive clip, and the countable pointwise-dense skeleton of Assumption 4, which reduces each supremum over Π to a supremum over the countable Π_0; these integrability side conditions are therefore established rather than assumed. The nuisance measurability and the $L^2(P_X)$ memberships of the foldwise nuisance errors are the hypotheses collected in the statement's measurability and L^2 regularity condition, and the L^2 memberships enter through the Cauchy–Schwarz step of the drift bound invoked below.

Objects. Write $q = q_n$ and set the envelope

$$B = \frac{36}{q},$$

so $B \geq 72 \geq 1$ because $q \leq 1/2$. For each fold k, let

$$g_{k,\pi}(O) = \{\mathbf{1}\{\pi(X) = 1\} - \mathbf{1}\{\pi^\star_P(X) = 1\}\}\, \Gamma_q\Big(O; \widehat{\eta}^{(-k)}\Big), \qquad \Delta_k(\pi) = \int_{\mathcal{X}} \{\mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi^\star_P(x) = 1\}\}\, b_q^{(k)}(x$$

where $b_q^{(k)}$ is the clipped-score drift of Lemma 12 formed from the fold-k nuisance triple $\widehat{\eta}^{(-k)}$. Let $w_k = |I_k|/n$, so $w_k \geq 0$ and $\sum_k w_k = 1$, and let

$$G_{\text{cf}}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \left\{ g_{k(i),\pi}(O_i) - \int g_{k(i),\pi}(O)\, dP(O) \right\}$$

be the pooled cross-fit centered process built from these increments.

Step 1: the offset bound at envelope B. By Lemma 6, the clipped AIPW score admits an almost-sure envelope proportional to $1/q$; the truncation performed below needs one explicit admissible constant, which we now record from Definitions 8 and 9 and the boundedness assumptions. On the bounded-outcome event, the clipped propensity $e_q(X; \widehat{\eta}^{(-k)})$ lies between q and $1-q$. Since $q \leq 1/2$, the two inverse denominators are at most $1/q$; since Y, $\widehat{\mu}_0^{(-k)}$, and $\widehat{\mu}_1^{(-k)}$ lie in $[-1, 1]$, and the binary treatment indicators are bounded by one,

$$\left|\widehat{\mu}_1^{(-k)}(X) - \widehat{\mu}_0^{(-k)}(X)\right| \leq 2, \qquad \left|Y - \widehat{\mu}_a^{(-k)}(X)\right| \leq 2, \quad a \in \{0, 1\}.$$

Thus the three terms in the clipped AIPW score satisfy

$$\left|\Gamma_q\left(O; \widehat{\eta}^{(-k)}\right)\right| \leq 2 + \frac{2}{q} + \frac{2}{q} \leq \frac{6}{q} \leq \frac{36}{q} = B$$

almost surely. Consequently truncating the score at level B changes nothing almost surely, and the pooled and foldwise offset suprema built from g and from its B-truncation coincide almost surely and therefore have equal expectations. Write $g^B_{k,\pi}$ for the increment obtained from $g_{k,\pi}$ by truncating the score at level B. The increments $g^B_{k,\pi}$ are policy-compatible, measurable, bounded by B, and vanish off D_π, so

$$\int \left(g^B_{k,\pi}(O)\right)^2 dP(O) \leq B^2 P_X(D_\pi), \qquad \pi \in \Pi.$$

Hence Lemma 11 applies at sample size n with this envelope and, transporting back to the untruncated increments, gives

$$\mathbb{E}_P \sup_{\pi \in \Pi} \left\{2|G_{\mathrm{cf}}(\pi)| - \frac{R_P(\pi)}{4}\right\}_+ \leq C_{\mathrm{off}} \left(\frac{B^2}{n}\right)^{A_\alpha} (\log n)^p.$$

Moreover, since $0 \leq A_\alpha \leq 1$, $1/n \leq 1$, $B \geq 1$, $\log n \geq 1$, and $C_{\mathrm{off}} \geq 1$,

$$\frac{1}{n} \leq \left(\frac{1}{n}\right)^{A_\alpha} \leq \left(\frac{B^2}{n}\right)^{A_\alpha} \leq C_{\mathrm{off}} \left(\frac{B^2}{n}\right)^{A_\alpha} (\log n)^p.$$

Step 2 ($\gamma > 0$): the samplewise selection inequality. Assume $\gamma > 0$, $0 < u_n \leq u_0$, and $q_n \leq c_o u_n^\gamma$; write $u = u_n$ and $R = R_P(\widehat{\pi}_n)$. For every $\pi \in \Pi$, applying Lemma 12 foldwise with the bounded measurable test function $\varphi(x) = \mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi^\star_P(x) = 1\}$ gives the foldwise drift decomposition. The contrast part is identified from the definitions of welfare, oracle policy, and regret in Definition 1:

$$\int_{\mathcal{X}} \varphi(x) \tau_P(x)\, dP_X(x) = V_P(\pi) - V_P(\pi^\star_P) = -R_P(\pi).$$

Combining these foldwise identities with the definition of G_{cf} gives the exact empirical-welfare identity

$$\widehat{V}_{n,q}(\pi) - \widehat{V}_{n,q}(\pi^\star_P) = G_{\mathrm{cf}}(\pi) - R_P(\pi) + \sum_{k=1}^K w_k \Delta_k(\pi).$$

By Assumption 5 the oracle $\pi_P^\star$ lies in Π, so Lemma 5 applies with comparator $\pi^b = \pi_P^\star$ and gives $\widehat{V}_{n,q}(\pi_P^\star) - \widehat{V}_{n,q}(\widehat{\pi}_n) \le 1/n$. Evaluating the identity at $\pi = \widehat{\pi}_n$ and bounding each term by its absolute value yields the samplewise selection bound

$$R \le |G_{\mathrm{cf}}(\widehat{\pi}_n)| + \left| \sum_{k=1}^{K} w_k \Delta_k(\widehat{\pi}_n) \right| + \frac{1}{n}.$$

The weighted drift sum is controlled foldwise. Each fold's nuisance triple satisfies the same $L^2(P_X)$ error bounds $r_{\mu,n}, r_{e,n}$, so Lemma 7 applies to each $\Delta_k(\widehat{\pi}_n)$ with the localization window u and the admissible clip $q \le c_o u^\gamma$; since $w_k \ge 0$ and $\sum_k w_k = 1$, averaging preserves the bound:

$$\left| \sum_{k=1}^{K} w_k \Delta_k(\widehat{\pi}_n) \right| \le \sum_{k=1}^{K} w_k \left| \Delta_k(\widehat{\pi}_n) \right| \le C_d \left\{ \frac{r_{\mu,n} r_{e,n}}{q} + r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)} + r_{\mu,n} \left(\frac{R}{u} \right)^{1/2} \right\}.$$

The last summand still involves R, and is absorbed by Young's inequality $2xy \le x^2 + y^2$ applied with $x = \sqrt{R}$ and $y = C_d r_{\mu,n}/\sqrt{u}$:

$$C_d r_{\mu,n} \left(\frac{R}{u} \right)^{1/2} \le \frac{R}{2} + \frac{C_d^2 r_{\mu,n}^2}{2u}.$$

Substituting the last two displays into the selection bound and moving $R/2$ to the left,

$$\frac{R}{2} \le |G_{\mathrm{cf}}(\widehat{\pi}_n)| + C_d \left\{ \frac{r_{\mu,n} r_{e,n}}{q} + r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)} \right\} + \frac{C_d^2 r_{\mu,n}^2}{2u} + \frac{1}{n},$$

that is,

$$R_P(\widehat{\pi}_n) \le 2\,|G_{\mathrm{cf}}(\widehat{\pi}_n)| + \delta, \qquad \delta = 2 \left[C_d \left\{ \frac{r_{\mu,n} r_{e,n}}{q} + r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)} \right\} + \frac{C_d^2 r_{\mu,n}^2}{2u} + \frac{1}{n} \right] \quad (\ge 0).$$

Step 3 ($\gamma > 0$): self-bounding and algebra. The samplewise inequality just proved is exactly the input of Lemma 10, applied with envelope $B \ge 1$, offset bound $C_{\mathrm{off}}(B^2/n)^{A_\alpha}(\log n)^p$ from the offset step, and slack δ. It gives

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \le \frac{8}{3} \left\{ C_{\mathrm{off}} \left(\frac{B^2}{n} \right)^{A_\alpha} (\log n)^p + \delta \right\}.$$

Since $B = 36/q$, the envelope factor is exactly the clipped empirical-process term:

$$\left(\frac{B^2}{n} \right)^{A_\alpha} = \left(\frac{36^2}{nq^2} \right)^{A_\alpha} = K_{\mathrm{pos}}\,(nq^2)^{-A_\alpha}.$$

Now abbreviate the five nonnegative target terms

$$T_1 = n^{-r_\star(\alpha,\gamma)}, \quad T_2 = (nq^2)^{-A_\alpha}, \quad T_3 = \frac{r_{\mu,n} r_{e,n}}{q}, \quad T_4 = r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)}, \quad T_5 = \frac{r_{\mu,n}^2}{u},$$

and write $L = (\log n)^p \ge 1$. Then $r_{\mu,n}^2/(2u) = T_5/2$, and the large-n bound above reads $1/n \le C_{\mathrm{off}} K_{\mathrm{pos}} T_2 L$. Expanding δ, the previous bound becomes

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \le \frac{8}{3} \left\{ C_{\mathrm{off}} K_{\mathrm{pos}} T_2 L + 2 \left(C_d (T_3 + T_4) + C_d^2 \frac{T_5}{2} + \frac{1}{n} \right) \right\} \le 8\, C_{\mathrm{off}} K_{\mathrm{pos}} T_2 L + \frac{16}{3} C_d T_3 + \frac{16}{3} C_d T_4 + \frac{8}{3} C_d^2 T_5.$$

Each coefficient $8C_{\text{off}}K_{\text{pos}}$, $\frac{16}{3}C_d$, $\frac{8}{3}C_d^2$ is at most C by the definition of C; using $L \geq 1$ to insert the factor L into the last three terms, and adding the nonnegative term CT_1L, gives

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C\,(T_1 + T_2 + T_3 + T_4 + T_5)\,L = C\left\{n^{-r_\star(\alpha,\gamma)} + (nq_n^2)^{-A_\alpha} + \frac{r_{\mu,n}r_{e,n}}{q_n} + r_{\mu,n}u_n^{\alpha/2}q_n^{1/(2\gamma)} + \frac{r_{\mu,n}^2}{u_n}\right\}(\log$$

This proves the positive-γ branch.

Step 4 ($\gamma = 0$). Now assume $\gamma = 0$ and $q_n \leq \underline{p}/2$; on the eventual set considered, $q = q_n = q_{\text{fix}}$ is the fixed clip, so $B = 36/q_{\text{fix}}$ is fixed in n. The empirical-welfare identity and Lemma 5 give the same samplewise selection bound as above, and the strict-overlap branch of Lemma 7, averaged over the fold weights exactly as before, gives

$$\left|\sum_{k=1}^{K} w_k \Delta_k(\widehat{\pi}_n)\right| \leq C_q\, r_{\mu,n}r_{e,n}.$$

Hence, using $|G_{\text{cf}}| \leq 2|G_{\text{cf}}|$,

$$R_P(\widehat{\pi}_n) \leq 2\,|G_{\text{cf}}(\widehat{\pi}_n)| + \delta, \qquad \delta = 2\left(C_q r_{\mu,n}r_{e,n} + \frac{1}{n}\right),$$

and Lemma 10 gives

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq \frac{8}{3}\left\{C_{\text{off}}\left(\frac{B^2}{n}\right)^{A_\alpha}(\log n)^p + \delta\right\}.$$

With the fixed clip, the envelope factor is a constant times the parametric-envelope rate,

$$\left(\frac{B^2}{n}\right)^{A_\alpha} = \left(\left(\frac{36}{q_{\text{fix}}}\right)^2\right)^{A_\alpha} n^{-A_\alpha} = K_{\text{zero}}\, n^{-A_\alpha}.$$

Writing $Z_1 = n^{-A_\alpha} \geq 0$ and $Z_2 = r_{\mu,n}r_{e,n} \geq 0$, the large-$n$ bound above gives $1/n \leq C_{\text{off}}K_{\text{zero}}Z_1L$, so

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq \frac{8}{3}\left\{C_{\text{off}}K_{\text{zero}}Z_1L + 2\left(C_qZ_2 + \frac{1}{n}\right)\right\} \leq 8\,C_{\text{off}}K_{\text{zero}}Z_1L + \frac{16}{3}C_qZ_2.$$

Since $8C_{\text{off}}K_{\text{zero}} \leq C$ and $\frac{16}{3}C_q \leq C$, and $L \geq 1$,

$$\mathbb{E}_P R_P(\widehat{\pi}_n) \leq C\left\{n^{-A_\alpha} + r_{\mu,n}r_{e,n}\right\}(\log n)^p,$$

which is the strict-overlap branch. □

Proof of Lemma 7. Fix the evaluation fold and abbreviate the foldwise estimators by $(\widehat{\mu}_0, \widehat{\mu}_1, \widehat{e})$. Write

$$\bar{e}_q(x) = \min\{1-q, \max\{q, \widehat{e}(x)\}\}, \qquad \Delta_a(x) = \widehat{\mu}_a(x) - \mu_a(x), \qquad \Delta_e(x) = \widehat{e}(x) - e_P(x),$$

and, for a policy π,

$$T_\pi(q) = D_\pi \cap \{x : p_P(x) \leq q\}, \qquad P[(\pi - \pi_P^\star)b_q] = \int_{\mathcal{X}} \{\mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi_P^\star(x) = 1\}\}\, b_q(x)\, dP_X(x).$$

By the exact drift identity of Lemma 12, the conditional-mean drift defined in the statement has the closed form

$$b_q(x) = \{\bar{e}_q(x) - e_P(x)\}\left\{\frac{\Delta_1(x)}{\bar{e}_q(x)} + \frac{\Delta_0(x)}{1-\bar{e}_q(x)}\right\}.$$

Step 1: a pointwise majorization. We bound the two factors of b_q separately.

First, the clipping map $z \mapsto \min\{1-q, \max\{q,z\}\}$ is 1-Lipschitz, being a composition of the 1-Lipschitz maps $z \mapsto \max\{q,z\}$ and $z \mapsto \min\{1-q,z\}$; applying it to $\widehat{e}(x)$ and to $e_P(x)$ therefore gives

$$|\bar{e}_q(x) - \min\{1-q, \max\{q, e_P(x)\}\}| \le |\Delta_e(x)|.$$

Second, clipping the true propensity moves it only inside the weak-overlap region: if $p_P(x) > q$, then $q < e_P(x) < 1-q$ and the clip is inactive, so the displacement is zero; and in all cases the displacement is at most q, because $e_P(x) \in [0,1]$ and the clipped value lies in $[q, 1-q]$. Hence

$$|\min\{1-q, \max\{q, e_P(x)\}\} - e_P(x)| \le q\,\mathbf{1}\{p_P(x) \le q\},$$

and by the triangle inequality

$$|\bar{e}_q(x) - e_P(x)| \le |\Delta_e(x)| + q\,\mathbf{1}\{p_P(x) \le q\}.$$

Third, since $q \le \bar{e}_q(x) \le 1-q$, both denominators are at least $q > 0$, so

$$\left|\frac{\Delta_1(x)}{\bar{e}_q(x)} + \frac{\Delta_0(x)}{1-\bar{e}_q(x)}\right| \le \frac{|\Delta_1(x)| + |\Delta_0(x)|}{q}.$$

Multiplying the last two displays,

$$|b_q(x)| \le \{|\Delta_e(x)| + q\,\mathbf{1}\{p_P(x) \le q\}\} \cdot \frac{|\Delta_1(x)| + |\Delta_0(x)|}{q}.$$

Finally, the policy factor satisfies $|\mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi^\star_P(x) = 1\}| \le 1$, and it vanishes off D_π; consequently

$$|\mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi^\star_P(x) = 1\}| \cdot \mathbf{1}\{p_P(x) \le q\} \le \mathbf{1}\{x \in T_\pi(q)\}.$$

Combining the last two displays and cancelling the factor q against q^{-1} in the indicator term gives the pointwise majorization

$$|\{\mathbf{1}\{\pi(x) = 1\} - \mathbf{1}\{\pi^\star_P(x) = 1\}\}\, b_q(x)| \le \frac{|\Delta_e(x)||\Delta_1(x)| + |\Delta_e(x)||\Delta_0(x)|}{q} + \mathbf{1}\{x \in T_\pi(q)\}\,\{|\Delta_1(x)| + |\Delta_0(x)|\}.$$

Step 2: integrate. All four majorant terms are P_X-integrable, being products of $L_2(P_X)$ functions (the assumed L_2 memberships of the three nuisance errors, and the indicator, which is bounded). Integrating and applying Cauchy–Schwarz with the assumed $L_2(P_X)$ rates gives, for $a = 0, 1$,

$$\int |\Delta_e|\,|\Delta_a|\,dP_X \le \left(\int \Delta_e^2\,dP_X\right)^{1/2}\left(\int \Delta_a^2\,dP_X\right)^{1/2} \le r_{e,n} r_{\mu,n},$$

and, again by Cauchy–Schwarz with the indicator in the first slot,

$$\int \mathbf{1}_{T_\pi(q)} \, |\Delta_a| \, dP_X \leq P_X(T_\pi(q))^{1/2} \, r_{\mu,n}.$$

Summing the four contributions yields the basic drift bound

$$|P[(\pi - \pi_P^\star)b_q]| \leq 2 \, \frac{r_{\mu,n} r_{e,n}}{q} + 2 \, r_{\mu,n} \, P_X(T_\pi(q))^{1/2} \, .$$

Step 3: the case $\gamma > 0$. Assume $\gamma > 0$, $0 < u \leq u_0$, and $q \leq c_o u^\gamma$, and write $r = R_P(\pi)$ and $C_{\mathrm{reg}} = \max\{C_o, 1\} > 0$. By Lemma 13,

$$P_X(T_\pi(q)) \leq C_{\mathrm{reg}} \, u^\alpha q^{1/\gamma} + \frac{r}{u}.$$

Both summands are nonnegative, so taking square roots and using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, together with the exponent identity

$$\sqrt{C_{\mathrm{reg}} \, u^\alpha q^{1/\gamma}} = C_{\mathrm{reg}}^{1/2} \, u^{\alpha/2} \, q^{1/(2\gamma)},$$

gives

$$P_X(T_\pi(q))^{1/2} \leq C_{\mathrm{reg}}^{1/2} \, u^{\alpha/2} q^{1/(2\gamma)} + \left(\frac{r}{u}\right)^{1/2} .$$

Substituting into Step 2 and abbreviating the four nonnegative quantities

$$J_1 = \frac{r_{\mu,n} r_{e,n}}{q}, \qquad J_2 = u^{\alpha/2} q^{1/(2\gamma)}, \qquad J_3 = \left(\frac{r}{u}\right)^{1/2}, \qquad S = C_{\mathrm{reg}}^{1/2} = \sqrt{\max\{C_o, 1\}},$$

we get $|P[(\pi - \pi_P^\star)b_q]| \leq 2J_1 + 2r_{\mu,n}(SJ_2 + J_3) \leq 4\,(J_1 + r_{\mu,n} S J_2 + r_{\mu,n} J_3)$, all terms being nonnegative. Since $J_1 \leq (1+S)J_1$, $S r_{\mu,n} J_2 \leq (1+S) r_{\mu,n} J_2$, and $r_{\mu,n} J_3 \leq (1+S) r_{\mu,n} J_3$,

$$4\,(J_1 + r_{\mu,n} S J_2 + r_{\mu,n} J_3) \leq 4 \left(1 + \sqrt{\max\{C_o, 1\}}\right) (J_1 + r_{\mu,n} J_2 + r_{\mu,n} J_3) \, ,$$

that is, with $C_0 = 4\{1 + \sqrt{\max\{C_o, 1\}}\}$ exactly as in the statement,

$$|P[(\pi - \pi_P^\star)b_q]| \leq C_0 \left\{ \frac{r_{\mu,n} r_{e,n}}{q} + r_{\mu,n} u^{\alpha/2} q^{1/(2\gamma)} + r_{\mu,n} \left(\frac{r}{u}\right)^{1/2} \right\} .$$

Step 4: the case $\gamma = 0$. Assume $\gamma = 0$ and $q \leq \underline{p}/2$. Assumption 10 gives $\underline{p} > 0$ and $p_P(X) \geq \underline{p}$ P_X-almost surely. Since $q \leq \underline{p}/2 < \underline{p}$, the set $\{p_P \leq q\}$ is P_X-null, hence so is its subset $T_\pi(q)$:

$$P_X(T_\pi(q)) = 0.$$

The second term of the basic drift bound of Step 2 therefore vanishes, leaving

$$|P[(\pi - \pi_P^\star)b_q]| \leq 2 \, \frac{r_{\mu,n} r_{e,n}}{q} = \frac{2}{q} \, r_{\mu,n} r_{e,n} \leq C_1 \, r_{\mu,n} r_{e,n}, \qquad C_1 = \max\{1, 2/q\},$$

the last step because $2/q \leq \max\{1, 2/q\}$ and $r_{\mu,n} r_{e,n} \geq 0$. This is the asserted strict-overlap bound. □

References

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021. doi: 10.3982/ECTA15732.

Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *The Annals of Statistics*, 35(2):608–633, 2007. doi: 10.1214/009053606000001217.

Bartlett. Local rademacher complexities, 2005. Bartlett, P. L., Bousquet, O., and Mendelson, S. (2005). Local Rademacher complexities. Annals of Statistics 33(4), 1497-1537.

Eli Ben-Michael and Luke Keele. Using balancing weights to target the treatment effect on the treated when overlap is poor, 2022. URL https://arxiv.org/abs/2210.01763.

Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, 1993.

Minshuo Chen, Hao Liu, Wenjing Liao, and Tuo Zhao. Doubly robust off-policy learning on low-dimensional manifolds by deep neural networks, 2020. URL https://arxiv.org/abs/2011.01797.

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.

Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins. Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535, 2022. doi: 10.3982/ECTA16294.

Alexander D'Amour, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. Overlap in observational studies with high-dimensional covariates, 2017. URL https://arxiv.org/abs/1711.02582.

Miroslav Dudik, John Langford, and Lihong Li. Doubly robust policy evaluation and learning, 2011. URL https://arxiv.org/abs/1103.4601.

Samuel Girard, Aurelien Bibaut, Arthur Gretton, Nathan Kallus, and Houssam Zenati. Fast best-in-class regret for contextual bandits, 2025. URL https://arxiv.org/abs/2510.15483.

Jinyong Hahn. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331, 1998. doi: 10.2307/2998560.

Jonathan B. Hill and Saraswata Chaudhuri. Heavy tail robust estimation and inference for average treatment effects, 2024. URL https://arxiv.org/abs/2412.08458.

Keisuke Hirano, Guido W. Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003. doi: 10.1111/1468-0262.00442.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge, 2015. doi: 10.1017/CBO9781139025751.

Nathan Kallus. Balanced policy evaluation and learning, 2017. URL https://arxiv.org/abs/1705.07384.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018. doi: 10.3982/ECTA13288.

Toru Kitagawa, Weining Wang, and Mengshan Xu. Policy choice in time series by empirical welfare maximization, 2022. URL https://arxiv.org/abs/2205.03970.

Lucien Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York, 1986.

Fan Li, Kari Lock Morgan, and Alan M. Zaslavsky. Balancing covariates via propensity score weighting, 2016. URL https://arxiv.org/abs/1609.07494.

Jingren Liu, Hanzhang Qin, Junyi Liu, Mabel C. Chou, and Jong-Shi Pang. Offline policy learning with weight clipping and heaviside composite optimization, 2026. URL https://arxiv.org/abs/2601.12117.

Alexander Luedtke and Antoine Chambaz. Faster rates for policy learning, 2017. URL https://arxiv.org/abs/1704.06431.

Charles F. Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72 (4):1221–1246, 2004. doi: 10.1111/j.1468-0262.2004.00530.x.

Charles F. Manski. *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA, 2009.

Pascal Massart and Élodie Nédélec. Risk bounds for statistical learning. *The Annals of Statistics*, 34(5):2326–2366, 2006. doi: 10.1214/009053606000000786.

Whitney K. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, 62 (6):1349–1382, 1994. doi: 10.2307/2951752.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2017.

Min Qian and Susan A. Murphy. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210, 2011. doi: 10.1214/10-AOS864.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818.

Shosei Sakaguchi. Policy learning for optimal dynamic treatment regimes with observational data, 2024. URL https://arxiv.org/abs/2404.00221.

Yuya Sasaki and Takuya Ura. Welfare analysis via marginal treatment effects, 2020. URL https://arxiv.org/abs/2012.07624.

Jörg Stoye. Minimax regret treatment choice with finite samples. *Econometrica*, 77(1):223–251, 2009.

Liyang Sun. Empirical welfare maximization with constraints, 2021. URL https://arxiv.org/abs/2103.15298.

Herbert P. Susmann, Alec McClean, and Iván Díaz. Non-overlap average treatment effect bounds, 2025. URL https://arxiv.org/abs/2509.20206.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback, 2015. URL https://arxiv.org/abs/1502.02362.

Tsybakov. Optimal aggregation of classifiers in statistical learning, 2004. Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. Annals of Statistics 32(1), 135-166.

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

Mark J. van der Laan and Sherri Rose. *Targeted Learning*. Springer, New York, 2011. doi: 10.1007/978-1-4419-9782-1.

Ruohan Zhan, Zhimei Ren, Susan Athey, and Zhengyuan Zhou. Policy learning with adaptively collected data, 2021. URL https://arxiv.org/abs/2105.02344.

Baqun Zhang, Anastasios A. Tsiatis, Marie Davidian, Min Zhang, and Eric B. Laber. Estimating optimal treatment regimes from a classification perspective. *Stat*, 1(1):103–114, 2012. doi: 10.1002/sta4.11.

Pan Zhao, Antoine Chambaz, Julie Josse, and Shu Yang. Positivity-free policy learning with observational data, 2023. URL https://arxiv.org/abs/2310.06969.

Yingqi Zhao, Donglin Zeng, A. John Rush, and Michael R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012. doi: 10.1080/01621459.2012.695674.