FOUR-CERTIFICATE MINIMAX RATES OVER PUBLIC CLASS PAIRS FOR BOUNDED PARTIALLY LINEAR MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study estimation of a bounded partially linear coefficient from two public nuisance classes with separate approximation and stochastic complexity certificates. In the strict envelope interior, the worst-public-pair expected absolute risk for $(a, s, b, t) \in [0, 1]^4$ has rate $n^{-1/2} + (a + s^2)(b + t^2) + (s \wedge t)^2 + t(a \wedge t) + s(b \wedge s)$. Matching directional lower bounds close the gap between the Gu–Yin–Cai–Fan lower envelope and TAME's calibration upper. We also construct one rule attaining this rate without any of the four budget inputs. Double calibration removes approximation-dependent orientation, and empirical local Rademacher fixed points calibrate the stochastic radii. The general rate persists over full finite dictionary ℓ_1 balls and finite unions of rank-one balls. Full interval boxes instead have sharp rate $n^{-1/2} + (a + s^2)(b + t^2)$, and full affine Hilbert balls have rate $n^{-1/2} + ab$; both are attained without any budget input. The model requires only positive average residual variance and bounded envelopes. Results include all sample sizes and zero budgets; a uniform upper connects to saturation. The guarantees are statistical; effective optimization requires additional interfaces.

1 INTRODUCTION

Estimating a partially linear coefficient involves two sources of nuisance error: approximation error from the choice of function class and stochastic error from learning within that class. Each nuisance contributes one of each, giving four error budgets. We determine their joint minimax rate and construct a single estimator that attains it without knowing any of the budgets.

The existing lower and upper bounds differ in their treatment of directional interactions. Write $q = n^{-1/2}$, $A = a + s^2$, $B = b + t^2$, and $M = (s \wedge t)^2$, where a, b are approximation budgets and s, t are stochastic budgets. Gu–Yin–Cai–Fan establish the lower envelope $q + AB + M$ (Gu et al., 2026). TAME's two-learner calibration method gives the leading upper bound

$$q + W, \qquad W = \min\{a(b + t) + (a + s)^2,\ b(a + s) + (b + t)^2\}$$

(Gu, 2026). We close the gap for bounded public classes by proving lower bounds for both directional terms

$$O = t(a \wedge t), \qquad P = s(b \wedge s).$$

Since $W \asymp AB + M + O + P$, the two-branch rate is minimax optimal in this experiment. TAME selects its branch using the public budgets (Gu, 2026, equation (3.6)); our adaptive estimator removes that requirement.

The gap is polynomial: on the O ray $(a, s, b, t) = (r, 0, 0, r)$, the previous nuisance lower bound is r^3, whereas the matching rate is of order r^2. Taking $r = n^{-1/8}$ makes $q = o(r^3)$. Here the aggregate-error product $xy = (a + s)(b + t)$ also equals r^2, so this example isolates the missing lower-bound term.

Our second main result is simultaneous adaptation to all four budgets (Section 5). A single rule attains $q + W$ for every represented M1 class pair. Two comparisons with the true propensity yield the two directional bounds for this rule, while an independent covariate block estimates the complexity radii. Building on double calibration and empirical local complexity (Liu et al., 2023; Bartlett et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2005), we establish a uniform expected-loss bound. The matching lower bound also holds for every preassigned atomless covariate marginal.

The directional terms persist under convexity: full finite dictionary ℓ_1 balls have the same minimax rate. In contrast, full interval boxes and affine Hilbert balls have sharp rates $q + AB$ and $q + ab$, attained without budget inputs (Section 6). On the ray above, these three worst-class risks are $q + r^2$, $q+r^3$, and q under the same budget constraints. The envelope explains why a Hilbert enclosure cannot transfer the dictionary lower bound unchanged: a full Hilbert ball containing $\{0, \pm h\phi_j : j \le p\}$, with orthonormal sign functions ϕ_j, must have pointwise radius at least $h\sqrt{p}$.

2 Experiment and four public certificates

Fix a nonempty infinite standard Borel space $(\mathcal{X}, \mathcal{A})$, shared by every law and public class below. Observe independent halves $\mathcal{D}_1, \mathcal{D}_2$, each containing n iid copies of $Z = (X, T, Y)$, with

$$T = \pi_0(X) + u, \qquad Y = \beta_0 T + \mu_0(X) + \varepsilon, \tag{1}$$

where $\mu_0, \pi_0 : \mathcal{X} \to \mathbb{R}$ are measurable, $\mathbb{E}(u \mid X) = 0$, $\mathbb{E}(\varepsilon \mid X, T) = 0$, and $v(P) := \mathbb{E}u^2 \in [v_-, v_+]$. Fix $\rho = (B_0, B_{\mathcal{G}}, v_-, v_+, c_{\mathrm{rad}})$, with $B_0 \ge 3$, $B_{\mathcal{G}} > 0$, $0 < v_- \le v_+$, and $c_{\mathrm{rad}} > 0$. Require

$$|Y|, |T|, |\mu_0(X)|, |\pi_0(X)|, |u|, |\varepsilon| \le B_0, \qquad \beta_0 \in [-3, 3].$$

Positive average residual variance identifies the coefficient: $\mathbb{E}(uY) = \beta_0 \mathbb{E}u^2$. Conditional treatment variance may vary with X and vanish on a set of positive mass.

A public instance $\Gamma = (\mathcal{G}_\mu, \mathcal{G}_\pi, \mathbf{g}_\mu, \mathbf{g}_\pi)$ supplies two classes and ordered countable cores with measurable evaluations. Every full-class and represented function obeys $|g(x)| \le B_{\mathcal{G}}$ for every x. Our sequential M1 convention requires every full-class element to be the pointwise-everywhere limit of a sequence from its core. The common envelope transfers population approximation infima by dominated convergence and continuous finite-sample objectives by pointwise convergence.

For $\|f\|_2 = \|f\|_{L_2(P_X)}$, the approximation certificates are

$$\inf_{g \in \mathcal{G}_\mu} \|g - \mu_0\|_2 \le a, \qquad \inf_{g \in \mathcal{G}_\pi} \|g - \pi_0\|_2 \le b. \tag{2}$$

Define $\Delta\mathcal{G} = \{g - g' : g, g' \in \mathcal{G}\}$ and the full-class localized complexity

$$\mathfrak{R}^*_{n,P_X}(r; \mathcal{H}) = \mathbb{E}^*_{X_{1:n}, \epsilon_{1:n}} \sup_{h \in \mathcal{H} : \|h\|_2 \le r} \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i h(X_i) \right|.$$

Here the ϵ_i are independent Rademacher signs, independent of $X_{1:n} \sim P_X^n$, and outer expectation is the infimum of expectations of measurable majorants. The stochastic certificates are

$$\sup_{r \ge r_\nu,\, r > 0} \frac{\mathfrak{R}^*_{n,P_X}(r; \Delta\mathcal{G}_\nu)}{3r} \le c_{\mathrm{rad}} r_\nu, \qquad (\nu, r_\nu) \in \{(\mu, s), (\pi, t)\}.$$

At zero budget, complexity vanishes at every positive radius; the definition uses only positive-radius quotients. The represented difference core is countable, so its supremum is measurable and its complexity is bounded by the full outer complexity. This inequality is sufficient for the upper bounds; M1 need not give equality.

Let $\mathcal{P}(\Gamma) = \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$ be all probability laws on $\mathcal{X} \times \mathbb{R}^2$ satisfying these model, envelope, approximation, and complexity conditions, and let $\mathfrak{G}_{\rho,n}(a, s, b, t)$ contain the represented pairs with $\mathcal{P}(\Gamma) \ne \varnothing$. For $(a, s, b, t) \in [0, 1]^4$, define

$$R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) = \sup_{\Gamma \in \mathfrak{G}_{\rho,n}(a,s,b,t)} \inf_{\delta_\Gamma \text{ Borel}} \sup_{P \in \mathcal{P}(\Gamma)} \mathbb{E}_P |\delta_\Gamma(Z_{1:2n}) - \beta(P)|. \tag{3}$$

In (3), the budgets are known upper bounds supplied before the data. The public pair may depend on n and the budget tuple, but is fixed before Nature chooses a compatible law or hidden label and generates the sample. The Borel rule knows Γ, ρ and the budgets, and is fixed before the law. Thus each pair has its own rule on the common ambient space. The results cover every $n \ge 1$ and all zero-budget boundaries; Appendix A gives the measurable conventions.

3 SHARP RATE AND THE ENVELOPE BOUNDARY

Put

$$q = n^{-1/2}, \qquad A = a + s^2, \qquad B = b + t^2,$$
$$M = (s \wedge t)^2, \qquad O = t(a \wedge t), \qquad P = s(b \wedge s), \tag{4}$$

and write $X = AB + M + O + P$ for the rate polynomial (distinct from the covariate).

Theorem 1 (Three-stratum minimax law)**.** *Fix a nonempty infinite standard Borel measurable space $(\mathcal{X}, \mathcal{A})$, and fix $B_0 \geq 3$, $B_{\mathcal{G}} > 0$, $0 < v_- \leq v_+$, and $c_{\mathrm{rad}} > 0$. Use the sequential M1 interface above.*

E. *If $v_- > B_0^2$, the experiment is empty and no minimax rate is assigned.*

I. *If $0 < v_- < B_0^2$, there are $0 < c_\rho \leq C_\rho < \infty$, depending only on fixed ρ, such that, for every $n \geq 1$ and budget tuple,*

$$c_\rho(q + X) \leq R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \leq C_\rho(q + X). \tag{5}$$

One valid lower constant is one tenth the minimum of the five component constants in Appendix G.

S. *If $v_- = B_0^2$, the upper bound holds uniformly for every legal nonempty public pair, while one explicit legal singleton pair witnesses the outer lower bound; consequently*

$$\frac{3}{128\sqrt{n}} \leq R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \leq \frac{1}{\sqrt{2n}}. \tag{6}$$

The matching constants in I are for fixed ρ; in particular the lower constants need not be uniform as $v_- \uparrow B_0^2$ or $c_{\mathrm{rad}} \downarrow 0$.

Average variance and the envelope boundary. For an explicit zero-variance region, put mass $p = v_-/B_0^2 \in (0, 1)$ on one of two covariate points, let $T = \pm B_0$ equiprobably there and $T = 0$ elsewhere, and take $\pi_0 = \mu_0 = \beta_0 = Y = 0$ and both classes $\{0\}$. Then $\mathbb{E}u^2 = v_-$, while conditional variance is zero on the second point.

Proposition 2 (Uniform one-sided envelope bound)**.** *Let $N = 2n$, let $P_N f = N^{-1} \sum_{i=1}^N f(Z_i)$ denote the empirical average over all observations, assume $0 < v_- \leq B_0^2$, and set $\omega = 1 - v_-/B_0^2$. For every legal public pair and compatible law, the fixed-denominator estimator*

$$\widehat{\beta}_{\mathrm{env}} = \Pi_{[-3,3]}\{P_N(TY)/B_0^2\} \quad \textit{satisfies} \quad \mathbb{E}_P|\widehat{\beta}_{\mathrm{env}} - \beta(P)| \leq N^{-1/2} + 4\omega.$$

The numerical constants are uniform, including in B_0; this estimator uses neither v_+, the four budgets, nor M1.

The envelope inequality $\mathbb{E}(u^2 \mid X) \leq B_0(B_0 - |\pi_0|)$ implies $B_0\mathbb{E}|\pi_0| \leq B_0^2 - v_-$; this controls the population bias by 4ω. Appendix H gives the proof and the saturation lower bound. In the strict interior, comparison of the two deterministic upper guarantees gives

$$R^{\mathrm{pair}}_{n,\rho} \leq \min\{C_\rho(q + X),\ q/\sqrt{2} + 4\omega\}.$$

This upper bound remains valid through saturation. The interior lower construction requires positive slack in $B_0 - \sqrt{v_-}$, so a matching joint transition rate remains open.

4 DIRECTIONAL LOWER BOUNDS

Work in $0 < v_- < B_0^2$. We vary the treatment mean while keeping its residual variance fixed. Set

$$m_\rho = \min\left\{\frac{1}{64}, \frac{v_-}{4B_0}, \frac{B_0 - \sqrt{v_-}}{4}\right\}, \quad \lambda_\rho = \min\left\{1, \frac{192c_{\mathrm{rad}}}{23}, 128B_{\mathcal{G}}\right\}, \quad L = \sqrt{v_-} + 2m_\rho.$$

For $|m| \leq m_\rho$, assign treatment masses

$$p_{\pm L}(m) = \frac{v_- + m^2 \pm Lm}{2L^2}, \qquad p_0(m) = 1 - \frac{v_- + m^2}{L^2}. \tag{7}$$

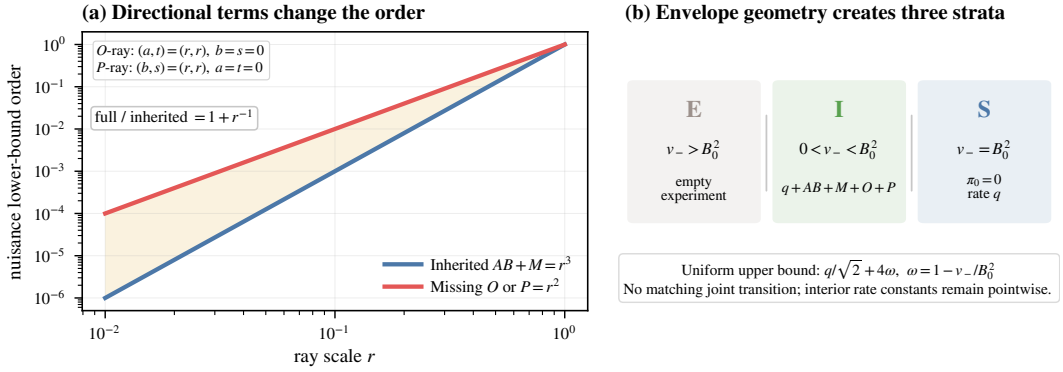


Figure 1: **Directional lower bounds and the envelope boundary.** On the O ray $(a, s, b, t) = (r, 0, 0, r)$ and its P mirror, the previous nuisance expression is r^3, whereas the matching expression is $r^3 + r^2$, a factor of $1 + r^{-1}$ larger. Taking $r = n^{-1/8}$ keeps $q = o(r^3)$. The product benchmark xy also has order r^2. The uniform upper bound $q/\sqrt{2} + 4\omega$, where $\omega = 1 - v_-/B_0^2$, reaches the parametric rate at saturation.

Indeed $Lm_\rho \le v_-/4$, $v_- + m_\rho^2 < L^2$, and $L + m_\rho < B_0$; thus every mass is positive and both treatment and residual envelopes hold. Direct summation gives

$$\mathbb{E}(T \mid X) = m, \qquad \mathbb{E}(T^2 \mid X) = v_- + m^2, \qquad \mathrm{Var}(T \mid X) = v_-.$$

Let $A_Y = B_0/2$. Under each label ℓ, use this treatment channel at $m = \pi_\ell(X)$ and set $Y = A_Y R$, with

$$\Pr_\ell(R = r \mid X, T = \zeta) = \frac{1}{2}\left[1 + r\frac{\beta_\ell \zeta + \mu_\ell(X)}{A_Y}\right], \qquad r \in \{-1, 1\}. \tag{8}$$

The two channels define a full observation law satisfying the PLM conditional-mean restriction.

Two public dictionaries and their labels. Let $N = 2n$, and initially suppose $n \ge 16$, $\nu^2 \ge q$, with $\nu = s$ for P and $\nu = t$ for O, and a positive directional term. Set $p = \lceil \exp(64N\nu^2) \rceil$. Embed a finite binary group with at least p nonzero Walsh characters into the fixed $\mathcal{X}$, let P_X be uniform on its image, and extend the selected characters ϕ_j by zero elsewhere. They are measurable, mean zero and orthonormal under P_X. Write $z = \sigma\phi_j(X)$.

For P, take

$$\eta = \frac{m_\rho(b \wedge s)}{4}, \quad \gamma = \frac{\lambda_\rho s}{256}, \quad \Delta = \frac{2\eta\gamma}{v_-}, \quad \kappa = \gamma + \Delta\eta.$$

For O, take

$$\gamma = \frac{\lambda_\rho m_\rho t}{4}, \quad \eta = \frac{a \wedge t}{256}, \quad \Delta = \frac{\gamma\eta}{v_- + \gamma^2}.$$

Each dictionary contains the listed functions for every j and both signs, and is fixed before a label is selected:

Branch	Label 0: (β, π, μ)	Label 1: (β, π, μ)	Public classes $(\mathcal{G}_\mu, \mathcal{G}_\pi)$
P	$(0, z\eta, z\gamma)$	$(\Delta, -z\eta, z\kappa)$	$(\{\pm\gamma\phi_j, \pm\kappa\phi_j\}, \{0\})$
O	$(0, z\gamma, z\eta)$	$(\Delta, z\gamma, 0)$	$(\{0\}, \{\pm\gamma\phi_j\})$

In P, the propensity approximation error is $\eta \le b$ and the outcome truth is represented. In O, the outcome approximation error is at most $\eta \le a$ and the propensity is represented. The remaining approximation distances and difference-class complexities are zero. The amplitudes give $\kappa \le 2\gamma$ and $\Delta L + \kappa < A_Y$ in P, and $\Delta L \le \eta/4 < A_Y$ in O. Consequently the response probabilities are positive, $|\varepsilon| \le B_0$, $\beta \in [-3, 3]$, and both public envelopes hold (Appendices C and D). Both labels have variance exactly v_-, including when $v_- = v_+$.

Atomwise cancellation. Sign averaging cancels the atomwise differences between the two labels. For either branch write $d = \eta$ in P and $d = \gamma$ in O, and define

$$\alpha_{\pm L} = \frac{v_- + d^2}{2L^2}, \quad \alpha_0 = 1 - \frac{v_- + d^2}{L^2}, \quad b_\zeta = \frac{d\zeta}{2L^2}.$$

Conditional on X, the six-atom masses from (7)–(8) are

$$\begin{aligned}
\text{P: } P_{0,z}(\zeta, r) &= \tfrac{1}{2}(\alpha_\zeta + z b_\zeta)(1 + rz\gamma/A_Y), \\
P_{1,z}(\zeta, r) &= \tfrac{1}{2}(\alpha_\zeta - z b_\zeta)(1 + r(\Delta\zeta + z\kappa)/A_Y); \\
\text{O: } P_{0,z}(\zeta, r) &= \tfrac{1}{2}(\alpha_\zeta + z b_\zeta)(1 + rz\eta/A_Y), \\
P_{1,z}(\zeta, r) &= \tfrac{1}{2}(\alpha_\zeta + z b_\zeta)(1 + r\Delta\zeta/A_Y).
\end{aligned}$$

Averaging over the sign gives the same reference under the two labels precisely when

$$\begin{aligned}
\text{P:} \quad & \Delta\alpha_\zeta\zeta = b_\zeta(\kappa + \gamma), \quad \Delta(v_- + \eta^2) = \eta(\kappa + \gamma), \\
\text{O:} \quad & \Delta\alpha_\zeta\zeta = b_\zeta\eta, \quad \Delta(v_- + \gamma^2) = \gamma\eta.
\end{aligned}$$

The second identity in each row proves the first on $\zeta = \pm L$; on $\zeta = 0$, both sides separately vanish. In P the identity follows by substituting $\kappa = \gamma + \Delta\eta$ and $\Delta v_- = 2\eta\gamma$. Thus the common positive conditional references are

$$Q_P(\zeta, r) = \tfrac{1}{2}(\alpha_\zeta + r b_\zeta\gamma/A_Y), \qquad Q_O(\zeta, r) = \tfrac{1}{2}(\alpha_\zeta + r b_\zeta\eta/A_Y).$$

Thus both labels have the same sign-averaged observation law, including the moment $\mathbb{E}(TY \mid X)$.

From cancellation to testing distance. Include the common marginal P_X in Q. Expanding the remaining sign-linear term gives

$$\frac{dP_{\ell,j,\sigma}}{dQ} = 1 + \sigma\phi_j(X)h_\ell(\zeta, r), \qquad \mathbb{E}_Q h_\ell = 0, \qquad K_\ell = \mathbb{E}_Q h_\ell^2.$$

For example, in O, with $e = \eta/A_Y$,

$$h_0 = \frac{b_\zeta + r\alpha_\zeta e}{\alpha_\zeta + r b_\zeta e}, \qquad h_1 = \frac{b_\zeta}{\alpha_\zeta}, \qquad K_0 \le t^2/9, \quad K_1 = \frac{\gamma^2}{v_- + \gamma^2} \le t^2/256.$$

The P expansion similarly gives $K_\ell \le s^2/196$ (Appendix C.6). Walsh orthogonality now yields the exact one-observation Gram entry

$$\mathbb{E}_Q\left[\frac{dP_{\ell,j,\sigma}}{dQ}\frac{dP_{\ell,j',\sigma'}}{dQ}\right] = 1 + \sigma\sigma'\mathbf{1}\{j = j'\}K_\ell.$$

First fix each component law, take its N-fold iid product, and only then mix the hidden coordinate and sign:

$$\overline{P}_\ell^N = \frac{1}{2p}\sum_{j=1}^{p}\sum_{\sigma=\pm1} P_{\ell,j,\sigma}^{\otimes N}.$$

Expanding the mixture square counts $2p$ equal-sign pairs, $2p$ opposite-sign pairs and $4p(p-1)$ different-coordinate pairs, giving

$$\chi^2(\overline{P}_\ell^N, Q^{\otimes N}) = \frac{\{(1 + K_\ell)^N + (1 - K_\ell)^N\}/2 - 1}{p} \le \frac{e^{NK_\ell}}{p} < e^{-511}. \tag{9}$$

The two mixtures therefore have TV distance below $1/2$. Since each label has one common target, the midpoint testing inequality gives maximum expected absolute error at least $\Delta/8$. Substitution yields $R_{n,\rho}^{\mathrm{pair}} \gtrsim_\rho P$ and $R_{n,\rho}^{\mathrm{pair}} \gtrsim_\rho O$.

Verifying the complexity certificates. In P the difference class has at most $16p^2$ elements and envelope 2κ. Since $\log(32p^2) \le 258ns^2$, the finite-class Rademacher bound, applied also to every localized subset, gives

$$\mathfrak{R}_n(r; \Delta\mathcal{G}_\mu) < 46\kappa s \le \frac{23\lambda_\rho}{64}s^2, \qquad \frac{\mathfrak{R}_n(r; \Delta\mathcal{G}_\mu)}{3r} \le \frac{23\lambda_\rho}{192}s \le c_{\mathrm{rad}}s$$

for every $r \geq s > 0$. In O, the corresponding bound is $\mathfrak{R}_n(r; \Delta\mathcal{G}_\pi) < 46\gamma t$, with normalized ratio at most $23\lambda_\rho t/384 \leq c_{\mathrm{rad}} t$. Thus amplitude times $\sqrt{\log p/n}$ is of order ν^2, while the likelihood energy is also of order ν^2; the exponential dictionary makes both requirements compatible.

When $\nu^2 < q$, the corresponding interaction is at most q. When $n < 16$, it is at most $1 \leq 4q$. These cases use the scalar lower experiment with zero nuisance classes; zero interactions need no construction. This supplies the all-n bounds without changing the active-family amplitudes.

Other components and assembly. The scalar experiment perturbs only β; the ab experiment uses two off-class truths; and the M experiment uses two represented stochastic dictionaries. Appendices C–F prove these three component bounds. The outer supremum in (3) dominates each component-specific pair. Put $S_0 = ab + M + O + P$. Expansion gives

$$S_0 \leq X \leq 2S_0, \tag{10}$$

since $at^2 \leq O$, $bs^2 \leq P$, and $s^2t^2 \leq M$ on $[0,1]^4$. Taking the maximum of the five lower bounds and then their sum up to a factor five proves $R^{\mathrm{pair}}_{n,\rho} \gtrsim_\rho q + X$.

Proposition 18 strengthens the construction by embedding all active P/O/ab/M families and the scalar family into one finite pair with a common P_X. Disjoint frequency bands preserve approximation distances, and the global complexity bound controls cross-band differences. The scalar family covers inactive terms.

A preassigned covariate distribution. For any atomless probability ν fixed before n and the budgets, restricting (3) to $P_X = \nu$ preserves rate $\asymp_\rho q + X$, with constants independent of ν, even when the learner knows it (Proposition 19). A Borel uniform coordinate transports the Walsh likelihoods. One pair still contains all specified families, but may depend on ν, ρ, n and the budgets. Atomlessness matters: a Dirac design permits order-q scalar regression.

5 Calibration without budget inputs

Write $x = a + s$, $y = b + t$, $T_\mu = ay + x^2$, $T_\pi = bx + y^2$, and $W = T_\mu \wedge T_\pi$. TAME gives both calibration branches and selects between them using the budgets (Gu, 2026). Appendix B implements this known-budget rule and proves its expected-loss bound.

Proposition 3 (Known-budget class-pair upper)**.** *That rule satisfies* $\sup_{P \in \mathcal{P}(\Gamma)} \mathbb{E}_P |\widehat{\beta}_\Gamma - \beta(P)| \leq C_\rho(q + X)$ *for every fixed legal pair, uniformly in* n, a, s, b, t.

The selector and rate algebra are

$$T_{\mathrm{selected}} \leq 2W \leq 48X, \qquad X/5 \leq W \leq 24X. \tag{11}$$

We remove all four inputs by combining double calibration (Liu et al., 2023, Section 4.1) and classical empirical local Rademacher fixed points (Bartlett et al., 2005), with a uniform guarantee under the original interface.

Theorem 4 (Adaptation to all four budgets)**.** *For every fixed* (Γ, ρ, n)*, one Borel rule* $\widehat{\beta}_{\mathrm{ad}}$*, using none of* a, s, b, t*, satisfies*

$$\forall (a,s,b,t) \in [0,1]^4, \qquad \sup_{P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)} \mathbb{E}_P |\widehat{\beta}_{\mathrm{ad}} - \beta(P)| \leq C_\rho(q + W)$$

whenever the compatible-law set is nonempty. The constant depends only on ρ*; no additional class geometry is imposed.*

Estimating the two complexity radii. Above a fixed sample threshold, split the $2n$ observations into three blocks of size $m = \lfloor 2n/3 \rfloor$, discarding at most two. Use only covariates from block $\mathcal{C}$ for calibration. Let $Q_\mathcal{C}$ be its empirical average, $L_0 = 1 \vee 2B_\mathcal{G}$, $\mathcal{F}_\nu = \Delta\mathcal{G}^{\mathrm{rep}}_\nu / L_0$, and compute

$$\widehat{\psi}_\nu(r) = 20\mathbb{E}_\epsilon \sup_{f \in \mathcal{F}_\nu} \left(1 \wedge \sqrt{\frac{2r}{Q_\mathcal{C} f^2}}\right) |Q_\mathcal{C}(\epsilon f)| + \frac{186 \log(em)}{m}.$$

Zero empirical norms contribute zero. This star-hull complexity with a positive floor has a unique positive fixed point $\widehat{r}_\nu$. Set $\widehat{s} = 1 \wedge C_0 L_0 \sqrt{\widehat{r}_\mu}$ and $\widehat{t} = 1 \wedge C_0 L_0 \sqrt{\widehat{r}_\pi}$, where $C_0 = 1 \vee (60 c_{\mathrm{rad}})^{-1}$. Appendix Q proves that, except with probability $8(em)^{-6}$, these are valid size-m full-class certificates with c_{rad} replaced by $2c_{\mathrm{rad}}$, and $\widehat{s} \le C_\rho(s+h)$, $\widehat{t} \le C_\rho(t+h)$, where $h^2 = \log(em)/m$. The star hull makes this calibration valid even when smaller localized balls contain only zero; the public classes themselves remain unchanged.

One calibrated rule for both directions. On the remaining main and fitting blocks, let P_1, P_2 be their empirical averages, $\|w\|_m^2 = P_1 w^2$, and $\tau = m^{-12}$. All constants below are fixed using $\rho' = (B_0, B_{\mathcal{G}}, v_-, v_+, 2c_{\mathrm{rad}})$. Fit $\widehat{\pi}$ and jointly $(\widehat{\beta}_0, \widehat{\mu})$ by represented least squares on the fitting block. Its internal train/validation split also supplies a fixed-anchor scalar pilot $\bar{\beta}$. Set $\kappa_\mu = K(\widehat{s}^2 + h^2)$, $\kappa_\pi = K(\widehat{t}^2 + h^2)$, and $A_{\widehat{t}} = K_A(\widehat{t}^2 + h^2)$. First approximately minimize $\|w - \widehat{\mu}_X\|_m^2$ over $w \in \mathbb{Q}^m$, subject to

$$\sup_{f \in \Delta \mathcal{G}_\pi^{\mathrm{rep}}} \{|P_2[(Y - \bar{\beta}T)f] - P_1(wf)| - P_1 f^2\} < \kappa_\pi, \qquad \|w\|_m < M_{\rho'}.$$

Call the result $\widehat{\eta}$, put $d = \widehat{\eta} - \widehat{\mu}_X$, and define $H_\nu(r) = \sup_{f \in \Delta \mathcal{G}_\nu^{\mathrm{rep}}} \{|P_1(rf)| - P_1 f^2\}$. Next minimize $\|z\|_m^2$ over $g \in \mathcal{G}_\pi^{\mathrm{rep}}$, $z \in \mathbb{Q}^m$. Writing $\widetilde{\pi} = g_X + z$, $r = T_X - \widetilde{\pi}$ and $D = P_1(Tr)$, impose

$$P_1(g - \widehat{\pi})^2 < A_{\widehat{t}}, \quad D > v_-/2, \quad \|r\|_m, \|\widetilde{\pi}\|_m < R_{\rho'}, \quad H_\nu(r) < \kappa_\nu \quad (\nu = \mu, \pi),$$
$$|P_1(rw)| < 8B_0 h \|w\|_m + \tau \quad \text{for } w \in \{d, \widetilde{\pi}, z\}.$$

The output is $\widehat{\beta}_{\mathrm{ad}} = \Pi_{[-3,3]}\{P_1[r(Y - \widehat{\eta})]/D\}$. Take the first feasible candidate within τ of each infimum; infeasibility or small n gives zero. The selections and fixed points are jointly Borel. Computational requirements are discussed in Section 8.

Proof of adaptation. Given a good calibration block, the remaining samples retain their iid law and the radii are valid fixed inputs. The pilot error $c = \beta_0 - \bar{\beta}$ has second moment at most $C_\rho \min\{a^2 + \widehat{s}^2 + h^2, b^2 + \widehat{t}^2 + h^2\}$. The outcome calibration uses fitting data and main covariates, with comparison $\eta_0 = \mu_0 + c\pi_0$. The propensity calibration also uses main treatments; main outcomes enter only the final ratio. Treatment noise is centered against directions fixed by fitting data and main covariates; outcome noise after additionally conditioning on main treatments.

On a further event of probability $1 - O_\rho(m^{-4})$, $g = \widehat{\pi}, z = \pi_0 - \widehat{\pi}$ represents the true propensity with zero class cost and strict feasibility margins. When $b \le \widehat{t} + h$, a fixed close core function g_0 supplies the second comparison $g = g_0, z = \pi_0 - g_0$. Rational approximation transfers these comparisons to the same minimum of $\|z\|_m^2$. Outcome calibration squares the represented fit error; propensity calibration exploits the smaller free residual. For larger b, the product bound for the same rule applies. Both cases analyze the estimator defined above; neither changes its inputs. Appendix P gives the resulting conditional risk bound

$$C_\rho \left[m^{-1/2} + \min\{a(b + \widehat{t}) + a^2 + \widehat{s}^2,\ b(a + \widehat{s}) + (b + \widehat{t})^2\} \right].$$

The radius bounds, $ah \le (a^2 + h^2)/2$, $bh \le (b^2 + h^2)/2$, and $h^2 \le 4q$ turn this into $C_\rho(q + W)$. Clipping pays for calibration failure. Appendix Q proves the transfers and zero-budget cases; only Γ, ρ, n are inputs.

6 Class geometry and budget adaptation

Class geometry changes the minimax rate. Restrict the outer supremum in (3) to full interval boxes or full affine Hilbert balls, and denote the resulting risks by $\mathcal{R}_{\mathrm{box}}$ and $\mathcal{R}_{\mathrm{Hilb}}$. Each rate below is attained without budget inputs.

Full interval boxes. For each $\nu \in \{\mu, \pi\}$, allow its own countable Borel partition and

$$\mathcal{G}_\nu = \left\{ c_\nu + \sum_j \theta_j \mathbf{1}_{C_{\nu j}} : |\theta_j| \le h_{\nu j} \right\}, \qquad |c_\nu(x)| + h_{\nu j} \le B_{\mathcal{G}} \quad (x \in C_{\nu j}).$$

All coordinate choices are included; the public centers may vary within cells. Recover c_ν, h_ν as the midpoint and half-range of its core evaluations. Group equal centered core profiles, which recover occupied positive-width cells. Set $N = 2n$, $\alpha = \min\{1/8, v_-/(8B_0^2)\}$, and $m = \lfloor \alpha N \rfloor$. On each side, trim the m largest-width singleton rows, or all if fewer exist, resolving ties by sample index. Let A_ν center repeated groups, vanish on trimmed rows, and retain other singletons. Define

$$\mathsf{B} = \mathsf{A}_\pi \mathsf{A}_\mu, \quad D = (T - c_\pi)' \mathsf{B} T/N, \quad U = (T - c_\pi)' \mathsf{B}(Y - c_\mu)/N,$$
$$S = T' \mathsf{A}_\mu T/N, \qquad V = T' \mathsf{A}_\mu (Y - c_\mu)/N.$$

$$\widehat{\beta}_{\mathrm{box}} = \begin{cases} \Pi_{[-3,3]}(U/D), & D \geq v_-/32, \\ \Pi_{[-3,3]}\{V/(S \vee (v_-/4))\}, & D < v_-/32. \end{cases} \tag{12}$$

Proposition 5 (Sharp box rate without budget inputs)**.** *For every fixed represented box pair and $n \geq 1$, the same Borel rule* (12) *satisfies*

$$\mathbb{E}_P |\widehat{\beta}_{\mathrm{box}} - \beta(P)| \leq C_\rho \{q + (a + s^2)(b + t^2)\}$$

simultaneously for all four certificate tuples and compatible laws. In the strict interior, $\mathcal{R}_{\mathrm{box}} \asymp_\rho q + AB$.

The estimator selectively retains singleton cells. For either side put $H = N^{-1} \sum_j h_j \mathbf{1}_{\{N_j = 1\}}$. The complexity certificate and a one-observation replacement bound give

$$\mathbb{E}H \leq C_\rho r^2, \qquad \mathbb{E}H^2 \leq C_\rho r^4 + 2B_{\mathcal{G}}^2/N.$$

After trimming, every retained width is at most H/α; hence each public residual has expected squared norm $O_\rho(r^4 + N^{-1})$. Writing $F_\nu = \|\mathsf{A}_\nu(\nu_0 - c_\nu)_X\|_N$, both the primary bias and the event-weighted fallback bias are controlled by $\mathbb{E}(F_\mu F_\pi) \leq C_\rho\{AB + q\}$. At most $2\alpha N$ diagonal entries of B lose the $1/4$ lower bound. A small cross denominator forces F_π above a fixed constant outside an $O_\rho(n^{-1})$ event.

For the matching lower, full boxes of widths proportional to s^2, t^2 have global difference complexity bounded by twice their width, regardless of host size. Rescaling the two cancellation channels above yields gaps ab, at^2, bs^2, s^2t^2; an exponential Walsh host hides their mixtures. Appendix M proves the upper bound and constructs worst-case box pairs attaining the lower bound.

Full affine Hilbert balls. Suppose $\mathcal{G}_\nu = c_\nu + R_\nu B_{\mathcal{H}_\nu}$ is an entire reproducing-kernel Hilbert ball obeying the same envelope and M1 interface. The two kernels may differ and have overlapping features or infinite rank. Polarization of centered-core evaluation suprema recovers the normalized kernels $\ell_\nu = R_\nu^2 k_\nu$ without extra inputs. Let M_c project off the two sampled center vectors and put

$$H = M_c[\ell_\mu(X_i, X_j) + \ell_\pi(X_i, X_j)]_{i,j} M_c/N, \quad \lambda = \frac{16 B_{\mathcal{G}}^2 B_0^2}{N v_-}, \quad A_c = M_c(I + H/\lambda)^{-1} M_c.$$

For $N \geq 16B_0^2/v_-$, use $\widehat{\beta}_{\mathrm{Hilb}} = \Pi_{[-3,3]}\{T' A_c Y/(T' A_c T \vee N v_-/4)\}$; otherwise output zero.

Proposition 6 (Hilbert-ball comparison)**.** *The same fixed-(Γ, ρ, n) rule satisfies $\mathbb{E}_P|\widehat{\beta}_{\mathrm{Hilb}} - \beta(P)| \leq C_\rho(q + ab)$ for every certificate tuple and compatible law. In the strict interior, $\mathcal{R}_{\mathrm{Hilb}} \asymp_\rho q + ab$, allowing zero-radius balls.*

In this large-sample branch, $\mathrm{tr}(I - A_c) \leq N v_-/(4B_0^2)$ and $\sup_{g \in \mathcal{G}_\mu \cup \mathcal{G}_\pi} \|A_c g_X\|_N \leq \sqrt{\lambda}/2$. These preserve average variance and leave $ab + O_\rho(q)$ bias. Appendix N proves the result; its outer lower already holds for the zero pair. The result extends classical penalized partialling-out (Heckman, 1986; Zhao et al., 2016) with a finite-sample guarantee under bounded misspecification. Finite versions require evaluable centers together with partition labels and widths, or kernel evaluations, respectively; M1 alone gives Borel existence.

Proposition 7 (Convex hulls and finite unions)**.** *For full dictionary ℓ_1 balls $\{\sum_{j=1}^k \theta_j d_j : \sum_j |\theta_j| \leq 1\}$, and for segment unions $\bigcup_{j=1}^k \{\tau d_j : |\tau| \leq 1\}$, both restricted outer risks are $\asymp_\rho q + X$ in the strict interior. Zero classes are allowed and k may grow with n and the budgets. Each segment is a full RKHS ball of rank at most one.*

Enlarging the same-pair dictionaries this way preserves global difference complexity (Bartlett & Mendelson, 2002, Theorem 12(2)); off-band truths remain orthogonal to the enlarged span. The global bound preserves all required certificates and testing lower bounds, regardless of changes in the localized complexities. Moreover, enclosing $\{0, \pm h\phi_j : j \le p\}$, for orthonormal sign functions ϕ_j, in a full affine Hilbert ball requires pointwise radius at least $h\sqrt{p}$, sharply (Appendix O). Thus both convex dictionary balls and finite segment unions retain the general rate.

7 Nuisance imbalance and budget allocation

Corollary 8 (Optimal calibration and certificate allocation)**.** *Fix ρ in the interior stratum. Then $R^{\mathrm{pair}}_{n,\rho} \asymp_\rho q + W$. On $0 < x \le y$,*

$$\frac{R^{\mathrm{pair}}_{n,\rho}}{xy} \asymp_\rho \frac{q}{xy} + \frac{a}{x} + \frac{x}{y}.$$

Along any such sequence with $q = o(xy)$, $R^{\mathrm{pair}}_{n,\rho} = o(xy)$ if and only if $a = o(x)$ and $x = o(y)$. For $0 < y < x$, interchange (a, x) and (b, y). For $0 < x \le y \le 1$, put $\mathcal{A}(x, y) = \{(a, s, b, t) \in [0, 1]^4 : a + s = x,\ b + t = y\}$. Then

$$\inf_{\mathcal{A}(x,y)} R^{\mathrm{pair}}_{n,\rho} \asymp_\rho q + x^2, \qquad \sup_{\mathcal{A}(x,y)} R^{\mathrm{pair}}_{n,\rho} \asymp_\rho q + xy.$$

Divide the selected-branch rate by xy. For allocation, $x^2/4 \le X \le 4xy$, with endpoints $(0, x, y, 0)$ and $(x, 0, y, 0)$; see Appendix B.9. TAME's sufficient imbalance conditions are therefore necessary in this experiment. Here xy is the aggregate-error benchmark; the corollary compares minimax rates with that benchmark, rather than evaluating a particular DML implementation.

8 Related work, limitations, and conclusion

Gu–Yin–Cai–Fan Theorem A.1 gives $q + AB + M$ in a fixed-covariate, constant-variance field (Gu et al., 2026). TAME Corollary 3.2 and equations (3.6)–(3.7) give branch selection, tuning, and the leading upper bound $q + W$ under the average residual-variance condition 2.1(b) (Gu, 2026). SADE Theorem 2.1 and Remark 2.1 give budget-free single-outcome-class adaptation (Gu et al., 2026); our results concern the sharp two-class risk and simultaneous adaptation to its four certificates.

Jin–Mackey–Syrgkanis Theorem 3.1 gives squared-radius and mixed-product obstructions around fixed estimates of $(\mathbb{E}[T \mid X], \mathbb{E}[Y \mid X])$ in binary-treatment PLMs (Jin et al., 2025). Because $\mathbb{E}[Y \mid X] = \mu_0 + \beta_0\pi_0$ changes with the target, those anchored radii are not our four certificates. Our cancellation constructions adapt these obstruction arguments to preserve all four certificates and a common variance. Related work studies anchored product bounds (Jin & Syrgkanis, 2026), calibrated inference (van der Laan et al., 2024), and fixed-pilot asymptotic risk (Liu et al., 2026).

Limitations. For fixed ρ, the upper bounds hold for each represented pair, and the lower bounds concern the worst admissible pair. The general adaptive rule is Borel; a finite-query implementation requires additional optimization interfaces. The lower-bound dictionaries can be exponential, and the faster box and Hilbert rates use the full-class geometry and uniform envelopes. Open problems include selection across public pairs, efficient optimization, and the joint transition to envelope saturation. Causal identification, confidence intervals, and empirical performance require separate analysis.

Conclusion. The four certificates determine the sharp worst-class risk: the O/P lower bounds establish the optimality of the two-branch calibration rate, and a single estimator attains it without the budgets. Dictionary ℓ_1 balls retain this rate, whereas full boxes and Hilbert balls allow faster estimation. The Hilbert enclosure bound quantifies the envelope cost separating these geometries.

Reproducibility statement

The appendices contain the model assumptions, complete proofs, constants, and a proof index. Figure 1 plots the theoretical expressions; its source is `figures/+generate_theory_summary.py`. Run `python3 figures/+generate_theory_summary.py` from the project root to regenerate the figure.

AI use statement

Generative AI assisted discovery, literature work, proof drafting and auditing, symbolic checks, figures, and editing. The authors reviewed all AI-assisted work and take responsibility for the final content, including text, mathematical claims, proofs, and artifacts produced with AI assistance. No AI-generated empirical evidence is reported.

References

Sivaraman Balakrishnan, Edward H. Kennedy, and Larry Wasserman. The fundamental limits of structure-agnostic functional estimation. *Statistical Science*, 41(3):659–670, 2026. doi: 10.1214/25-STS997. URL https://doi.org/10.1214/25-STS997.

Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002. URL https://jmlr.org/papers/volume3/bartlett02a/bartlett02a.pdf.

Peter L. Bartlett, Olivier Bousquet, and Shahar Mendelson. Local rademacher complexities. *The Annals of Statistics*, 33(4):1497–1537, 2005. doi: 10.1214/009053605000000282. URL https://doi.org/10.1214/009053605000000282.

Matias D. Cattaneo, Michael Jansson, and Whitney K. Newey. Alternative asymptotics and the partially linear model with many regressors. *Econometric Theory*, 34(2):277–301, 2018. doi: 10.1017/S026646661600013X. URL https://eml.berkeley.edu/~mjansson/Publications/Cattaneo-Jansson-Newey_2018_ET.pdf.

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. URL https://doi.org/10.1111/ectj.12097.

Dylan J. Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *The Annals of Statistics*, 51(3):879–908, 2023. doi: 10.1214/23-AOS2258. URL https://doi.org/10.1214/23-AOS2258.

Yihong Gu. Open problem: Structure-agnostic minimax risk for partial linear model. In *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pp. 6220–6224, 2025. URL https://proceedings.mlr.press/v291/gu25b.html.

Yihong Gu. Optimal use of a black-box learner in semiparametric estimation, 2026. URL https://arxiv.org/abs/2607.21541. Version 1, submitted 23 July 2026.

Yihong Gu, Qishuo Yin, Tianxi Cai, and Jianqing Fan. Optimally taming biases in black-box models for efficient semiparametric estimation, 2026. URL https://arxiv.org/abs/2606.06368. Version 1, 4 June 2026.

Nancy E. Heckman. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(2):244–248, 1986. doi: 10.1111/j.2517-6161.1986.tb01407.x. URL https://doi.org/10.1111/j.2517-6161.1986.tb01407.x.

Jikai Jin and Vasilis Syrgkanis. Sharp structure-agnostic lower bounds for general linear functional estimation, 2026. URL https://arxiv.org/abs/2512.17341. Version 2, 5 January 2026.

Jikai Jin, Lester Mackey, and Vasilis Syrgkanis. It's hard to be normal: The impact of noise on structure-agnostic estimation. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 164176–164245. Curran Associates, Inc., 2025. doi: 10.52202/085713-5477. URL https://proceedings.nips.cc/paper_files/paper/2025/hash/f0156a82b6af6a4e838923ce9c124424-Abstract-Conference.html.

Lin Liu, Xinbo Wang, and Yuhao Wang. Root-n consistent semiparametric learning with high-dimensional nuisance parameters under minimal sparsity. *arXiv preprint arXiv:2305.04174*, 2023. URL https://arxiv.org/abs/2305.04174v4. Version 4, revised 29 September 2024.

Lin Liu, Rajarshi Mukherjee, and James M. Robins. On the asymptotic inadmissibility of double machine learning estimators under structure-agnostic models, 2026. URL https://arxiv.org/abs/2606.22391. Version 1, submitted 21 June 2026.

James Robins, Eric Tchetgen Tchetgen, Lingling Li, and Aad van der Vaart. Semiparametric minimax rates. *Electronic Journal of Statistics*, 3:1305–1321, 2009. doi: 10.1214/09-EJS479. URL https://doi.org/10.1214/09-EJS479.

Peter M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954, 1988. doi: 10.2307/1912705. URL https://doi.org/10.2307/1912705.

Lars van der Laan, Alex Luedtke, and Marco Carone. Doubly robust inference via calibration, 2024. URL https://arxiv.org/abs/2411.02771. Version 3, revised 20 August 2026; first posted 5 November 2024.

Tianqi Zhao, Guang Cheng, and Han Liu. A partially linear framework for massive heterogeneous data. *The Annals of Statistics*, 44(4):1400–1437, 2016. doi: 10.1214/15-AOS1410. URL https://doi.org/10.1214/15-AOS1410.

A Full model, measurable access, and minimax quantifiers

A.1 Measurable experiment and statistical target

Fix a nonempty infinite standard Borel measurable space $(\mathcal{X}, \mathcal{A})$, shared by every public pair and law below. Equip

$$\mathcal{Z} = \mathcal{X} \times \mathbb{R} \times \mathbb{R}$$

with the product sigma-field $\mathcal{A} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. The coordinates $X : \mathcal{Z} \to \mathcal{X}$, $T, Y : \mathcal{Z} \to \mathbb{R}$ are measurable, and all conditional expectations are ordinary versions on this probability space. Public represented functions are jointly supplied with measurable evaluation maps $x \mapsto g_j(x)$. The ambient space is fixed before n, the budget tuple, the public pair, and the compatible law are chosen. The interior theorem covers both countably infinite and uncountable standard Borel spaces.

For a fixed regularity tuple

$$\rho = (B_0, B_{\mathcal{G}}, v_-, v_+, c_{\mathrm{rad}}), \qquad B_0 \geq 3, \quad B_{\mathcal{G}} > 0, \quad 0 < v_- \leq v_+, \quad c_{\mathrm{rad}} > 0,$$

a law in $\mathcal{P}(\Gamma)$ satisfies (1), $\mathbb{E}(u \mid X) = 0$, $\mathbb{E}(\varepsilon \mid X, T) = 0$, and $v := \mathbb{E}u^2 \in [v_-, v_+]$. The conditional variance $\mathrm{Var}(T \mid X) = \mathbb{E}(u^2 \mid X)$ may vary with X and may vanish on subsets of the covariate space. It also satisfies

$$|Y|, |T|, |\beta_0|, |\mu_0(X)|, |\pi_0(X)|, |u|, |\varepsilon| \leq B_0, \qquad \beta_0 \in [-3, 3].$$

The target is the coefficient β_0 in this conditional mean model. A causal interpretation requires additional identification assumptions.

Every admissible public pair also satisfies the fixed pointwise candidate envelope

$$\sup_{g \in \mathcal{G}_\mu \cup \mathcal{G}_\pi} |g(x)| \leq B_{\mathcal{G}} \qquad \text{for every } x \in \mathcal{X}. \tag{A.0}$$

Every function in the supplied countable representations obeys the same bound. Thus each public difference class has envelope $2B_{\mathcal{G}}$. The candidate envelope is a separate model assumption, imposed before taking the outer supremum over public pairs.

A.2 POSITIVE-RADIUS PUBLIC COMPLEXITY AND M1 ACCESS

For a public class $\mathcal{G}$, let $\Delta\mathcal{G} = \{g - g' : g, g' \in \mathcal{G}\}$. For any class $\mathcal{H}$ and $r \geq 0$, put

$$Z_{\mathcal{H}}(r) := \sup_{\substack{h\in\mathcal{H} \\ \|h\|_{L_2(P_X)}\leq r}} |\frac{1}{n}\sum_{i=1}^{n} \epsilon_i h(X_i)|, \qquad \mathfrak{R}^*_{n,P_X}(r;\mathcal{H}) := \mathbb{E}^*_{X_{1:n},\epsilon_{1:n}} Z_{\mathcal{H}}(r).$$

Here $X_{1:n} \sim P_X^n$, the ϵ_i's are independent Rademacher signs independent of the sample, and the outer expectation $\mathbb{E}^* Z$ is the infimum of $\mathbb{E}U$ over all measurable majorants $U \geq Z$. When $Z_{\mathcal{H}}(r)$ is measurable, outer and ordinary expectation agree. If $\mathcal{G}^{\mathrm{rep}} = \{g_j : j \in \mathbb{N}\}$, its countable represented difference core is $\Delta\mathcal{G}^{\mathrm{rep}} = \{g_j - g_k : j, k \in \mathbb{N}\}$. Its supremum is an ordinary measurable random variable, and we write its ordinary-expectation complexity as $\mathfrak{R}_{n,P_X}(r;\Delta\mathcal{G}^{\mathrm{rep}})$. Pointwise set inclusion and monotonicity of outer expectation give

$$\mathfrak{R}_{n,P_X}(r;\Delta\mathcal{G}^{\mathrm{rep}}) \leq \mathfrak{R}^*_{n,P_X}(r;\Delta\mathcal{G}).$$

The upper proof uses this inequality; sequential M1 need not give equality. At $r = 0$ the unquotiented quantities above remain defined. No star-shapedness or scaling closure is assumed or used. The stochastic constraints are exactly

$$\begin{aligned} \sup_{r\geq s,\, r>0} \frac{\mathfrak{R}^*_{n,P_X}(r;\Delta\mathcal{G}_\mu)}{3r} &\leq c_{\mathrm{rad}} s, \\ \sup_{r\geq t,\, r>0} \frac{\mathfrak{R}^*_{n,P_X}(r;\Delta\mathcal{G}_\pi)}{3r} &\leq c_{\mathrm{rad}} t. \end{aligned} \tag{A.1}$$

Every quotient in (A.1) is evaluated at a positive radius. When a declared stochastic budget is zero, (A.1) directly forces the corresponding full-class outer complexity, and hence every represented-core complexity, to vanish at each positive radius. The upper proof can therefore use base radius zero while keeping every quotient at a positive radius.

M1 means that the learner receives the represented public instance

$$\Gamma = (\mathcal{G}_\mu, \mathcal{G}_\pi, \mathbf{g}_\mu, \mathbf{g}_\pi), \qquad \mathbf{g}_\nu = (g_{\nu,j})_{j\in\mathbb{N}}, \quad \nu \in \{\mu, \pi\},$$

where $\mathbf{g}_\nu$ is an ordered countable representation of $\mathcal{G}_\nu$. It also receives the fixed tuple including $B_{\mathcal{G}}$, and the four declared budgets before observing data. The representation has measurable evaluations, obeys the same envelope, and has the following sequential property: for every full-class element $g \in \mathcal{G}$, there is a sequence g_{j_k} from the representation that converges pointwise everywhere to g on $\mathcal{X}$. The common envelope and dominated convergence make the represented and full-class population $L_2(P_X)$ approximation infima equal; finite-sample continuity gives the corresponding empirical fit and calibration transfers. Empirical approximate minimizers select the first represented index within n^{-12} of a countable infimum. Scalar coefficients use $\mathbb{Q} \cap [-3, 3]$; finite-dimensional convex objectives use objective-dense rational parameterizations. Countable infima, first qualifying indices, finite arithmetic, and clipping are measurable, so every fitted rule is Borel. M1 does not supply nuisance truths, population best approximants, realized errors, a hard label, or an optimization oracle.

Definition 9 (Indexed compatible-law class)**.** Fix ρ, n, a, s, b, t, and let $\Gamma = (\mathcal{G}_\mu, \mathcal{G}_\pi, \mathbf{g}_\mu, \mathbf{g}_\pi)$ be a represented public instance obeying the envelope and measurable sequential-M1 interface above. We call $(\mathcal{G}_\mu, \mathcal{G}_\pi)$ its class-pair projection. The indexed compatible-law class $\mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$ is the set of probability laws P on $\mathcal{Z}$ for which there are $\beta_0 \in [-3, 3]$, $\mathcal{A}/\mathcal{B}(\mathbb{R})$-measurable functions $\mu_0, \pi_0 : \mathcal{X} \to \mathbb{R}$, and residuals

$$u = T - \pi_0(X), \qquad \varepsilon = Y - \beta_0 T - \mu_0(X),$$

such that the PLM representation (1), the conditional-centering conditions $\mathbb{E}(u \mid X) = 0$ and $\mathbb{E}(\varepsilon \mid X, T) = 0$ hold P-almost surely, and the average residual-variance condition $v := \mathbb{E}_P u^2 \in [v_-, v_+]$ holds. The shared law and target envelopes are

$$|Y|, |T|, |\beta_0|, |\mu_0(X)|, |\pi_0(X)|, |u|, |\varepsilon| \leq B_0,$$

and the public classes and their representations retain the pointwise candidate envelope (A.0), measurable evaluation maps, and the pointwise-everywhere sequential-M1 approximation property. Under this law's covariate marginal P_X, the approximation-budget conditions are

$$\inf_{g\in\mathcal{G}_\mu} \|g - \mu_0\|_2 \leq a, \qquad \inf_{g\in\mathcal{G}_\pi} \|g - \pi_0\|_2 \leq b,$$

and the localized-complexity conditions are

$$\sup_{r\ge s,\, r>0} \frac{\mathfrak{R}^*_{n,P_X}(r;\Delta\mathcal{G}_\mu)}{3r} \le c_{\rm rad}\, s, \qquad \sup_{r\ge t,\, r>0} \frac{\mathfrak{R}^*_{n,P_X}(r;\Delta\mathcal{G}_\pi)}{3r} \le c_{\rm rad}\, t.$$

The target $\beta(P)$ is uniquely identified: if two such conditional-mean PLM representations describe the same law, their conditional-mean difference is $\gamma T + f(X) = 0$ almost surely. Conditional centering identifies $\pi_0(X) = \mathbb{E}(T \mid X)$, hence the same residual u, in both representations. Multiplying the difference by u and taking expectations gives $0 = \gamma \mathbb{E}u^2$, since $\mathbb{E}\{f(X)u\} = \mathbb{E}\{\pi_0(X)u\} = 0$. Because $\mathbb{E}u^2 \ge v_- > 0$, $\gamma = 0$, so the two coefficients agree. The definition collects the model assumptions and the four public certificates. In the main text and the rest of the appendices we suppress the fixed indices ρ, n, a, s, b, t and write $\mathcal{P}(\Gamma)$.

A.3 Minimax order and sample-size dependence

For each integer $n \ge 1$ and budget tuple, define $\mathfrak{G}_{\rho,n}(a, s, b, t)$ as the collection of represented public pairs on this same fixed ambient space satisfying the candidate-envelope and measurable sequential-M1 fields above and having $\mathcal{P}_{\rho,n,a,s,b,t}(\Gamma) \neq \varnothing$. The ambient space is not a coordinate of Γ and is not varied by the outer supremum. Q1 is the order printed in (3):

$$\sup_{\Gamma\in\mathfrak{G}_{\rho,n}(a,s,b,t)} \inf_{\delta_\Gamma \text{ Borel}} \sup_{P\in\mathcal{P}(\Gamma)} \mathbb{E}_P|\delta_\Gamma(Z_{1:2n}) - \beta(P)|.$$

The public pair and a lower hard family may depend on n and the declared budgets, but they are fixed before the data and before any latent label, coordinate, or sign is drawn. The pair is deterministic at each sample size and budget tuple, and may vary across tuples. For each fixed Γ, the decision rule may depend on Γ, ρ, a, s, b, t but not on P, proof-only comparators, or realized errors. Hence the outer public-pair supremum is taken only after the fixed-pair rule and its uniform law bound are established.

The total sample size is $N = 2n$, while $q = n^{-1/2}$; replacing q by $N^{-1/2}$ changes only a fixed factor. Risk is expected absolute error. All interior constants may depend on fixed ρ, including $B_\mathcal{G}$, but not on n, a, s, b, t, Γ, P. At $c_{\rm rad} = 0$, (A.1) collapses every legal difference class in $L_2(P_X)$, and the positive-budget nuisance lower is false. Thus the theorem is pointwise for $c_{\rm rad} > 0$, with constants that may degenerate as $c_{\rm rad} \downarrow 0$; it is not uniform at that boundary.

B Proof of the global upper bound

We define the estimator and prove Proposition 3. Throughout, P_j denotes the empirical average on half $\mathcal{D}_j$, $j = 1, 2$. Every unnamed proof constant depends only on the fixed tuple ρ. Write $v := \mathbb{E}u^2 \in [v_-, v_+]$. Conditional treatment variance may vary with X and may vanish. Put

$$\|f\|_{1,n} := (P_1 f^2)^{1/2}, \qquad \|f\|_{2,n} := (P_2 f^2)^{1/2}.$$

These half-indexed symbols are used for fitted functions of X. A vector norm whose sample half is already fixed by its definition retains the shorter notation $\|\cdot\|_n$. Put

$$C_f = 1 + \frac{B_0}{\sqrt{v_-}}, \qquad C_J = 1 + 3C_f + \frac{B_0}{c_{\rm rad}\sqrt{v_-}}.$$

Let $K_{1,\star}$ be the maximum of the radius constants in Lemma 12 for the seven applications summarized in Appendix B.7. They use only the two public difference cores and the joint class, on a specified sample half, with a specified multiplier or square process, before any fitted index is substituted. Fix

$$C_{\rm rad} := 1 + K_{1,\star}(1 + C_J), \qquad K_\lambda := 1, \qquad K_d := 8.$$

Here the model parameter $c_{\rm rad} > 0$ is distinct from the proof/estimator inflation constant $C_{\rm rad}$. Also put

$$B_\pi = B_0 + B_\mathcal{G}, \qquad B_{\rm all} = B_0 + 4B_\mathcal{G}, \qquad \vartheta_\rho = \frac{B_\pi}{8}, \qquad \delta_\rho = \frac{\vartheta_\rho}{C_{\rm rad}}.$$

These quantities are fixed before $n, (a, s, b, t), \Gamma, P$, and the data; they are never enlarged or redefined by a later proof constant. Finally set

$$h_n = \sqrt{\frac{\log(en)}{n}}, \quad \bar{s} = C_{\rm rad}(s \vee h_n), \quad \bar{t} = C_{\rm rad}(t \vee h_n), \quad \rho_n = n^{-6}, \quad \tau_n = n^{-12}.$$

B.1 Elementary concentration and countable optimization

Lemma 10 (Bounded Bernstein inequality). *Let $\mathcal{F}$ be a sigma-field. Suppose $W_1, \dots, W_n$ are conditionally independent given $\mathcal{F}$ and conditionally centered given $\mathcal{F}$, with $|W_i| \le B$, and $n^{-1}\sum_i \mathbb{E}(W_i^2 \mid \mathcal{F}) \le \sigma^2$, then, conditionally on $\mathcal{F}$, with probability at least $1 - 2e^{-z}$,*

$$|P_n W| \le \sqrt{\frac{2\sigma^2 z}{n}} + \frac{2Bz}{3n}. \tag{B.1}$$

Proof. For $0 < \lambda < 3/B$, the power series and $k! \ge 2\,3^{k-2}$ for $k \ge 2$ give

$$\mathbb{E}(e^{\lambda W_i} \mid \mathcal{F}) \le 1 + \frac{\lambda^2 \mathbb{E}(W_i^2 \mid \mathcal{F})}{2(1 - \lambda B/3)}.$$

Use $1 + x \le e^x$. Conditional independence given $\mathcal{F}$ permits multiplication of the conditional moment generating functions. Apply Markov's inequality and optimize in λ. This gives the upper tail in (B.1); apply the same argument to $-W_i$ for the lower tail. □

We also use the following weighted form, which does not require a bound on the largest weight. If $c_1, \dots, c_n$ are $\mathcal{F}$-measurable and, given $\mathcal{F}$, the ξ_i are independent, centered, and bounded by $|\xi_i| \le B$, then

$$|P_n(\xi c)| \le B\|c\|_n \sqrt{\frac{2z}{n}} \tag{B.1a}$$

with conditional failure at most $2e^{-z}$. Indeed, Hoeffding's one-variable exponential bound and conditional independence give $\mathbb{E}[\exp\{\lambda \sum_i c_i \xi_i\} \mid \mathcal{F}] \le \exp\{\lambda^2 B^2 \sum_i c_i^2/2\}$; optimize the two Chernoff tails.

Lemma 11 (Localized supremum concentration). *Let $\mathcal{W}$ be a countable class of centered functions of an iid observation, with $|W(w)| \le B$ and $\mathbb{E}W(w)^2 \le \sigma^2$, uniformly in w. Put*

$$\mathcal{R}_n(\mathcal{W}) = \mathbb{E}_{\mathrm{data},\epsilon} \sup_{w \in \mathcal{W}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i W_i(w) \right|.$$

If $Z = \sup_{w \in \mathcal{W}} |P_n W(w)|$, then

$$\mathbb{P}\left\{ Z > 4\mathcal{R}_n(\mathcal{W}) + \sqrt{\frac{2\sigma^2 z}{n}} + \frac{8Bz}{3n} \right\} \le 2e^{-z}. \tag{B.2}$$

Proof. Apply Theorem 2.1 of Bartlett et al. (2005, pp. 1503–1504). Enlarge $\mathcal{W}$ by its negatives and map the source theorem's $[a, b]$, variance parameter r, confidence x, and tuning α to $[-B, B]$, σ^2, z, and 1, respectively. Its one-sided supremum is then exactly Z; its Rademacher term is $\mathcal{R}_n(\mathcal{W})$; and its range term is $2B(1/3 + 1)z/n = 8Bz/(3n)$. The source probability is $1 - e^{-z}$, so the displayed $2e^{-z}$ is conservative. The source assumes measurable suprema; countability supplies this here. This proves the stated concentration bound. □

Lemma 12 (Localized multiplier and square processes). *Let $\mathcal{H}$ be a countable class with envelope B, containing zero, and suppose*

$$\mathfrak{R}_{n,P_X}(r; \mathcal{H}) \le c_{\mathrm{rad}} r_0 r, \qquad r \ge r_0, \quad r > 0. \tag{B.2a}$$

Let ξ be bounded and satisfy $\mathbb{E}(\xi \mid X) = 0$. For $\delta_z = K_1(r_0 \vee h_n) + \sqrt{z/n}$ and $1 \le z \le n$, an event of probability at least $1 - K_2 e^{-z}$ satisfies, simultaneously for every $f \in \mathcal{H}$,

$$|(P_n - P)(\xi f)| \le K_3\{\delta_z \|f\|_2 + \delta_z^2\}, \tag{B.3}$$

$$|(P_n - P) f^2| \le K_3\{\delta_z \|f\|_2 + \delta_z^2\}. \tag{B.4}$$

The same conclusion holds when ξ is a fixed bounded multiplier, with the centered process in (B.3).

Proof. Write $L = \|\xi\|_\infty$. Define the inner ball and dyadic shells by

$$\mathcal{H}_0 = \{f : \|f\|_2 \le \delta_z\}, \qquad \mathcal{H}_j = \{f : 2^{j-1}\delta_z < \|f\|_2 \le \min(2^j\delta_z, B)\}, \quad j \ge 1,$$

and set

$$J = \lceil \log_2(B/\delta_z) \rceil_+ .$$

The envelope gives $\mathcal{H} = \bigcup_{j=0}^{J} \mathcal{H}_j$; if $\delta_z \ge B$, then $J = 0$. Since $\delta_z \ge h_n$,

$$J \le \lceil \log_2(B/h_n) \rceil_+ \le C_B \log(en), \qquad \sqrt{j/n} \le C_B h_n, \qquad j/n \le C_B h_n^2. \tag{B.2b}$$

For $j \ge 1$, put $R_j = 2^j\delta_z$, and put $R_0 = \delta_z$. On $\mathcal{H}_j$, conditional Rademacher contraction bounds the multiplier mean by $L\mathfrak{R}_n(R_j; \mathcal{H})$. If the multiplier is fixed rather than conditionally centered, centering adds at most $LR_j/\sqrt{n}$. Thus its centered Rademacher mean is at most $C(r_0 + h_n)R_j \le C\delta_z R_j$, its variance is at most $L^2R_j^2$, and its centered range is at most $2LB$. Boundedness supplies these multiplier bounds for heteroskedastic observations.

For the square process,

$$\mathrm{Var}(f^2) \le \mathbb{E}f^4 \le B^2R_j^2.$$

The map $u \mapsto u^2$ is $2B$-Lipschitz on $[-B, B]$, so symmetrization and contraction bound the uncentered Rademacher term by $CB\mathfrak{R}_n(R_j; \mathcal{H})$. By localization and the envelope, the centering contribution satisfies

$$\sup_{f \in \mathcal{H}_j} Pf^2 \, \mathbb{E}_\epsilon \Big| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i \Big| \le \frac{\min\{R_j^2, B^2\}}{\sqrt{n}} \le \frac{BR_j}{\sqrt{n}} \le Bh_nR_j.$$

Its centered mean is therefore at most $C\delta_z R_j$, and its centered range is at most $2B^2$.

For each process and every $0 \le j \le J$, apply Lemma 11 with $x_j = z + 2j + \log 8$. The multiplier and square-process failures form the geometric sum

$$4\sum_{j=0}^{J} e^{-x_j} \le \frac{1}{2}e^{-z}\sum_{j=0}^{\infty} e^{-2j} < e^{-z}. \tag{B.2c}$$

The shellwise variance and range terms satisfy, respectively,

$$R_j\sqrt{x_j/n} \le CR_j\{\sqrt{z/n} + \sqrt{j/n} + n^{-1/2}\} \le C\delta_z R_j,$$

and

$$Cx_j/n \le C\{z/n + j/n + 1/n\} \le C\delta_z^2. \tag{B.2d}$$

For $j \ge 1$, $R_j < 2\|f\|_2$; on the inner ball, $R_0 = \delta_z$. These inequalities prove (B.3)–(B.4). Finally, $\delta_z \ge r_0$, so the sub-r_0 region is contained in the actual positive-radius δ_z-ball. No function is rescaled, and no star-shapedness or scaling closure is used. The same calculation covers $n = 1$, zero stochastic budgets, $z = 1$, the maximum-radius shell, and the case in which the inner ball is the whole class. Countability makes every supremum measurable. □

Apply Lemma 12 before inserting fitted indices. We apply it to the following seven processes:

1. $\mathsf{P}^{u}_{\pi,2}(z)$: on $\mathcal{D}_2$, the countable represented core of $\Delta\mathcal{G}_\pi$, with multiplier u and with the square process;
2. $\mathsf{P}^{\mathrm{sq}}_{\mu,2}(z)$: on $\mathcal{D}_2$, the countable represented core of $\Delta\mathcal{G}_\mu$, retaining the square process;
3. $\mathsf{P}^{\varepsilon}_{J,2}(z)$: on $\mathcal{D}_2$, the joint class $\mathcal{J}_\mu$ defined below, with multiplier ε and with the square process;
4. $\mathsf{P}^{u}_{\mu,1}(z)$: on $\mathcal{D}_1$, the countable represented core of $\Delta\mathcal{G}_\mu$, with multiplier u and with the square process;
5. $\mathsf{P}^{\xi}_{\pi,2}(z)$: on $\mathcal{D}_2$, the countable represented core of $\Delta\mathcal{G}_\pi$, with multiplier $\xi = Y - m_0(X)$;
6. $\mathsf{P}^{m_0}_{\pi,2}(z)$: on $\mathcal{D}_2$, the countable represented core of $\Delta\mathcal{G}_\pi$, with the fixed multiplier $m_0(X)$; and

7. $\mathsf{P}^{m_0}_{\pi,1}(z)$: on $\mathcal{D}_1$, the countable represented core of $\Delta\mathcal{G}_\pi$, with the fixed multiplier $m_0(X)$ and with the square process.

The model-field certificate (A.1) is an outer-expectation bound for the full difference class. The pointwise inclusion in Appendix A therefore gives the ordinary countable-core bound needed in (B.2a). The seven calls use the following explicit core base radii:

$$\begin{array}{ll}
\mathsf{P}^{u}_{\pi,2}(z) & 3t \\
\mathsf{P}^{\mathrm{sq}}_{\mu,2}(z) & 3s \\
\mathsf{P}^{\varepsilon}_{J,2}(z) & C_J(s \vee h_n) \\
\mathsf{P}^{u}_{\mu,1}(z) & 3s \\
\mathsf{P}^{\xi}_{\pi,2}(z) & 3t \\
\mathsf{P}^{m_0}_{\pi,2}(z) & 3t \\
\mathsf{P}^{m_0}_{\pi,1}(z) & 3t
\end{array}$$

Indeed, (A.1) gives (B.2a) with $r_0 = 3s$ or $r_0 = 3t$ for each simple core. Consequently

$$K_1(3s \vee h_n) \le 3K_1(s \vee h_n), \qquad K_1(3t \vee h_n) \le 3K_1(t \vee h_n).$$

Since $C_J \ge 4$, the fixed C_{rad} absorbs both $3K_{1,\star}$ and $K_{1,\star}C_J$. If $s = 0$ or $t = 0$, the relevant call uses $r_0 = 3s = 0$ or $r_0 = 3t = 0$ and the zero full-class certificate directly. There is no radius-zero quotient. Repeated downstream uses reuse these uniform events; they are not new calls at fitted random indices. Let

$$\mathcal{F}_2 = \sigma(\mathcal{D}_2), \quad \mathcal{F}_X = \mathcal{F}_2 \vee \sigma(X_{1:n}), \quad \mathcal{F}_T = \mathcal{F}_X \vee \sigma(T_{1:n}).$$

These seven bounds apply to the unconditional iid empirical processes. The nested sigma-fields are used later for the fitted weighted inequality (B.1a) and, in the case of $\mathcal{F}_2$, for the R denominator: given $\mathcal{F}_X$, the u_i's remain independent and centered, and given $\mathcal{F}_T$, the ε_i's do. Let $\mathcal{E}^{\mathrm{proc}}_z$ be the finite intersection of the seven listed uniform process events. Use the same admissible confidence parameter z for all seven bounds; a finite union bound absorbs the seven fixed failure multipliers into K_4 and gives

$$\mathbb{P}\{(\mathcal{E}^{\mathrm{proc}}_z)^c\} \le K_4 e^{-z}, \qquad 1 \le z \le n. \tag{B.5}$$

For each fixed z, this event contains the seven process bounds. We define the comparator, scalar-Bernstein, denominator, and fitted weighted events separately below.

Lemma 13 (Sequential M1 transfer)**.** *Let $\mathcal{G}^{\mathrm{rep}} = \{g_j : j \in \mathbb{N}\} \subset \mathcal{G}$ obey the common envelope and suppose that, for every $g \in \mathcal{G}$, some represented sequence converges pointwise everywhere to g. Then, for every P_X, bounded truth, and finite sample: (i) full-class and represented population infima coincide; (ii) full-class and represented empirical infima coincide for every continuous fit objective used here; (iii) full and represented difference-class suprema coincide for the calibration objectives; and (iv) the ordered first-within-tolerance represented selector is Borel.*

Proof. The common envelope dominates squared differences, so pointwise convergence and dominated convergence give $L_2(P_X)$ convergence and hence (i). Convergence of the finite evaluation vectors, followed by continuity of finite sums, products, absolute values, and squares, gives (ii). Pair representing sequences for g, g' to approximate $g - g'$; the same finite-vector argument gives (iii). Finally, if $L_j(D)$ are the countable Borel objectives, then $L_*(D) = \inf_j L_j(D)$ is Borel and the least index with $L_j(D) \le L_*(D) + \tau_n$ has Borel level sets, proving (iv). □

For every empirical infimum over a public class, use its supplied countable representation and select the first index within τ_n. A countable infimum of Borel functions and its first qualifying index are Borel. The set is nonempty by the definition of infimum, even without attainment. Lemma 13 transfers all full-class population and finite-sample objectives used below to that representation.

B.2 Initial fits

On $\mathcal{D}_2$, let $\widehat{\pi}$ be the first τ_n-minimizer of $P_2(T-g)^2$ over $\mathcal{G}_\pi$, and let $(\widehat{\beta}_0, \widehat{\mu})$ be the first τ_n-minimizer of $P_2(Y - \beta T - g)^2$ over the rational representation of $[-3,3] \times \mathcal{G}_\mu$. For arbitrary $\varepsilon_{\mathrm{cmp},\pi} > 0$,

Lemma 13 supplies a deterministic represented comparator g_π satisfying $\|g_\pi - \pi_0\|_2 \le b + \varepsilon_{\mathrm{cmp},\pi}$. With $h = \widehat{\pi} - g_\pi$ and $r_\pi = g_\pi - \pi_0$, the basic inequality is

$$P_2 h^2 \le 2P_2(uh) - 2P_2\{(g_\pi - \pi_0)h\} + \tau_n. \tag{B.6}$$

For this fixed comparator, define the one-sided scalar-Bernstein event

$$\mathcal{E}^{\mathrm{cmp}}_{\pi,z}(\varepsilon_{\mathrm{cmp},\pi}) := \left\{ P_2 r_\pi^2 \le \|r_\pi\|_2^2 + B_\pi \|r_\pi\|_2 \sqrt{2z/n} + \frac{2B_\pi^2 z}{3n} \right\}.$$

Indeed, $r_\pi^2 \le B_\pi^2$ and $\mathrm{Var}(r_\pi^2) \le B_\pi^2 \|r_\pi\|_2^2$, so its failure is at most $2e^{-z}$, uniformly in the positive slack. On $\mathcal{E}^{\mathrm{proc}}_z \cap \mathcal{E}^{\mathrm{cmp}}_{\pi,z}(\varepsilon_{\mathrm{cmp},\pi})$, put $w_z = \sqrt{z/n}$ and

$$\delta_{\pi,z} := K_{1,\star}(3t \vee h_n) + w_z \le \bar{t} + w_z.$$

Because h belongs to the represented treatment difference core and $P(uh) = 0$, equations (B.3)–(B.4) give

$$\begin{aligned} |P_2(uh)| &\le C_\rho\{\delta_{\pi,z}\|h\|_2 + \delta_{\pi,z}^2\}, \\ P_2 h^2 &\ge \|h\|_2^2 - C_\rho\{\delta_{\pi,z}\|h\|_2 + \delta_{\pi,z}^2\}. \end{aligned} \tag{B.6a}$$

Writing $R_\pi = \|r_\pi\|_{2,n}$, the fixed-comparator event and $\|r_\pi\|_2 \le b + \varepsilon_{\mathrm{cmp},\pi}$ imply

$$R_\pi \le C_\rho\{b + \varepsilon_{\mathrm{cmp},\pi} + w_z\}. \tag{B.6b}$$

Indeed, this follows by taking square roots in the displayed Bernstein bound and using $\sqrt{x^2 + xy + y^2} \le x + y$ after absorbing the fixed envelope coefficients into C_ρ. The upper side of (B.4) similarly gives the population-to-empirical square-process conversion

$$\|h\|_{2,n} \le C_\rho(\|h\|_2 + \delta_{\pi,z}). \tag{B.6c}$$

Thus (B.6), (B.6a), empirical Cauchy–Schwarz, and (B.6c) yield

$$\begin{aligned} \|h\|_2^2 &\le C_\rho\{\delta_{\pi,z}\|h\|_2 + \delta_{\pi,z}^2\} + 2R_\pi\|h\|_{2,n} + \tau_n \\ &\le C_\rho\{(\delta_{\pi,z} + R_\pi)\|h\|_2 + (\delta_{\pi,z} + R_\pi)^2 + \tau_n\}. \end{aligned} \tag{B.6d}$$

Specifically, Young's inequality gives

$$C_\rho(\delta_{\pi,z} + R_\pi)\|h\|_2 \le \frac{1}{2}\|h\|_2^2 + C_\rho(\delta_{\pi,z} + R_\pi)^2,$$

so the quadratic term is absorbed into the left-hand side. Using $\sqrt{\tau_n} = \rho_n$, (B.6b), and $\delta_{\pi,z} \le \bar{t} + w_z$ gives

$$\|h\|_2 \le C_\rho\{b + \varepsilon_{\mathrm{cmp},\pi} + \bar{t} + w_z + \rho_n\}. \tag{B.6e}$$

Finally, the triangle inequality $\|\widehat{\pi} - \pi_0\|_2 \le \|h\|_2 + \|r_\pi\|_2$ gives, with a failure bound independent of $\varepsilon_{\mathrm{cmp},\pi}$,

$$\|\widehat{\pi} - \pi_0\|_2 \le K_5\{b + t + \bar{t} + \sqrt{z/n} + \rho_n\} + C_\rho\varepsilon_{\mathrm{cmp},\pi}.$$

The resulting scalar tail inequality is uniform in the positive comparator slack. Letting $\varepsilon_{\mathrm{cmp},\pi} \downarrow 0$ by continuity of the scalar tail probabilities gives

$$\|\widehat{\pi} - \pi_0\|_2 \le K_5\{b + t + \bar{t} + \sqrt{z/n} + \rho_n\}. \tag{B.7}$$

with failure at most $K_{\pi,2}e^{-z}$. This limit does not intersect the comparator-specific events.

For the joint fit first record the population curvature. For every $\gamma \in \mathbb{R}$ and measurable $f : \mathcal{X} \to \mathbb{R}$, conditional centering gives

$$P(\gamma T + f)^2 = \gamma^2 \mathbb{E}u^2 + \|\gamma\pi_0 + f\|_2^2 \ge \frac{v_-}{2(1 + B_0)^2}\{\gamma^2 + \|f\|_2^2\}. \tag{B.8}$$

Indeed, if $q = \|\gamma\pi_0 + f\|_2$, then

$$\gamma^2 + \|f\|_2^2 \le (1 + 2B_0^2)\gamma^2 + 2q^2.$$

For $c = v_-/\{2(1+B_0)^2\}$, one has $c(1+2B_0^2) \le v_- \le v$ and $2c < 1$, so $c\{\gamma^2 + \|f\|_2^2\} \le v\gamma^2 + q^2$, proving (B.8). The transfer to a joint class requires a separate localization argument. Let $\Delta\mathcal{G}_\mu^{\mathrm{rep}}$ be the supplied countable difference core and define

$$\mathcal{J}_\mu = \{\gamma T + f(X) : \gamma \in \mathbb{Q} \cap [-6, 6],\ f \in \Delta\mathcal{G}_\mu^{\mathrm{rep}}\}.$$

This class is countable, contains zero, and has the fixed envelope $B_{\mathcal{J}} = 6B_0 + 2B_{\mathcal{G}}$. If $j = \gamma T + f(X)$ and $\|j\|_2 \le R$, then the orthogonal identity

$$\|j\|_2^2 = v\gamma^2 + \|\gamma\pi_0 + f\|_2^2$$

implies $|\gamma| \le R/\sqrt{v_-}$ and $\|f\|_2 \le C_f R$. Hence, for $R \ge r_{0,\mu} := C_J(s \vee h_n)$,

$$\begin{aligned}
\mathfrak{R}_{n,P}(R; \mathcal{J}_\mu) &\le \frac{R}{\sqrt{v_-}} \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i T_i\right| \\
&\quad + \mathfrak{R}_{n,P_X}(C_f R; \Delta\mathcal{G}_\mu^{\mathrm{rep}}) \\
&\le \left\{\frac{B_0}{\sqrt{v_-}} h_n + 3c_{\mathrm{rad}} C_f s\right\} R \\
&\le c_{\mathrm{rad}} r_{0,\mu} R,
\end{aligned}$$

where the last inequality follows from the definition of C_J. This uses the public outcome-class budget only at admissible positive radii; when $s = 0$, the floor $C_J h_n$ remains positive. Lemma 12, with base observation (X, T) and conditionally centered multiplier ε, therefore supplies both the multiplier and square-process bounds uniformly over $\mathcal{J}_\mu$.

For completeness, here is the full basic-inequality step. For arbitrary $\varepsilon_{\mathrm{cmp},\mu} > 0$, Lemma 13 supplies deterministic represented comparators $\beta^\circ \in \mathbb{Q} \cap [-3, 3]$ and $g^\circ \in \mathcal{G}_\mu$ satisfying $|\beta^\circ - \beta_0| \le \varepsilon_{\mathrm{cmp},\mu}$ and $\|g^\circ - \mu_0\|_2 \le a + \varepsilon_{\mathrm{cmp},\mu}$, and put

$$e^\circ = (\beta^\circ - \beta_0)T + (g^\circ - \mu_0), \qquad j = (\widehat{\beta}_0 - \beta^\circ)T + (\widehat{\mu} - g^\circ).$$

The approximate empirical minimization gives exactly

$$P_2 j^2 \le 2P_2(\varepsilon j) - 2P_2(e^\circ j) + \tau_n.$$

For this fixed pair of comparators, let $B_{\mathrm{cmp},\mu} := 7B_0 + B_{\mathcal{G}}$ and define

$$\mathcal{E}_{\mu,z}^{\mathrm{cmp}}(\varepsilon_{\mathrm{cmp},\mu}) := \left\{P_2(e^\circ)^2 \le \|e^\circ\|_2^2 + B_{\mathrm{cmp},\mu}\|e^\circ\|_2\sqrt{2z/n} + \frac{2B_{\mathrm{cmp},\mu}^2 z}{3n}\right\}.$$

Here $|e^\circ| \le B_{\mathrm{cmp},\mu}$ and $\mathrm{Var}\{(e^\circ)^2\} \le B_{\mathrm{cmp},\mu}^2\|e^\circ\|_2^2$, so its failure is at most $2e^{-z}$, uniformly in the positive slack. Put $w_z = \sqrt{z/n}$ and $\delta_{\mu,z} = C_{\mathrm{rad}}(s \vee h_n) + w_z$. Since j belongs to the countable joint class $\mathcal{J}_\mu$, the multiplier and square parts of $\mathsf{P}_{J,2}^{\varepsilon}(z)$, before insertion of the fitted index, give, using $P(\varepsilon j) = 0$,

$$\begin{aligned}
|P_2(\varepsilon j)| + |(P_2 - P)j^2| &\le C_\rho\{\delta_{\mu,z}\|j\|_2 + \delta_{\mu,z}^2\}, \\
\|j\|_{2,n} &\le C_\rho(\|j\|_2 + \delta_{\mu,z}).
\end{aligned} \tag{B.8a}$$

The deterministic comparator residual obeys

$$\|e^\circ\|_2 \le B_0|\beta^\circ - \beta_0| + \|g^\circ - \mu_0\|_2 \le a + (1 + B_0)\varepsilon_{\mathrm{cmp},\mu}.$$

Consequently the fixed-comparator Bernstein event gives, for $E_\mu := \|e^\circ\|_{2,n}$,

$$E_\mu \le C_\rho\{a + (1 + B_0)\varepsilon_{\mathrm{cmp},\mu} + w_z\}. \tag{B.8b}$$

The joint basic inequality, the lower square-process comparison in (B.8a), and empirical Cauchy–Schwarz now give

$$\begin{aligned}
\|j\|_2^2 &\le |(P_2 - P)j^2| + 2|P_2(\varepsilon j)| + 2E_\mu\|j\|_{2,n} + \tau_n \\
&\le C_\rho\{(\delta_{\mu,z} + E_\mu)\|j\|_2 + (\delta_{\mu,z} + E_\mu)^2 + \tau_n\}.
\end{aligned} \tag{B.8c}$$

Here the explicit quadratic absorption is

$$C_\rho(\delta_{\mu,z}+E_\mu)\|j\|_2 \le \frac{1}{2}\|j\|_2^2 + C_\rho(\delta_{\mu,z}+E_\mu)^2.$$

Together with $\sqrt{\tau_n}=\rho_n$, it yields

$$\|j\|_2 \le C_\rho\{a+(1+B_0)\varepsilon_{\mathrm{cmp},\mu}+\bar{s}+w_z+\rho_n\}. \tag{B.8d}$$

Set $\Gamma=\widehat{\beta}_0-\beta_0$ and $F=\widehat{\mu}-\mu_0$. Since $\Gamma T+F=e^\circ+j$, the population comparator bound, the triangle inequality, and (B.8) imply

$$\begin{aligned}|\Gamma|+\|F\|_2 &\le \sqrt{2}\{\Gamma^2+\|F\|_2^2\}^{1/2}\\ &\le \frac{2(1+B_0)}{\sqrt{v_-}}\|\Gamma T+F\|_2 \\ &\le C_\rho\{a+(1+B_0)\varepsilon_{\mathrm{cmp},\mu}+\bar{s}+w_z+\rho_n\}.\end{aligned} \tag{B.8e}$$

Equations (B.8a)–(B.8e) use only the already listed joint process call and the fixed-comparator event; they add no joint-complexity certificate. Their failure bound is independent of the positive comparator slack. Letting $\varepsilon_{\mathrm{cmp},\mu}\downarrow 0$ by continuity of the scalar tail probabilities yields

$$|\widehat{\beta}_0-\beta_0|+\|\widehat{\mu}-\mu_0\|_2 \le K_6\{a+s+\bar{s}+\sqrt{z/n}+\rho_n\}. \tag{B.9}$$

with failure at most $K_{\mu,2}e^{-z}$. The limit is taken in scalar tail probabilities, with the estimator fixed. Countable approximate minimization supplies the rule for arbitrary represented classes.

Empirical initial-fit radii. Let K_{sq} be the maximum, after the preceding radius rescaling, of the square-process coefficients for the represented treatment and outcome difference cores, and put

$$c_{\mathrm{sq}}=K_{\mathrm{sq}}+\sqrt{K_{\mathrm{sq}}},\qquad L_\pi=K_5+c_{\mathrm{sq}}+B_\pi+2,\qquad L_\mu=K_6+c_{\mathrm{sq}}+B_\pi+2.$$

The square-process event is uniform over each countable difference core before the fitted index is inserted and implies, on either named half, $\|f\|_{j,n}\le\|f\|_2+c_{\mathrm{sq}}\Delta$. For a deterministic comparator residual r, bounded by B_π, scalar Bernstein gives $\|r\|_{j,n}\le\|r\|_2+C_\rho B_\pi\sqrt{z/n}$. Applying this first on the fitting half $\mathcal{D}_2$, for every positive comparator slack,

$$\|\widehat{\pi}-\pi_0\|_{2,n}\le L_\pi\{b+t+\bar{t}+\sqrt{z/n}+\rho_n\}+2\varepsilon_{\mathrm{cmp},\pi},$$

and

$$|\widehat{\beta}_0-\beta_0|+\|\widehat{\mu}-\mu_0\|_{2,n}\le L_\mu\{a+s+\bar{s}+\sqrt{z/n}+\rho_n\}+2\varepsilon_{\mathrm{cmp},\mu}.$$

For each display the failure bound is independent of the slack. Taking the corresponding slack down to zero in the scalar tail probabilities, rather than intersecting comparator-specific events, yields the same displays without the final slack terms. Define $\mathcal{E}^{\mathrm{fit},2}_{\pi,z}$ to be the event on which (B.7) and the exact $\mathcal{D}_2$ empirical treatment-fit display both hold, and define $\mathcal{E}^{\mathrm{fit},2}_{\mu,z}$ analogously using (B.9). Their failure bounds are

$$\mathbb{P}\{(\mathcal{E}^{\mathrm{fit},2}_{\pi,z})^c\}\le K_{\pi,\mathrm{fit},2}e^{-z},\qquad \mathbb{P}\{(\mathcal{E}^{\mathrm{fit},2}_{\mu,z})^c\}\le K_{\mu,\mathrm{fit},2}e^{-z}.$$

These are events defined by the displayed exact inequalities, not intersections over comparator choices.

The downstream edits instead require $\mathcal{D}_1$ empirical norms. Condition on $\mathcal{F}_2=\sigma(\mathcal{D}_2)$. Then each fitted represented index and hence $h=\widehat{\pi}-g_\pi$ and $f=\widehat{\mu}-g^\circ$ is conditionally fixed, while $\mathcal{D}_1$ remains iid. Before conditioning, the uniform square parts of $\mathsf{P}^{m_0}_{\pi,1}(z)$ and $\mathsf{P}^{u}_{\mu,1}(z)$ were already formed over their entire represented difference cores. Thus they may be evaluated at these conditionally fixed indices and give

$$\|h\|_{1,n}\le\|h\|_2+c_{\mathrm{sq}}\{C_{\mathrm{rad}}(t\vee h_n)+\sqrt{z/n}\},\qquad \|f\|_{1,n}\le\|f\|_2+c_{\mathrm{sq}}\{C_{\mathrm{rad}}(s\vee h_n)+\sqrt{z/n}\}.$$

For each fixed positive proof slack, apply scalar Bernstein on $\mathcal{D}_1$ to the fixed comparator residuals $r_\pi=g_\pi-\pi_0$ and $r_\mu=g^\circ-\mu_0$. Denote the resulting events by $\mathcal{E}^{\mathrm{cmp},1}_{\pi,z}(\varepsilon_{\mathrm{cmp},\pi})$ and $\mathcal{E}^{\mathrm{cmp},1}_{\mu,z}(\varepsilon_{\mathrm{cmp},\mu})$; each has failure at most $2e^{-z}$, $1\le z\le n$, and supplies the preceding fixed-residual norm conversion.

Combining these events with (B.7)–(B.9), the triangle inequality, and then taking each proof slack down to zero by scalar tail-probability continuity gives

$$\|\widehat{\pi} - \pi_0\|_{1,n} \le L_\pi\{b + t + \bar{t} + \sqrt{z/n} + \rho_n\}, \tag{B.9a}$$

$$|\widehat{\beta}_0 - \beta_0| + \|\widehat{\mu} - \mu_0\|_{1,n} \le L_\mu\{a + s + \bar{s} + \sqrt{z/n} + \rho_n\}. \tag{B.9b}$$

No infinite comparator-event intersection is used, and neither slack is an estimator input. Let $\mathcal{E}^{\mathrm{tr},1}_{\pi,z}$ and $\mathcal{E}^{\mathrm{tr},1}_{\mu,z}$ be the exact events defined respectively by (B.7),(B.9a) and by (B.9),(B.9b). The finite unions just described give

$$\mathbb{P}\{(\mathcal{E}^{\mathrm{tr},1}_{\pi,z})^c\} \le K_{\pi,\mathrm{tr},1}e^{-z}, \qquad \mathbb{P}\{(\mathcal{E}^{\mathrm{tr},1}_{\mu,z})^c\} \le K_{\mu,\mathrm{tr},1}e^{-z}, \qquad 1 \le z \le n.$$

These are population-to-$\mathcal{D}_1$ transfer events; they are distinct from the $\mathcal{D}_2$ fit events above and change only fixed failure multipliers.

B.3 The O edit: definition, normalization, and calibration

Put $x = a + s$, $y = b + t$, and recall the fixed envelope B_π. Put

$$d_\pi = \min\{B_\pi, 8(y + \bar{s} + \rho_n)\}, \qquad \lambda_O = (\bar{s}/d_\pi)^2.$$

The cap is fixed because $|\widehat{\pi} - \pi_0| \le B_{\mathcal{G}} + B_0 = B_\pi$ pointwise. On the O-safe cell

$$x \le y, \qquad \bar{s} \le \vartheta_\rho,$$

both arguments in the minimum are at least $8\bar{s}$. Hence

$$d_\pi \ge 8\bar{s}, \qquad 0 < \lambda_O \le 1/64 < 1/16.$$

Define

$$\widetilde{D}_O = (v_-/2) \vee P_1\{T(T - \widehat{\pi})\}, \qquad \widetilde{a}_i = \frac{T_i - \widehat{\pi}(X_i)}{\widetilde{D}_O}.$$

For $f \in \Delta\mathcal{G}_\mu$, $\gamma \in \mathbb{R}$, and $a \in \mathbb{R}^n$, let

$$\begin{aligned} \Phi_\mu(f, \gamma, a) &= |P_1\{(\gamma T + f)a\} - \gamma| - P_1 f^2, \\ H_\mu(a) &= \sup_{\gamma \in \mathbb{R},\, f \in \Delta\mathcal{G}_\mu} \Phi_\mu(f, \gamma, a), \\ F_O(a) &= \lambda_O \|a - \widetilde{a}\|_n^2 + H_\mu(a). \end{aligned}$$

Here and below, the estimator takes the supremum over the countable represented core. On a finite sample, Lemma 13(iii) identifies its value with the displayed full-class supremum. Because $0 \in \Delta\mathcal{G}_\mu$, $H_\mu(a) < \infty$ if and only if $P_1(Ta) = 1$. On this hyperplane,

$$H_\mu(a) = \sup_f\{|P_1(fa)| - P_1 f^2\}. \tag{B.10}$$

For a, a' on this normalization hyperplane, the uniform envelope gives

$$|H_\mu(a) - H_\mu(a')| \le \sup_{f \in \Delta\mathcal{G}_\mu} |P_1\{f(a - a')\}| \le 2B_{\mathcal{G}}\|a - a'\|_n.$$

Hence the reduced objective is continuous on the hyperplane; no continuity of the extended-valued H_μ on all of $\mathbb{R}^n$ is asserted. If $T_{1:n} = 0$, set

$$\widehat{a} := 0, \qquad \widehat{\beta}_O^{\mathrm{raw}} := 0.$$

On $T_{1:n} \ne 0$, choose the first largest-magnitude coordinate j, enumerate rational values of the other coordinates, and set

$$a_j = \frac{n - \sum_{i \ne j} T_i a_i}{T_j}. \tag{B.11}$$

This is a Borel dense chart of the exact hyperplane, even for irrational T_j. The objective is continuous and coercive there. The first point within τ_n of its countable infimum is a Borel τ_n-minimizer $\widehat{a}$ satisfying $P_1(T\widehat{a}) = 1$, and we put

$$\widehat{\beta}_O^{\mathrm{raw}} = P_1\{(Y - \widehat{\mu})\widehat{a}\}.$$

Finally define

$$\widehat{\beta}_O = \Pi_{[-3,3]}(\widehat{\beta}_O^{\mathrm{raw}}).$$

With $p_\rho = v_-/B_0^2$,

$$\mathbb{E}T^2 = \mathbb{E}\pi_0(X)^2 + v \geq v_-, \qquad \Pr(T \neq 0) \geq \frac{\mathbb{E}T^2}{B_0^2} \geq p_\rho.$$

Independence of the D1 observations therefore gives

$$\Pr(T_{1:n} = 0) \leq e^{-p_\rho n},$$

and projection bounds the loss on this event by three. Therefore

$$3\Pr(T_{1:n} = 0) \leq 3e^{-p_\rho n} \leq \frac{3}{\sqrt{2ep_\rho}}\, n^{-1/2}.$$

The O proof below is on $T_{1:n} \neq 0$; this display accounts separately for the zero-vector branch.

For the proof-only comparator let $D_0 = P_1(Tu)$ and $a_0 = u/D_0$. Since $\mathbb{E}(Tu) = v$, $|Tu| \leq B_0^2$, and $\mathrm{Var}(Tu) \leq \mathbb{E}(T^2u^2) \leq B_0^2 v$, define

$$\mathcal{E}_{D_0,z} := \left\{ |D_0 - v| \leq B_0 \sqrt{\frac{2vz}{n}} + \frac{4B_0^2 z}{3n} \right\}.$$

Lemma 10 gives $\mathbb{P}(\mathcal{E}_{D_0,z}^{\mathrm{c}}) \leq 2e^{-z}$: the centered range satisfies $|Tu - v| \leq 2B_0^2$, which gives the displayed $4/3$ linear coefficient. With $c_0 = v_-/(32B_0^2)$, the square-root and linear deviations are at most $v/4$ and $v/24$, respectively, whenever $1 \leq z \leq c_0 n$. Hence $D_0 \geq 17v/24 \geq v_-/2$ on $\mathcal{E}_{D_0,z}$, and $P_1(Ta_0) = 1$. On $\mathcal{E}_z^{\mathrm{proc}} \cap \mathcal{E}_{D_0,z}$, apply the localized multiplier and square bounds to (B.10). Because $\mathbb{E}\{f(X)u\} = 0$, with $w_z = \sqrt{z/n}$ they give

$$|P_1(fa_0)| \leq C_\rho\{(\bar{s} + w_z)\|f\|_2 + (\bar{s} + w_z)^2\}, \quad P_1 f^2 \geq \|f\|_2^2 - C_\rho\{(\bar{s} + w_z)\|f\|_2 + (\bar{s} + w_z)^2\}.$$

Maximizing the resulting concave quadratic gives the all-z benchmark

$$H_\mu(a_0) \leq K_7\{\bar{s}^2 + z/n\}. \tag{B.12}$$

Projection of the fitted denominator onto $[v_-/2, \infty)$ is one-Lipschitz and leaves D_0 fixed. With $e_\pi = \widehat{\pi} - \pi_0$,

$$|\widetilde{D}_O - D_0| \leq |P_1(Te_\pi)| \leq B_0\|e_\pi\|_{1,n}.$$

On $D_0 \geq v_-/2$, put

$$C_u := \frac{2}{v_-} + \frac{4B_0^2}{v_-^2}.$$

The same denominator algebra gives

$$\|\widetilde{a} - a_0\|_n \leq C_u\|\widehat{\pi} - \pi_0\|_{1,n}.$$

If the cap in d_π is inactive, $\mathcal{E}_{\pi,z}^{\mathrm{tr},1}$ and (B.9a), $\bar{t} \leq C_{\mathrm{rad}} y + \bar{s}$ bound the deterministic part by a fixed multiple of d_π. If the cap is active, the pointwise envelope gives $|\widehat{\pi} - \pi_0| \leq B_\pi = d_\pi$. Put

$$K_8^* := C_u \max\left\{1, L_\pi, \frac{L_\pi(1 + C_{\mathrm{rad}})}{8}\right\}.$$

Thus, including cap equality,

$$\|\widetilde{a} - a_0\|_n \leq K_8^*\{d_\pi + \sqrt{z/n}\}. \tag{B.13}$$

This covers active and inactive caps.

Compare the approximate objective at $\widehat{a}$ and a_0. Since $H_\mu \geq 0$,

$$\|\widehat{a} - \widetilde{a}\|_n \leq K_9\{d_\pi + \sqrt{z/n} + (d_\pi/\bar{s})\sqrt{z/n} + \sqrt{\tau_n/\lambda_O}\}, \tag{B.14}$$

$$H_\mu(\widehat{a}) \leq K_9\{\bar{s}^2 + z/n + \lambda_O(d_\pi + \sqrt{z/n})^2 + \tau_n\}. \tag{B.15}$$

Indeed, approximate optimality first bounds $\lambda_O\|\widehat{a} - \widetilde{a}\|_n^2 + H_\mu(\widehat{a})$ by $K_\rho\{\lambda_O(d_\pi + w_z)^2 + \bar{s}^2 + w_z^2 + \tau_n\}$; only the norm bound divides by λ_O, and $\sqrt{\lambda_O} \asymp_\rho \bar{s}/d_\pi$. This is the source of the amplified term in (B.14). Exact normalization and (B.10) now give, for every represented f,

$$|P_1(f\widehat{a})| \leq P_1 f^2 + H_\mu(\widehat{a}). \tag{B.16}$$

This establishes the one-sided O calibration inequality. Equations (B.12)–(B.16) hold on the finite intersection $\mathcal{E}_z^{\mathrm{proc}} \cap \mathcal{E}_{D_0,z} \cap \mathcal{E}_{\pi,z}^{\mathrm{tr},1}$, for $1 \leq z \leq (c_0 \wedge 1)n$; each fitted substitution is made only after the corresponding uniform or exact-fit event has been formed.

B.4 THE O ALL-Z ORACLE INEQUALITY AND EXPECTATION

Recall the piecewise raw O output and its clipped version defined above. Choose a deterministic represented g_μ, available by Lemma 13, such that $\|g_\mu - \mu_0\|_2 \le a + \rho_n$, and put $w = u/v$ and $r_\mu = \mu_0 - g_\mu$. Exact normalization gives the exhaustive identity

$$\begin{aligned}\widehat{\beta}_O^{\mathrm{raw}} - \beta_0 = P_1(\varepsilon w) + P_1\{\varepsilon(\widehat{a} - w)\} + P_1\{(\mu_0 - g_\mu)w\} \\ + P_1\{(\mu_0 - g_\mu)(\widehat{a} - w)\} + P_1\{(g_\mu - \widehat{\mu})\widehat{a}\}.\end{aligned} \tag{B.17}$$

Let $r = \bar{s}$, $d = d_\pi$, $w_z = \sqrt{z/n}$, and $\ell = \sqrt{\tau_n/\lambda_O}$. On $\mathcal{E}_{D_0,z}$, the preceding deviation bound and $\|u\|_n \le B_0$ give $\|a_0 - w\|_n \le K_\rho w_z$, hence (B.13)–(B.14) imply

$$\|\widehat{a} - w\|_n \le K\{d + w_z + (d/r)w_z + \ell\}.$$

For this fixed comparator, let

$$\mathcal{E}_{O,z}^{\mathrm{cmp}} := \{\|r_\mu\|_{1,n} \le K_{O,\mathrm{cmp}}(a + \rho_n + w_z)\}.$$

Bounded scalar Bernstein applied to $r_\mu(X)^2$ gives $\mathbb{P}\{(\mathcal{E}_{O,z}^{\mathrm{cmp}})^c\} \le 2e^{-z}$, after one fixed enlargement of $K_{O,\mathrm{cmp}}$. No fitted index occurs in this event. Define the fitted weighted event

$$\begin{aligned}\mathcal{E}_{O,z}^{\mathrm{wt}} := \Big\{ &|P_1(\varepsilon w)| \le K_{O,\mathrm{wt}} w_z, \\ &|P_1\{\varepsilon(\widehat{a} - w)\}| \le K_{O,\mathrm{wt}} \|\widehat{a} - w\|_n w_z, \\ &|P_1(r_\mu w)| \le K_{O,\mathrm{wt}} \|r_\mu\|_{1,n} w_z \Big\}.\end{aligned}$$

The first two inequalities follow from (B.1a) conditional on $\mathcal{F}_T$; the last follows from (B.1a) conditional on $\mathcal{F}_X$, because $\mathbb{E}(u \mid X) = 0$. Taking expectations of each conditional failure over its own conditioning field and then applying an unconditional union bound gives failure at most $6e^{-z}$.

On the finite event

$$\mathcal{E}_{O,z} := \mathcal{E}_z^{\mathrm{proc}} \cap \mathcal{E}_{\pi,z}^{\mathrm{tr},1} \cap \mathcal{E}_{\mu,z}^{\mathrm{tr},1} \cap \mathcal{E}_{D_0,z} \cap \mathcal{E}_{O,z}^{\mathrm{cmp}} \cap \mathcal{E}_{O,z}^{\mathrm{wt}}, \qquad \mathbb{P}(\mathcal{E}_{O,z}^c) \le K_{O,0} e^{-z},$$

Equation (B.1a), the exact fit events, and (B.15)–(B.16) give the five separate bounds

$$\begin{aligned}|P_1(\varepsilon w)| &\le K w_z, \\ |P_1\{\varepsilon(\widehat{a} - w)\}| &\le K\{d + w_z + (d/r)w_z + \ell\} w_z, \\ |P_1\{(\mu_0 - g_\mu)w\}| &\le K(a w_z + w_z^2), \\ |P_1\{(\mu_0 - g_\mu)(\widehat{a} - w)\}| &\le K(a + w_z)\{d + w_z + (d/r)w_z + \ell\}, \\ |P_1\{(g_\mu - \widehat{\mu})\widehat{a}\}| &\le K\{(x + r + w_z)^2 + r^2 + w_z^2 + \lambda_O(d + w_z)^2 + \tau_n\}.\end{aligned}$$

Adding these inequalities gives the displayed bound first for the raw output. Projection cannot increase distance to the legal target $\beta_0 \in [-3, 3]$, so the same tail bound holds for the clipped $\widehat{\beta}_O$. Retaining a common nonnegative amplification envelope for the subsequent layer-cake calculation gives

$$\begin{aligned}|\widehat{\beta}_O - \beta_0| \le K_{10} \Big[&ad + x^2 + r^2 + \lambda_O d^2 + \tau_n + \ell(a + r) \\ &+ (1 + a + d + x + r)w_z + \frac{ad}{r} w_z + \left(1 + \frac{d}{r}\right) w_z^2 + w_z^3 + \ell w_z \Big]\end{aligned} \tag{B.18}$$

on $\mathcal{E}_{O,z}$. Choose the fixed endpoint $c_2 \le c_0 \wedge 1$ and $K_{11} \ge K_{O,0}$; then its failure is at most $K_{11}e^{-z}$, $1 \le z \le c_2 n$. The displayed dw_z is the edited-noise contribution. The amplified terms $(ad/r)w_z$ and $(d/r)w_z^2$ come from multiplying the norm-bound term $(d/r)w_z$ in (B.14) by a and w_z, respectively, in the five-term identity. Equation (B.13) accounts for the denominator perturbation.

Lemma 14 (O tail integration)**.** *On the O-safe cell, every monomial in (B.18), integrated against its pointwise tail, is bounded by $K(q + ay + x^2)$. The complete accounting is*

tail monomial	*layer-cake scale*	*deterministic control*
$(1 + a + d + x + r)w_z$	q	Kq
$\frac{ad}{r} w_z$	$\frac{ad}{r} q$	$K(q + ay + x^2)$
w_z^2	q^2	Kq
$\frac{d}{r} w_z^2$	$\frac{d}{nr}$	Kq
w_z^3	q^3	Kq
ℓw_z	ℓq	$K(q + ay + x^2)$.

(B.19)

The deterministic terms $ad, x^2, r^2, \lambda_O d^2, \tau_n$, and $\ell(a+r)$ obey the same target bound. This statement includes zero budgets, both cap branches, and cap equality. It is pointwise in z, not simultaneous in z; unsafe cells are handled only by the later fallback.

Proof. Use

$$h_n^2 \le 2q, \quad r \ge C_{\mathrm{rad}} h_n, \quad d \le B_\pi, \quad d \le K_d(y + r + \rho_n), \quad \lambda_O d^2 = K_\lambda r^2.$$

Because $s \le x \le y$, $r^2 \le K(x^2 + q)$, $\rho_n \le q \le Kr$, and $ad \le K(ay + x^2 + q)$: when $h_n \le y$, send ah_n to ay; when $y < h_n$, use $a \le x \le y < h_n$ and $ah_n \le h_n^2 \le 2q$. Since $q/r \le K$, the same split proves $(ad/r)q \le K(q + ay + x^2)$, while

$$\frac{d}{nr} \le \frac{B_\pi}{C_{\mathrm{rad}} \sqrt{n \log(en)}} \le Kq, \qquad \ell = \frac{\rho_n}{\sqrt{\lambda_O}} = \frac{\rho_n d}{r}.$$

Thus $\ell(a+r) + \ell q$ has the claimed bound. The remaining rows use $q^2, q^3 \le q$ and the fixed envelopes. The positive floor $r = C_{\mathrm{rad}}(s \vee h_n)$ makes every ratio defined when $s = 0$, and the two inequalities for d cover the active, inactive, and equality cases of the cap. Finally, $\int_1^\infty z^{p/2} e^{-z}\, dz \le \Gamma(1 + p/2)$ gives the listed scales. The interval below $z = 1$ uses the $z = 1$ threshold; clipping handles $z > c_2 n$; and if $c_2 n < 1$, then $q > \sqrt{c_2}$ absorbs the bounded loss. This uses a separate event for each z, never one simultaneous event. □

Applying Lemma 14 to (B.18) and adding the zero-vector contribution $3/\sqrt{2ep_\rho}\, n^{-1/2}$ to the fixed O constant, we obtain

$$\sup_{P \in \mathcal{P}(\Gamma)} \mathbb{E}_P |\widehat{\beta}_O - \beta(P)| \le C_O\{q + ay + x^2\}. \tag{B.20}$$

B.5 The R edit and two-sample calibration

We prove the R bound directly from the public-class process lemma and the initial-fit bounds. Let $m_0(X) = \mathbb{E}(Y \mid X) = \beta_0 \pi_0(X) + \mu_0(X)$, and recall the fixed envelope B_{all}. Put

$$d_{\mathrm{all}} = \min\{B_{\mathrm{all}}, 8(x + y + \bar{t} + \rho_n)\}, \qquad \lambda_R = (\bar{t}/d_{\mathrm{all}})^2.$$

Here $|m_0| \le B_0$ by conditional expectation of the bounded response, while $|\widehat{\beta}_0 \widehat{\pi} + \widehat{\mu}| \le 4B_{\mathcal{G}}$. Thus B_{all} is a fixed envelope for their discrepancy. On the R-safe cell $y < x$, $\bar{t} \le \vartheta_\rho$, both arguments in the minimum are at least $8\bar{t}$, so

$$d_{\mathrm{all}} \ge 8\bar{t}, \qquad 0 < \lambda_R \le 1/64 < 1/16.$$

On D1 covariates set $\widetilde{m}_i = \widehat{\beta}_0 \widehat{\pi}(X_i) + \widehat{\mu}(X_i)$. For $f \in \Delta\mathcal{G}_\pi$, define

$$H_\pi(m) = \sup_f \{|P_2(Yf) - P_1(mf)| - P_1 f^2\}.$$

For every $m, m' \in \mathbb{R}^n$, the uniform envelope similarly gives

$$|H_\pi(m) - H_\pi(m')| \le \sup_{f \in \Delta\mathcal{G}_\pi} |P_1\{f(m - m')\}| \le 2B_{\mathcal{G}} \|m - m'\|_n.$$

Thus H_π is continuous on $\mathbb{R}^n$; together with the positive quadratic penalty and coercivity below, this supports first-within-tolerance selection on the rational dense chart. The implemented supremum is

over the represented difference core and equals the full-class display by Lemma 13(iii). Over the rational enumeration of $\mathbb{R}^n$, select the first τ_n-minimizer $\widehat{m}$ of $\lambda_R\|m-\widetilde{m}\|_n^2 + H_\pi(m)$. The positive quadratic is coercive, so this Borel selection is well defined.

Put $\xi = Y - m_0(X)$ and $\delta_z = K(\bar{t} + \sqrt{z/n})$. On $\mathcal{E}_z^{\mathrm{proc}}$, uniformly in $f \in \Delta\mathcal{G}_\pi$, (B.3)–(B.4) give

$$\begin{aligned}|P_2(\xi f)| + |(P_2 - P)(m_0 f)| + |(P_1 - P)(m_0 f)| + |(P_1 - P)f^2| \\ \le K\{\delta_z \|f\|_2 + \delta_z^2\}.\end{aligned} \tag{B.21}$$

Inserting $m = m_0(X_{1:n})$ and maximizing $K\delta_z r - r^2/2$ proves

$$H_\pi(m_0(X_{1:n})) \le K\delta_z^2. \tag{B.22}$$

The exact decomposition

$$\widetilde{m} - m_0 = (\widehat{\beta}_0 - \beta_0)\widehat{\pi} + \beta_0(\widehat{\pi} - \pi_0) + (\widehat{\mu} - \mu_0)$$

and $\mathcal{E}_{\pi,z}^{\mathrm{tr},1} \cap \mathcal{E}_{\mu,z}^{\mathrm{tr},1}$, together with (B.9a)–(B.9b) and $\bar{s} \le C_{\mathrm{rad}} x + \bar{t}$ on $y < x$, give the following bound after putting

$$L_R := (B_{\mathcal{G}} + 1)L_\mu(1 + C_{\mathrm{rad}}) + 3L_\pi :$$

$$\|\widetilde{m} - m_0\|_{1,n} \le L_R\{x + y + \bar{t} + \rho_n + \sqrt{z/n}\}.$$

If the cap is inactive, the deterministic base is a fixed multiple of d_{all}; if active, the pointwise envelope gives $|\widetilde{m} - m_0| \le B_{\mathrm{all}} = d_{\mathrm{all}}$. Put $K_m^* := 1 + L_R$. Hence, including equality,

$$\|\widetilde{m} - m_0\|_{1,n} \le K_m^*\{d_{\mathrm{all}} + \sqrt{z/n}\}.$$

Objective comparison therefore yields

$$H_\pi(\widehat{m}) \le K\{\delta_z^2 + \lambda_R(d_{\mathrm{all}} + \delta_z)^2 + \tau_n\}, \tag{B.23}$$

$$\|\widehat{m} - m_0\|_n \le K\{d_{\mathrm{all}} + \delta_z + \delta_z/\sqrt{\lambda_R} + \sqrt{\tau_n/\lambda_R}\}. \tag{B.24}$$

For every represented f, the identity

$$P_1\{(m_0 - \widehat{m})f\} = [P_2(Yf) - P_1(\widehat{m}f)] - [P_2(Yf) - P_1(m_0 f)]$$

and (B.21)–(B.23) imply

$$|P_1\{(m_0 - \widehat{m})f\}| \le K\{\|f\|_2^2 + \delta_z^2 + \lambda_R(d_{\mathrm{all}} + \delta_z)^2 + \tau_n\}. \tag{B.25}$$

The event is uniform before substitution, so (B.25) applies legally to $f = \widehat{\pi} - g_\pi$. The edit uses D2 outcomes and D1 covariates only; conditioning has not frozen the D1 treatments. Equations (B.21)–(B.25) hold on $\mathcal{E}_z^{\mathrm{proc}} \cap \mathcal{E}_{\pi,z}^{\mathrm{tr},1} \cap \mathcal{E}_{\mu,z}^{\mathrm{tr},1}$.

B.6 The R denominator, oracle inequality, and expectation

Let $e_\pi = \widehat{\pi} - \pi_0$, $r_m = \widehat{m} - m_0$, and $D = P_1(T - \widehat{\pi})^2 = P_1(u - e_\pi)^2$. Conditional on $\mathcal{F}_2$, the function e_π is fixed and the variables

$$W_i = (u_i - e_\pi(X_i))^2 = (T_i - \widehat{\pi}(X_i))^2$$

are conditionally iid. Their conditional center is

$$m_\pi := \mathbb{E}(W_i \mid \mathcal{F}_2) = v + \|e_\pi\|_2^2, \tag{B.26}$$

where sample-split independence and $\mathbb{E}(u \mid X) = 0$ were used, with $v = \mathbb{E}u^2$. Put $B_D := B_0 + B_{\mathcal{G}} = B_\pi$. Since $0 \le W_i \le B_D^2$, also $|W_i - m_\pi| \le B_D^2$, and

$$\mathrm{Var}(W_i \mid \mathcal{F}_2) \le \mathbb{E}(W_i^2 \mid \mathcal{F}_2) \le B_D^2 m_\pi.$$

Lemma 10, conditionally on $\mathcal{F}_2$, therefore uses $B = B_D^2$, $\sigma^2 = B_D^2 m_\pi$, and defines the event

$$\mathcal{E}_{R,z}^{\mathrm{den}} := \left\{ |D - m_\pi| \le B_D\sqrt{\frac{2m_\pi z}{n}} + \frac{2B_D^2 z}{3n} \right\}, \qquad \mathbb{P}\{(\mathcal{E}_{R,z}^{\mathrm{den}})^c \mid \mathcal{F}_2\} \le 2e^{-z}.$$

The same failure bound holds unconditionally. With $c_{\mathrm{den}} := v_-/(32B_D^2)$, the square-root term is at most $m_\pi/4$, and the linear term is at most $v_-/48 \le m_\pi/48$, whenever $1 \le z \le c_{\mathrm{den}} n$. Consequently, on this event,

$$v_-/2 \le m_\pi/2 \le D \le K_{\mathrm{den}} := B_D^2, \qquad 1 \le z \le c_{\mathrm{den}} n. \tag{B.27}$$

The upper bound is also immediate from $D = P_1 W \le B_D^2$. Thus the floor $(v_-/2) \vee D$ is inactive on $\mathcal{E}_{R,z}^{\mathrm{den}}$ and defines the estimator on its complement.

The raw R estimator is

$$\widehat{\beta}_R^{\mathrm{raw}} = \frac{P_1\{(Y - \widehat{m})(T - \widehat{\pi})\}}{(v_-/2) \vee P_1(T - \widehat{\pi})^2}.$$

Define its clipped version by

$$\widehat{\beta}_R = \Pi_{[-3,3]}(\widehat{\beta}_R^{\mathrm{raw}}).$$

On (B.27), direct substitution gives the exact signed identity

$$\begin{aligned} D(\widehat{\beta}_R^{\mathrm{raw}} - \beta_0) = P_1\{\varepsilon(u - e_\pi)\} - P_1\{r_m(u - e_\pi)\} \\ + \beta_0\{P_1(ue_\pi) - P_1 e_\pi^2\}. \end{aligned} \tag{B.28}$$

The identity follows by substituting the model into the R ratio on the denominator event. With $r = \bar{t}$, $d = d_{\mathrm{all}}$, $w_z = \sqrt{z/n}$, and $\ell = \sqrt{\tau_n/\lambda_R}$, (B.9a), (B.24), and $\delta_z = K(r + w_z)$ give

$$\|e_\pi\|_{1,n} \le K(y + r + w_z), \qquad \|r_m\|_{1,n} \le K\{d + r + w_z + (d/r)w_z + \ell\}.$$

Choose a deterministic represented g_π, available by Lemma 13, such that $\|g_\pi - \pi_0\|_2 \le b + \rho_n$. Its fixed-comparator event is

$$\mathcal{E}_{R,z}^{\mathrm{cmp}} := \{\|g_\pi - \pi_0\|_{1,n} \le K_{R,\mathrm{cmp}}(b + \rho_n + w_z)\}.$$

Bounded scalar Bernstein applied to $(g_\pi - \pi_0)^2$ gives $\mathbb{P}\{(\mathcal{E}_{R,z}^{\mathrm{cmp}})^c\} \le 2e^{-z}$. Define also

$$\begin{aligned} \mathcal{E}_{R,z}^{\mathrm{wt}} := \Big\{ & |P_1\{\varepsilon(u - e_\pi)\}| \le K_{R,\mathrm{wt}} \|u - e_\pi\|_{1,n} w_z, \\ & |P_1(r_m u)| \le K_{R,\mathrm{wt}} \|r_m\|_{1,n} w_z, \\ & |P_1(ue_\pi)| \le K_{R,\mathrm{wt}} \|e_\pi\|_{1,n} w_z \Big\}. \end{aligned}$$

The first inequality follows from (B.1a) conditional on $\mathcal{F}_T$; the other two follow from (B.1a) conditional on $\mathcal{F}_X$. Taking expectations of each conditional failure over its own conditioning field and then applying an unconditional union bound gives failure at most $6e^{-z}$. The remaining fitted propensity calibration is supplied by (B.25). Decomposing $e_\pi = (\widehat{\pi} - g_\pi) + (g_\pi - \pi_0)$, (B.1a), (B.25), and empirical Cauchy–Schwarz yield

$$\begin{aligned} |P_1\{\varepsilon(u - e_\pi)\}| &\le K(1 + y + r + w_z)w_z, \\ |P_1(r_m u)| &\le K\{d + r + w_z + (d/r)w_z + \ell\}w_z, \\ |P_1(r_m e_\pi)| &\le K\{(y + r + w_z)^2 + (r + w_z)^2 + \lambda_R(d + r + w_z)^2 + \tau_n \\ &\qquad + (b + w_z)[d + r + w_z + (d/r)w_z + \ell]\}, \\ |P_1(ue_\pi)| + P_1 e_\pi^2 &\le K\{(y + r + w_z)w_z + (y + r + w_z)^2\}. \end{aligned}$$

On the finite event

$$\mathcal{E}_{R,z} := \mathcal{E}_z^{\mathrm{proc}} \cap \mathcal{E}_{\pi,z}^{\mathrm{tr},1} \cap \mathcal{E}_{\mu,z}^{\mathrm{tr},1} \cap \mathcal{E}_{R,z}^{\mathrm{den}} \cap \mathcal{E}_{R,z}^{\mathrm{cmp}} \cap \mathcal{E}_{R,z}^{\mathrm{wt}}, \qquad \mathbb{P}(\mathcal{E}_{R,z}^c) \le K_{R,0} e^{-z},$$

the uniform events precede every fitted substitution, so no random D1 index enters a pointwise event. Adding the four displays and dividing by the floor $D \ge v_-/2$ gives the complete tail for the raw output. Projection cannot increase distance to $\beta_0 \in [-3, 3]$, so the same tail holds for the clipped $\widehat{\beta}_R$:

$$\begin{aligned} |\widehat{\beta}_R - \beta_0| \le K_{12} \Big[& bd + y^2 + r^2 + \lambda_R d^2 + \tau_n + \ell(b + r) \\ & + (1 + b + d + y + r)w_z + \frac{bd}{r} w_z + \left(1 + \frac{d}{r}\right) w_z^2 + w_z^3 + \ell w_z \Big], \end{aligned} \tag{B.29}$$

Choose $c_3 \le c_{\mathrm{den}} \wedge 1$ and $K_{13} \ge K_{R,0}$. Then (B.29) has failure at most $K_{13} e^{-z}$, for $1 \le z \le c_3 n$.

Lemma 15 (R tail integration). *On the R-safe cell, the complete stochastic accounting for (B.29) is*

tail monomial	*layer-cake scale*	*deterministic control*
$(1+b+d+y+r)w_z$	q	Kq
$\frac{bd}{r}w_z$	$\frac{bd}{r}q$	$K(q+bx+y^2)$
w_z^2	q^2	Kq
$\frac{d}{r}w_z^2$	$\frac{d}{nr}$	Kq
w_z^3	q^3	Kq
ℓw_z	ℓq	$K(q+bx+y^2)$.

The deterministic terms $bd, y^2, r^2, \lambda_R d^2, \tau_n$, and $\ell(b+r)$ obey the same bound, including zero budgets, both cap branches, and cap equality. The events are pointwise in z, not simultaneous in z; unsafe cells use the public fallback.

Proof. From

$$r \geq C_{\mathrm{rad}} h_n, \quad d \leq B_{\mathrm{all}}, \quad d \leq K_d(x+y+r+\rho_n), \quad \lambda_R d^2 = K_\lambda r^2,$$

we get

$$bd \leq K(bx+y^2+q), \qquad \frac{bd}{r}q \leq K(bx+y^2+q), \qquad \frac{d}{nr} \leq \frac{B_{\mathrm{all}}}{C_{\mathrm{rad}}\sqrt{n\log(en)}} \leq Kq.$$

Indeed, $h_n \leq x$ sends bh_n to bx, whereas $x < h_n$ and $b \leq y < x$ send it to $h_n^2 \leq 2q$. Also $r^2 \leq K(y^2+q)$ and $\ell = \rho_n d/r$, so the same split controls $\ell(b+r)+\ell q$. The other rows follow from fixed envelopes and $q^2, q^3 \leq q$. The floor $r = C_{\mathrm{rad}}(t \vee h_n) > 0$ covers $t = 0$, and the two bounds on d cover both cap branches and equality. The layer-cake moments are the same gamma integrals as in Lemma 14. Below $z = 1$ use the $z = 1$ threshold; above $c_3 n$ use clipping; if $c_3 n < 1$, then $q > \sqrt{c_3}$ absorbs the bounded loss. No simultaneous-in-z event is asserted. □

Lemma 15 and (B.29) prove

$$\sup_{P \in \mathcal{P}(\Gamma)} \mathbb{E}_P|\widehat{\beta}_R - \beta(P)| \leq C_R\{q+bx+y^2\}. \tag{B.30}$$

B.7 Summary of probability events

The following summary records the events used in the proof, their conditioning fields, failure probabilities, admissible z-ranges, and applications. All bounds are for each fixed z.

1. $\mathsf{P}^u_{\pi,2}(z)$ is the simultaneous (B.3)–(B.4) event from Lemma 12 for the represented treatment difference core on $\mathcal{D}_2$, with random multiplier u. It is unconditional, concerns the iid $\mathcal{D}_2$ observations, has failure $K^u_{\pi,2}e^{-z}$ for $1 \leq z \leq n$, and is used in (B.6)–(B.7) and the treatment empirical-radius conversion.
2. $\mathsf{P}^{\mathrm{sq}}_{\mu,2}(z)$ is the (B.4) event for the represented outcome difference core on $\mathcal{D}_2$. It is unconditional, concerns the $\mathcal{D}_2$ covariates, has failure $K^{\mathrm{sq}}_{\mu,2}e^{-z}$ for $1 \leq z \leq n$, and is used in the outcome empirical-radius conversion following (B.9).
3. $\mathsf{P}^{\varepsilon}_{J,2}(z)$ is the simultaneous (B.3)–(B.4) event for $\mathcal{J}_\mu$ on $\mathcal{D}_2$, with multiplier ε. It is unconditional, concerns the (X,T,Y) variables in $\mathcal{D}_2$, has failure $K^{\varepsilon}_{J,2}e^{-z}$ for $1 \leq z \leq n$, and is used in (B.8)–(B.9).
4. $\mathsf{P}^u_{\mu,1}(z)$ is the simultaneous (B.3)–(B.4) event for the represented outcome difference core on $\mathcal{D}_1$, with multiplier u. It is unconditional, concerns the iid $\mathcal{D}_1$ observations, has failure $K^u_{\mu,1}e^{-z}$ for $1 \leq z \leq n$, and its square part is used in (B.9b), while its multiplier and square parts are used in (B.12)–(B.16).
5. $\mathsf{P}^{\xi}_{\pi,2}(z)$ is the (B.3) event for the represented treatment difference core on $\mathcal{D}_2$, with $\xi = Y - m_0(X)$. It is unconditional, concerns $\mathcal{D}_2$, has failure $K^{\xi}_{\pi,2}e^{-z}$ for $1 \leq z \leq n$, and is used in (B.21)–(B.22).

6. $\mathsf{P}^{m_0}_{\pi,2}(z)$ is the centered (B.3) event for the same represented core on $\mathcal{D}_2$, with fixed multiplier $m_0(X)$. It is unconditional, concerns $\mathcal{D}_2$, has failure $K^{m_0}_{\pi,2}e^{-z}$ for $1 \le z \le n$, and is used in (B.21)–(B.22).
7. $\mathsf{P}^{m_0}_{\pi,1}(z)$ is the simultaneous fixed-multiplier (B.3) and square (B.4) event for the represented treatment difference core on $\mathcal{D}_1$. It is unconditional, concerns $\mathcal{D}_1$, has failure $K^{m_0}_{\pi,1}e^{-z}$ for $1 \le z \le n$, and is used in (B.9a) and (B.21)–(B.25).
8. $\mathcal{E}^{\mathrm{proc}}_z$ is exactly the finite intersection of the preceding seven process events. It is unconditional, involves both sample halves as just itemized, satisfies (B.5) for $1 \le z \le n$, and is used by the initial fits, O calibration, and R calibration. It contains none of the scalar, comparator, denominator, or fitted weighted events.
9. For each fixed positive proof slack, $\mathcal{E}^{\mathrm{cmp}}_{\pi,z}(\varepsilon_{\mathrm{cmp},\pi})$ is the displayed one-sided bound on $P_2 r_\pi^2$. It is unconditional, involves only the fixed comparator residual on $\mathcal{D}_2$, has failure at most $2e^{-z}$ for $1 \le z \le n$, and is used in (B.6)–(B.7). The exact treatment-fit event $\mathcal{E}^{\mathrm{fit},2}_{\pi,z}$ is then defined by (B.7) and its $\mathcal{D}_2$ empirical analogue after scalar tail-probability continuity; it is unconditional, concerns $\widehat{\pi}$ and $\mathcal{D}_2$, and has failure $K_{\pi,\mathrm{fit},2}e^{-z}$ for $1 \le z \le n$.
10. For each fixed positive proof slack, $\mathcal{E}^{\mathrm{cmp}}_{\mu,z}(\varepsilon_{\mathrm{cmp},\mu})$ is the displayed one-sided bound on $P_2(e^\circ)^2$. It is unconditional, involves the fixed joint comparator on $\mathcal{D}_2$, has failure at most $2e^{-z}$ for $1 \le z \le n$, and is used in (B.8)–(B.9). The exact joint-fit event $\mathcal{E}^{\mathrm{fit},2}_{\mu,z}$ is defined by (B.9) and its $\mathcal{D}_2$ empirical analogue after scalar tail-probability continuity; it is unconditional, concerns $(\widehat{\beta}_0, \widehat{\mu})$ and $\mathcal{D}_2$, and has failure $K_{\mu,\mathrm{fit},2}e^{-z}$ for $1 \le z \le n$.
11. For each fixed positive slack, the D1 fixed-residual events $\mathcal{E}^{\mathrm{cmp},1}_{\pi,z}(\varepsilon_{\mathrm{cmp},\pi})$ and $\mathcal{E}^{\mathrm{cmp},1}_{\mu,z}(\varepsilon_{\mathrm{cmp},\mu})$ are scalar Bernstein events for $g_\pi - \pi_0$ and $g^\circ - \mu_0$ on $\mathcal{D}_1$, each with failure at most $2e^{-z}$. Conditional on $\mathcal{F}_2$, the fitted represented indices are fixed, while the D1 square-process events are uniform before insertion. Combining these facts with (B.7)–(B.9), then using scalar tail-probability continuity rather than an infinite event intersection, defines $\mathcal{E}^{\mathrm{tr},1}_{\pi,z}$ and $\mathcal{E}^{\mathrm{tr},1}_{\mu,z}$. They are unconditional, have failures $K_{\pi,\mathrm{tr},1}e^{-z}$ and $K_{\mu,\mathrm{tr},1}e^{-z}$ for $1 \le z \le n$, and supply (B.9a)–(B.9b) at every D1 use site: (B.13), (B.18)–(B.20), (B.23)–(B.25), and (B.29)–(B.30).
12. $\mathcal{E}_{D_0,z}$ is the displayed scalar Bernstein event for $D_0 = P_1(Tu)$. It is unconditional, involves (T, u) on $\mathcal{D}_1$, has failure at most $2e^{-z}$, and for $1 \le z \le c_0 n$ supplies the normalization and denominator bounds used in (B.12)–(B.18).
13. $\mathcal{E}^{\mathrm{cmp}}_{O,z}$ is the displayed empirical-norm event for the fixed $r_\mu = \mu_0 - g_\mu$. It is unconditional, involves the $\mathcal{D}_1$ covariates, has failure at most $2e^{-z}$ for $1 \le z \le n$, and is used in the third and fourth terms of (B.17). The event $\mathcal{E}^{\mathrm{wt}}_{O,z}$ is the displayed intersection of three weighted inequalities: the first two are conditional on $\mathcal{F}_T$ and the third on $\mathcal{F}_X$; they involve respectively εw, $\varepsilon(\widehat{a} - w)$, and $r_\mu w$. After integrating each conditional bound, their unconditional union failure is at most $6e^{-z}$ for $1 \le z \le n$; they are used in (B.17)–(B.18).
14. $\mathcal{E}_{O,z}$ is exactly the displayed six-event finite intersection of $\mathcal{E}^{\mathrm{proc}}_z$, the two D1 transfer events, $\mathcal{E}_{D_0,z}$, and the O comparator and weighted events. It is unconditional after integrating the conditional bounds, has failure $K_{O,0}e^{-z}$ for $1 \le z \le c_2 n$, and supports (B.18), whose layer-cake integration gives (B.20).
15. $\mathcal{E}^{\mathrm{den}}_{R,z}$ is the displayed conditional Bernstein event for $W_i = (T_i - \widehat{\pi}(X_i))^2$, conditional on $\mathcal{F}_2$. Because e_π is $\mathcal{F}_2$-measurable, conditioning fixes it and leaves the $\mathcal{D}_1$ variables W_i conditionally iid, with center $m_\pi = v + \|e_\pi\|_2^2$, envelope $0 \le W_i \le B_D^2$, and variance proxy $\mathrm{Var}(W_i \mid \mathcal{F}_2) \le B_D^2 m_\pi$. The event has conditional and unconditional failure at most $2e^{-z}$ for $1 \le z \le c_{\mathrm{den}} n$; the choices $c_{\mathrm{den}} = v_-/(32B_D^2)$ and $K_{\mathrm{den}} = B_D^2$ imply (B.27) and supply the denominator in (B.28)–(B.29).
16. $\mathcal{E}^{\mathrm{cmp}}_{R,z}$ is the displayed empirical-norm event for the fixed $g_\pi - \pi_0$. It is unconditional, involves the $\mathcal{D}_1$ covariates, has failure at most $2e^{-z}$ for $1 \le z \le n$, and is used after the decomposition following (B.28). The event $\mathcal{E}^{\mathrm{wt}}_{R,z}$ is the displayed intersection of three weighted inequalities: the first is conditional on $\mathcal{F}_T$ and the other two on $\mathcal{F}_X$; they involve $\varepsilon(u - e_\pi)$, $r_m u$, and $u e_\pi$. After integrating each conditional bound, their unconditional union failure is at most $6e^{-z}$ for $1 \le z \le n$; they are used in (B.29).

17. $\mathcal{E}_{R,z}$ is exactly the displayed six-event finite intersection of $\mathcal{E}_z^{\mathrm{proc}}$, the two D1 transfer events, and the R denominator, comparator, and weighted events. It is unconditional after integrating the conditional bounds, has failure $K_{R,0}e^{-z}$ for $1 \le z \le c_3 n$, and supports (B.29), whose layer-cake integration gives (B.30).

B.8 Public four-cell selector and Q1 lift

Define $T_\mu = ay + x^2$, $T_\pi = bx + y^2$, and

$$I_O = \mathbf{1}\{x \le y,\ \bar{s} \le \vartheta_\rho\}, \qquad I_R = \mathbf{1}\{y < x,\ \bar{t} \le \vartheta_\rho\}.$$

Define

$$\widehat{\beta}_\Gamma = I_O \widehat{\beta}_O + I_R \widehat{\beta}_R.$$

When both indicators vanish the value is zero. Ties go to O and safe equalities remain safe. Since the indicators are disjoint and zero belongs to $[-3, 3]$, branchwise clipping equals clipping after selection.

Let $W = T_\mu \wedge T_\pi$. If $x \le y$, then $T_\mu \le 2xy \le 2T_\pi$; the mirror holds when $y < x$. Put $V = \min\{a + s^2, xy, T_\mu, T_\pi\}$. Directly,

$$V \le W \le 4V, \qquad X/5 \le V \le 6X. \tag{B.31}$$

For the last inequality, use the following six exhaustive weak-order cells, with the stated boundary convention:

$(a \ge t, b \le s)$	$T_\pi \le 6X$
$(a \le t, b \ge s)$	$T_\mu \le 6X$
$(a \ge t, b \ge s, a \ge b)$	$T_\pi \le 6X$
$(a \ge t, b \ge s, a < b)$	$T_\mu \le 6X$
$(a \le t, b \le s, s \le t)$	$T_\mu \le 6X$
$(a \le t, b \le s, t < s)$	$T_\pi \le 6X$.

Weak equalities are assigned by the displayed first matching row, so the cells cover all zeros and ties without changing the bound. This proves

$$T_{\mathrm{selected}} \le 2W \le 48X. \tag{B.32}$$

For the zero fallback use the fixed δ_ρ. If $h_n > \delta_\rho$, then $\delta_\rho^2 < h_n^2 \le 2q$. If $h_n \le \delta_\rho$ and the O-oriented cell is unsafe, then $s > \delta_\rho$. Either $t \ge s/2$ and $M \ge s^2/4$, or $t < s/2$ and $x \le y$ imply $P > s^2/2$. The R-oriented argument is the exact mirror. Since clipped zero loses at most three, on every unsafe cell

$$|0 - \beta(P)| \le 12\delta_\rho^{-2}(q + X) \tag{B.33}$$

on every unsafe cell.

Equations (B.20), (B.30), (B.32), and (B.33) prove, uniformly for every fixed legal Γ,

$$\sup_{P \in \mathcal{P}(\Gamma)} \mathbb{E}_P |\widehat{\beta}_\Gamma - \beta(P)| \le C_\rho(q + X).$$

Every fit, edit, indicator, fallback, and clipping operation is Borel by the countable and finite-dimensional constructions above. Therefore this rule is a legal competitor inside the fixed-Γ infimum. Taking the outer supremum of this uniform fixed-pair bound proves the upper half of Theorem 1 in the stated minimax order.

B.9 Proof of the allocation corollary

Proof. Theorem 1 and $W \le T_{\mathrm{selected}} \le 2W$, $X/5 \le W \le 24X$, give $R_{n,\rho}^{\mathrm{pair}} \asymp_\rho q + T_{\mathrm{selected}}$. Division by $xy > 0$ gives the first assertion, including necessity because all three terms are nonnegative. For the allocation bounds, $X \le 4xy$: each of AB, M, O, P is at most xy. Also $X \ge x^2/4$: one of a, s is at least $x/2$, and one of b, t is at least $y/2$; the four combinations force respectively $ab, O, P, M \ge x^2/4$. Finally the allocations $(0, x, y, 0)$ and $(x, 0, y, 0)$ have $X = x^2(1 + y)$ and $X = xy$, respectively. □

C TESTING TOOLS AND THE PARAMETRIC AND P LOWER BOUNDS

C.1 FINITE TESTING TOOLS

Lemma 16 (Expected-loss testing, chi-square, and finite Gram mixtures)**.** *Let Q_0, Q_1 be laws whose component parameters equal θ_0, θ_1, respectively, with $\Delta = |\theta_1 - \theta_0|$. Then every decision rule satisfies*

$$\max_{\ell=0,1} \mathbb{E}_{Q_\ell} |\widehat{\theta} - \theta_\ell| \geq \frac{\Delta}{4}\{1 - \mathrm{TV}(Q_0, Q_1)\}. \tag{C.1}$$

If $P \ll Q$, then

$$\mathrm{TV}(P, Q) \leq \frac{1}{2}\sqrt{\chi^2(P, Q)}. \tag{C.2}$$

Finally, suppose $P_{\ell,j,\sigma}$, $j \in [p]$, $\sigma \in \{-1, 1\}$, have common reference Q, and their one-observation likelihood Gram entries are $1 + K_\ell$ for the same coordinate/sign, $1 - K_\ell$ for the same coordinate/opposite sign, and one otherwise. For the uniform N-sample mixture $\overline{P}_\ell^N$,

$$\chi^2(\overline{P}_\ell^N, Q^N) = \frac{\{(1 + K_\ell)^N + (1 - K_\ell)^N\}/2 - 1}{p}. \tag{C.3}$$

Proof. Threshold $\widehat{\theta}$ at the midpoint. On a testing error its absolute loss is at least $\Delta/2$. The sum of the two testing errors is at least $1 - \mathrm{TV}(Q_0, Q_1)$, proving (C.1). For (C.2), write $L = dP/dQ$ and apply Cauchy–Schwarz to $\frac{1}{2}\mathbb{E}_Q|L - 1|$.

For (C.3), expand the square of the uniform mixture likelihood. Product independence raises every one-observation Gram entry to the N-th power. Among the $(2p)^2$ ordered pairs there are $2p$ equal-sign diagonal pairs, $2p$ opposite-sign same-coordinate pairs, and $4p(p-1)$ off-coordinate pairs. Divide their sum by $4p^2$ and subtract one. □

Lemma 17 (Finite-class complexity)**.** *If a finite class $\mathcal{H}$ has cardinality M and empirical envelope $\max_{h\in\mathcal{H}}\|h\|_n \leq R$, then*

$$\mathbb{E}_\epsilon \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^n \epsilon_i h(X_i) \leq R\sqrt{\frac{2\log M}{n}}. \tag{C.4}$$

Proof. For $\lambda > 0$, the maximum is bounded by the log-sum-exp divided by λ. Conditional independence of the signs and $\cosh x \leq e^{x^2/2}$ bound its conditional expectation by $\log M/\lambda + \lambda R^2/(2n)$. Optimize at $\lambda = \sqrt{2n\log M}/R$. A localized subclass only decreases the supremum, so (C.4) is valid at every radius. □

Every application below takes $\mathcal{H}$ to be a symmetric finite difference class; hence $\mathcal{H} = -\mathcal{H}$, and the one-sided supremum in (C.4) equals its absolute-value version.

C.2 THE EXACT THREE-POINT TREATMENT CHANNEL

Fix the strict lower-endpoint slack stratum, write $v_0 = v_-$, and put

$$m_\rho = \min\left\{\frac{1}{64}, \frac{v_0}{4B_0}, \frac{B_0 - \sqrt{v_0}}{4}\right\}, \qquad \lambda_\rho = \min\left\{1, \frac{192c_{\mathrm{rad}}}{23}, 128B_{\mathcal{G}}\right\}, \qquad L = \sqrt{v_0} + 2m_\rho. \tag{C.5}$$

Then $m_\rho > 0$, $0 < \lambda_\rho \leq 1$, $v_0 + m_\rho^2 < L^2$, $Lm_\rho \leq v_0/4$, and $L + m_\rho < B_0$. For $|m| \leq m_\rho$, define on $\{-L, 0, L\}$

$$p_L(m) = \frac{v_0 + m^2 + Lm}{2L^2}, \qquad p_{-L}(m) = \frac{v_0 + m^2 - Lm}{2L^2},$$
$$p_0(m) = 1 - \frac{v_0 + m^2}{L^2}. \tag{C.6}$$

All masses are positive: the smaller side numerator is at least $v_0 - Lm_\rho \geq 3v_0/4$, and the middle numerator is positive by (C.5). They sum to one, and direct calculation gives

$$\mathbb{E}T = m, \qquad \mathbb{E}T^2 = v_0 + m^2, \qquad \mathrm{Var}(T) = v_0. \tag{C.7}$$

Moreover $|T| < B_0$ and $|T - m| < B_0$. This is the only treatment channel used in the four nonparametric witnesses.

C.3 A NATIVE PARAMETRIC LOWER FOR EVERY BUDGET CELL

Place a point-mass covariate law on the fixed ambient space (so X is constant), let both public classes be $\{0\}$, take $T = \pm\sqrt{v_0}$ equiprobably, and set

$$d_\rho^\star = \min\{3, B_0/16, B_0/(16\sqrt{v_-})\}, \qquad \delta_n = d_\rho^\star/\sqrt{n}.$$

For $\theta \in \{-\delta_n, \delta_n\}$, let $Y = (B_0/2)R$, where $R \in \{-1, 1\}$ and

$$\mathbb{P}_\theta(R = 1 \mid T = t) = \frac{1}{2} + \frac{\theta t}{B_0}. \tag{C.8}$$

Then $\mathbb{E}(Y \mid T) = \theta T$, so $\beta = \theta$, both nuisances vanish, and $\varepsilon = Y - \theta T$ is conditionally centered. The probabilities lie in $[7/16, 9/16]$;

$$|\beta| = |\theta| \leq d_\rho^\star \leq 3, \qquad |\varepsilon| \leq B_0/2 + |\theta|\sqrt{v_0} \leq 9B_0/16.$$

Thus target membership and all envelopes hold for the raw law before any estimator clipping; the treatment residual variance is exactly v_0.

The scalar experiment also has a common-reference representation. On its four (t, r) atoms,

$$P_\theta(t, r) = \frac{1}{4}\left(1 + \frac{2r\theta t}{B_0}\right), \qquad Q(t, r) = \frac{1}{4}.$$

Thus $dP_{\pm\delta_n}/dQ = 1 \pm h$, where $h(t, r) = 2r\delta_n t/B_0$, $\mathbb{E}_Q h = 0$, and

$$K_q = \mathbb{E}_Q h^2 = \frac{4(d_\rho^\star)^2 v_0}{nB_0^2} \leq \frac{1}{64n}.$$

The one-observation diagonal and cross Gram entries are $1 + K_q$ and $1 - K_q$. Product independence gives the exact second moment

$$\chi^2(P_{\pm\delta_n}^N, Q^N) = (1 + K_q)^N - 1 < \frac{1}{31} < \frac{1}{16}, \qquad N = 2n.$$

Indeed $(1 + K_q)^N \leq e^{1/32} < 32/31$, where $e^x < 1/(1 - x)$ for $0 < x < 1$. Hence each label is within TV $1/(2\sqrt{31}) < 3/32$ of Q^N, and the cross-label TV is below $3/16 < 1/4$. This independently verifies positivity, likelihood ratios, both Gram entries, product second moment, and TV conversion for the native q experiment. Lemma 16 yields

$$R_{n,\rho}^{\mathrm{pair}}(a, s, b, t) \geq c_{q,\rho} q, \qquad c_{q,\rho} = 3d_\rho^\star/8 > 0, \tag{C.9}$$

for every declared budget cell. The public singleton pair is fixed before the label and is a finite M1 pair.

C.4 P FAMILY: OBSERVATION LAWS AND MODEL CONDITIONS

For every Walsh construction in Appendices C–F, preselect a measurable injection

$$\iota : \bigsqcup_{m \geq 1} (\mathbb{Z}/2\mathbb{Z})^m \longrightarrow \mathcal{X}$$

into the same fixed ambient space. Such an injection exists because the domain is countable and every singleton in an infinite standard Borel space is measurable. When p directions are required, set $m = \lceil \log_2(p + 1) \rceil$, let S_m be the finite image of the m-th block, and take P_X uniform on S_m. Choose p distinct nonzero characters and pull each back through ι_m^{-1} on S_m, extending it by zero off

S_m. The extended characters are measurable, bounded by one, and under this P_X have mean zero, unit $L_2(P_X)$ norm, and pairwise orthogonal Gram matrix. Their finite public classes are their own sequential M1 representations. Thus only the finite support law and public pair vary; $(\mathcal{X}, \mathcal{A})$ never does.

Assume first $n \geq 16$, $s^2 \geq q$, and $s(b \wedge s) > 0$. Let

$$N = 2n, \quad w = b \wedge s, \quad p = \left\lceil e^{64Ns^2} \right\rceil, \quad A = B_0/2,$$

and use the p characters from the convention above. They are Rademacher-valued, mean zero, and orthonormal. For $z = \sigma\phi_j$, define

$$\eta = \frac{m_\rho w}{4}, \quad \gamma = \frac{\lambda_\rho s}{256}, \quad \Delta = \frac{2\eta\gamma}{v_0}, \quad \kappa = \gamma + \Delta\eta. \tag{C.10}$$

The two labels are

$$(\beta_0, \pi_0, \mu_0) = (0, z\eta, z\gamma), \qquad (\beta_1, \pi_1, \mu_1) = (\Delta, -z\eta, z\kappa). \tag{C.11}$$

Use (C.6) with treatment means $\pm z\eta$. Given $X, T = t$, let $Y = AR$, $R \in \{-1, 1\}$, and

$$\mathbb{P}_\ell(R = r \mid X, T = t) = \frac{1}{2}\left[1 + r\frac{\beta_\ell t + \mu_\ell(X)}{A}\right]. \tag{C.12}$$

The scales satisfy $\eta \leq 1/256$, $\gamma \leq 1/256$, $\Delta \leq 1/(2048B_0)$, and $\kappa \leq 2\gamma$. Also $\Delta L + \kappa < A$, and $\kappa \leq 2\gamma \leq \lambda_\rho/128 \leq B_\mathcal{G}$. Thus every conditional probability is strict; all variables, nuisances, target, u, and ε obey the B_0 envelope; and every public candidate obeys the $B_\mathcal{G}$ envelope.

For $t \in \{-L, 0, L\}$, write

$$a_{\pm L} = \frac{v_0 + \eta^2}{2L^2}, \quad a_0 = 1 - \frac{v_0 + \eta^2}{L^2}, \quad b_t = \frac{\eta t}{2L^2}.$$

Suppressing X, the complete six-atom masses are

$$\begin{aligned} P_{0,z}(t, r) &= \tfrac{1}{2}(a_t + zb_t)(1 + rz\gamma/A), \\ P_{1,z}(t, r) &= \tfrac{1}{2}(a_t - zb_t)(1 + r(\Delta t + z\kappa)/A). \end{aligned} \tag{C.13}$$

Their factors are positive, and summing first over r, then t, gives one. Equation (C.7) gives exact residual variance v_0 under both labels. Moreover

$$\mathbb{E}_0(Y \mid X) = \mathbb{E}_1(Y \mid X) = z\gamma, \qquad \mathbb{E}_0(TY \mid X) = \mathbb{E}_1(TY \mid X) = \eta\gamma,$$

because $\kappa = \gamma + \Delta\eta$ and $\Delta v_0 = 2\eta\gamma$. Hence (C.13) defines observation laws satisfying all PLM restrictions.

C.5 P PUBLIC PAIR AND COMPLEXITY BOUNDS

Before choosing the label, coordinate, or sign, publish

$$\mathcal{G}_\pi = \{0\}, \qquad \mathcal{G}_\mu = \{z_{j,\sigma}\gamma, z_{j,\sigma}\kappa : j \in [p],\ \sigma \in \{-1, 1\}\}. \tag{C.14}$$

Both propensity truths have distance $\eta \leq b$ from $\mathcal{G}_\pi$, whose difference class has zero complexity. Both outcome truths belong to $\mathcal{G}_\mu$, so their outcome approximation error is zero. The outcome difference class has at most $16p^2$ elements and envelope 2κ. Because $Ns^2 \geq 8$,

$$\log(32p^2) \leq 129Ns^2 = 258ns^2.$$

Lemma 17, used globally and hence on every localized subset, gives, for every $r' \geq s$,

$$\mathfrak{R}_{n,P_X}(r'; \Delta\mathcal{G}_\mu) \leq 2\kappa\sqrt{\frac{2\log(32p^2)}{n}} < 46\kappa s \leq \frac{23\lambda_\rho}{64}s^2.$$

Consequently the normalized all-radii ratio is at most

$$\frac{23\lambda_\rho}{192}s \leq c_{\mathrm{rad}}s$$

for every $r' \geq s$ with $r' > 0$. At $s = 0$, the class is the singleton zero class and its complexity is zero at every positive radius. The covariate support and public classes are finite inside the fixed infinite ambient space, so listing the classes is a sequential M1 representation. The exponentially large dictionary is public; the realized (j, σ, ℓ) is not learner input.

C.6 P COMMON REFERENCE AND LIKELIHOOD KERNELS

Averaging (C.13) over $z = \pm 1$, label zero has reference

$$Q_0(t, r) = \tfrac{1}{2}(a_t + r b_t \gamma / A).$$

The label-one average equals this atomwise exactly when

$$\Delta a_t t = b_t(\kappa + \gamma). \tag{C.15}$$

For $t = \pm L$, multiplying by $2L^2/t$ turns (C.15) into $\Delta(v_0 + \eta^2) = \eta(\kappa + \gamma)$, which follows from (C.10). At $t = 0$, both sides are separately zero. Thus

$$Q(t, r) = \tfrac{1}{2}(a_t + r b_t \gamma / A) \tag{C.16}$$

is common to both labels. It is strictly positive because $|b_t|/a_t \le s/16$ and $\gamma/A \le s/384$.

Put $c = \gamma/A$, $k = \kappa/A$, and $d_t = \Delta t/A$. Dividing the raw masses by (C.16) gives

$$\frac{dP_{0,z}}{dQ} = 1 + z h_0, \quad h_0 = \frac{b_t + r a_t c}{a_t + r b_t c},$$

$$\frac{dP_{1,z}}{dQ} = 1 + z h_1, \quad h_1 = \frac{r a_t k - b_t(1 + r d_t)}{a_t + r b_t c}. \tag{C.17}$$

Both coefficients are centered under Q: summing $Q h_0$ over r leaves b_t, while summing $Q h_1$ leaves $-b_t$, and $\sum_t b_t = 0$. The bounds

$$|h_0| < s/15, \qquad |h_1| < s/14$$

follow from $|b_t|/a_t \le s/16$, $c \le s/384$, $k \le s/192$, and $|d_t| \le s/3072$. Hence

$$0 \le K_0 = \mathbb{E}_Q h_0^2 \le s^2/225, \qquad 0 \le K_1 = \mathbb{E}_Q h_1^2 \le s^2/196. \tag{C.18}$$

This likelihood calculation has an independent raw-mass check. Write $P_{\ell,z} = Q + z C_\ell$, where

$$C_0(t, r) = \tfrac{1}{2}(b_t + r a_t c), \qquad C_1(t, r) = \tfrac{1}{2}\{r a_t k - b_t(1 + r d_t)\}.$$

Then $\sum C_\ell = 0$, and direct multiplication gives

$$\sum_{t,r} \frac{P_{\ell,z} P_{\ell,z'}}{Q} = 1 + z z' \sum_{t,r} \frac{C_\ell^2}{Q}. \tag{C.19}$$

The last sum is exactly K_ℓ in (C.18), without using the likelihood formula. Restoring $z = \sigma \phi_j(X)$, Walsh orthogonality gives Gram entries $1 + K_\ell$, $1 - K_\ell$, and one for equal signs, opposite signs, and distinct coordinates, respectively.

C.7 P MIXTURE DISTANCE, TESTING, AND ALL-N CLOSURE

Let $\overline{P}_\ell^N$ be the uniform mixture over the same $2p$ coordinate/sign choices within label ℓ. Lemma 16 and (C.18) give the exact identity

$$\chi^2(\overline{P}_\ell^N, Q^N) = \frac{\{(1 + K_\ell)^N + (1 - K_\ell)^N\}/2 - 1}{p} \le e^{-(64 - 1/196) N s^2} < e^{-511}. \tag{C.20}$$

Thus each mixture is within TV $e^{-511/2}/2$ of the same Q^N, and $\mathrm{TV}(\overline{P}_0^N, \overline{P}_1^N) < 1/2$. Every component in label zero has target zero, and every component in label one has

$$\Delta = \frac{\lambda_\rho m_\rho}{512 v_0} s(b \wedge s).$$

For each rule tailored to the fixed public pair, its law supremum dominates the finite Bayes average. Lemma 16, followed by the inner infimum and only then the outer pair supremum, gives

$$R_{n,\rho}^{\mathrm{pair}}(a, s, b, t) \ge \frac{\lambda_\rho m_\rho}{4096 v_0} s(b \wedge s)$$

on the nonparametric branch.

The remaining cases use only the already proved scalar risk (C.9), by a maximum of two valid lower bounds. If $s^2 < q$, then $s(b \wedge s) < q$. If $n < 16$, then $s(b \wedge s) \le 1 \le 4q$. At $s = 0$ or $b = 0$ the desired term is zero. These cases are exhaustive, including $n = 1$ and equality at $n = 16$. Therefore, for all $n \ge 1$ and all budgets,

$$R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \ge c_{P,\rho} s(b \wedge s), \qquad c_{P,\rho} = \min\left\{ \frac{\lambda_\rho m_\rho}{4096 v_0}, \frac{3 d^\star_\rho}{32} \right\} > 0. \tag{C.21}$$

All laws in the nonparametric prior have exact variance $v_0 = v_-$, so (C.21) also covers singleton variance intervals. No Bayes risks are added; Appendix L embeds this active native P witness with q and the other active native families; inactive component inequalities reuse the common q splice before the fixed-pair maximum.

D THE OUTCOME-ORIENTED O LOWER BOUND

Assume first $n \ge 16$, $t^2 \ge q$, and $t(a \wedge t) > 0$. Let $w = a \wedge t$, $N = 2n$,

$$\begin{aligned} \lambda_\rho &= \min\left\{1, \frac{192 c_{\mathrm{rad}}}{23}, 128 B_{\mathcal{G}}\right\}, \quad \gamma = \frac{\lambda_\rho m_\rho t}{4}, \qquad \eta = \frac{w}{256}, \\ \Delta &= \frac{\gamma\eta}{v_0 + \gamma^2}, \qquad\qquad\qquad p = \left\lceil e^{64 N t^2} \right\rceil, \quad A = B_0/2. \end{aligned} \tag{D.1}$$

Use the fixed-space, finite-support Walsh convention in Appendix C with this value of p, and write $z = \sigma\phi_j$. Both labels have the common propensity $\pi = z\gamma$ and use the channel (C.6) at mean $z\gamma$. Their target/outcome-nuisance pairs are

$$(\beta_0, \mu_0) = (0, z\eta), \qquad (\beta_1, \mu_1) = (\Delta, 0). \tag{D.2}$$

Given $X, T = u$, set $Y = AR$ with

$$\mathbb{P}_\ell(R = r \mid X, T = u) = \frac{1}{2}\left[1 + r\frac{\beta_\ell u + \mu_\ell(X)}{A}\right]. \tag{D.3}$$

D.1 RAW LAWS, PLM, VARIANCE, AND ENVELOPES

For $u \in \{-L, 0, L\}$, write

$$a_{\pm L} = \frac{v_0 + \gamma^2}{2L^2}, \quad a_0 = 1 - \frac{v_0 + \gamma^2}{L^2}, \quad b_u = \frac{\gamma u}{2L^2}.$$

The complete conditional masses are

$$\begin{aligned} P_{0,z}(u, r) &= \tfrac{1}{2}(a_u + z b_u)(1 + r z\eta/A), \\ P_{1,z}(u, r) &= \tfrac{1}{2}(a_u + z b_u)(1 + r\Delta u/A). \end{aligned} \tag{D.4}$$

There are no omitted atoms. The treatment factors are positive by (C.5)–(C.7), while $\eta < A$ and $\Delta L < A$, so the outcome factors are positive. Summing over r and then u proves normalization for each label and sign.

The channel gives $\mathbb{E}(T \mid X) = z\gamma$ and $\mathrm{Var}(T \mid X) = v_0$ under both labels. Equation (D.3) gives $\mathbb{E}(Y \mid X, T) = \beta_\ell T + \mu_\ell(X)$, so the PLM residual is conditionally centered. In particular,

$$\mathbb{E}_0(TY \mid X) = \gamma\eta = \Delta(v_0 + \gamma^2) = \mathbb{E}_1(TY \mid X). \tag{D.5}$$

The scales satisfy $\gamma \le m_\rho/4$, $\eta \le 1/256$, $\Delta L \le \eta/4$, and $\Delta \le 1/(4096 B_0)$. Hence $T, u, Y, \varepsilon, \pi, \mu, \beta$ obey the original law envelope. Moreover $\gamma \le \lambda_\rho/256 \le B_{\mathcal{G}}$, so every public candidate obeys the declared class envelope. The observation laws and public classes therefore satisfy the model and envelope conditions.

D.2 One public pair and both budget types

Before the label, coordinate, or sign is selected, publish

$$\mathcal{G}_\mu = \{0\}, \qquad \mathcal{G}_\pi = \{\sigma\gamma\phi_j : j \in [p],\ \sigma \in \{-1, 1\}\}. \tag{D.6}$$

Label zero has outcome approximation error $\eta \le a$; label one has zero error. The outcome stochastic complexity is zero. Both propensity truths lie in $\mathcal{G}_\pi$, so propensity approximation is zero. The propensity difference class has at most $4p^2$ elements and envelope 2γ. Since $\log(32p^2) \le 258nt^2$, Lemma 17 gives, for every $r' \ge t$,

$$\mathfrak{R}_{n,P_X}(r'; \Delta\mathcal{G}_\pi) < 46\gamma t, \qquad \frac{\mathfrak{R}_{n,P_X}(r'; \Delta\mathcal{G}_\pi)}{3r'} \le \frac{23\lambda_\rho}{384}t \le \frac{c_{\mathrm{rad}}}{2}t \le c_{\mathrm{rad}}t. \tag{D.7}$$

The inequalities hold for every $r' \ge t$ with $r' > 0$. At $t = 0$, the class is $\{0\}$ and its complexity is zero at every positive radius. The finite pair is M1. The dictionary is public, while the realized coordinate, sign, and label remain latent.

D.3 Common reference and two independent Gram routes

The sign average of label zero is

$$Q_0(u, r) = \tfrac{1}{2}(a_u + rb_u\eta/A),$$

while label one has $Q_1(u, r) = \frac{1}{2}a_u(1 + r\Delta u/A)$. They agree exactly when

$$\Delta a_u u = b_u\eta. \tag{D.8}$$

For $u = \pm L$, (D.8) is $\Delta(v_0 + \gamma^2) = \gamma\eta$; at $u = 0$, both sides vanish. Thus their common reference is

$$Q(u, r) = \tfrac{1}{2}(a_u + rb_u\eta/A). \tag{D.9}$$

It is positive because $|b_u|/a_u \le t/4$ and $\eta/A \le t/384$.

Put $e = \eta/A$ and $d_u = \Delta u/A$. Direct division yields

$$h_0(u, r) = \frac{b_u + ra_ue}{a_u + rb_ue}, \qquad h_1(u, r) = \frac{b_u}{a_u}. \tag{D.10}$$

The second formula keeps the label-one outcome factor: it cancels exactly between numerator and the positive reference. Both coefficients are centered because summing Qh_ℓ over r leaves b_u, and $\sum_u b_u = 0$. Therefore

$$\begin{aligned} K_0 &= \sum_{u,r} \frac{(b_u + ra_ue)^2}{2(a_u + rb_ue)} \le t^2/9, \\ K_1 &= \sum_u \frac{b_u^2}{a_u} = \frac{\gamma^2}{v_0 + \gamma^2} \le t^2/256. \end{aligned} \tag{D.11}$$

Independently, write $P_{\ell,z} = Q + zC_\ell$, where

$$C_0 = \tfrac{1}{2}(b_u + ra_ue), \qquad C_1 = \tfrac{1}{2}b_u(1 + rd_u).$$

Direct raw-mass multiplication gives

$$\sum_{u,r} \frac{P_{\ell,z}P_{\ell,z'}}{Q} = 1 + zz' \sum_{u,r} \frac{C_\ell^2}{Q}. \tag{D.12}$$

The two sums are exactly those in (D.11); for label one the positive factor $1 + rd_u$ cancels only after multiplication. Walsh orthogonality then gives the full Gram entries $1 + K_\ell$, $1 - K_\ell$, and one according as coordinates/signs agree, signs oppose, or coordinates differ.

D.4 Mixture distance, testing, and all-n closure

Mix uniformly over the same $2p$ coordinate/sign choices within each label. Lemma 16 and (D.11) give

$$\chi^2(\overline{P}_\ell^N, Q^N) \le \frac{e^{NK_\ell}}{p} \le e^{-(64-1/9)Nt^2} < e^{-511}. \tag{D.13}$$

Therefore the cross-label TV is below $1/2$. The target separation obeys

$$\Delta = \frac{\lambda_\rho m_\rho t(a \wedge t)}{1024(v_0 + \gamma^2)} \ge \frac{\lambda_\rho m_\rho}{1024(v_0 + m_\rho^2/16)} t(a \wedge t). \tag{D.14}$$

Applying (C.1) inside this one fixed public pair, then the learner infimum and outer pair supremum, gives the nonparametric O lower with one eighth of (D.14).

If $t^2 < q$, then $t(a \wedge t) < q$, and (C.9) supplies the claim. If $n < 16$, then $t(a \wedge t) \le 1 \le 4q$. If $a = 0$ or $t = 0$, the right-hand side is zero. The cases cover every $n \ge 1$, including $n = 16, t^2 = q$, so

$$R_{n,\rho}^{\mathrm{pair}}(a, s, b, t) \ge c_{O,\rho} t(a \wedge t), \quad c_{O,\rho} = \min\left\{ \frac{\lambda_\rho m_\rho}{8192(v_0 + m_\rho^2/16)}, \frac{3d_\rho^\star}{32} \right\} > 0. \tag{D.15}$$

The exact variance is v_0 under all components, so singleton intervals are covered. Unlike the P family, this construction uses the same propensity under both labels.

E The approximation-product lower bound

Assume first $n \ge 16$, $S = a^2 + b^2 \ge q$, and $ab > 0$. Let

$$N = 2n, \quad p = \left\lceil e^{64NS} \right\rceil, \quad A = B_0/2, \quad \eta = \frac{m_\rho b}{4}, \quad \gamma = \frac{a}{512},$$
$$\Delta = \frac{2\eta\gamma}{v_0} = \frac{m_\rho ab}{1024 v_0}, \qquad \kappa = \gamma + \Delta\eta. \tag{E.1}$$

Use the fixed-space, finite-support Walsh convention in Appendix C with this value of p, and write $z = \sigma\phi_j$. Before any latent choice, publish the singleton pair

$$\mathcal{G}_\mu = \mathcal{G}_\pi = \{0\}. \tag{E.2}$$

The labels are

$$(\beta_0, \pi_0, \mu_0) = (0, z\eta, z\gamma), \qquad (\beta_1, \pi_1, \mu_1) = (\Delta, -z\eta, z\kappa). \tag{E.3}$$

Treatment uses (C.6) at means $\pm z\eta$, and the binary response channel is (C.12).

E.1 Observation laws and model conditions

For $t \in \{-L, 0, L\}$, set

$$\alpha_{\pm L} = \frac{v_0 + \eta^2}{2L^2}, \quad \alpha_0 = 1 - \frac{v_0 + \eta^2}{L^2}, \quad \xi_t = \frac{\eta t}{2L^2}.$$

With $c = \gamma/A$, $k = \kappa/A$, and $d_t = \Delta t/A$, the six raw masses are

$$P_{0,z}(t, r) = \tfrac{1}{2}(\alpha_t + z\xi_t)(1 + rzc),$$
$$P_{1,z}(t, r) = \tfrac{1}{2}(\alpha_t - z\xi_t)(1 + r(d_t + zk)). \tag{E.4}$$

The three-point and outcome factors are strictly positive; summing the conditional factors proves normalization. The treatment residual has exact variance v_0 under both labels, and $\mathbb{E}(Y \mid X, T) = \beta_\ell T + \mu_\ell(X)$, so the PLM equations hold. The inequalities

$$\eta \le b, \quad \gamma \le a, \quad \kappa = \gamma(1 + 2\eta^2/v_0) \le 129a/65536 \le a \tag{E.5}$$

certify both approximation budgets relative to (E.2). Both difference classes are $\{0\}$, so every localized stochastic complexity is zero for all s, t, including zero radii. Repeating zero gives the countable M1 representation. The Walsh functions parameterize the off-class truths; the public information remains the singleton pair (E.2).

The scales also give $\Delta \le 1/(4096B_0)$, $\kappa < 1$, and $\Delta L + \kappa < 1 < A$. Hence all outcome probabilities and every authoritative envelope are legal.

E.2 Common reference and G1/G2 calculations

Put $D_{t,r} = \alpha_t + r\xi_t c$. The sign average of label zero is $D_{t,r}/2$. The label-one average is the same precisely when

$$\Delta\alpha_t t = \xi_t(\kappa + \gamma). \tag{E.6}$$

At $t = \pm L$, this is $\Delta(v_0 + \eta^2) = \eta(\kappa + \gamma)$, which follows from (E.1); at $t = 0$, both sides vanish. Thus

$$Q(t, r) = D_{t,r}/2 \tag{E.7}$$

is common and positive. Indeed, $(|\xi_t|/\alpha_t)c < 1/12288$ on side atoms and is zero in the middle.

Direct division of (E.4) gives centered coefficients

$$h_0 = \frac{\xi_t + r\alpha_t c}{\alpha_t + r\xi_t c}, \qquad h_1 = \frac{r\alpha_t k - \xi_t(1 + rd_t)}{\alpha_t + r\xi_t c}. \tag{E.8}$$

Summing Qh_0 over r leaves ξ_t, while summing Qh_1 leaves $-\xi_t$; both sum to zero over t. Since $|\xi_t|/\alpha_t < b/16$, $c \le a/768$, $k < 2c$, and $|d_t| \le c/8$,

$$K_0 = \mathbb{E}_Q h_0^2 \le S/32, \qquad K_1 = \mathbb{E}_Q h_1^2 \le S/8. \tag{E.9}$$

For an independent raw derivation, let

$$E_0(t, r) = \xi_t + r\alpha_t c, \qquad E_1(t, r) = r\alpha_t k - \xi_t(1 + rd_t).$$

Since $P_{\ell,+} - P_{\ell,-} = E_\ell$ and $(P_{\ell,+} + P_{\ell,-})/2 = Q$, raw multiplication gives

$$\sum_{t,r} \frac{P_{\ell,z} P_{\ell,z'}}{Q} = 1 + zz' \sum_{t,r} \frac{E_\ell(t, r)^2}{4Q(t, r)}. \tag{E.10}$$

The energy in (E.10) equals K_ℓ in (E.9) atom by atom, without assuming the likelihood calculation. Walsh orthogonality again gives entries $1 + K_\ell$, $1 - K_\ell$, and one.

E.3 Mixture distance, testing, and all-n closure

Mix uniformly over $2p$ latent coordinate/sign pairs within each label. Equations (C.3) and (E.9) give

$$\chi^2(\overline{P}_\ell^N, Q^N) \le \frac{e^{NK_\ell}}{p} \le e^{-(64-1/8)NS} \le e^{-511}, \tag{E.11}$$

because $NS \ge 2\sqrt{n} \ge 8$. Hence cross-label TV is below $1/4$. All label-zero components have target zero; all label-one components have the gap in (E.1). Applying (C.1) inside the one public singleton pair gives

$$R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \ge \frac{3m_\rho}{16384 v_0} ab \tag{E.12}$$

on this branch.

If $ab = 0$, the desired lower is zero. If $n < 16$, then $ab \le 1 \le 4q$; if $n \ge 16$ and $a^2 + b^2 < q$, then $ab < q/2$. In both cases the scalar bound (C.9) gives the required lower bound. Thus, for every $n \ge 1$,

$$R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \ge c_{ab,\rho} ab, \qquad c_{ab,\rho} = \min\left\{\frac{3m_\rho}{16384 v_0}, \frac{3d_\rho^\star}{32}\right\} > 0. \tag{E.13}$$

The construction covers $a = t$, $b = s$, all zero stochastic budgets, and singleton variance intervals. Its public pair is literally (E.2), so no latent label or dictionary is exposed.

F The common-stochastic M lower bound

Assume first $n \ge 16$, $w = s \wedge t > 0$, and $w^2 \ge q$. Put

$$\begin{aligned} &N = 2n, && p = \left\lceil e^{64Nw^2} \right\rceil, \quad A = B_0/2, \\ &\lambda_\rho = \min\left\{1, \frac{192 c_{\mathrm{rad}}}{23}, 128 B_{\mathcal{G}}\right\}, && \gamma = \frac{\lambda_\rho m_\rho w}{4}, \qquad \eta = \frac{\lambda_\rho w}{256}, \\ &\Delta = \frac{\gamma\eta}{v_0 + \gamma^2}. \end{aligned} \tag{F.1}$$

Use the fixed-space, finite-support Walsh convention in Appendix C with this value of p, and write $z = \sigma\phi_j$. The two labels share $\pi = z\gamma$ and the treatment channel (C.6). They have

$$(\beta_0, \mu_0) = (0, z\eta), \qquad (\beta_1, \mu_1) = (\Delta, 0). \tag{F.2}$$

The conditional binary response channel is (D.3).

F.1 Raw laws, PLM, and one public M1 pair

For $u \in \{-L, 0, L\}$, let

$$\alpha_{\pm L} = \frac{v_0 + \gamma^2}{2L^2}, \quad \alpha_0 = 1 - \frac{v_0 + \gamma^2}{L^2}, \quad \xi_u = \frac{\gamma u}{2L^2}.$$

The complete raw masses are

$$\begin{aligned} P_{0,z}(u, r) &= \tfrac{1}{2}(\alpha_u + z\xi_u)(1 + rz\eta/A), \\ P_{1,z}(u, r) &= \tfrac{1}{2}(\alpha_u + z\xi_u)(1 + r\Delta u/A). \end{aligned} \tag{F.3}$$

Every factor is positive by (C.5)–(C.7), $\eta < A$, and $\Delta L < A$; summing proves normalization. Both labels have exact treatment residual variance v_0, and the response definition gives the PLM conditional mean. The inequalities $|T|, |u| < B_0$, $|Y| = B_0/2$, $\eta \le 1/256$, and $\Delta L < 1/4096$ certify every envelope. In addition, $\gamma, \eta \le \lambda_\rho/256 \le B_{\mathcal{G}}$, so both public dictionaries satisfy the fixed candidate envelope.

Before the latent variables, publish

$$\mathcal{G}_\pi = \{\sigma\gamma\phi_j : j \in [p],\ \sigma \in \{-1, 1\}\}, \qquad \mathcal{G}_\mu = \{0\} \cup \{\sigma\eta\phi_j : j \in [p],\ \sigma \in \{-1, 1\}\}. \tag{F.4}$$

Both labels have zero approximation error. The pair is finite M1 and is identical under both labels.

F.2 Separate stochastic-budget certificates

Since $p \le 2e^{64Nw^2}$ and $nw^2 \ge 4$,

$$\log(32p^2) < 258nw^2. \tag{F.5}$$

The outcome difference class has fewer than $8p^2$ elements and envelope 2η. Lemma 17 gives, for every $r' \ge s > 0$,

$$\frac{\mathfrak{R}_n(r'; \Delta\mathcal{G}_\mu)}{3r'} < \frac{46\eta w}{3s} = \frac{23\lambda_\rho w^2}{384s} \le \frac{c_{\mathrm{rad}}}{2}s \le c_{\mathrm{rad}}s. \tag{F.6}$$

The propensity difference class has at most $4p^2$ elements and envelope 2γ, so, for every $r' \ge t > 0$,

$$\frac{\mathfrak{R}_n(r'; \Delta\mathcal{G}_\pi)}{3r'} < \frac{46\gamma w}{3t} = \frac{23\lambda_\rho m_\rho w^2}{6t} \le \frac{23\lambda_\rho}{384}t \le \frac{c_{\mathrm{rad}}}{2}t \le c_{\mathrm{rad}}t. \tag{F.7}$$

The inequalities $w \le s$ and $w \le t$ verify the two complexity budgets separately. All displayed quotients use positive radii. If $w = 0$, both classes collapse to zero and both complexities vanish at every positive radius.

F.3 Common reference and two Gram calculations

Put $e = \eta/A$, $d_u = \Delta u/A$. The sign averages of (F.3) are

$$Q_0(u, r) = \tfrac{1}{2}(\alpha_u + r\xi_u e), \qquad Q_1(u, r) = \tfrac{1}{2}\alpha_u(1 + rd_u).$$

They agree atomwise because

$$\Delta\alpha_u u = \xi_u\eta : \tag{F.8}$$

at $u = \pm L$, this is $\Delta(v_0 + \gamma^2) = \gamma\eta$; at zero both sides vanish. Hence

$$Q(u, r) = \tfrac{1}{2}(\alpha_u + r\xi_u e) \tag{F.9}$$

is common. It is positive since $|\xi_u|/\alpha_u < w/16$ and $e \le w/384$.

The centered likelihood coefficients are

$$h_0 = \frac{\xi_u + r\alpha_u e}{\alpha_u + r\xi_u e}, \qquad h_1 = \frac{\xi_u}{\alpha_u}. \tag{F.10}$$

Summing Qh_ℓ over r leaves ξ_u, whose sum is zero. The denominator margin gives

$$K_0 = \mathbb{E}_Q h_0^2 < w^2/9, \qquad K_1 = \sum_u \frac{\xi_u^2}{\alpha_u} = \frac{\gamma^2}{v_0 + \gamma^2} < w^2/256. \tag{F.11}$$

Independently, write $P_{\ell,z} = Q + zC_\ell$, with

$$C_0 = \tfrac{1}{2}(\xi_u + r\alpha_u e), \qquad C_1 = \tfrac{1}{2}\xi_u(1 + rd_u).$$

Raw multiplication gives

$$\sum_{u,r} \frac{P_{\ell,z}P_{\ell,z'}}{Q} = 1 + zz' \sum_{u,r} \frac{C_\ell^2}{Q}. \tag{F.12}$$

The two raw sums equal (F.11) exactly. Restoring $z = \sigma\phi_j(X)$ produces Gram entries $1 + K_\ell$, $1 - K_\ell$, and one by Walsh orthogonality.

F.4 Mixture distance, testing, and all-n closure

For uniform coordinate/sign mixtures, (C.3) and (F.11) yield

$$\chi^2(\overline{P}_\ell^N, Q^N) \le e^{-(64-1/9)Nw^2} < e^{-511}. \tag{F.13}$$

Thus cross-label TV is below $1/4$. The target gap satisfies

$$\Delta = \frac{\lambda_\rho^2 m_\rho w^2}{1024(v_0 + \gamma^2)} \ge \frac{\lambda_\rho^2 m_\rho w^2}{1024(v_0 + m_\rho^2/16)}. \tag{F.14}$$

The expected-loss test within the common public pair gives one eighth of (F.14).

If $w = 0$, the claim is trivial. If $n < 16$, then $w^2 \le 1 \le 4q$; if $n \ge 16, w^2 < q$, use (C.9). These branches are exhaustive and their lower bounds are compared by a maximum. Therefore

$$R_{n,\rho}^{\mathrm{pair}}(a,s,b,t) \ge c_{M,\rho}(s \wedge t)^2, \quad c_{M,\rho} = \min\left\{ \frac{\lambda_\rho^2 m_\rho}{8192(v_0 + m_\rho^2/16)}, \frac{c_{q,\rho}}{4} \right\}. \tag{F.15}$$

This covers $s = t$, either imbalanced ordering, both zero faces, every n, and singleton variance intervals. A learner may scan the entire public dictionary: the TV bound already quantifies over every Borel rule.

G Lower assembly, deterministic equivalence, and constants

Appendix L constructs, for every fixed strict-interior ρ, n, a, s, b, t, one finite public pair $\Gamma^\star$ and one $P_X^\star$ supporting the q family and every active native P/O/ab/M subfamily. An inactive component inequality reuses the common q splice, so all five component lower inequalities hold in this pair. The pair may depend on the public cell, but it is fixed before data and before the component, coordinate, sign, or label. Define

$$R^\star = \inf_{\delta_{\Gamma^\star}\ \mathrm{Borel}} \sup_{P \in \mathcal{P}(\Gamma^\star)} \mathbb{E}_P |\delta_{\Gamma^\star} - \beta(P)|.$$

Every active component mixture, together with the common q-splice mixture used by inactive components, is supported on $\mathcal{P}(\Gamma^\star)$, and every testing inequality quantifies over the same arbitrary rule after the full pair has been disclosed. Taking the infimum only after the five inequalities gives

$$R_{n,\rho}^{\mathrm{pair}} \ge R^\star \ge \max\{c_{q,\rho}q, c_{ab,\rho}ab, c_{M,\rho}M, c_{O,\rho}O, c_{P,\rho}P\}. \tag{G.1}$$

Equation (G.1) takes the maximum of five lower bounds for the same rule on the same public pair.

Let $S_0 = ab + M + O + P$. Since $AB \ge ab$, one has $S_0 \le X$. Conversely,

$$AB = ab + at^2 + bs^2 + s^2t^2.$$

For all budgets in $[0, 1]$,

$$at^2 \le t(a \wedge t) = O, \qquad bs^2 \le s(b \wedge s) = P, \qquad s^2t^2 \le (s \wedge t)^2 = M.$$

Indeed, in the first inequality, $a \le t$ gives $at^2 \le at$, while $a > t$ gives $at^2 \le t^2$; the second is its mirror. The third follows because the larger of s, t is at most one. Therefore

$$S_0 \le X \le 2S_0. \tag{G.2}$$

This proof uses no division and covers all zeros and ties, including $a = t$ and $b = s$.

Set

$$c_{\min,\rho} = \min\{c_{q,\rho}, c_{ab,\rho}, c_{M,\rho}, c_{O,\rho}, c_{P,\rho}\} > 0.$$

Since the maximum of five nonnegative numbers is at least their sum divided by five, (G.1)–(G.2) imply

$$R^{\mathrm{pair}}_{n,\rho} \ge \frac{c_{\min,\rho}}{5}(q + S_0) \ge \frac{c_{\min,\rho}}{10}(q + X). \tag{G.3}$$

One explicit compatible set of constants is

$$\begin{aligned}
d^\star_\rho &= \min\{3, B_0/16, B_0/(16\sqrt{v_-})\},\\
m_\rho &= \min\{1/64, v_-/(4B_0), (B_0 - \sqrt{v_-})/4\},\\
\lambda_\rho &= \min\{1, 192c_{\mathrm{rad}}/23, 128B_{\mathcal{G}}\},\\
c_{q,\rho} &= 3d^\star_\rho/8,\\
c_{P,\rho} &= \min\{\lambda_\rho m_\rho/(4096v_-), 3d^\star_\rho/32\},\\
c_{O,\rho} &= \min\{\lambda_\rho m_\rho/[8192(v_- + m^2_\rho/16)], 3d^\star_\rho/32\},\\
c_{ab,\rho} &= \min\{3m_\rho/(16384v_-), 3d^\star_\rho/32\},\\
c_{M,\rho} &= \min\{\lambda^2_\rho m_\rho/[8192(v_- + m^2_\rho/16)], c_{q,\rho}/4\}.
\end{aligned}$$

They are positive for every fixed $0 < v_- < B_0^2$, $B_{\mathcal{G}} > 0$, and $c_{\mathrm{rad}} > 0$, but no uniform lower bound is claimed as the tuple approaches the variance boundary, $B_{\mathcal{G}} \downarrow 0$, or $c_{\mathrm{rad}} \downarrow 0$. Equation (G.3) proves the interior lower half of Theorem 1.

H Complete proof of the empty and saturation strata

Because $|u| \le B_0$, every legal law satisfies $\mathbb{E}u^2 \le B_0^2$. Thus no law exists when $v_- > B_0^2$. The theorem records this regime as an empty experiment.

Now let $v_- = B_0^2$. Since $\mathbb{E}u^2 \ge B_0^2$ and $0 \le B_0^2 - u^2$, we have $\mathbb{E}(B_0^2 - u^2) = 0$. Thus $u^2 = B_0^2$ almost surely. Since $\mathbb{E}(u \mid X) = 0$,

$$\mathbb{P}(u = B_0 \mid X) = \mathbb{P}(u = -B_0 \mid X) = 1/2. \tag{H.1}$$

Both values $T = \pi_0(X) \pm B_0$ must lie in $[-B_0, B_0]$. The intersection of these two constraints is $\pi_0(X) = 0$. Hence $T = u$, $\mathbb{E}T^2 = B_0^2$, and conditional centering gives

$$\mathbb{E}(T\mu_0(X)) = 0, \qquad \mathbb{E}(T\varepsilon) = 0, \qquad \beta_0 = \frac{\mathbb{E}(TY)}{B_0^2}. \tag{H.2}$$

For every nonempty public pair, the public Borel estimator

$$\widehat{\beta} = \Pi_{[-3,3]}\left\{\frac{1}{2nB_0^2}\sum_{i=1}^{2n} T_iY_i\right\} \tag{H.3}$$

therefore satisfies, by Cauchy–Schwarz and $|TY| \le B_0^2$,

$$\mathbb{E}|\widehat{\beta} - \beta_0| \le \sqrt{\mathrm{Var}\left(\frac{1}{2nB_0^2}\sum_{i=1}^{2n} T_iY_i\right)} \le \frac{1}{\sqrt{2n}}. \tag{H.4}$$

Projection cannot increase loss because $\beta_0 \in [-3, 3]$.

For the outer lower bound, place a point-mass covariate law on the same fixed ambient space (so X is constant), take $\mathcal{G}_\mu = \mathcal{G}_\pi = \{0\}$, and $T = \pm B_0$ equiprobably. Put

$$\delta_n = \frac{3}{32\sqrt{n}}, \qquad \theta \in \{-\delta_n, \delta_n\},$$

and let $Y = (B_0/2)R$, where $R \in \{-1, 1\}$ and

$$\mathbb{P}_\theta(R = 1 \mid T = t) = \frac{1}{2} + \frac{\theta t}{B_0}. \tag{H.5}$$

Then $\mathbb{E}(Y \mid T) = \theta T$, so $\beta = \theta$, both nuisances vanish, and the outcome residual is conditionally centered and bounded. The public pair is finite M1 and legal for every declared budget.

For one observation the Bernoulli shift magnitude is δ_n, so the cross-label Bhattacharyya coefficient (Hellinger affinity) is $\sqrt{1 - 4\delta_n^2}$, while its square is $1 - 4\delta_n^2$. Affinities tensorize, and Cauchy–Schwarz gives $\mathrm{TV}^2(P, Q) \le 1 - \{\int \sqrt{dP\, dQ}\}^2$. Consequently,

$$\mathrm{TV}^2 \le 4N\delta_n^2 = 72/1024 < 1/4.$$

Thus TV is below $1/2$. The target separation is $2\delta_n = 3/(16\sqrt{n})$. Lemma 16 therefore yields

$$R^{\mathrm{pair}}_{n,\rho}(a, s, b, t) \ge \frac{3}{128\sqrt{n}}. \tag{H.6}$$

H.1 Uniform one-sided envelope bound

We prove Proposition 2. Put $h = B_0^2 - v_-$. On $\pi_0 \ge 0$, the bounds on u, T give $-B_0 \le u \le B_0 - \pi_0$. Taking conditional expectations in $(u + B_0)(B_0 - \pi_0 - u) \ge 0$, and treating $\pi_0 < 0$ symmetrically, gives $\mathbb{E}(u^2 \mid X) \le B_0(B_0 - |\pi_0|)$. Thus $B_0\mathbb{E}|\pi_0| \le h$. Also $v_- \le \mathbb{E}T^2 = \mathbb{E}u^2 + \mathbb{E}\pi_0^2 \le B_0^2$ and $\mathbb{E}(TY) = \beta_0\mathbb{E}T^2 + \mathbb{E}(\pi_0\mu_0)$, so $|\mathbb{E}(TY) - \beta_0 B_0^2| \le 3h + h$. Since $|TY/B_0^2| \le 1$, the expected absolute centered empirical error is at most $N^{-1/2}$. Clipping cannot increase loss. At saturation $h = 0$, this is exactly (H.4).

The upper bound (H.4) is pairwise for every legal nonempty pair; (H.6) needs only this one pair because the pair supremum is outermost. The saturation lower bound uses the separate construction above. Proposition 2 gives the uniform one-sided bound for this same fixed-denominator estimator before saturation; it uses only the integrated bound $B_0\mathbb{E}|\pi_0| \le B_0^2 - v_-$, not a pointwise bound on π_0.

I Assumptions, measurability, and scope

The measurable representation specifies access to the public classes without imposing smoothness or convexity. The underlying covariate space is the one fixed nonempty infinite standard Borel space $(\mathcal{X}, \mathcal{A})$ specified in Appendix A; the observation space carries the corresponding product sigma-field, and the coordinates X, T, Y and all represented evaluation maps are measurable. Countably infinite and uncountable standard Borel hosts contain the measurable finite blocks needed by every lower construction. Finite ambient spaces can have smaller minimax risk, as the singleton example below shows. The E and S arguments themselves remain valid on every nonempty standard Borel space, but the matching interior theorem requires an infinite ambient space. Conditional expectations are versions on this declared probability space.

The residual restriction is marginal: $\mathbb{E}u^2 \in [v_-, v_+]$. Conditional treatment-residual variance may vary and vanish pointwise. The upper curvature uses $\mathbb{E}u^2$, while the lower witnesses in Appendices C–G retain constant conditional variance. All declared pointwise envelopes remain in force.

For any preassigned atomless standard Borel marginal ν, Appendix L.6 and Proposition 19 give the same-pair lower and matching minimax bounds with constants independent of ν. The public pair may depend on ν, n and the budgets; the rule may know ν. The transported saturation witness retains the original constants.

Proposition 2 gives the public rule $\Pi_{[-3,3]}[P_N(TY)/B_0^2]$, whose expected error is at most $1/\sqrt{2n} + 4\omega$, where $\omega = 1 - v_-/B_0^2$. Within the bounded model, this uniform bound holds without the budgets, M1, or v_+.

For every full-class function, sequential M1 supplies a represented sequence converging pointwise everywhere inside the fixed $B_{\mathcal{G}}$ envelope. It preserves bounded empirical objectives exactly in the limit and preserves population L_2 infima by dominated convergence against the B_0-bounded truths. This sequential approximation does not make an arbitrary full-class exact localized supremum measurable. Accordingly, the model-field certificate uses outer expectation on each full difference class, while the upper proof uses only the ordinary expectation of measurable countable represented difference cores, dominated pointwise by the full-class supremum. For the finite lower classes the supremum is measurable, so outer expectation and ordinary expectation are equal. The proof uses domination of the core complexity by the full-class complexity. Countable infima, first-within-tolerance indices, finite arithmetic, and clipping are Borel. The branch objectives over rational finite-dimensional parameters are objective-dense because the relevant bounded convex functions are continuous and coercive on the reduced domain. M1 means that these countable representations and evaluation access are public. Q1 means that for each n and declared budget cell the pair is fixed before data and any latent hard label, after which a pair-specific Borel rule is chosen; neither realized-data classes nor one pair fixed across all n are required.

I.1 The role of the ambient space and sequential approximation

The finite-space exclusion is substantive. On a singleton covariate space, pair the observations across the two independent halves and put $D_i = T_{1i} - T_{2i}$, $E_i = Y_{1i} - Y_{2i}$, and $S_n = \sum_i D_i^2$. The clipped ratio

$$\widehat{\beta} = \Pi_{[-3,3]}\left[\frac{\sum_i D_i E_i}{S_n}\mathbf{1}\{S_n > 0\}\right]$$

is Borel and satisfies, uniformly over every compatible law and public pair,

$$\mathbb{E}|\widehat{\beta} - \beta_0| \le B_0\sqrt{\frac{2}{nv_-}} + 6\exp\left\{-\frac{nv_-^2}{8B_0^4}\right\}. \tag{I.1}$$

Indeed, differencing removes both constant nuisances, the residual numerator is conditionally centered with conditional second moment at most $2B_0^2 S_n$, and Hoeffding gives $\Pr(S_n < nv_-) \le \exp\{-nv_-^2/(8B_0^4)\}$. Yet the legal zero singleton pair in the cell $(a, s, b, t) = (0, 1/2, 0, 1/2)$ has $X = 1/16 + 1/4 = 5/16$. Thus a positive fixed multiple of $q + X$ cannot be a uniform lower bound on that finite ambient space.

The sequential premise is also substantive. On $[0, 1]$ with Lebesgue law, let R consist of $g_B = \mathbf{1}_{[0,1]\setminus B}$, where B ranges over finite unions of rational-endpoint intervals of measure at most $1/2$, and let $\mathcal{G} = R \cup \{0\}$. Every finite-coordinate neighborhood of zero meets R, but

$$\inf_{r\in R}\|r\|_2^2 \ge \frac{1}{2} \quad \text{whereas} \quad \inf_{g\in\mathcal{G}}\|g\|_2^2 = 0. \tag{I.2}$$

No sequence in R converges pointwise to zero: dominated convergence would force $P(B_k) \to 1$, contradicting $P(B_k) \le 1/2$. Thus sequential approximation is essential for the comparators in Appendix B.2 and for transferring calibration suprema.

The baseline minimax risk permits known pre-data upper certificates for the four budgets, an arbitrary Borel decision rule, expected absolute loss, $2n$ iid observations, and constants depending on fixed ρ, including fixed $B_{\mathcal{G}} > 0$ and $c_{\mathrm{rad}} > 0$. The upper bounds are pair-specific, and the lower bounds concern the worst admissible pair. Finite-query optimization, efficient algorithms, confidence intervals, and empirical performance require further analysis. In particular, constants may degenerate as $B_{\mathcal{G}} \downarrow 0$ or $c_{\mathrm{rad}} \downarrow 0$, where the class constraints degenerate and the interior nuisance lower bound no longer applies. The finite lower instances satisfy the same M1 interface and do not use inaccessible latent labels as learner inputs. The target is a statistical PLM coefficient; a causal or potential-outcomes interpretation requires additional identification assumptions. Theorem 4 and Appendix Q strengthen the upper quantifiers: for each fixed (Γ, ρ, n) one rule attains $q + W$ simultaneously over all four unknown budgets. Empirical local Rademacher fixed points on an independent covariate block supply the stochastic inputs to the double-calibrated rule of Appendix P. A positive-radius margin transfers the core's linear certificate to the original full outer complexity; the public classes are unchanged. Separately, Proposition 5 and Appendix M give an all-budget-independent rule and the matching restricted outer-pair rate $q + (a + s^2)(b + t^2)$ for full interval boxes on possibly different

partitions around arbitrary public center functions. Theorem 23 uses one pair depending on ρ, n, s, t, uniform over a, b. Proposition 6 and Appendix N give the smaller matching outer-pair rate $q + ab$ for full affine Hilbert balls, again without budget inputs; zero-radius balls witness its lower. Both statements use geometry beyond M1 and characterize the worst risk over their respective geometric subclasses. Finite implementations require explicitly evaluable centers and the stated partition/width or kernel interface.

Antecedents include partially linear and orthogonal estimation (Robinson, 1988; Chernozhukov et al., 2018; Foster & Syrgkanis, 2023), structure-agnostic product bounds (Balakrishnan et al., 2026), two-mixture testing (Robins et al., 2009), and Gu's four-budget question (Gu, 2025).

J PROOF ROADMAP

Appendix A defines the model and minimax order; Appendix I explains the measurable representation. The main proof dependencies are as follows.

Upper arguments. Lemma 13 transfers full-class comparators to represented cores. Lemma 12 and Appendix B.2 provide localization and initial fits. The O and R arguments are in Appendices B.3–B.4 and B.5–B.6; Appendix B.8 gives the selector and Q1 lift. The curvature is the marginal moment $\mathbb{E}u^2$, and conditional variance need not be constant or pointwise positive. Proposition 2 separately proves the uniform $1/\sqrt{2n} + 4(1 - v_-/B_0^2)$ bound for the clipped empirical TY/B_0^2 rule, without budgets, M1, or v_+.

Lower arguments and marginal transport. Appendices C–G give the constant-conditional-variance witnesses and their testing bounds. Proposition 18 and Appendix L place the active native components and common scalar splice inside one cell-dependent public pair before taking a maximum of lower bounds. Appendix L.6 and Proposition 19 transport the same-pair lower to each preassigned atomless standard Borel marginal ν, with constants independent of ν, even when the rule knows ν. The pair may depend on ν, n and the budgets. Appendix H proves the empty and saturation strata; the fixed-marginal transport preserves the saturation constants.

Adaptation. The matching results retain expected absolute loss, $N = 2n$, pre-data upper certificates, sequential M1, and fixed-ρ constants. The baseline risk permits knowledge of the certificates; Appendices P–Q construct one rule that uses none of them. The model allows heteroskedasticity and pointwise zero conditional variance. Appendix I discusses the model boundaries and implementation requirements.

K PROOF INDEX AND DEPENDENCIES

Result	Proof location
Model, M1, and minimax order	Appendix A and Appendix I.
Sequential full-to-represented transfer	Lemma 13; applications in Appendices B.2, B.3, and B.5.
Localized processes, joint class, and initial fits	Appendix A's outer-expectation full-class certificate and countable-core localization; Lemmas 10 and 12, including the maximum-shell/confidence/range argument; joint localization and initial fits in Appendix B.2.
O definition/calibration/oracle	Appendices B.3–B.4.
R definition/calibration/oracle	Appendices B.5–B.6.
Selector, fallback, and Q1 lift	Appendix B.8.
Testing, finite Gram, Massart	Lemmas 16 and 17; all lower amplitudes include the fixed positive λ_ρ where required.
Fixed-space Walsh embedding	Appendix C.4, opening convention; Appendices C–F use its measurable finite blocks and zero extensions.

Result	Proof location (continued)
Equal-variance channel and target-capped q	Appendices C.2 and C.3.
P certificate	Appendices C.4–C.7.
O certificate	Appendix D, equations (D.1)–(D.15).
ab certificate	Appendix E, equations (E.1)–(E.13).
M certificate	Appendix F, equations (F.1)–(F.15).
Same-public-pair lift	Proposition 18 and Appendix L: common host, full classes, cross-band complexity, membership, arbitrary-rule reduction, and all-n splice.
Lower assembly	Appendices G and L.
Uniform envelope bridge	Proposition 2 and its proof in Appendix H.1: clipped empirical TY/B_0^2, risk at most $1/\sqrt{2n} + 4(1 - v_-/B_0^2)$, without budgets, M1, or v_+.
Fixed atomless marginal	Appendix L.6 and Proposition 19: same-pair $q + X$ lower and matching minimax law, constants independent of the preassigned ν, known ν permitted; saturation constants preserved.
E/S boundary	Appendix H, equations (H.1)–(H.6).

Sharp box adaptation: Proposition 5 and Appendix M. Complexity controls the first two singleton-mass moments; trimming reduces each public residual to $O_\rho(r^2+q)$. A cross moment and a controlled fallback yield $q + AB$ without budgets. Theorem 23 matches this by four rescaled cancellation families and the scalar family on one pair.

Hilbert adaptation: Proposition 6 and Appendix N. Core support functions recover the normalized kernels; center projection and ridge residualization preserve average variance and yield $q + ab$. The zero pair witnesses its restricted outer lower. Both structured results allow distinct nuisance geometries and give faster rates than the general all-budget-adaptive rule.

Convex and union comparison: Proposition 7 and Appendix O. Convexification preserves the global difference support function, while orthogonality preserves approximation distances. The same five lower families yield $q + X$ for full dictionary ℓ_1 balls and finite unions of rank-at-most-one balls. Lemma 26 gives the sharp enclosing-ball envelope cost. Theorem 4 supplies the upper without any budget inputs.

General-class approximation adaptation: Theorem 27 in Appendix P. The fixed-anchor pilot has two second-moment bounds. Conditional transfer gives a known local propensity cap; two real-vector comparisons establish feasibility and the free-residual bounds. Double calibration, the pilot bounds and a proof-only budget partition give both directional rates for the same Borel rule, without a, b. Appendix Q then removes s, t: Lemma 30 controls the star hull, Lemma 31 calibrates both radii, and Lemma 32 transfers their linear certificates to the full classes. A three-block split, capped inputs and joint Borel selection prove Theorem 4.

The main dependencies are

$$\left.\begin{array}{c}\text{outer-expectation full-class certificate} \longrightarrow \text{countable-core localization} \\ \longrightarrow \text{positive-radius localized process} \\ \text{sequential M1 transfer}\end{array}\right\}$$
$$\longrightarrow \text{joint localization} \longrightarrow \text{initial fits}$$
$$\longrightarrow \begin{cases}\text{O calibration} \longrightarrow \text{O tail}, \\ \text{R calibration} \longrightarrow \text{R tail},\end{cases}$$
$$\{\text{O tail}, \text{R tail}, \text{fallback}, \text{selector}\}$$
$$\longrightarrow \text{fixed-pair upper} \longrightarrow \text{Q1 upper},$$
$$\text{fixed-space Walsh embedding} + \text{four disjoint frequency bands}$$
$$+\text{target-capped scalar channel} + \text{three-point channel}$$
$$+\text{testing tools} \longrightarrow \{q, \text{every active native P/O/}ab\text{/M family}\} \subset \mathcal{P}(\Gamma^\star)$$
$$\text{inactive named components} \longrightarrow \text{common q splice}$$
$$\text{five component inequalities} \longrightarrow \text{fixed-pair maximum lower} \longrightarrow \text{interior characterization},$$
$$\text{fixed atomless marginal transport} \longrightarrow \text{fixed-marginal lower},$$
$$\text{envelope slack} \longrightarrow \text{uniform envelope bridge},$$
$$\text{envelope equality} \longrightarrow \text{propensity identification}$$
$$\longrightarrow \text{saturation theorem}.$$

The references above provide the proofs of each step.

L A SINGLE PUBLIC PAIR FOR THE FIVE COMPONENT LOWER INEQUALITIES

Proposition 18 (Common-pair lower bound)**.** *The displayed quantifiers are*

$$\forall \rho \text{ in the strict-interior field } 0 < v_- < B_0^2, \quad \forall n \geq 1, \quad \forall (a,s,b,t) \in [0,1]^4, \quad \exists (\Gamma^\star, P_X^\star),$$

where $\Gamma^\star = \Gamma^\star_{\rho,n,a,s,b,t}$ *is one finite sequential-M1 public pair and* $P_X^\star$ *is one covariate law, such that*

$$\inf_{\delta_{\Gamma^\star} \text{ Borel}} \sup_{P \in \mathcal{P}(\Gamma^\star)} \mathbb{E}_P |\delta_{\Gamma^\star}(Z_{1:2n}) - \beta(P)| \geq c_\rho^\star(q + X)$$

for a constant $c_\rho^\star > 0$ *depending only on fixed* ρ*. Its compatible-law family contains* q *and every active native P/O/ab/M family; inactive component inequalities use the common* q *splice. Thus the five inequalities hold on one covariate support and marginal* $P_X^\star$*, without an explicit component tag. The pair is fixed before Nature chooses component, coordinate, sign, label, or data, and the inequality covers every arbitrary Borel rule that knows the entire pair.*

This appendix proves Proposition 18. We place the observation laws from Appendices C–F and a scalar q family on a common covariate support. We then verify the enlarged pair's M1 representation and four budgets, establish compatibility of the component laws, and apply the testing bound to every rule that knows this pair.

L.1 HOST, ACTIVE BRANCHES, AND PUBLIC PAIR

Fix $0 < v_- < B_0^2$, $n \geq 1$, and $(a,s,b,t) \in [0,1]^4$. Write $N = 2n$, $v_0 = v_-$, and

$$m_\rho = \min\left\{\frac{1}{64}, \frac{v_0}{4B_0}, \frac{B_0 - \sqrt{v_0}}{4}\right\}, \qquad \lambda_\rho = \min\left\{1, \frac{192 c_{\mathrm{rad}}}{23}, 128 B_{\mathcal{G}}\right\}.$$

Both constants are strictly positive and depend only on fixed ρ. Declare a large P, O, ab, or M branch, respectively, when

branch	large-branch predicate	dictionary size
P	$n \geq 16,\ s^2 \geq q,\ s(b \wedge s) > 0$	$p_P = \lceil \exp(64 N s^2) \rceil$
O	$n \geq 16,\ t^2 \geq q,\ t(a \wedge t) > 0$	$p_O = \lceil \exp(64 N t^2) \rceil$
ab	$n \geq 16,\ a^2 + b^2 \geq q,\ ab > 0$	$p_{ab} = \lceil \exp(64 N (a^2 + b^2)) \rceil$
M	$n \geq 16,\ (s \wedge t)^2 \geq q,\ s \wedge t > 0$	$p_M = \lceil \exp(64 N (s \wedge t)^2) \rceil$.

An inactive branch has count zero. Put

$$K = p_P + p_O + p_{ab} + p_M, \qquad m_\star = \max\{1, \lceil \log_2(K+1) \rceil\}.$$

The group $(\mathbb{Z}/2\mathbb{Z})^{m_\star}$ has at least K distinct nonzero Walsh characters. Partition K of them into four disjoint frequency sets $\mathcal{J}_P, \mathcal{J}_O, \mathcal{J}_{ab}, \mathcal{J}_M$ of the displayed sizes. By the fixed-infinite-space convention of Appendix A, preselect a measurable injection of this finite group into $\mathcal{X}$, extend every character by zero outside the image, and let $P_X^\star$ be uniform on the whole image. Thus every allocated character ϕ_j is Rademacher-valued, mean zero, and unit-norm under $P_X^\star$, and distinct frequencies are orthogonal. The support and $P_X^\star$ are identical for every component, coordinate, sign, and label. No component coordinate or disjoint-support event is observed.

For an active P branch, let $w_P = b \wedge s$ and set

$$\eta_P = \frac{m_\rho w_P}{4}, \quad \gamma_P = \frac{\lambda_\rho s}{256}, \quad \Delta_P = \frac{2\eta_P \gamma_P}{v_0}, \quad \kappa_P = \gamma_P + \Delta_P \eta_P.$$

For active O and M branches, put

$$w_O = a \wedge t, \quad \gamma_O = \frac{\lambda_\rho m_\rho t}{4}, \qquad \eta_O = \frac{w_O}{256}, \qquad \Delta_O = \frac{\gamma_O \eta_O}{v_0 + \gamma_O^2},$$
$$w_M = s \wedge t, \quad \gamma_M = \frac{\lambda_\rho m_\rho w_M}{4}, \quad \eta_M = \frac{\lambda_\rho w_M}{256}, \quad \Delta_M = \frac{\gamma_M \eta_M}{v_0 + \gamma_M^2}.$$

For an active ab branch, retain

$$\eta_{ab} = \frac{m_\rho b}{4}, \quad \gamma_{ab} = \frac{a}{512}, \quad \Delta_{ab} = \frac{2\eta_{ab}\gamma_{ab}}{v_0}, \quad \kappa_{ab} = \gamma_{ab} + \Delta_{ab}\eta_{ab}.$$

Terms from inactive branches are omitted from the following definitions.

$$\mathcal{G}_\mu^\star = \{0\} \cup \{\sigma\gamma_P\phi_j, \sigma\kappa_P\phi_j : j \in \mathcal{J}_P,\ \sigma \in \{-1,1\}\} \cup \{\sigma\eta_M\phi_j : j \in \mathcal{J}_M,\ \sigma \in \{-1,1\}\}.$$
$$\mathcal{G}_\pi^\star = \{0\} \cup \{\sigma\gamma_O\phi_j : j \in \mathcal{J}_O,\ \sigma \in \{-1,1\}\} \cup \{\sigma\gamma_M\phi_j : j \in \mathcal{J}_M,\ \sigma \in \{-1,1\}\}.$$

Fix ordered countable representations $\mathbf{g}_\mu^\star = (g_{\mu,j}^\star)_{j\in\mathbb{N}}$ and $\mathbf{g}_\pi^\star = (g_{\pi,j}^\star)_{j\in\mathbb{N}}$ by listing every element of the corresponding finite class and then repeating zero, and set

$$\Gamma^\star = (\mathcal{G}_\mu^\star, \mathcal{G}_\pi^\star, \mathbf{g}_\mu^\star, \mathbf{g}_\pi^\star).$$

Each class is a finite union, not a linear span. Taking spans would permit additive combinations across bands, change the pointwise envelope, and make the following finite-cardinality complexity certificate inapplicable. The ab band is deliberately absent from both public classes because its size is controlled by $a^2 + b^2$, not by s^2 or t^2.

L.2 Sequential-M1, envelopes, and stochastic budgets

These deterministic enumerations are sequential M1 representations; their measurable evaluation maps are well-defined because the embedded finite-block characters are measurable. Pointwise convergence and the common candidate envelope are immediate. The component bounds in Appendices C and F give

$$\kappa_P \le 2\gamma_P \le \frac{\lambda_\rho s}{128} \le B_{\mathcal{G}}, \qquad \eta_M, \gamma_O, \gamma_M \le \frac{\lambda_\rho}{256} \le B_{\mathcal{G}}.$$

Because the classes are unions rather than sums, their envelopes are the largest displayed amplitudes.

Suppose the outcome class is nontrivial and put $p_s = \lceil \exp(64Ns^2) \rceil$. Active M implies $p_M \le p_s$, while active P gives $p_P = p_s$. Counting the full difference class, including cross-band differences, yields

$$|\mathcal{G}_\mu^\star| \le 1 + 4p_P + 2p_M \le 7p_s, \qquad |\Delta\mathcal{G}_\mu^\star| \le 49p_s^2 \le 196\exp(128Ns^2).$$

Nontriviality implies $Ns^2 \ge 8$; hence $\log|\Delta\mathcal{G}_\mu^\star| < 129Ns^2 = 258ns^2$. The difference-class envelope is at most twice $H_\mu = \max\{\kappa_P, \eta_M\} \le \lambda_\rho s/128$. The finite-class Massart inequality from Lemma 17, applied globally and hence to every localized subset, gives

$$\mathfrak{R}_n(r; \Delta\mathcal{G}_\mu^\star) < 46H_\mu s \le \frac{23\lambda_\rho}{64}s^2, \qquad \frac{\mathfrak{R}_n(r; \Delta\mathcal{G}_\mu^\star)}{3r} \le \frac{23\lambda_\rho s}{192} \le c_{\mathrm{rad}} s$$

for every $r \geq s > 0$.

Similarly, with $p_t = \lceil \exp(64Nt^2) \rceil$,

$$|\mathcal{G}_\pi^\star| \leq 1 + 2p_O + 2p_M \leq 5p_t, \qquad |\Delta\mathcal{G}_\pi^\star| \leq 25p_t^2 \leq 100\exp(128Nt^2).$$

Nontriviality implies $\log|\Delta\mathcal{G}_\pi^\star| < 258nt^2$, and the difference envelope is twice $H_\pi \leq \lambda_\rho t/256$. Therefore

$$\mathfrak{R}_n(r;\Delta\mathcal{G}_\pi^\star) < 46H_\pi t \leq \frac{23\lambda_\rho}{128}t^2, \qquad \frac{\mathfrak{R}_n(r;\Delta\mathcal{G}_\pi^\star)}{3r} \leq \frac{23\lambda_\rho t}{384} \leq c_{\mathrm{rad}}t$$

for every $r \geq t > 0$. If $s = 0$, neither P nor M is active, so $\mathcal{G}_\mu^\star = \{0\}$ when $s = 0$; if $t = 0$, neither O nor M is active, so $\mathcal{G}_\pi^\star = \{0\}$ when $t = 0$. The corresponding difference class is then exactly zero at every positive radius, and no radius-zero quotient is used.

L.3 Approximation and raw-law membership

For an off-class truth $c\phi_j$ and a nonzero public candidate $d\phi_k$ from another band,

$$\|c\phi_j - d\phi_k\|_2^2 = c^2 + d^2 > c^2 = \|c\phi_j\|_2^2.$$

Consequently the added bands preserve the P/O/ab approximation distances exactly. In detail, q has two zero truths; P has its outcome truth in $\mathcal{G}_\mu^\star$ and propensity distance $\eta_P \leq b$; O has its propensity truth in $\mathcal{G}_\pi^\star$ and outcome distance $\eta_O \leq a$; ab uses $\mathcal{J}_{ab}$ off both public classes and has the original distances bounded by a, b; M has both truths in the corresponding public classes.

We next verify the observation laws for the q/P/O/ab/M families. For q, use $X \sim P_X^\star$ independently of

$$T \in \{-L, 0, L\}, \qquad L = \sqrt{v_0} + 2m_\rho, \qquad \mathbb{E}[T] = 0, \quad \mathbb{E}[T^2] = v_0,$$

with explicit masses

$$\Pr(T = L) = \Pr(T = -L) = \frac{v_0}{2L^2}, \qquad \Pr(T = 0) = 1 - \frac{v_0}{L^2}.$$

Because $L^2 \geq v_0$, these probabilities are nonnegative. They sum to one since $v_0/L^2 + 1 - v_0/L^2 = 1$, and

$$\mathbb{E}T = L\,v_0/(2L^2) - L\,v_0/(2L^2) = 0, \qquad \mathbb{E}T^2 = L^2 v_0/L^2 = v_0.$$

Thus $\mathbb{E}T = 0$ and $\mathbb{E}T^2 = v_0$ exactly. Set the target $\theta = \pm d_\rho^\star/\sqrt{n}$, where

$$d_\rho^\star = \min\left\{3, \frac{B_0}{16}, \frac{B_0}{16\sqrt{v_0}}\right\}.$$

Conditional on $(X, T) = (x, t)$, let $R \in \{-1, +1\}$ satisfy

$$\Pr(R = r \mid X, T = t) = \frac{1}{2}\left(1 + \frac{2r\theta t}{B_0}\right), \qquad r \in \{-1, +1\},$$

and set $Y = B_0 R/2$. This is the three-point common-host variant of the scalar q experiment in Appendix C.3. The bounds defining m_ρ imply

$$L \leq \frac{B_0 + \sqrt{v_0}}{2} < B_0, \qquad \frac{L}{\sqrt{v_0}} \leq 1 + \frac{\sqrt{v_0}}{2B_0} < \frac{3}{2}, \qquad |\frac{2\theta T}{B_0}| \leq \frac{3}{16}.$$

Thus the two response probabilities lie in $[13/32, 19/32]$, $\mathbb{E}(R \mid X, T) = 2\theta T/B_0$, and $\mathbb{E}(Y \mid X, T) = \theta T$. With zero nuisances, $\varepsilon = Y - \theta T$ is conditionally centered and obeys $|\varepsilon| \leq B_0/2 + 3B_0/32 < B_0$; the target, treatment, outcome, and treatment residual obey the same shared envelope, and the latter has constant variance v_0. With this displayed response likelihood held fixed, the likelihood-energy calculation's dependence on the treatment marginal uses only this second moment $\mathbb{E}T^2 = v_0$. Consequently K_q, the testing constant, and the rate are unchanged.

For P, O, ab, and M, use respectively the raw conditional masses from Appendices C.4, D, E, and F, replacing only their Walsh coordinate by a coordinate in $\mathcal{J}_P, \mathcal{J}_O, \mathcal{J}_{ab}, \mathcal{J}_M$. A nonzero character on the larger host is still a mean-zero unit Rademacher variable, so every atom probability, normalization identity, conditional-centering identity, response-centering identity, target range, treatment/response/

nuisance/residual envelope, and constant conditional variance calculation is unchanged. In particular the common reference laws remain valid because

$$\Delta(v_0+\eta^2)=\eta(\kappa+\gamma) \quad \text{for P and } ab, \qquad \Delta(v_0+\gamma^2)=\gamma\eta \quad \text{for O and M.}$$

Thus the q family and every active native P/O/ab/M hard subfamily lie in the single $\mathcal{P}(\Gamma^\star)$. An inactive component inequality reuses the common q splice; hence all five component lower inequalities hold within the pair even though the named subfamilies need not be five distinct laws at every cell. The proof uses strict lower-endpoint slack only through $m_\rho > 0$ and the component atom envelopes; constant conditional variance is inherited from the raw laws; the fixed infinite ambient space is used only to embed the finite host; and known pre-data upper certificates are used only to choose the bands and amplitudes before data.

L.4 Likelihood distance and arbitrary-rule reduction

For two different coordinates in any component band, $\mathbb{E}_{P_X^\star}\phi_j\phi_k = 0$; equal coordinates with equal or opposite signs give the same $1 \pm K_\ell$ likelihood inner products as in the component appendices. Hence their exact finite Gram-mixture χ^2/Hellinger bounds, and therefore their product-law total-variation bounds, are unchanged. For q, the exact one-sample likelihood energy is

$$K_q = \frac{4(d_\rho^\star)^2 v_0}{nB_0^2} \le \frac{1}{64n},$$

so the original $2n$-sample testing threshold also remains strict.

Fix the fully disclosed $\Gamma^\star$, frequency allocation, $P_X^\star$, and budgets. Any allowed $\delta_{\Gamma^\star}$ is now an ordinary Borel statistic of $Z_{1:2n}$. The two-mixture expected-loss inequality applies to every arbitrary Borel rule after disclosure of the full pair:

$$\max_{\ell\in\{0,1\}} \mathbb{E}_{\Pi_{i,\ell}}|\delta_{\Gamma^\star}-\theta_{i,\ell}| \ge \frac{\Delta_i}{4}\{1-\mathrm{TV}(\Pi_{i,0},\Pi_{i,1})\}.$$

Every $\Pi_{i,\ell}$ is supported on a subfamily of the same $\mathcal{P}(\Gamma^\star)$. Evaluating unused public functions at observed X is deterministic post-processing of X; it is already available to the arbitrary statistic and creates no label-dependent side channel. The same testing-distance bound therefore applies with the enlarged public dictionary.

L.5 All-sample-size splice and fixed-pair assembly

When a large branch is inactive, its named splice subfamily is the same q subfamily inside the same pair. If $s^2 < q$, then $P = s(b\wedge s) \le s^2 < q$; if $n < 16$, then $P \le 1 \le 4q$. The O statement is the exact mirror. If $a^2+b^2 < q$, then $ab \le (a^2+b^2)/2 < q/2$; if $n < 16$, then $ab \le 1 \le 4q$. If $(s\wedge t)^2 < q$, then $M < q$; if $n < 16$, then $M \le 1 \le 4q$. The equality boundaries belong to the large branch by definition, and zero terms require no division or limiting argument.

Use $c_{\min,\rho}$ from Appendix G, the minimum of the q testing constant, the four fixed large-branch constants, and all four q-splice constants. It depends only on fixed ρ, not on $n, a, s, b, t, \Gamma^\star$, or a compatible law. Because all five inequalities hold for every rule on the same pair, the infimum is taken only after their maximum:

$$\begin{aligned}\inf_{\delta_{\Gamma^\star}\ \text{Borel}} \sup_{P\in\mathcal{P}(\Gamma^\star)} \mathbb{E}_P|\delta_{\Gamma^\star}-\beta(P)| &\ge c_{\min,\rho}\max\{q, ab, M, O, P\} \\ &\ge \frac{c_{\min,\rho}}{5}(q+S_0) \ge \frac{c_{\min,\rho}}{10}(q+X),\end{aligned}$$

where $S_0 = ab + M + O + P$ and $S_0 \le X \le 2S_0$ by (10). Setting $c_\rho^\star = c_{\min,\rho}/10$ proves Proposition 18. Taking the outer supremum in (3) now proves the interior lower bound in Theorem 1.

L.6 Every prescribed atomless covariate marginal

Let ν be a probability measure on the fixed standard Borel space $\mathcal{X}$, prescribed before the public pair is chosen. Define

$$\mathcal{P}_\nu(\Gamma) = \{P\in\mathcal{P}(\Gamma) : P_X = \nu\}, \qquad \mathfrak{G}_\nu = \{\Gamma \text{ legal} : \mathcal{P}_\nu(\Gamma) \ne \varnothing\},$$

and define the fixed-marginal experiment by

$$\mathfrak{R}_{\nu,n}(a,s,b,t) = \sup_{\Gamma\in\mathfrak{G}_\nu} \inf_{\delta_{\Gamma,\nu}\ \mathrm{Borel}} \sup_{P\in\mathcal{P}_\nu(\Gamma)} \mathbb{E}_P\left|\delta_{\Gamma,\nu}(Z_{1:2n}) - \beta(P)\right|. \tag{13}$$

The learner may know ν as well as the entire represented pair. Nonemptiness here is with respect to this prescribed marginal, rather than compatibility with some other covariate law.

Proposition 19 (Same-pair lower for every prescribed atomless marginal)**.** *For each fixed strict-interior ρ, there are constants $0 < c_\rho \le C_\rho < \infty$, independent of ν, such that*

$$\forall \nu \text{ atomless on } \mathcal{X}, \quad \forall n \ge 1, \quad \forall (a,s,b,t) \in [0,1]^4, \quad \exists \Gamma^\star_{\nu,\rho,n,a,s,b,t} \in \mathfrak{G}_\nu$$

with finite sequential-M1 classes and

$$\inf_{\delta_{\Gamma^\star,\nu}\ \mathrm{Borel}} \sup_{P\in\mathcal{P}_\nu(\Gamma^\star)} \mathbb{E}_P|\delta_{\Gamma^\star,\nu} - \beta(P)| \ge c_\rho(q + X). \tag{14}$$

This single compatible-law set contains the q family and all active native P/O/ab/M families, and supplies the inactive inequalities by the q splice. Consequently

$$c_\rho(q+X) \le \mathfrak{R}_{\nu,n}(a,s,b,t) \le C_\rho(q+X).$$

At saturation $v_- = B_0^2$, for every prescribed marginal ν, including atomic marginals,

$$\frac{3}{128\sqrt{n}} \le \mathfrak{R}_{\nu,n}(a,s,b,t) \le \frac{1}{\sqrt{2n}}.$$

If $v_- > B_0^2$, every $\mathcal{P}_\nu(\Gamma)$ is empty.

Proof. We first construct measurable Walsh functions for the prescribed atomless ν. A standard Borel space has a countable Borel family $(A_k)_{k\ge1}$ separating its points: take the pullbacks of a countable base under a Borel realization in a Polish space. The map

$$e(x) = \sum_{k=1}^{\infty} 2\,3^{-k}\mathbf{1}_{A_k}(x) \in [0,1]$$

is Borel as a limit of Borel partial sums and is injective. Indeed, at the first differing digit k, the difference $2\,3^{-k}$ exceeds the largest possible opposing tail $\sum_{j>k} 2\,3^{-j} = 3^{-k}$. Thus $\mu = e_\#\nu$ has no atoms: every fiber of e is empty or a singleton. Its distribution function $F(z) = \mu((-\infty, z])$ is continuous, since the jump at z equals $\mu\{z\} = 0$. Set

$$U(x) = F(e(x)).$$

This is Borel and $U_\#\nu$ is uniform on $[0,1]$. To see the probability integral transform directly, for $0 < t < 1$ continuity and the limits of F give a point with $F(z) = t$. The right endpoint $b_t = \sup\{z : F(z) \le t\}$ is finite, satisfies $F(b_t) = t$, and $\{z : F(z) \le t\} = (-\infty, b_t]$. Hence $\nu\{U \le t\} = F(b_t) = t$; the endpoints follow by limits.

For $0 \le u < 1$ define

$$b_k(u) = \lfloor 2^k u\rfloor - 2\lfloor 2^{k-1}u\rfloor, \qquad r_k(u) = (-1)^{b_k(u)},$$

and set $b_k(1) = 0$ and $r_k(1) = 1$. These Borel bits are defined even at dyadic endpoints; our convention changes no distributional identity. The first m bits have each binary pattern on one half-open interval of length 2^{-m}, so their vector is uniform on $\{0,1\}^m$. Enumerate the nonempty finite subsets $H \subset \mathbb{N}$ and put

$$w_H(u) = \prod_{k\in H} r_k(u), \qquad \phi_H(x) = w_H(U(x)).$$

Every ϕ_H is Borel, everywhere sign-valued, centered under ν, and has squared norm one. For distinct H, J, $\phi_H\phi_J = \phi_{H\triangle J}$, so $\mathbb{E}_\nu\phi_H\phi_J = 0$. These characters are orthogonal; we do not require them to be jointly independent.

Use precisely the active predicates, band counts, $m_\star$, amplitudes, and finite unions already defined in this appendix. Allocate the bands among the nonempty subsets of $\{1,\dots,m_\star\}$, identifying such a

subset with its nonzero binary frequency. Replace each finite-host character by the corresponding ϕ_H. The entire finite vector of allocated characters under ν then has exactly the same joint law as under the earlier uniform finite host, because the first $m_\star$ bits are uniform. Define $\Gamma^\star_{\nu,\rho,n,a,s,b,t}$ using the same ordered finite lists followed by repetitions of zero. This is a public choice made before the component, coordinate, sign, and label.

All functions are measurable, and these representations satisfy sequential M1 with their finite classes. Their pointwise envelopes remain $H_\mu \le \lambda_\rho s/128$ and $H_\pi \le \lambda_\rho t/256$. Orthogonality gives exactly the same off-band approximation calculation $\|c\phi_H - d\phi_J\|^2_{L^2(\nu)} = c^2 + d^2$ for $H \ne J$. Thus the P/O/ab distances are preserved, the M truths remain in class, and the q truths remain zero. The complete difference classes, including cross-band differences, still obey

$$|\Delta\mathcal{G}^\star_\mu| \le 49p_s^2, \qquad |\Delta\mathcal{G}^\star_\pi| \le 25p_t^2$$

whenever the corresponding class is nontrivial. The same global Massart calculation therefore gives, at every positive radius in the stated ranges,

$$\frac{\mathfrak{R}_n(r;\Delta\mathcal{G}^\star_\mu)}{3r} \le \frac{23\lambda_\rho s}{192} \le c_{\mathrm{rad}} s \quad (r \ge s > 0),$$
$$\frac{\mathfrak{R}_n(r;\Delta\mathcal{G}^\star_\pi)}{3r} \le \frac{23\lambda_\rho t}{384} \le c_{\mathrm{rad}} t \quad (r \ge t > 0).$$

Zero stochastic budgets again give singleton zero classes, with zero complexity at every positive radius. There is no dependence on ν in these bounds.

For each active component, retain the original conditional kernel of (T, Y) given its signed Walsh value, and let $X \sim \nu$. This is a Borel probability kernel because it has finitely many atoms with Borel masses. At each x, the signed value is either -1 or 1, so the original atom positivity, normalization, target range, all shared envelopes, both conditional-centering identities, and exact constant treatment-residual variance v_0 hold pointwise. The transport preserves the full conditional observation laws. Together with the class calculations it proves membership in $\mathcal{P}_\nu(\Gamma^\star)$.

We check testing on the full observation space. For each active component $i \in \{P, O, ab, M\}$, let Q_i be its native finite-atom reference law on (T, Y). A different component may use a different Q_i, but the two labels within component i use the same Q_i, by the reference identities verified above. On $\mathcal{X} \times \mathbb{R}^2$ take $\widetilde{Q}_i = \nu \otimes Q_i$. The full one-observation likelihood is exactly

$$\frac{dP_{i,\ell,j,\sigma}}{d\widetilde{Q}_i}(x,t,y) = 1 + \sigma\phi_j(x)h_{i,\ell}(t,y), \qquad \mathbb{E}_{Q_i}h_{i,\ell} = 0, \quad K_{i,\ell} = \mathbb{E}_{Q_i}h^2_{i,\ell}.$$

Consequently, for a fixed label, its full Gram entries are

$$\mathbb{E}_{\widetilde{Q}_i}[(1+\sigma\phi_j h_{i,\ell})(1+\tau\phi_k h_{i,\ell})] = 1 + \sigma\tau K_{i,\ell}\mathbb{E}_\nu(\phi_j\phi_k).$$

They equal 1 off the coordinate diagonal and $1 \pm K_{i,\ell}$ on it. For the uniform coordinate/sign mixture of $N = 2n$ independent observations, Lemma 16 gives exactly

$$\chi^2(\overline{P}^N_{i,\ell}, \widetilde{Q}_i^N) = \frac{\{(1+K_{i,\ell})^N + (1-K_{i,\ell})^N\}/2 - 1}{p_i}.$$

The native energy bounds and all ensuing Hellinger and total-variation bounds are therefore unchanged. This calculation is for the full X, not a projection onto its first bits. Extra digits and evaluations of all public functions are measurable information already included in this experiment. Thus an arbitrary Borel learner knowing both ν and the full pair is covered by the same testing inequality.

For q use the three-point variant proved in this appendix, independently of $X \sim \nu$, with $\theta = \pm d^\star_\rho/\sqrt{n}$. The zero truths belong to the enlarged classes. Its full reference is $\nu \otimes Q_q$, where Q_q has the displayed three-point treatment marginal and an independent uniform response sign. Its likelihood is $1 \pm 2d^\star_\rho rt/(B_0\sqrt{n})$, with unchanged energy $K_q \le 1/(64n)$. Hence the same q lower constant holds. This also proves $\mathcal{P}_\nu(\Gamma^\star) \ne \varnothing$ for every budget cell. All inactive-branch comparisons with q are numerical and unchanged, including $n < 16$, zero budgets, and equality boundaries. Taking the maximum of the five within-pair inequalities before taking the infimum over Borel rules gives

$$\inf_\delta \sup_{P \in \mathcal{P}_\nu(\Gamma^\star)} \mathbb{E}_P|\delta - \beta(P)| \ge \frac{c_{\min,\rho}}{10}(q + X).$$

This proves (14) with the same constant as for the finite host. The global upper of Proposition 3 applies to each restricted compatible-law set, and its pair-specific Borel estimator is allowed in (13); the upper constant is likewise independent of ν.

For saturation, the pairwise upper (H.1)–(H.4) requires no assumption on the marginal. In the lower experiment of Appendix H, replace only its constant covariate by an independent $X \sim \nu$. Keep the singleton zero classes, $T = \pm B_0$ equiprobably, $\delta_n = 3/(32\sqrt{n})$, and exactly the response probabilities in (H.5). All raw-law and zero-budget checks remain valid. The common factor ν contributes affinity one, so the product-law calculation still gives $\mathrm{TV}^2 \le 72/1024 < 1/4$, and (H.6) gives $3/(128\sqrt{n})$. These saturation constants are not obtained by continuing the interior constants. Finally the envelope contradiction when $v_- > B_0^2$ holds for every marginal. □

Atomlessness is essential to this extension. For example, if ν is a Dirac mass, the model reduces to bounded scalar regression with an intercept, and $\beta = \mathrm{Cov}(T, Y)/\mathrm{Var}(T)$ with $\mathrm{Var}(T) \ge v_-$. Clipping the empirical covariance divided by the empirical variance truncated below at $v_-/2$ gives uniform expected error $O_\rho(q)$: bounded empirical first and second moments have expected error $O_\rho(q)$, the truncated denominator is bounded away from zero, and truncation cannot increase its error from the true variance. The q family gives the matching lower. Thus fixed positive nuisance budgets cannot force the interior nonparametric terms for this marginal. The original finite-host proposition remains valid on every fixed infinite standard Borel host, including countably infinite hosts that admit no atomless probability measure.

L.7 Scope of the strengthening

All component laws share the same X marginal and support, and the public complexity bound includes cross-band differences. The constructed pair $\Gamma^\star$ may depend on n and the budget tuple; the result is a lower bound for this pair. The finite-host construction applies on countably infinite ambient spaces, while transport to a preassigned marginal requires atomlessness. The lower-bound families retain constant conditional treatment variance.

M Sharp adaptation on full interval boxes

This appendix proves Proposition 5 for full interval boxes with separate countable partitions and arbitrary bounded Borel centers. Throughout, $N = 2n$, $q = n^{-1/2}$, $A = a + s^2$, $B = b + t^2$, and all model and sequential-M1 conventions are those of Appendix A. The covariate law is arbitrary and unknown. The estimator is fixed for each (Γ, ρ, n) before choosing any of the four certificates or a compatible law, and attains $q + AB$.

M.1 Public centers and separate represented groupings

For each $\nu \in \{\mu, \pi\}$, let $(C_{\nu j})_{j \ge 1}$ be a countable Borel partition, and let c_ν be a bounded Borel function. The two partitions can differ, and a center need not be constant on its partition cells. Consider the full interval boxes

$$\mathcal{G}_\nu = \left\{ c_\nu + \sum_j \theta_j \mathbf{1}_{C_{\nu j}} : \theta_j \in [-h_{\nu j}, h_{\nu j}] \text{ for every } j \right\}, \qquad h_{\nu j} \ge 0, \qquad |c_\nu(x)| + h_{\nu j} \le B_{\mathcal{G}} \quad (x \in C_{\nu j}). \tag{15}$$

Each class contains *every independent* coordinate choice displayed in (15); inclusion in a box would not suffice. These classes admit sequential-M1 cores: approximate the normalized coordinates on the first k cells by rational numbers in $[-1, 1]$, and use zero normalized coordinates elsewhere. Any supplied sequential-M1 core may be used in the construction.

Neither the centers nor the partitions are required as additional inputs. The original cores recover both centers and widths by

$$\begin{aligned} c_\nu(x) &= \tfrac{1}{2}\left\{\sup_{\ell\ge 1} g_{\nu,\ell}(x) + \inf_{\ell\ge 1} g_{\nu,\ell}(x)\right\}, \\ h_\nu(x) &= \tfrac{1}{2}\left\{\sup_{\ell\ge 1} g_{\nu,\ell}(x) - \inf_{\ell\ge 1} g_{\nu,\ell}(x)\right\}. \end{aligned} \tag{16}$$

These countable extrema are Borel and finite. Pointwise sequential M1 transfers the full-class extrema $c_\nu(x) \pm h_{\nu j}$ to the core, which proves the equality. Define a separate equivalence relation for each side using its *centered* core:

$$i \sim_\nu k \quad\Longleftrightarrow\quad g_{\nu,\ell}(X_i) - c_\nu(X_i) = g_{\nu,\ell}(X_k) - c_\nu(X_k) \quad \text{for every } \ell \ge 1. \tag{17}$$

Each relation is Borel. Equality of all centered core evaluations implies equality of every centered full-class evaluation, by sequential M1. Two different cells on side ν are separated whenever either has positive width: vary that coordinate while holding the other fixed. Thus its groups are its occupied partition cells, except that its zero-width cells merge. Every centered public function on that side vanishes on the merged cells. No common refinement is taken.

Set the public constants

$$\alpha = \min\left\{\frac{1}{8}, \frac{v_-}{8B_0^2}\right\} > 0, \qquad m = \lfloor \alpha N \rfloor. \tag{18}$$

For each side, let J_ν contain the m singleton rows with largest widths $h_\nu(X_i)$, or every singleton row if fewer than m exist. Resolve equal widths by increasing sample index. Let A_ν have block $I_k - \mathbf{1}_k\mathbf{1}_k'/k$ on a group of size $k \ge 2$, block $[0]$ on a row in J_ν, and block $[1]$ on every other singleton. Then

$$\mathsf{A}_\nu' = \mathsf{A}_\nu, \qquad \mathsf{A}_\nu^2 = \mathsf{A}_\nu, \qquad 0 \preceq \mathsf{A}_\nu \preceq I_N, \qquad (\mathsf{A}_\nu)_{ii} \ge \tfrac{1}{2}\mathbf{1}_{\{i\notin J_\nu\}}. \tag{19}$$

Every matrix entry is Borel: if $m_{\nu i} = \sum_k \mathbf{1}_{\{i\sim_\nu k\}}$, it is $(\mathsf{A}_\nu)_{ik} = \mathbf{1}_{\{i=k\}}\mathbf{1}_{\{i\notin J_\nu\}} - \mathbf{1}_{\{i\sim_\nu k\}}\mathbf{1}_{\{m_{\nu i}\ge 2\}}/m_{\nu i}$. Also, $(\mathsf{A}_\nu)_{ik} \le 0$ when $i \ne k$, and $\|\mathsf{A}_\nu z\|_\infty \le 2\|z\|_\infty$. For a function f, write $f_X = (f(X_1), \dots, f(X_N))'$, $\|z\|_N^2 = z'z/N$, and $\|z\|_{N,1} = N^{-1}\sum_i |z_i|$. For a public function $g_\nu = c_\nu + z_\nu$, its centered part obeys $|z_\nu| \le h_{\nu j}$ on $C_{\nu j}$, and $\mathsf{A}_\nu(z_\nu)_X$ vanishes on repeated groups and trimmed rows, retaining its original values on the remaining singleton rows. In particular $|J_\nu| \le \alpha N$.

M.2 Localized complexity controls singleton amplitudes

The occupancy bound below controls the total singleton amplitude. Trimming converts this control into a squared residual bound of order $r^4 + N^{-1}$.

Lemma 20 (An occupancy consequence of the full-box certificate)**.** *Fix either class in* (15)*, suppress its index (including in $\mathsf{A} = \mathsf{A}_\nu$), and let $r \in [0,1]$ be its stochastic upper certificate in (A.1). Set $p_j = P_X(C_j)$ and*

$$K_\rho = \max\left\{\frac{3\sqrt{3}}{2}c_{\mathrm{rad}}, 27B_{\mathcal{G}}c_{\mathrm{rad}}^2\right\}.$$

Then, for every $n \ge 1$, including $r = 0$,

$$\sum_j p_j h_j (1-p_j)^{2n-1} \le \sum_j p_j h_j e^{-np_j} \le K_\rho r^2. \tag{20}$$

Consequently every fixed centered public function z satisfies

$$\mathbb{E}\|\mathsf{A}z_X\|_{N,1} \le K_\rho r^2, \qquad \mathbb{E}\|\mathsf{A}z_X\|_N^2 \le B_{\mathcal{G}}K_\rho r^2. \tag{21}$$

Proof. Put $\lambda_j = np_j$ and $L = \sum_j p_j h_j e^{-\lambda_j}$. Independent interval coordinates imply that $\Delta\mathcal{G}$ contains the damped sign cube

$$f_\sigma = 2\sum_j \sigma_j h_j e^{-\lambda_j/2}\mathbf{1}_{C_j}, \qquad \sigma_j \in \{-1, 1\}.$$

Each such function has squared population norm at most $4B_{\mathcal{G}}L$. For the sample of size n used in the complexity certificate, let $S_j = \sum_{i=1}^{n} \epsilon_i \mathbf{1}_{\{X_i \in C_j\}}$. Then

$$\mathbb{E}S_j^2 = \lambda_j, \qquad \mathbb{E}S_j^4 = np_j + 3n(n-1)p_j^2 \le \lambda_j + 3\lambda_j^2.$$

Hölder's inequality in the form $\mathbb{E}S_j^2 \le (\mathbb{E}|S_j|)^{2/3}(\mathbb{E}S_j^4)^{1/3}$ gives, for $\lambda_j > 0$,

$$\mathbb{E}|S_j| \ge \frac{(\mathbb{E}S_j^2)^{3/2}}{(\mathbb{E}S_j^4)^{1/2}} \ge \frac{\lambda_j}{\sqrt{1+3\lambda_j}};$$

the lower bound is also valid when $\lambda_j = 0$. Maximizing the signs coordinatewise produces the measurable random variable $2n^{-1}\sum_j h_j e^{-\lambda_j/2}|S_j|$, which is bounded above by the full localized supremum at radius $2\sqrt{B_{\mathcal{G}}L}$. Since $e^{\lambda/2}/\sqrt{1+3\lambda} \ge 1/\sqrt{3}$ for $\lambda \ge 0$, monotonicity of outer expectation and Tonelli's theorem give a complexity lower bound $2L/\sqrt{3}$.

For $r > 0$, evaluate (A.1) at $R = r \vee 2\sqrt{B_{\mathcal{G}}L}$ to obtain

$$\frac{2L}{\sqrt{3}} \le 3c_{\mathrm{rad}} r \max\{r, 2\sqrt{B_{\mathcal{G}}L}\}.$$

If the maximum is r, this bounds L by $(3\sqrt{3}/2)c_{\mathrm{rad}}r^2$; otherwise, squaring after division by $\sqrt{L}$ gives $L \le 27 B_{\mathcal{G}} c_{\mathrm{rad}}^2 r^2$. If $r = 0$ and $L > 0$, the same cube at its positive containing radius contradicts the zero-budget requirement in (A.1), so $L = 0$. No radius-zero quotient is used.

Since $2n - 1 \ge n$, $(1-p_j)^{2n-1} \le e^{-np_j}$. The expected empirical amplitude on cells occupied exactly once in the full sample is exactly

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} h(X_i)\mathbf{1}_{\{\#\{k:X_k \in C(X_i)\}=1\}}\right] = \sum_j p_j h_j (1-p_j)^{2n-1},$$

where $h(x) = h_j$ on C_j and $C(x)$ denotes the cell of x. Merged zero-width cells contribute zero. This proves the first residual bound; the second follows pointwise from $|z|^2 \le B_{\mathcal{G}}|z|$ on the surviving singleton groups. The unknown probabilities are used only in this proof. □

Lemma 21 (Squared residual bound after trimming)**.** *Under Lemma 20, put*

$$H = \frac{1}{N}\sum_{i=1}^{N} h(X_i)\mathbf{1}_{\{\textit{the own cell of } X_i \textit{ occurs once}\}}, \qquad R_N(r) = \alpha^{-1}\sqrt{K_\rho^2 r^4 + \frac{2B_{\mathcal{G}}^2}{N}}.$$

Then

$$\mathbb{E}H^2 \le K_\rho^2 r^4 + \frac{2B_{\mathcal{G}}^2}{N}, \qquad \mathbb{E}\|\mathsf{A}z_X\|_N^2 \le R_N(r)^2 \tag{22}$$

for every fixed centered public function z.

Proof. If N_j denotes the number of observations in cell j, then $H = N^{-1}\sum_j h_j \mathbf{1}_{\{N_j=1\}}$. Replacing one observation changes only its source and destination cells' contributions, each by at most its width. Thus the replacement changes H by at most $2B_{\mathcal{G}}/N$. Efron–Stein and Lemma 20 give

$$\mathrm{Var}(H) \le 2B_{\mathcal{G}}^2/N, \qquad \mathbb{E}H \le K_\rho r^2,$$

proving the first assertion. Zero-width cells contribute zero, so merging them in the represented equivalence relation does not change H.

If any singleton remains after trimming, let $h_{\max}$ be its largest width. The m trimmed rows and a remaining row attaining $h_{\max}$ all have widths at least $h_{\max}$. Therefore

$$(m+1)h_{\max} \le NH, \qquad h_{\max} \le \frac{NH}{m+1} \le \frac{H}{\alpha},$$

because $m + 1 > \alpha N$. This also covers $m = 0$. If every singleton was trimmed, the residual is zero. Otherwise, $\mathsf{A}z_X$ vanishes on all repeated groups and trimmed rows, and has absolute entries at most $h_{\max}$ on its remaining rows. Hence $\|\mathsf{A}z_X\|_N^2 \le H^2/\alpha^2$, proving the second assertion. At $r = 0$, $\mathbb{E}H = 0$ also implies $H = 0$ almost surely. □

M.3 One stabilized rule for the two partitions

Put $\mathsf{B} = \mathsf{A}_\pi \mathsf{A}_\mu$. This matrix need not be symmetric or positive semidefinite. Its operator norm is at most one, and the off-diagonal entries of each factor are nonpositive. Thus

$$\mathsf{B}_{ii} = \sum_k (\mathsf{A}_\pi)_{ik}(\mathsf{A}_\mu)_{ki} \geq \frac{1}{4}\mathbf{1}_{\{i \notin J_\pi \cup J_\mu\}}, \qquad \mathsf{B}_{ii} = 0 \quad (i \in J_\pi \cup J_\mu). \tag{23}$$

For sample vectors T, Y, c_μ, c_π, define

$$\begin{aligned} D &= (T - c_\pi)'\mathsf{B}T/N, & U &= (T - c_\pi)'\mathsf{B}(Y - c_\mu)/N, \\ S_\mu &= T'\mathsf{A}_\mu T/N, & V_\mu &= T'\mathsf{A}_\mu(Y - c_\mu)/N. \end{aligned} \tag{24}$$

Proposition 22 (All-four-certificate adaptation on full boxes)**.** *For every fixed sequential-M1 pair of the form* (15)*, the Borel rule*

$$\widehat{\beta}_{\mathrm{box}} = \begin{cases} \Pi_{[-3,3]}(U/D), & D \geq v_-/32, \\ \Pi_{[-3,3]}\{V_\mu/(S_\mu \vee (v_-/4))\}, & D < v_-/32 \end{cases} \tag{25}$$

uses only (Γ, ρ, n) *and satisfies, with a constant depending only on* ρ*, simultaneously for every* $(a, s, b, t) \in [0, 1]^4$ *and every* $P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$*,*

$$\mathbb{E}_P|\widehat{\beta}_{\mathrm{box}} - \beta(P)| \leq C_\rho\{q + (a + s^2)(b + t^2)\}. \tag{26}$$

This includes every $n \geq 1$*, every zero-budget boundary, and every regularity tuple with a nonempty compatible-law class.*

Proof. Write $\mu_X = (\mu_0)_X$, $\pi_X = (\pi_0)_X$, and define the proof quantities

$$d_\mu = \mu_X - c_\mu, \qquad d_\pi = \pi_X - c_\pi, \qquad F_\nu = \|\mathsf{A}_\nu d_\nu\|_N, \qquad J = \mathbb{E}(F_\mu F_\pi).$$

For any $\eta > 0$, take fixed approximation comparators $g_\mu = c_\mu + z_\mu$, $g_\pi = c_\pi + z_\pi$ with errors at most $a + \eta$, $b + \eta$. Contraction and Minkowski's inequality, applied in $L_2(P^N)$ to their empirical residual norms, give

$$(\mathbb{E}F_\mu^2)^{1/2} \leq a + \eta + R_N(s), \qquad (\mathbb{E}F_\pi^2)^{1/2} \leq b + \eta + R_N(t).$$

The rule does not use these comparators. Letting $\eta \downarrow 0$ handles unattained approximation infima and yields

$$J \leq \{a + R_N(s)\}\{b + R_N(t)\}. \tag{27}$$

Primary bias and noise. The conditional-mean part of the primary score satisfies

$$|d_\pi'\mathsf{B}d_\mu|/N = |\langle \mathsf{A}_\pi d_\pi, \mathsf{A}_\mu d_\mu\rangle_N| \leq F_\pi F_\mu. \tag{28}$$

Its full decomposition is

$$U - \beta_0 D = d_\pi'\mathsf{B}d_\mu/N + u'\mathsf{B}d_\mu/N + (T - c_\pi)'\mathsf{B}\varepsilon/N.$$

Conditional on $X_{1:N}$, independence and centering of the u_i give

$$\mathbb{E}\{(u'\mathsf{B}d_\mu/N)^2 \mid X_{1:N}\} \leq B_0^2\|\mathsf{B}d_\mu\|^2/N^2 \leq B_0^2(B_0 + B_\mathcal{G})^2/N.$$

Conditional on $(X_{1:N}, T_{1:N})$, the independent centered ε_i give the same bound for the other noise term, using $\|\mathsf{B}'(T - c_\pi)\| \leq \|T - c_\pi\|$. Consequently

$$\mathbb{E}|U - \beta_0 D| \leq J + 2B_0(B_0 + B_\mathcal{G})/\sqrt{N}. \tag{29}$$

On the primary branch the clipped loss is at most $32|U - \beta_0 D|/v_-$.

Denominators under average conditional variance. Put $v_i = \mathbb{E}(u_i^2 \mid X_i)$, $\overline{v} = N^{-1}\sum_i v_i$, and $\Omega = \mathrm{diag}(v_i)$. Since $0 \le v_i \le B_0^2$, $\mathbb{E}\overline{v} \ge v_-$, and $|J_\nu| \le \alpha N$,

$$\mathbb{P}(\overline{v} < 3v_-/4) \le \frac{16B_0^4}{Nv_-^2}, \tag{30}$$

$$\mathrm{tr}(\mathsf{B}\Omega)/N \ge \tfrac{1}{4}(\overline{v} - 2\alpha B_0^2),$$
$$\mathrm{tr}(\mathsf{A}_\mu\Omega)/N \ge \tfrac{1}{2}(\overline{v} - \alpha B_0^2). \tag{31}$$

In particular, on $\overline{v} \ge 3v_-/4$,

$$\mathrm{tr}(\mathsf{B}\Omega)/N \ge v_-/8, \qquad \mathbb{E}(S_\mu \mid X_{1:N}) \ge 5v_-/16. \tag{32}$$

The second inequality also uses $\pi_X'\mathsf{A}_\mu\pi_X \ge 0$. Neither calculation assumes a pointwise lower bound on v_i.

For any symmetric covariate-measurable matrix H, independence and centering give

$$\begin{aligned}\mathrm{Var}(u'Hu \mid X_{1:N}) &= \sum_i H_{ii}^2 \,\mathrm{Var}(u_i^2 \mid X_i) + 4\sum_{i<k} H_{ik}^2 v_i v_k \\ &\le 2B_0^4 \|H\|_{\mathrm{F}}^2.\end{aligned}$$

The random part of NS_μ is $u'\mathsf{A}_\mu u + 2u'\mathsf{A}_\mu\pi_X$. The two variances are at most $2B_0^4N$ and $4B_0^4N$. Applying $\mathrm{Var}(Q+L) \le 2\,\mathrm{Var}\,Q + 2\,\mathrm{Var}\,L$, which does not assume vanishing third moments, therefore gives

$$\mathrm{Var}(S_\mu \mid X_{1:N}) \le 12B_0^4/N. \tag{33}$$

For the nonsymmetric cross moment, put

$$d_X = \mathbb{E}(D \mid X_{1:N}) = \mathrm{tr}(\mathsf{B}\Omega)/N + d_\pi'\mathsf{B}\pi_X/N.$$

Let $H = (\mathsf{B} + \mathsf{B}')/2$ and $\ell_X = \mathsf{B}\pi_X + \mathsf{B}'d_\pi$. The random part of ND is $u'Hu + u'\ell_X$, with

$$\|H\|_{\mathrm{op}} \le 1, \qquad \|H\|_{\mathrm{F}}^2 \le N, \qquad \|\ell_X\| \le (2B_0 + B_{\mathcal{G}})\sqrt{N}.$$

The same calculation yields

$$\mathrm{Var}(D \mid X_{1:N}) \le C_D/N, \qquad C_D = 4B_0^4 + 2B_0^2(2B_0 + B_{\mathcal{G}})^2. \tag{34}$$

Define

$$\mathcal{E} = \{\overline{v} < 3v_-/4\} \cup \{S_\mu < v_-/4\} \cup \{|D - d_X| > v_-/32\}.$$

On $\overline{v} \ge 3v_-/4$, the conditional-mean gap in the second event is at least $v_-/16$. Conditional Chebyshev and the preceding bounds, counting the low-$\overline{v}$ event only once, give

$$\mathbb{P}(\mathcal{E}) \le \frac{3088B_0^4 + 1024C_D}{Nv_-^2}. \tag{35}$$

Clipping limits its contribution to loss to six times this probability.

Fallback bias is controlled by the same product. On $\mathcal{E}^c$, choosing fallback implies $d_X < v_-/16$. Equation (32) then forces

$$v_-/16 \le |d_\pi'\mathsf{B}\pi_X|/N \le B_0F_\pi.$$

Thus the remaining fallback event is contained in $\{F_\pi \ge \kappa\}$, where $\kappa = v_-/(16B_0)$.

The fallback score decomposes as

$$V_\mu - \beta_0 S_\mu = \pi_X'\mathsf{A}_\mu d_\mu/N + u'\mathsf{A}_\mu d_\mu/N + T'\mathsf{A}_\mu\varepsilon/N.$$

Its bias is pointwise at most B_0F_μ, and conditioning as above bounds its total expected absolute noise by $B_0(2B_0 + B_{\mathcal{G}})/\sqrt{N}$. Therefore

$$\mathbb{E}\big[B_0F_\mu\mathbf{1}_{\{\text{fallback}\}\cap\mathcal{E}^c}\big] \le B_0\mathbb{E}\{F_\mu\mathbf{1}_{\{F_\pi\ge\kappa\}}\} \le (B_0/\kappa)J.$$

Absolute noise terms can be bounded without their event indicators; branch selection need not be independent of the scores. Divide by $S_\mu \geq v_-/4$ on $\mathcal{E}^c$, and combine with the primary branch and exceptional-event bounds to obtain

$$\mathbb{E}|\widehat{\beta}_{\mathrm{box}} - \beta_0| \leq \left(\frac{32}{v_-} + \frac{64B_0^2}{v_-^2}\right) J + \frac{6(3088B_0^4 + 1024C_D)}{Nv_-^2} + \left\{\frac{64B_0(B_0 + B_{\mathcal{G}})}{v_-} + \frac{4B_0(2B_0 + B_{\mathcal{G}})}{v_-}\right\} N^{-1/2}. \tag{36}$$

Let $C_* = \max\{1, K_\rho/\alpha, \sqrt{2}B_{\mathcal{G}}/\alpha\}$. Then $R_N(r) \leq C_*(r^2 + N^{-1/2})$, and (27), with all four budgets at most one, gives

$$J \leq C_*^2\{(a + s^2)(b + t^2) + 5N^{-1/2}\}.$$

Since $N = 2n$, this proves (26).

Countable extrema, countable centered-core equalities, finite group sizes, width comparisons with the specified tie rule, matrix operations, and clipping are Borel. The trimming fraction uses only ρ; no certificate, comparator, or covariate probability enters the rule. The floor in (18) and Lemma 21 include all small samples and zero budgets. This proves simultaneous adaptation. □

Given evaluable partition labels, centers, and widths, the construction reduces to finite block operations and sorting. Under M1 alone it defines a Borel rule. The occupancy argument uses the independent coordinate choices available in a full box.

Block residualization is classical (Cattaneo et al., 2018). Our analysis uses the full-box complexity certificate to bound the widths of retained singleton rows and control the cross-moment fallback.

M.4 Matching lower bound over full-box pairs

Let $\mathfrak{G}^{\mathrm{box}}_{\rho,n}(a, s, b, t)$ be the members of $\mathfrak{G}_{\rho,n}(a, s, b, t)$ whose class-pair projection has the full-box form (15). Define the restricted outer-pair minimax risk by

$$\mathcal{R}^{\mathrm{box}}_{\rho,n}(a, s, b, t) = \sup_{\Gamma \in \mathfrak{G}^{\mathrm{box}}_{\rho,n}(a,s,b,t)} \inf_{\delta_\Gamma \text{ Borel}} \sup_{P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)} \mathbb{E}_P|\delta_\Gamma(Z_{1:2n}) - \beta(P)|. \tag{37}$$

The infimum here can use the declared budgets. The upper rule in Proposition 22 has the stronger property of being fixed before all four budgets.

Theorem 23 (Restricted full-box minimax rate)**.** *On the strict lower-endpoint slack stratum $0 < v_- < B_0^2$, there are constants $c_{\mathrm{box},\rho}, C_{\mathrm{box},\rho} > 0$ such that*

$$c_{\mathrm{box},\rho}\{q + (a + s^2)(b + t^2)\} \leq \mathcal{R}^{\mathrm{box}}_{\rho,n}(a, s, b, t) \leq C_{\mathrm{box},\rho}\{q + (a + s^2)(b + t^2)\} \tag{38}$$

for every $n \geq 1$ and every $(a, s, b, t) \in [0, 1]^4$. For each fixed (ρ, n, s, t), the lower bound is witnessed by one represented full-box pair and one covariate marginal that work for every $(a, b) \in [0, 1]^2$. The lower bound is attained by the box pairs constructed below. At $v_- = B_0^2$, the restricted minimax rate is q; at $v_- > B_0^2$ the compatible-law field is empty.

Proof. The upper bound follows from Proposition 22. For the lower, write $v_0 = v_-$ and use m_ρ, L from (C.5). Put

$$\epsilon_\rho = \min\{1, c_{\mathrm{rad}}, B_{\mathcal{G}}\} > 0, \qquad p = 2^{N+6}, \qquad d = N + 7,$$
$$H_\mu = \epsilon_\rho s^2/128, \qquad H_\pi = \epsilon_\rho t^2/128.$$

Embed $(\mathbb{Z}/2\mathbb{Z})^d$ measurably into the fixed infinite standard Borel space, as in Appendix C.4; denote its image by S and let P_X^* be uniform on S. There are $2^d - 1 \geq p$ nonzero Walsh characters, so choose p distinct characters $\phi_1, \ldots, \phi_p$, extended by zero off S.

Publish the zero-centered full interval boxes with width H_ν on every singleton $\{x\}$, $x \in S$, and width zero on $\mathcal{X} \setminus S$. Rational normalized coordinates give a countable pointwise-everywhere dense representation. This fixes $\Gamma^* = \Gamma^*(\rho, n, s, t)$ before any approximation budget, hard-family component, label, coordinate, or sign is selected.

Both public complexity certificates. For either side with stochastic budget $r \in \{s, t\}$, the full difference class has pointwise envelope $2H_\nu$. Its localized outer Rademacher complexity is therefore at most $2H_\nu$ at every radius. When $r > 0$ and $R \geq r$,

$$\frac{\mathfrak{R}^*_{n,P^*_X}(R; \Delta\mathcal{G}_\nu)}{3R} \leq \frac{2H_\nu}{3r} = \frac{\epsilon_\rho r}{192} \leq c_{\mathrm{rad}} r.$$

When $r = 0$, that side is the singleton zero class, and its complexity vanishes at every positive radius. Also $H_\nu \leq \epsilon_\rho/128 \leq B_{\mathcal{G}}$. These are certificates for both full boxes in every family below, regardless of which nuisance truth is represented.

Four legal families inside the same pair. Write $z = \sigma\phi_j$. We use the exact treatment channel (C.6), which has the specified conditional mean and residual variance exactly v_0 by (C.7). Given (X, T), use the binary response channel (C.12) with response amplitude $B_0/2$.

Two label templates suffice. The P template has parameters $\eta, \gamma \geq 0$ and

$$\Delta = \frac{2\eta\gamma}{v_0}, \qquad \kappa = \gamma + \Delta\eta,$$
$$(\beta, \pi, \mu)_0 = (0, z\eta, z\gamma), \quad (\beta, \pi, \mu)_1 = (\Delta, -z\eta, z\kappa).$$

The O template has parameters $\eta, \gamma \geq 0$ and

$$\Delta = \frac{\gamma\eta}{v_0 + \gamma^2},$$
$$(\beta, \pi, \mu)_0 = (0, z\gamma, z\eta), \quad (\beta, \pi, \mu)_1 = (\Delta, z\gamma, 0).$$

Their complete conditional masses are respectively (C.13) and (D.4). Choose their amplitudes as follows:

term	template	η	γ
ab	P	$m_\rho b/4$	$a/512$
bs^2	P	$m_\rho b/4$	$\epsilon_\rho s^2/256$
at^2	O	$a/256$	$\epsilon_\rho m_\rho t^2/4$
s^2t^2	O	$\epsilon_\rho s^2/256$	$\epsilon_\rho m_\rho t^2/4$

(39)

In both P rows, $2\eta^2/v_0 \leq m_\rho^2/(8v_0) < 1$, so $\kappa \leq 2\gamma$. In the ab row, the zero candidates give approximation errors at most $\kappa \leq a/256 \leq a$ and $\eta \leq b$. In the bs^2 row, both outcome truths belong to the outcome box because $\kappa \leq H_\mu$, while zero gives propensity error at most b. In both O rows, $\gamma \leq \epsilon_\rho t^2/256 \leq H_\pi$, so propensity is represented. The at^2 row has outcome error at most $\eta \leq a$ using zero; in the last row both outcome truths are represented since $\eta \leq H_\mu$. Exact approximation distances are unnecessary.

Every treatment mean has absolute value at most $m_\rho/4$. The channel inequalities (C.5)–(C.7) ensure strict treatment probabilities and $|T|, |u| < B_0$. Uniformly over these four rows,

$$0 \leq \Delta \leq \frac{1}{2048B_0}, \qquad |\mu| \leq \frac{1}{128}, \qquad |\beta T + \mu| \leq \frac{17}{2048} < \frac{1}{64}.$$

The binary response probabilities are consequently strict, and $|\varepsilon| \leq B_0/2 + 1/64 < B_0$. The remaining target, nuisance, and response envelopes follow from these same bounds. Thus every component is a legal PLM, with exact residual variance v_0, not merely a moment-matched array.

Common reference and mixture distance for every sample size. For the P template, the sign averages of the two labels coincide atomwise by (C.15)–(C.16): their required identity is

$$\Delta(v_0 + \eta^2) = \eta(\kappa + \gamma),$$

which follows from $\kappa = \gamma + \Delta\eta$ and $\Delta v_0 = 2\eta\gamma$. For the O template, (D.8)–(D.9) give the same conclusion from $\Delta(v_0 + \gamma^2) = \gamma\eta$. Let Q denote the resulting common law of the treatment and binary response sign. It is strictly positive because it averages strictly positive component masses. Only the component laws are required to have exact residual variance v_0.

For either label, its two sign laws can be written

$$P_{\ell,z} = Q(1 + zh_\ell), \qquad h_\ell = \frac{P_{\ell,+} - P_{\ell,-}}{2Q}, \qquad \mathbb{E}_Q h_\ell = 0, \quad |h_\ell| < 1.$$

Put $K_\ell = \mathbb{E}_Q h_\ell^2 \le 1$ and $\mathbb{Q} = P_X^* \otimes Q$. A full one-observation likelihood is $1 + \sigma\phi_j(X)h_\ell(T,Y)$, where the binary sign is identified with $Y/(B_0/2)$. Walsh orthogonality gives Gram entries $1 + K_\ell$, $1 - K_\ell$, and one, as in Lemma 16.

Choose (j,σ) once uniformly from $[p] \times \{-1,1\}$ and then sample N iid observations from that component. For each label's resulting mixture, (C.3) gives

$$\chi^2(\overline{P}_\ell^{\,N}, \mathbb{Q}^N) = \frac{\{(1+K_\ell)^N + (1-K_\ell)^N\}/2 - 1}{p} \le \frac{2^{N-1}-1}{2^{N+6}} < \frac{1}{128}.$$

The cross-label total variation is at most $1/\sqrt{128} < 1/4$. Equation (C.1) therefore yields a lower risk of at least $\Delta/8$ for every row of (39), for all $n \ge 1$. A zero term has zero separation and causes no exception. The public pair is already fixed, and all public evaluations are deterministic functions of X, so disclosing the entire represented pair does not change this full-observation comparison.

The four gaps are bounded below by their corresponding terms times

$$k_\rho = \frac{\epsilon_\rho^2 m_\rho}{1024(v_0 + m_\rho^2/16)} > 0.$$

Indeed the P gaps are $m_\rho ab/(1024 v_0)$ and $\epsilon_\rho m_\rho bs^2/(512 v_0)$; the O gaps are $\epsilon_\rho m_\rho at^2/\{1024(v_0 + \gamma^2)\}$ and $\epsilon_\rho^2 m_\rho s^2 t^2/\{1024(v_0+\gamma^2)\}$.

Parametric term and assembly. Keep the same P_X^* and make it independent of the scalar experiment (C.8), with both nuisances zero, $T = \pm\sqrt{v_0}$ equiprobably, and $\beta = \pm d_\rho^\star/\sqrt{n}$. The zero truths belong to both published boxes, whose complexity certificates have already been verified. The independent covariate factor leaves the calculation leading to (C.9) unchanged, so this same fixed pair has lower risk $3d_\rho^\star q/8$.

For every rule on Γ^*, its compatible-law supremum is therefore at least

$$\frac{1}{8}\min\{3d_\rho^\star, k_\rho\}\max\{q, ab, bs^2, at^2, s^2t^2\}.$$

Taking the learner infimum and using that the maximum of five nonnegative numbers is at least their sum divided by five proves the lower bound with

$$c_{\mathrm{box},\rho} = \frac{1}{40}\min\{3d_\rho^\star, k_\rho\} > 0.$$

The construction fixes the same Γ^* and P_X^* for every (a,b); only the compatible hard laws vary. The outer supremum in (37) then gives the stated result.

At $v_- = B_0^2$, (H.3)–(H.4) give the q upper bound for every nonempty pair, and the zero singleton pair used in (H.5)–(H.6) is a full interval box and supplies the q lower bound. When $v_- > B_0^2$, the law field is empty by Appendix H; no numerical risk is assigned to that empty field. □

N Adaptation on full affine Hilbert balls

The geometry in this appendix allows overlapping features and infinite rank. Let $B_{\mathcal{H}}$ denote the closed unit ball of a real RKHS. Assume that, for $\nu \in \{\mu, \pi\}$,

$$\mathcal{G}_\nu = c_\nu + R_\nu B_{\mathcal{H}_\nu}, \qquad R_\nu \ge 0, \tag{40}$$

with the pointwise envelope and supplied sequential-M1 cores of Appendix A. These are full affine balls; the two RKHSs and their centers can differ. A zero radius gives a singleton. The centers, radii, and kernels specify the geometric subclass, but need not be additional inputs to the estimator.

Theorem 24 (Adaptation on full affine Hilbert balls)**.** *For every represented pair satisfying* (40)*, there is one Borel rule $\delta_{\Gamma,\rho,n} : \mathcal{Z}^{2n} \to [-3,3]$ such that, simultaneously for all $(a,s,b,t) \in [0,1]^4$ and every $P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$,*

$$\mathbb{E}_P|\delta_{\Gamma,\rho,n} - \beta(P)| \le C_\rho(q + ab) \le C_\rho(q + W). \tag{41}$$

The rule uses none of the four certificates. Its bound is independent of the dimensions, ranks, and overlap of the two feature spaces. In fact, the upper bound does not use the localized-complexity inequalities.

Proof. Write $N = 2n$, $f_X = (f(X_1), \dots, f(X_N))'$, and $\|z\|_N^2 = z'z/N$. First recover the geometry from the supplied cores. For each side set

$$\begin{aligned} \widetilde{c}_\nu(x) &= \tfrac{1}{2}\{\sup_j g_{\nu,j}(x) + \inf_j g_{\nu,j}(x)\}, \qquad z_{\nu,j} = g_{\nu,j} - \widetilde{c}_\nu, \\ u_{\nu,\pm}(x,y) &= \sup_j\{z_{\nu,j}(x) \pm z_{\nu,j}(y)\}, \qquad \ell_\nu(x,y) = \tfrac{1}{4}\{u_{\nu,+}(x,y)^2 - u_{\nu,-}(x,y)^2\}. \end{aligned} \tag{42}$$

Every extremum is finite and Borel. Pointwise sequential M1 preserves the supremum of every finite linear combination of evaluations. For a kernel k_ν representing $\mathcal{H}_\nu$, symmetry of its ball gives

$$\widetilde{c}_\nu = c_\nu, \qquad u_{\nu,\pm}(x,y) = R_\nu \|k_\nu(x,\cdot) \pm k_\nu(y,\cdot)\|_{\mathcal{H}_\nu}.$$

Polarization therefore shows that

$$\ell_\nu(x,y) = R_\nu^2 k_\nu(x,y), \qquad |c_\nu(x)| + \sqrt{\ell_\nu(x,x)} \le B_{\mathcal{G}}. \tag{43}$$

These statements also hold when $R_\nu = 0$. Thus the recovered ℓ_ν is a Borel positive semidefinite kernel, whether or not a kernel was supplied separately.

If $N < 16B_0^2/v_-$, define $\delta = 0$. Its loss is at most three, which is absorbed by $C_\rho q$. Otherwise form the matrices

$$\begin{aligned} H_\nu &= [\ell_\nu(X_i, X_j)/N]_{i,j}, \qquad C_X = [c_{\mu,X}, c_{\pi,X}], \\ M_X &= I_N - C_X(C_X'C_X)^\dagger C_X', \qquad H = M_X(H_\mu + H_\pi)M_X, \\ \lambda_N &= \frac{16B_{\mathcal{G}}^2 B_0^2}{N v_-}, \qquad A_X = M_X(I_N + H/\lambda_N)^{-1} M_X. \end{aligned} \tag{44}$$

Here † denotes the Moore–Penrose inverse. The center projection is Borel, for example by taking the limit of $C_X(C_X'C_X + k^{-1}I_2)^{-1}C_X'$ as $k \to \infty$. All other operations are Borel and the displayed ordinary inverse always exists. For $D_A = T'A_XT/N$, the rule is

$$\delta = \Pi_{[-3,3]}\left\{\frac{T'A_XY/N}{D_A \vee (v_-/4)}\right\}. \tag{45}$$

It is defined on every sample, including singular designs and zero kernels. In particular, it does not estimate the approximation errors.

Since $H = M_XHM_X$, the matrices M_X and H commute. Hence A_X is symmetric, $0 \preceq A_X \preceq I_N$, and it annihilates both center vectors. The envelope in (43) gives $\operatorname{tr} H \le 2B_{\mathcal{G}}^2$, so, on every sample,

$$\begin{aligned} \operatorname{tr}(I_N - A_X) &= \operatorname{rank}(I_N - M_X) + \operatorname{tr}\{H(H + \lambda_N I_N)^{-1}\} \\ &\le 2 + 2B_{\mathcal{G}}^2/\lambda_N \le N v_-/(4B_0^2). \end{aligned} \tag{46}$$

For either side,

$$A_XH_\nu A_X = A_XM_XH_\nu M_XA_X \preceq A_XHA_X \preceq (\lambda_N/4)I_N, \tag{47}$$

where the last inequality is $x/(1+x)^2 \le 1/4$ for $x \ge 0$. The evaluation operator $f \mapsto R_\nu f_X/\sqrt{N}$ has Gram matrix H_ν. Because $A_Xc_{\nu,X} = 0$, it follows that, with $e_N = \sqrt{\lambda_N}/2$,

$$\sup_{g \in \mathcal{G}_\mu} \|A_Xg_X\|_N \le e_N, \qquad \sup_{h \in \mathcal{G}_\pi} \|A_Xh_X\|_N \le e_N. \tag{48}$$

No population covariance inverse or eigenvalue-decay condition enters these inequalities.

For the risk calculation, first take fixed full-class comparators g, h with $\|\mu_0 - g\|_2 \le a$ and $\|\pi_0 - h\|_2 \le b$. Write $r_\mu = \mu_0 - g$, $r_\pi = \pi_0 - h$ and suppress the subscript on A_X. Expanding $\pi_{0,X}'A\mu_{0,X}$ into the four comparator and remainder terms, empirical Cauchy–Schwarz followed by expectation Cauchy–Schwarz and (48) gives

$$\mathbb{E}|\pi_{0,X}'A\mu_{0,X}/N| \le ab + (a+b)e_N + B_{\mathcal{G}}e_N. \tag{49}$$

For example, the public–public term is at most $\|h_X\|_N\|Ag_X\|_N \le B_{\mathcal{G}}e_N$; the remainder–remainder term has expectation at most ab.

Conditional on $X_{1:N}$, the u_i are independent and centered. Conditional on $(X_{1:N}, T_{1:N})$, the ε_i are likewise independent and centered. Boundedness and contraction therefore imply

$$\mathbb{E}\{(u'A\mu_{0,X}/N)^2 \mid X_{1:N}\} \le B_0^4/N, \qquad \mathbb{E}\{(T'A\varepsilon/N)^2 \mid X_{1:N}, T_{1:N}\} \le B_0^4/N.$$

Thus $Q_A = T'A(Y - \beta_0 T)/N$ satisfies

$$\mathbb{E}|Q_A| \le ab + (a + b + B_{\mathcal{G}})e_N + 2B_0^2/\sqrt{N}. \tag{50}$$

It remains to control the denominator under only average residual variance. Put $v_i = \mathbb{E}(u_i^2 \mid X_i)$, $\overline{v} = N^{-1}\sum_i v_i$, and $\Omega = \mathrm{diag}(v_i)$. Since $0 \le v_i \le B_0^2$, (46) yields

$$\mathbb{E}(D_A \mid X_{1:N}) = \{\pi_{0,X}' A \pi_{0,X} + \mathrm{tr}(A\Omega)\}/N \ge \overline{v} - v_-/4.$$

Moreover $\mathbb{E}\overline{v} = v \ge v_-$, so $\mathbb{P}(\overline{v} < 3v_-/4) \le 16B_0^4/(Nv_-^2)$. Conditional independence and centering give

$$\mathrm{Var}(u'Au \mid X_{1:N}) = \sum_i A_{ii}^2 \mathrm{Var}(u_i^2 \mid X_i) + 4\sum_{i<j} A_{ij}^2 v_i v_j \le 2B_0^4 \,\mathrm{tr}(A^2) \le 2B_0^4 N,$$

$$\mathrm{Var}(2u'A\pi_{0,X} \mid X_{1:N}) \le 4B_0^4 N.$$

Using $\mathrm{Var}(U+V) \le 2\,\mathrm{Var}\,U + 2\,\mathrm{Var}\,V$ therefore gives $\mathrm{Var}(D_A \mid X_{1:N}) \le 12B_0^4/N$; no assumption about third moments being zero is needed. On $\{\overline{v} \ge 3v_-/4\}$ the conditional mean is at least $v_-/2$. Another Chebyshev bound consequently proves

$$\mathbb{P}(D_A < v_-/4) \le 208B_0^4/(Nv_-^2). \tag{51}$$

On the complementary event the floor is inactive and clipping cannot increase the error, so $|\delta - \beta_0| \le 4|Q_A|/v_-$. On the bad event the clipped loss is at most six. Combining (50)–(51) gives the explicit finite-sample bound

$$\mathbb{E}|\delta - \beta_0| \le \frac{4}{v_-}\left\{ab + (a + b + B_{\mathcal{G}})e_N + \frac{2B_0^2}{\sqrt{N}}\right\} + \frac{1248B_0^4}{Nv_-^2}. \tag{52}$$

For unattained approximation infima, use fixed comparators with errors $a + \eta, b + \eta$ throughout and let $\eta \downarrow 0$; neither comparator is an algorithm input. Finally $e_N = 2B_{\mathcal{G}}B_0/\sqrt{Nv_-}$, $a, b \le 1$, and $ab \le W$. This proves (41), including every zero-budget face. □

For a finite implementation, let the centers c_ν and features $\phi_\nu : \mathcal{X} \to \mathbb{R}^{d_\nu}$ be explicitly evaluable Borel functions, and let Q_ν be a public positive semidefinite matrix. The full ellipsoid

$$\mathcal{G}_\nu = c_\nu + \{\phi_\nu' Q_\nu^{1/2}\theta : \|\theta\|_2 \le 1\}$$

has kernel $\ell_\nu(x, y) = \phi_\nu(x)'Q_\nu\phi_\nu(y)$, and its envelope condition is exactly $|c_\nu(x)| + \sqrt{\phi_\nu(x)'Q_\nu\phi_\nu(x)} \le B_{\mathcal{G}}$. The rule then uses finite matrix operations, even if $d_\nu > N$ or the features are linearly dependent. Arbitrary overlap within and between the two feature maps is permitted. Bounded-diagonal RKHS balls give the corresponding infinite-rank examples, without any spectral-decay assumption. The same upper proof also applies to classes merely *contained* in such balls if the enclosing centers and kernels, with the displayed diagonal bound, are supplied as additional public geometry; the recovery in (42) is asserted only for the full balls. An arbitrary bounded ℓ_1 dictionary does not automatically have this Hilbert-ball envelope.

The construction uses classical penalized partialling-out: the centers are unpenalized nuisance directions and H supplies the remaining ridge smoother. Partial smoothing splines and partially linear kernel ridge regression are established methods (Heckman, 1986; Zhao et al., 2016); series residualization also uses product approximation-bias arguments (Cattaneo et al., 2018). We give these methods a bounded finite-sample guarantee with unknown approximation errors and recover the required geometry from the represented full balls.

For completeness, restrict the outer public-pair supremum in (3) to the full affine Hilbert balls above, including radius zero. On the strict lower-endpoint slack stratum $0 < v_- < B_0^2$, its minimax risk is $\asymp_\rho q + ab$. The upper bound is Theorem 24. For the lower bound, Appendices C.3 and E both use the identical singleton pair $\mathcal{G}_\mu = \mathcal{G}_\pi = \{0\}$, with zero difference complexities and repeated-zero M1 cores. Their bounds (C.9) and (E.13) therefore hold in this restricted outer class for every s, t; $\max(c_{q,\rho}q, c_{ab,\rho}ab) \ge \frac{1}{2}\min(c_{q,\rho}, c_{ab,\rho})(q + ab)$ gives the claim. The zero-radius pair thus attains the lower bound for the restricted outer risk.

SADE also achieves budget-free adaptation: Theorem 2.1 and Remark 2.1 of Gu et al. (2026) give a fixed-outcome-class high-probability bound of order $\sqrt{\tau/n} + a + s^2 + \lambda$, without those budget inputs. At fixed tail parameter and $\lambda \lesssim q$, its leading scale is $q + a + s^2$. Our result instead controls expected-L_1 error for two full Hilbert classes at rate $q + ab$.

O CONVEX DICTIONARY CLASSES AND UNIONS OF HILBERT BALLS

This appendix sharpens the class-geometry comparison in Proposition 7. Throughout, all public classes retain the model, pointwise envelope, certificates, and sequential-M1 interface of Appendix A. The general upper bound applies to every such pair. Appendix Q provides the budget-free upper bound; the closure argument below preserves the matching lower bound.

O.1 TWO RESTRICTED OUTER MINIMAX FIELDS

A finite dictionary ℓ_1 ball is a full class

$$\mathcal{K}(d_1, \dots, d_k) = \left\{ \sum_{j=1}^{k} \theta_j d_j : \sum_{j=1}^{k} |\theta_j| \le 1 \right\},$$

where the d_j are bounded Borel functions and the whole class satisfies the candidate envelope. We allow $k = 0$, giving $\{0\}$. For the witnesses below, all nonzero generators can be chosen linearly independent, so this is also an injective image of a finite-dimensional ℓ_1 unit ball. Define a segment-union class by

$$\mathcal{U}(d_1, \dots, d_k) = \bigcup_{j=1}^{k} \{\tau d_j : |\tau| \le 1\},$$

with the same zero-class convention. Each nonzero segment is the full unit ball of the rank-one kernel $d_j(x)d_j(y)$; the zero class is a degenerate ball. Thus these are finite unions of full centered RKHS balls of rank at most one, with no uniform bound on their number.

Let $\mathcal{R}_{\ell_1}$ and $\mathcal{R}_{\mathrm{union}}$ be the risk (3) with its outer public-pair supremum restricted, respectively, to pairs of these two forms. The representations are part of the public instance as before.

Theorem 25 (The general rate persists under convexification and finite unions)**.** *For fixed ρ with $0 < v_- < B_0^2$, every $n \ge 1$ and every $(a, s, b, t) \in [0, 1]^4$,*

$$\mathcal{R}_{\ell_1} \asymp_\rho q + X, \qquad \mathcal{R}_{\mathrm{union}} \asymp_\rho q + X.$$

For either field, one public pair and one covariate marginal in each budget cell support the five component lower inequalities of Proposition 18. The same conclusions hold with every preassigned atomless marginal as in Proposition 19, with constants independent of that marginal. The dictionaries may depend on n and the budget cell. At $v_- = B_0^2$ both restricted risks have rate q; above that boundary the compatible-law field is empty.

Proof. The general upper of Proposition 3 applies to every fixed represented pair in either restriction. For the lower, denote the finite symmetric dictionaries of Appendix L by D_μ, D_π; both contain zero. Replace each D_ν either by $K_\nu = \mathrm{conv}(D_\nu) = \mathrm{absconv}(D_\nu)$ or by $U_\nu = \bigcup_{d \in D_\nu} [-d, d]$. Then

$$D_\nu \subset U_\nu \subset K_\nu.$$

A redundant smaller P-band amplitude can be discarded from the generators: outcome needs only $\kappa_P \phi_j$ on $\mathcal{J}_P$ and $\eta_M \phi_j$ on $\mathcal{J}_M$; propensity needs $\gamma_O \phi_j$ on $\mathcal{J}_O$ and $\gamma_M \phi_j$ on $\mathcal{J}_M$. All allocated frequencies are distinct and orthogonal, giving the claimed independence.

Representation and envelope. Enumerating rational coefficient vectors in the ℓ_1 unit ball gives a countable core for K_ν, and enumerating rational parameters on its finitely many segments gives one for U_ν. For the zero class, repeat zero. Coefficient convergence implies pointwise-everywhere convergence, so both cores satisfy sequential M1. Every convex combination of D_ν obeys its same pointwise envelope $H_\nu \le B_{\mathcal{G}}$. These representations are fixed before Nature chooses a law or hard label.

Full-class complexity. For every linear functional L on this finite span,

$$\Delta K_\nu = \mathrm{conv}(D_\nu - D_\nu), \qquad \sup_{f \in \Delta K_\nu} |Lf| = \max_{d, d' \in D_\nu} |L(d - d')|. \tag{53}$$

This is the standard convex-hull invariance of global Rademacher complexity (Bartlett & Mendelson, 2002, Theorem 12(2)): a convex combination cannot exceed the largest absolute functional value, and the extreme generating differences are included. In particular, for every positive radius r the full localized supremum for ΔK_ν, and hence for ΔU_ν, is bounded by the finite measurable *global* supremum on the right of (53), with $Lf = n^{-1}\sum_i \epsilon_i f(X_i)$. Monotonicity of outer expectation supplies the same bound on localized complexity.

The calculations in Appendix L already apply Massart's inequality to exactly this global finite difference class, including cross-band differences. Their constants, radii, and all zero-budget cases therefore remain valid without change. No equality of *localized* complexities is used: localization need not commute with convexification.

Approximation and compatible laws. Every original in-class nuisance truth remains in both enlarged classes. Every off-class truth $c\phi_j$ is orthogonal to $\mathrm{span}(D_\nu)$ on that side, because its frequency is absent from that dictionary. For every $f \in K_\nu$,

$$\|c\phi_j - f\|_2^2 = c^2 + \|f\|_2^2, \qquad \inf_{f\in K_\nu} \|c\phi_j - f\|_2 = \inf_{f\in U_\nu} \|c\phi_j - f\|_2 = |c|. \tag{54}$$

The equalities use $0 \in U_\nu$. Thus the P propensity, O outcome, and both ab approximation distances are unchanged; q and M have their original represented truths.

Consequently all the original raw laws, including the common scalar family, remain compatible with each new pair. Neither the observation laws nor their mixture distances change. The testing inequality applies to any Borel rule knowing the whole new pair, since its public evaluations are fixed functions of the observations and reveal no hidden label. The component inequalities and inactive-component q splice can therefore be combined inside the same new pair exactly as in Appendix L.

The fixed-marginal construction in Appendix L.6 uses the same finite orthogonal bands under the preassigned atomless law. The support-function identity is pointwise, and (54) uses only that preserved orthogonality. Hence the same enlargement proves its two restricted versions, without adding marginal-dependent constants.

Finally, the saturation upper is pairwise and the zero pair belongs to both restrictions. Appendix H thus supplies the matching q lower at saturation and emptiness above it. □

Convexification and finite unions preserve the hard families and hence the general minimax rate. The comparison allows the dictionary size or number of components to grow with the sample size and budgets. Characterizing the dependence on a separately fixed number of components requires a finer analysis; the lower bound here concerns the worst dictionary in the stated class.

O.2 The envelope cost of an enclosing Hilbert ball

The following elementary geometric calculation explains why replacing a dictionary by an enclosing Hilbert ball can change the fixed regularity field.

Lemma 26 (Exact enclosure cost for orthogonal sign dictionaries)**.** *Let $\phi_1, \dots, \phi_p$ be Borel functions orthonormal in $L_2(\nu)$, with $|\phi_j(x)| \le 1$ everywhere and $|\phi_j| = 1$ ν-almost surely, and let $h_j \ge 0$. Suppose a full affine RKHS ball $c + RB_{\mathcal{H}}$ contains $\{0, \pm h_j\phi_j : 1 \le j \le p\}$. Assume its kernel k is Borel, put $\ell = R^2 k$, and suppose $\int \ell(x,x)\, d\nu(x) < \infty$. Then*

$$\int \ell(x,x)\, d\nu(x) \ge \sum_{j=1}^p h_j^2, \qquad \sup_x \sqrt{\ell(x,x)} \ge \left(\sum_{j=1}^p h_j^2\right)^{1/2}. \tag{55}$$

The normalized kernel $\ell_0(x,y) = \sum_j h_j^2 \phi_j(x)\phi_j(y)$ and zero center attain both bounds. For equal amplitudes h, the smallest enclosing Hilbert pointwise radius is $h\sqrt{p}$, whereas the dictionary and its ℓ_1 hull have pointwise envelope h.

Proof. If $R = 0$, containment of zero makes $c = 0$, and containment of the signed functions forces every $h_j = 0$; the claim is immediate. Otherwise use the normalized RKHS $\mathcal{H}_\ell$, whose unit ball is

$RB_{\mathcal{H}}$. The half-difference of the two residuals $h_j\phi_j - c$ and $-h_j\phi_j - c$ shows that $h_j\phi_j$ belongs to this unit ball, regardless of the center.

Every function in the RKHS of a Borel kernel is Borel: each function is the norm limit of a sequence of finite linear combinations of kernel sections, hence a pointwise limit. The diagonal integral makes the embedding into $L_2(\nu)$ bounded. Let r_j be the Riesz representer of the continuous linear functional $f \mapsto \int \phi_j f \, d\nu$ on $\mathcal{H}_\ell$. Testing at $h_j\phi_j$ gives $\|r_j\|_{\mathcal{H}_\ell} \geq h_j$. Choose a finite orthonormal basis $e_1, \ldots, e_m$ for $\operatorname{span}\{r_1, \ldots, r_p\}$. Two finite Bessel inequalities give

$$\sum_j \|r_j\|^2_{\mathcal{H}_\ell} = \sum_{a=1}^{m} \sum_j \left| \int \phi_j e_a \, d\nu \right|^2 \leq \int \sum_{a=1}^{m} e_a(x)^2 \, d\nu(x) \leq \int \ell(x,x) \, d\nu(x).$$

No separability of the whole RKHS or vector-valued integral is needed. This proves (55).

For ℓ_0, the nonzero functions $h_j\phi_j$ are unit-norm orthogonal kernel features. Its unit ball contains the signed dictionary and, by convexity, its whole ℓ_1 hull. Its diagonal is at most $\sum_j h_j^2$ everywhere and equals that value ν-almost surely, proving sharpness, including zero amplitudes. □

Thus the full-ball envelope of Appendix N is a substantive geometric condition. For the active M family, the number of equal-amplitude orthogonal features is exponential in $n(s \wedge t)^2$. A fixed dictionary envelope and a legal localized-complexity certificate do not supply an enclosing Hilbert ball with the same fixed $B_{\mathcal{G}}$: for fixed $s \wedge t > 0$ and sufficiently large n, the M-band sum of squared amplitudes exceeds $B_{\mathcal{G}}^2$.

P Adaptation to unknown approximation budgets

We construct a single estimator that adapts to both approximation budgets in the represented-pair model of Appendix A. Here the stochastic certificates s, t remain known; Appendix Q subsequently removes those inputs. The construction uses double calibration with a scalar correction direction, following Liu et al. (2023). We establish the required bounds for arbitrary M1 classes under a positive average residual variance. A represented propensity candidate is restricted to an empirical neighborhood of the initial fit, and a free vector is chosen by constrained minimum distance.

Theorem 27 (Adaptation to both approximation budgets)**.** *Fix (Γ, ρ, n, s, t) as in Definition 9. There is one Borel rule $\delta_{\Gamma,\rho,n,s,t} : \mathcal{Z}^{2n} \to [-3,3]$ such that, simultaneously for every $(a,b) \in [0,1]^2$ and $P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$,*

$$\begin{aligned} \mathbb{E}_P|\delta_{\Gamma,\rho,n,s,t} - \beta(P)| &\leq C_\rho \left[q + \min\{a(b+t) + a^2 + s^2, \; b(a+s) + (b+t)^2\} \right] \\ &\leq C_\rho(q + W). \end{aligned} \tag{56}$$

The constant is independent of Γ, n, s, t, a, b, P. The rule does not use a, b, a population comparator, or the covariate law. It uses the known s, t. Theorem 29 supplies the separate extension that omits all four certificates.

Throughout the proof, P_1, P_2 denote the empirical averages on the two original halves $\mathcal{D}_1, \mathcal{D}_2$, each of size n. Write $f_X = (f(X_i) : i \in \mathcal{D}_1)$ and $\|w\|_n^2 = n^{-1} \sum_{i \in \mathcal{D}_1} w_i^2$. A product inside P_1 can involve such vectors. Put

$$q = n^{-1/2}, \qquad h^2 = \frac{\log(en)}{n}, \qquad \tau = n^{-12}, \qquad \zeta = n^{-6}. \tag{57}$$

In particular $h^2 \leq 2q$, and $\zeta, \sqrt{\tau} \leq q$. Every unspecified constant below depends only on ρ.

P.1 Initial fits and the scalar pilot

The initial treatment fit on a sample is the first represented τ-approximate minimizer of the empirical squared loss for T. The initial outcome fit is the represented component of the first τ-approximate minimizer of the squared loss for $Y - \gamma T - g$, over represented g and $\gamma \in \mathbb{Q} \cap [-3,3]$. These objectives use neither approximation budget. The fits on the whole of $\mathcal{D}_2$ are denoted by $\widehat{\pi}, \widehat{\mu}$.

The initial-fit bounds (B.7) and (B.9) in Appendix B.2, used at $z = 4\log(en)$, imply

$$\begin{aligned}\mathbb{E}\|\widehat{\pi} - \pi_0\|_2^2 &\le C_\rho(b^2 + t^2 + h^2),\\ \mathbb{E}\|\widehat{\mu} - \mu_0\|_2^2 &\le C_\rho(a^2 + s^2 + h^2).\end{aligned} \tag{58}$$

Indeed their squared bounds at this one confidence level have these orders, their failures are $O_\rho(n^{-4})$, and the fitted functions and the true functions have fixed envelopes. The contribution of the complements is therefore $O_\rho(n^{-4}) \le C_\rho h^2$. A single confidence level therefore suffices for these second-moment bounds.

For the scalar pilot alone, split $\mathcal{D}_2$ deterministically into an internal training block of size $k = \lfloor n/2 \rfloor$ and a validation block of size $m = n - k$. Fit $g^{\mathrm{tr}}, h^{\mathrm{tr}}$ on the internal training block by the preceding rules. Let P_{va} be the internal validation average, and define

$$\begin{aligned}\widehat{D}_{\mathrm{va}} &= P_{\mathrm{va}}T(T - h^{\mathrm{tr}}), & \widehat{A}_{\mathrm{va}} &= P_{\mathrm{va}}(T - h^{\mathrm{tr}})(Y - g^{\mathrm{tr}}),\\ \widehat{V}_{\mathrm{va}} &= P_{\mathrm{va}}T^2, & \widehat{B}_{\mathrm{va}} &= P_{\mathrm{va}}T(Y - g^{\mathrm{tr}}).\end{aligned}$$

With $\Pi = \Pi_{[-3,3]}$, the pilot is

$$\bar{\beta} = \begin{cases}\Pi(\widehat{A}_{\mathrm{va}}/\widehat{D}_{\mathrm{va}}), & \widehat{D}_{\mathrm{va}} \ge v_-/2,\\ \Pi\{\widehat{B}_{\mathrm{va}}/(\widehat{V}_{\mathrm{va}} \vee (v_-/2))\}, & \widehat{D}_{\mathrm{va}} < v_-/2.\end{cases} \tag{59}$$

It is $\mathcal{D}_2$-measurable. The full-half fits may reuse these data.

Lemma 28 (Two second-moment bounds for one pilot)**.** *For n larger than a constant depending only on ρ, the pilot* (59) *satisfies, with $c = \beta_0 - \bar{\beta}$,*

$$|c| \le 6, \qquad \mathbb{E}c^2 \le C_\rho \min\{a^2 + s^2 + h^2,\ b^2 + t^2 + h^2\}. \tag{60}$$

Proof. First take two fixed bounded anchors g, h_0 and a fresh sample of size m. Let $a_0 = \|\mu_0 - g\|_2$, $b_0 = \|\pi_0 - h_0\|_2$. Write D, A, V, B for the population counterparts of the four empirical moments in (59). Conditional centering gives

$$\begin{aligned}D &= v + \mathbb{E}\{\pi_0(\pi_0 - h_0)\}, & A &= \beta_0 D + \mathbb{E}\{(\pi_0 - h_0)(\mu_0 - g)\},\\ V &\ge v_-, & B &= \beta_0 V + \mathbb{E}\{T(\mu_0 - g)\}.\end{aligned}$$

On the primary branch the clipped error is at most

$$\frac{2}{v_-}\{|\widehat{A} - A| + 3|\widehat{D} - D| + a_0 b_0\}.$$

On the fallback it is at most

$$\frac{2}{v_-}\{|\widehat{B} - B| + 3|\widehat{V} - V| + B_0 a_0\},$$

because $|V - (\widehat{V} \vee (v_-/2))| \le |V - \widehat{V}|$. Every centered empirical moment has second moment at most C_ρ/m. If $D < 3v_-/4$, then $B_0 b_0 \ge v_-/4$, so the squared fallback bias is bounded by $C_\rho a_0^2 b_0^2$. If $D \ge 3v_-/4$, Chebyshev gives $\Pr(\widehat{D} < v_-/2) \le C_\rho/m$; clipped squared loss on this event is at most 36. Squaring the pointwise branch inequalities thus proves

$$\mathbb{E}|\delta_{g,h_0} - \beta_0|^2 \le C_\rho\{m^{-1} + a_0^2 b_0^2\}. \tag{61}$$

No independence of the branch selection and empirical moments is required.

Condition now on the internal training block and insert its anchors. Both anchor errors are bounded by $B_0 + B_{\mathcal{G}}$, whence $a_0^2 b_0^2 \le (B_0 + B_{\mathcal{G}})^2 \min(a_0^2, b_0^2)$. The initial-fit moment bounds at size k, proved at the single confidence level $4\log(ek)$ exactly as above, complete the proof of (60) after averaging (61).

The original certificates also control the size-k fits. For a countable localized difference core, adding independent signed observations and using conditional Jensen gives

$$\mathfrak{R}_{k,P_X}(r; \Delta\mathcal{G}^{\mathrm{rep}}) \le \frac{n}{k}\mathfrak{R}_{n,P_X}(r; \Delta\mathcal{G}^{\mathrm{rep}}).$$

For $n \ge 4$, $n/k \le 3$, so the original core inequalities imply the size-k inequalities with a changed fixed process constant. The joint-fit localization in Appendix B.2 then applies with the same changed constant. Moreover $\log(ek)/k \le Ch^2$ and $m^{-1} \le Ch^2$. The tolerance $\tau = n^{-12} \le k^{-12}$ is no larger than the one permitted there. All comparisons above use the measurable countable-core suprema. □

P.2 Comparison costs and the order of fixed constants

For a fixed compatible law choose a represented propensity comparator h_0 with $\|\pi_0 - h_0\|_2 \le b + \zeta$. Its existence follows from Lemma 13; it is only a proof device. Define

$$B_{\mathrm{cmp}} = P_1(h_0 - \widehat{\pi})^2.$$

There is a fixed C_{cmp}, depending only on ρ, such that

$$\Pr\{B_{\mathrm{cmp}} > C_{\mathrm{cmp}}((b+t)^2 + h^2)\} \le C_\rho n^{-4}, \qquad \mathbb{E}B_{\mathrm{cmp}} \le C_\rho((b+t)^2 + h^2). \tag{62}$$

To prove the transfer bound, proceed as follows. The $\mathcal{D}_2$ treatment-fit event at $4\log(en)$ has $\|\widehat{\pi} - \pi_0\|_2^2 \le C_\rho(b^2 + t^2 + h^2)$. Given $\mathcal{D}_2$, the function $f = \widehat{\pi} - h_0$ is fixed with envelope $2B_{\mathcal{G}}$. Conditional scalar Bernstein, Lemma 10, gives $P_1 f^2 \le 2\|f\|_2^2 + C_\rho h^2$, except with conditional probability $2(en)^{-4}$. Combine this with $\|f\|_2^2 \le 2\|\widehat{\pi} - \pi_0\|_2^2 + 2(b+\zeta)^2$. This proves the tail assertion because $\zeta \le h$. For its expectation assertion use $\mathbb{E}P_1 f^2 = \mathbb{E}\|f\|_2^2$ and (58). The selected $\mathcal{D}_2$ index is fixed under this conditioning, so no fitted main-half index is inserted into a pointwise empirical-process event.

Choose the constants in the following order. First fix the initial-fit constants, C_{cmp}, and the constants of Lemma 12 for the two difference cores and the bounded multipliers $u, \varepsilon, \mu_0, \pi_0$ used below. These constants depend only on their fixed envelopes and ρ. Take

$$R = 2(B_0 + B_{\mathcal{G}} + 1), \qquad M = 8B_0 + B_{\mathcal{G}} + 1, \qquad L = 8B_0.$$

Choose a fixed K large enough that the uniform true-benchmark process bounds in (68) and (71) below are at most half their displayed tolerances. Their proofs use only the process constants already fixed, the bound $|c| \le 6$, and quadratic absorption. Choose $K_A \ge 12C_{\mathrm{cmp}}$ and put

$$\kappa_\mu = K(s^2 + h^2), \qquad \kappa_\pi = K(t^2 + h^2), \qquad A_t = K_A(t^2 + h^2). \tag{63}$$

Finally choose $n_\rho \ge 4$ so that all the confidence levels $4\log(en)$, and their size-k counterparts, are in the process tail ranges, and the scalar denominator event below has its stated margin for $n \ge n_\rho$. This last choice changes none of the other constants. All these choices precede the budgets, pair, law, and data. For $n < n_\rho$ define the rule to be zero, whose loss is absorbed by $C_\rho q$. In what follows $n \ge n_\rho$.

P.3 The two calibration programs

Let

$$\mathcal{F}_X = \sigma(\mathcal{D}_2, X_i : i \in \mathcal{D}_1), \qquad \mathcal{F}_T = \sigma(\mathcal{F}_X, T_i : i \in \mathcal{D}_1).$$

The first program constructs an outcome correction without using the main-half treatments or outcomes. For an n-vector m define

$$H_\pi^{\mathrm{pil}}(m) = \sup_{f \in \Delta\mathcal{G}_\pi^{\mathrm{rep}}} \{|P_2[(Y - \bar{\beta}T)f] - P_1(mf)| - P_1 f^2\}.$$

Retain $m \in \mathbb{Q}^n$ satisfying

$$H_\pi^{\mathrm{pil}}(m) < \kappa_\pi, \qquad \|m\|_n < M. \tag{64}$$

If this set is nonempty, let $\widehat{\eta}$ be its first point whose $\|m - \widehat{\mu}_X\|_n^2$ is less than the feasible countable infimum plus τ. Otherwise output zero and set $\widehat{\eta} = 0$ for analytical vector conventions. Write $d = \widehat{\eta} - \widehat{\mu}_X$. Both vectors are $\mathcal{F}_X$-measurable.

For the second program write $(h_j)_{j \ge 1}$ for the supplied propensity core. Enumerate all pairs $(j, z) \in \mathbb{N} \times \mathbb{Q}^n$. For each pair set

$$\pi^c = (h_j)_X + z, \qquad r = T_X - \pi^c, \qquad D = P_1(Tr), \qquad B_j = P_1(h_j - \widehat{\pi})^2.$$

For $\nu \in \{\mu, \pi\}$ define $H_\nu(r) = \sup_{f \in \Delta\mathcal{G}_\nu^{\mathrm{rep}}}\{|P_1(rf)| - P_1 f^2\}$. Retain exactly the pairs satisfying

$$\begin{aligned} B_j < A_t, \qquad D > v_-/2, \qquad \|r\|_n < R, \qquad \|\pi^c\|_n < R, \\ H_\mu(r) < \kappa_\mu, \qquad H_\pi(r) < \kappa_\pi, \\ |P_1(rd)| < Lh\|d\|_n + \tau, \\ |P_1(r\pi^c)| < Lh\|\pi^c\|_n + \tau, \\ |P_1(rz)| < Lh\|z\|_n + \tau. \end{aligned} \tag{65}$$

Only the free-vector objective $\|z\|_n^2$ is minimized: choose the first retained pair whose objective is less than its countable feasible infimum plus τ. Denote the selected candidate function by $\widehat{h}$, and its other quantities by $\widehat{z}, \widehat{\pi}^c, \widehat{r}, \widehat{D}$. If either program is empty, output zero; define $\widehat{r} = 0$ there for the conditional noise arguments. Otherwise output

$$\delta = \Pi_{[-3,3]}\left[\bar{\beta} + \frac{P_1\{\widehat{r}(Y - \bar{\beta}T - \widehat{\eta})\}}{\widehat{D}}\right] = \Pi_{[-3,3]}\frac{P_1\{\widehat{r}(Y - \widehat{\eta})\}}{\widehat{D}}. \tag{66}$$

The denominator is positive on every feasible sample. The second equality follows directly from the denominator definition. Neither program uses a, b.

Each program is Borel. The cores are countable; all suprema in the constraints are countable and are continuous in the calibrated vector by the common difference envelope. All other constraints are continuous in the finite-dimensional vector, for a fixed represented index. Countable feasible-set decisions, countable infima, and first strict approximate minimizers are consequently Borel. The strict inequalities and the scalar $+\tau$ terms provide open neighborhoods even when a calibrated direction is zero. Thus a strictly feasible real vector can be approximated by retained rational vectors with objective converging to its own objective. No exact minimizer is required. The initial fits and the scalar pilot are Borel by their countable selection conventions and explicit ratios, respectively. In particular $\widehat{r}$ is $\mathcal{F}_T$-measurable: no main-half outcome is used to choose it.

P.4 Feasibility, margins, and two comparisons for the same program

Fix a law and represented comparators g, h_0 whose approximation errors are at most $a + \zeta, b + \zeta$, respectively. Write

$$\eta_0 = (\mu_0 + c\pi_0)_X, \quad r_\eta = \widehat{\eta} - \eta_0, \quad e_\pi = (\widehat{\pi} - \pi_0)_X, \quad r_\pi = (\pi_0 - h_0)_X.$$

The random function defining η_0 depends on $\mathcal{D}_2$ only through the bounded scalar c; its vector also uses the main covariates. Before inserting that scalar, form uniform process events using the exact identity

$$\begin{aligned} P_2[(Y - \bar{\beta}T)f] - P_1(\eta_0 f) = {} & P_2(\varepsilon f) + cP_2(uf) \\ & + (P_2 - P_1)(\mu_0 f) + c(P_2 - P_1)(\pi_0 f). \end{aligned} \tag{67}$$

Lemma 12, applied to the fixed bounded multipliers displayed here and to the square process, gives a finite intersection of events, of failure $O_\rho(n^{-4})$, on which the absolute value of the right side minus $P_1 f^2$ is bounded by

$$C_\rho(t + h)\|f\|_2 - \|f\|_2^2 + C_\rho(t^2 + h^2) \le C_\rho(t^2 + h^2)$$

uniformly over the propensity difference core. The chosen K therefore supplies the event

$$G_\eta = \{H_\pi^{\mathrm{pil}}(\eta_0) \le \kappa_\pi/2\} \in \mathcal{F}_X, \qquad \Pr(G_\eta^c) \le C_\rho n^{-4}. \tag{68}$$

Also $\|\eta_0\|_n \le 7B_0 < M$, with strict norm margin. On G_η the first program is consequently nonempty and its comparison with η_0 yields

$$\begin{aligned} \|d\|_n \le \|\eta_0 - \widehat{\mu}_X\|_n + \sqrt{\tau}, \qquad \|r_\eta\|_n \le 2\|\eta_0 - \widehat{\mu}_X\|_n + \sqrt{\tau}, \\ |P_1(r_\eta f)| \le 2P_1 f^2 + 3\kappa_\pi/2 \quad (f \in \Delta\mathcal{G}_\pi^{\mathrm{rep}}). \end{aligned} \tag{69}$$

The vector norms of d, r_η are bounded by fixed constants even outside this event, using the norm constraint and the fallback convention. Moreover (58), the first pilot moment bound, and independence of the main covariates give

$$\mathbb{E}1_{G_\eta}(\|d\|_n^2 + \|r_\eta\|_n^2) \le C_\rho(a^2 + s^2 + h^2). \tag{70}$$

Independence between $\bar{\beta}$ and the full-half fits is unnecessary. The same moment bound also holds without 1_{G_η}: both vector norms are uniformly bounded on its complement, whose probability is $O_\rho(n^{-4}) \le C_\rho h^2$.

The same process lemma and quadratic absorption, now with multiplier u on the main half, give

$$H_\mu(u_X) \le \kappa_\mu/2, \qquad H_\pi(u_X) \le \kappa_\pi/2 \tag{71}$$

except with probability $C_\rho n^{-4}$. For scalar constraints, conditionally on $\mathcal{F}_X$ and any of the four vectors

$$w = d, \quad w = (\pi_0)_X, \quad w = -e_\pi = (\pi_0 - \widehat{\pi})_X, \quad w = r_\pi,$$

weighted Hoeffding gives, at confidence $4\log(en)$,

$$|P_1(uw)| \le \sqrt{8} B_0 h \|w\|_n \le (L/2) h \|w\|_n, \tag{72}$$

with conditional failure at most $2(en)^{-4}$ for each vector. These statements also hold for the random vector $-e_\pi$: conditioning fixes the full-half fit and main covariates, and leaves the u_i independent, centered, and bounded. No coordinate bound on w is needed for weighted Hoeffding. If $w = 0$, the added $\tau > 0$ in the program still makes its scalar constraint strict.

Finally bounded scalar Bernstein and $\mathbb{E}(Tu) = v \ge v_-$ yield $P_1(Tu) > 3v_-/4$ except with probability $C_\rho n^{-4}$, by the choice of n_ρ. Let G be the finite intersection of G_η, the two events in (71), the four scalar events (72), this denominator event, and the comparison-cost event in (62). Then

$$G \in \mathcal{F}_T, \qquad \Pr(G^c) \le C_\rho n^{-4}. \tag{73}$$

All uniform class events precede every fitted-index substitution.

On G, consider the following two representations of the same real vector $(\pi_0)_X$ in the second program. First take the represented candidate $h_j = \widehat{\pi}$ and $z = -e_\pi$. Its comparison cost is zero, and $\pi^c = (\pi_0)_X$, $r = u_X$. All its constraints have the strict margins just proved, including $\|u\|_n, \|\pi_0\|_n \le B_0 < R$. Thus the second program is nonempty for every a, b.

For the other comparison suppose only in the proof that $b \le t + h$. Then

$$(b + t)^2 + h^2 \le (2t + h)^2 + h^2 \le 6(t^2 + h^2).$$

The cost event and $K_A \ge 12 C_{\mathrm{cmp}}$ give

$$B_{\mathrm{cmp}} \le A_t/2. \tag{74}$$

Take the fixed candidate $h_j = h_0$ and $z = r_\pi$. Again $\pi^c = (\pi_0)_X$, $r = u_X$, and every constraint is strictly feasible. In each comparison the represented index is a legitimate index of the program, and its real free vector can be approximated by rational vectors. Minimization of the same free-vector objective therefore gives, on G,

$$\begin{aligned} &\|\widehat{z}\|_n \le \|e_\pi\|_n + \sqrt{\tau} \quad \text{for all } b, \\ &\|\widehat{z}\|_n \le \|r_\pi\|_n + \sqrt{\tau} \quad \text{if } b \le t + h. \end{aligned} \tag{75}$$

Both comparisons concern the same program; the condition is used only to analyze its objective value.

Put $\Delta = \widehat{\pi}^c - (\pi_0)_X$. The empirical neighborhood constraint and the first line of (75) imply

$$\|\Delta\|_n \le \|e_\pi\|_n + \|(\widehat{h} - \widehat{\pi})_X\|_n + \|\widehat{z}\|_n \le 2\|e_\pi\|_n + \sqrt{A_t} + \sqrt{\tau}.$$

Consequently

$$\mathbb{E} 1_G \|\Delta\|_n^2 \le C_\rho (b^2 + t^2 + h^2). \tag{76}$$

On every feasible sample, $\|\widehat{z}\|_n \le \|\widehat{\pi}^c\|_n + \|\widehat{h}_X\|_n < R + B_{\mathcal{G}}$; this uniform bound is useful for scalar terms.

P.5 Two outcome analyses of the same rule

Conditional on $\mathcal{F}_T$, the ε_i are independent and centered, and $\|\widehat{r}\|_n < R$ on feasible samples. Thus

$$\mathbb{E}|P_1(\varepsilon \widehat{r})| \le C_\rho q. \tag{77}$$

Likewise conditioning on $\mathcal{F}_X$ gives $C_\rho q$ bounds for $P_1(u r_\eta)$, $P_1[u(\mu_0 - \widehat{\mu})]$, and $P_1[u(\mu_0 - g)]$. Their empirical coefficient norms are uniformly bounded. Good-event indicators can first be dropped from nonnegative absolute values. The denominator can be bounded below pointwise by $v_-/2$ before applying these bounds. These conditioning fields leave the main-half outcome noise centered.

On G the raw, unclipped output has the exact identity

$$\widehat{D}(\delta^{\mathrm{raw}} - \beta_0) = P_1(\varepsilon\widehat{r}) + P_1\{(\mu_0 - \widehat{\mu})\widehat{r}\} - P_1(d\widehat{r}). \tag{78}$$

Split $\mu_0 - \widehat{\mu} = (\mu_0 - g) + (g - \widehat{\mu})$ and use $\widehat{r} = u_X - \Delta$. The centered fixed-comparator term costs $C_\rho q$. Cauchy–Schwarz and (76) bound the other fixed-comparator term by $C_\rho(a + \zeta)(b + t + h)$ in expectation. The outcome-class constraint in (65), uniform before the selected represented index is inserted, gives

$$|P_1\{(g - \widehat{\mu})\widehat{r}\}| \le P_1(g - \widehat{\mu})^2 + \kappa_\mu.$$

Its expectation is at most $C_\rho(a^2 + s^2 + h^2)$ by the initial fit moment bound and the fixed-comparator error. The correction constraint and (70) give $\mathbb{E}1_G|P_1(d\widehat{r})| \le C_\rho h(a + s + h) + \tau$. Use $ah \le (a^2 + h^2)/2$, $sh \le (s^2 + h^2)/2$, and $h^2 \le 2q$. Clipping cannot increase the error, and clipped loss on G^c is at most six. We obtain

$$\mathbb{E}|\delta - \beta_0| \le C_\rho\{q + a(b + t) + a^2 + s^2\}. \tag{79}$$

Keeping $\mu_0 - \widehat{\mu}$ together instead in (78) gives a second bound for this very same rule. The centered term again costs $C_\rho q$; Cauchy–Schwarz bounds the Δ term by $C_\rho(a + s + h)(b + t + h)$, and the correction term by $C_\rho h(a + s + h) + \tau$. Therefore

$$\mathbb{E}|\delta - \beta_0| \le C_\rho\{q + (a + s + h)(b + t + 2h)\}. \tag{80}$$

The stochastic floor remains h in this bound.

P.6 The propensity analysis and the proof of the theorem

If $b > t + h$, the product bound already gives the required propensity rate: $b + t + 2h \le 4b$ and $bh \le (b^2 + h^2)/2$. Thus

$$\mathbb{E}|\delta - \beta_0| \le C_\rho\{q + b(a + s) + (b + t)^2\} \qquad (b > t + h). \tag{81}$$

It remains to analyze $b \le t + h$, where the second comparison in (75) is available.

Because $\mu_0 - \widehat{\eta} = -r_\eta - c\pi_0$, direct substitution in (66) gives

$$\widehat{D}(\delta^{\mathrm{raw}} - \beta_0) = P_1(\varepsilon\widehat{r}) - P_1(r_\eta\widehat{r}) - cP_1(\pi_0\widehat{r}). \tag{82}$$

Let $f_1 = (\widehat{h} - \widehat{\pi})_X$ and $f_0 = (\widehat{\pi} - h_0)_X$. Both are represented propensity differences, and $\Delta = f_1 + \widehat{z} + f_0 - r_\pi$. The two calibration inequalities in (69) give

$$\begin{aligned} |P_1(r_\eta\Delta)| \le{}& 2A_t + 2B_{\mathrm{cmp}} + 3\kappa_\pi \\ &+ \|r_\eta\|_n(\|\widehat{z}\|_n + \|r_\pi\|_n). \end{aligned} \tag{83}$$

The second comparison for $\widehat{z}$, (70), and $\mathbb{E}\|r_\pi\|_n^2 \le (b + \zeta)^2$ bound its good-event expectation by

$$C_\rho\{(b + t)^2 + h^2 + (b + \zeta)(a + s + h) + \sqrt{\tau}\}. \tag{84}$$

The term $P_1(r_\eta\widehat{r}) = P_1(r_\eta u) - P_1(r_\eta\Delta)$ has only an additional $C_\rho q$ contribution, by the conditional noise bound.

For the final term in (82), use $\pi_0 = \widehat{\pi}^c - \Delta$. The propensity difference constraints and the two scalar constraints for $\widehat{\pi}^c, \widehat{z}$ give, pointwise on G,

$$\begin{aligned} |P_1(\pi_0\widehat{r})| \le{}& Lh\|\widehat{\pi}^c\|_n + \tau + A_t + B_{\mathrm{cmp}} + 2\kappa_\pi \\ &+ Lh\|\widehat{z}\|_n + \tau + R\|r_\pi\|_n. \end{aligned} \tag{85}$$

Multiply by $|c|$. Bounded $|c|$, the deterministic cap, and (62) control the cap, comparator, and tolerance terms by $C_\rho((b + t)^2 + h^2)$ in expectation. Uniform norm bounds reduce the two terms proportional to h to $C_\rho h\mathbb{E}|c|$. The second bound on the one pilot in (60) gives

$$h\mathbb{E}|c| \le C_\rho h(b + t + h) \le C_\rho((b + t)^2 + h^2). \tag{86}$$

Its first bound, together with Cauchy–Schwarz, gives

$$\mathbb{E}|c|\|r_\pi\|_n \le C_\rho(b + \zeta)(a + s + h). \tag{87}$$

The pilot and fits may be dependent; these bounds do not condition their shared fitting data on a fitted coefficient.

Combine (82)– (87), divide by the imposed denominator bound, and use $bh \le (b^2+h^2)/2$, $h^2 \le 2q$, and $\zeta, \sqrt{\tau} \le q$. Clipping and (73) add at most $6\Pr(G^c) = O_\rho(n^{-4})$. It follows that

$$\mathbb{E}|\delta - \beta_0| \le C_\rho\{q + b(a+s) + (b+t)^2\} \qquad (b \le t+h). \tag{88}$$

Equations (81) and (88) bound the same estimator in the two budget regimes. Together with (79), they prove (56) for all $n \ge n_\rho$. The zero output covers smaller n, since its loss is at most three and $q \ge n_\rho^{-1/2}$. Positive h makes the tolerances and empirical neighborhood well defined when $s = 0$ or $t = 0$. Fixed ζ-comparators exist even if an approximation infimum is not attained, including when $a = 0$ or $b = 0$; their remaining terms were explicitly absorbed above. This proves Theorem 27. □

Q ADAPTATION TO ALL FOUR BUDGETS

This appendix removes the stochastic-budget inputs from Theorem 27. An independent block of covariates supplies two empirical local Rademacher fixed points. The construction follows the empirical fixed-point method of Bartlett et al. (2005). Two additional steps are needed for the present experiment: controlling the star hull of an arbitrary represented difference core without multiplying its declared radius by a logarithm, and transferring a resulting linear core certificate to the full outer complexity under sequential M1. Both steps apply to arbitrary M1 classes.

Theorem 29 (Adaptation to all four budgets)**.** *Fix a represented public instance Γ, the regularity tuple ρ, and $n \ge 1$. There is one Borel rule $\delta_{\Gamma,\rho,n} : \mathcal{Z}^{2n} \to [-3,3]$ such that, simultaneously for every $(a,s,b,t) \in [0,1]^4$ and $P \in \mathcal{P}_{\rho,n,a,s,b,t}(\Gamma)$,*

$$\begin{aligned}\mathbb{E}_P|\delta_{\Gamma,\rho,n} - \beta(P)| &\le C_\rho\left[q + \min\{a(b+t) + a^2 + s^2,\ b(a+s) + (b+t)^2\}\right] \\ &\le C_\rho(q+W).\end{aligned} \tag{89}$$

The rule uses Γ, ρ, n and the observations, but none of a, s, b, t, the covariate law, or a population comparator. The constant is independent of the public instance, the budgets, the law, and n. Only the average lower bound $\mathbb{E}u^2 \ge v_-$ is required.

Q.1 A STAR-HULL BOUND FOR AN ARBITRARY COUNTABLE CLASS

For a countable symmetric class $\mathcal{H}$ containing zero, write

$$\operatorname{star}(\mathcal{H}) = \{\alpha f : f \in \mathcal{H},\ 0 \le \alpha \le 1\}.$$

Every localized supremum over this class equals the corresponding supremum with $\alpha \in \mathbb{Q} \cap [0,1]$. Indeed, any permitted real coefficient can be approximated from below by rational coefficients, preserving either a population or an empirical upper norm constraint and converging in every finite-sample linear process. Squares and population quadratic means are also continuous in the scaling coefficient, so the same argument makes the square-process suprema and ball-inclusion events below measurable. We may use the real star hull for its scaling property and the countable rational star hull for measurability, without changing a supremum.

Lemma 30 (Cost of the star hull above the logarithmic floor)**.** *Let $\mathcal{H}$ be a countable symmetric class containing zero, with envelope one. Let $m \ge 1$, $h_m = \sqrt{\log(em)/m}$, and $s \ge 0$. Suppose that, for a fixed $\kappa \ge 0$,*

$$\mathfrak{R}_{m,P_X}(r;\mathcal{H}) \le \kappa s r, \qquad r \ge s, \quad r > 0.$$

There is a constant $C_\star$, depending only on κ, such that

$$\mathfrak{R}_{m,P_X}(r;\operatorname{star}(\mathcal{H})) \le C_\star(s + h_m)r, \qquad r \ge s \vee h_m. \tag{90}$$

No closure of $\mathcal{H}$ under scaling is assumed.

Proof. Fix $s \vee h_m \le r < 1$ and put

$$J = \lceil \log_2(1/r) \rceil, \qquad R_j = 2^j r, \qquad Z_j = \sup_{\substack{f \in \mathcal{H} \\ \|f\|_2 \le R_j}} \left| \frac{1}{m}\sum_{i=1}^m \epsilon_i f(X_i) \right|, \quad 0 \le j \le J.$$

If αf has population norm at most r, a function with $\|f\|_2 \le r$ contributes at most Z_0. A function with $2^{j-1}r < \|f\|_2 \le 2^j r$ has $\alpha \le 2^{1-j}$. Consequently, pointwise in the sample and signs, its star-hull supremum is at most

$$\max\{Z_0,\ \max_{1\le j\le J} 2^{1-j} Z_j\}. \tag{91}$$

Apply Theorem 2.1 of Bartlett et al. (2005) to the signed observations (X_i, ϵ_i) and the centered symmetric class $\{(x,e) \mapsto ef(x) : \|f\|_2 \le R_j\}$. Its range is $[-1,1]$, its variance is at most R_j^2, and its Rademacher mean is exactly $\mathfrak{R}_{m,P_X}(R_j;\mathcal{H})$: multiplying ϵ_i by a second independent Rademacher sign preserves the joint sign distribution. With the source theorem's tuning parameter equal to one, it follows that

$$\Pr\left\{ Z_j > 4\kappa s R_j + R_j\sqrt{\frac{2x_j}{m}} + \frac{8x_j}{3m} \right\} \le e^{-x_j}.$$

Take $x_j = y + 2j + \log 4$, for $y \ge 0$, and sum these failures. After multiplying the outer-shell bounds by 2^{1-j}, the maximum in (91) is at most

$$C\kappa s r + Cr\sqrt{\frac{y+J+1}{m}} + \frac{C(y+1)}{m}$$

except with probability at most e^{-y}. Here $\sup_{j\ge1} 2^{-j}(y+2j+\log 4) \le C(y+1)$. Integrating this tail gives

$$\mathfrak{R}_{m,P_X}(r;\mathrm{star}(\mathcal{H})) \le C\kappa s r + Cr\sqrt{\frac{J+1}{m}} + \frac{C}{m}.$$

Since $r \ge h_m$, we have $J+1 \le C\log(em)$ and $m^{-1} \le r h_m$. This proves (90) for $r < 1$. For $r \ge 1$ the localized core and star-hull suprema are their global suprema, which are equal; the retained condition $r \ge s \vee h_m$ allows the original certificate to prove the assertion directly. The argument includes $s = 0$. It takes the expectation of a maximum using concentration; summing the means of the shells would give an unnecessary logarithmic factor. □

Q.2 An empirical radius and its population certificate

For an iid calibration sample $X_1, \dots, X_m$, let Q_m denote its empirical average. Auxiliary Rademacher signs below are averaged exactly over their 2^m possible values; no additional randomization is part of the estimating rule.

Lemma 31 (Measurable calibration of a countable difference core)**.** *Let $\mathcal{F}$ be a countable symmetric class containing zero, with envelope at most L, where $L \ge 1$ is known. Suppose that*

$$\mathfrak{R}_{m,P_X}(u;\mathcal{F}) \le \kappa s u, \qquad u \ge s, \quad u > 0, \tag{92}$$

for some $s \ge 0$ and fixed κ. Set $\overline{\mathcal{F}} = \{f/L : f \in \mathcal{F}\}$, $x = 6\log(em)$, and, for $r > 0$, define

$$\widehat{\Phi}(r) = 2^{-m} \sum_{e\in\{-1,1\}^m} \sup_{f\in\overline{\mathcal{F}}} \left(1 \wedge \sqrt{\frac{2r}{Q_m f^2}}\right) \left|\frac{1}{m}\sum_{i=1}^m e_i f(X_i)\right|, \tag{93}$$

$$\widehat{\psi}(r) = 20\widehat{\Phi}(r) + \frac{31x}{m}. \tag{94}$$

A summand with $Q_m f^2 = 0$ is defined to be zero. Let

$$\widehat{r} = \inf\{r \in \mathbb{Q} : r > 0,\ \widehat{\psi}(r) \le r\}, \qquad \widehat{R} = L\sqrt{\widehat{r}}. \tag{95}$$

This defines a finite positive Borel statistic, using neither s nor P_X. With probability at least $1 - 4e^{-x}$, simultaneously,

$$\mathfrak{R}_{m,P_X}(u;\mathcal{F}) \le \frac{\widehat{R}}{10}\,u \qquad \text{for every } u \ge \widehat{R}, \tag{96}$$

$$\widehat{R} \le C_{L,\kappa}(s + h_m). \tag{97}$$

Moreover $\widehat{r} \ge 31x/m$ on every sample.

Proof. Write $\mathcal{S} = \operatorname{star}(\overline{\mathcal{F}})$. The expression in (93) is exactly

$$\widehat{\Phi}(r) = \mathbb{E}_\epsilon \sup_{\substack{f \in \mathcal{S} \\ Q_m f^2 \le 2r}} \left| \frac{1}{m} \sum_i \epsilon_i f(X_i) \right|.$$

It is Borel for each $r > 0$, by its finite sign average and countable supremum. Each term before taking the supremum, as a function of $\sqrt{r}$, is Lipschitz with constant at most $\sqrt{2}$: when $Q_m f^2 > 0$, divide its linear process by $\sqrt{Q_m f^2}$ and use Cauchy–Schwarz. Thus $\widehat{\Phi}$ is continuous on $(0, \infty)$. It is nondecreasing and $\widehat{\Phi}(r)/\sqrt{r}$ is nonincreasing. The same properties hold for $\widehat{\psi}$, which is positive because of its deterministic additive term. It follows that $\widehat{\psi}$ has a unique positive fixed point and that $\widehat{\psi}(r) \le r$ exactly when r is at least this fixed point. In particular (95) equals the fixed point and is Borel. The bounds

$$\frac{31x}{m} \le \widehat{r} \le 20 + \frac{31x}{m}$$

follow from $0 \le \widehat{\Phi}(r) \le 1$. These observations also handle functions with zero empirical norm.

For the lower comparison, define the nonrandom function

$$\Phi(r) = \mathfrak{R}_{m, P_X}(\sqrt{r}; \mathcal{S}), \qquad \psi_1(r) = 10\Phi(r) + \frac{11x}{m},$$

and let r_1 be the unique positive fixed point of ψ_1. Scaling in $\mathcal{S}$ shows directly that Φ is nondecreasing and $\Phi(r)/\sqrt{r}$ is nonincreasing. These properties also give continuity on $(0, \infty)$ and the asserted fixed-point properties of ψ_1. We use three results of Bartlett et al. (2005): Corollary 2.2, Lemma 3.6, and Lemma A.4, always with envelope one. Their suprema are measurable by the rational-star observation above.

Corollary 2.2 at the fixed, population-dependent radius r_1 gives

$$\{f \in \mathcal{S} : P f^2 \le r_1\} \subseteq \{f \in \mathcal{S} : Q_m f^2 \le 2r_1\}$$

except with probability e^{-x}, because $r_1 = 10\Phi(r_1) + 11x/m$. Apply the lower comparison in Lemma A.4 with its tuning parameter $1/2$ to the fixed class on the left. Its range is $[-1, 1]$, so, except with probability e^{-x},

$$\Phi(r_1) \le 2\mathbb{E}_\epsilon \sup_{\substack{f \in \mathcal{S} \\ P f^2 \le r_1}} \left| \frac{1}{m} \sum_i \epsilon_i f(X_i) \right| + \frac{2x}{m}.$$

Symmetry makes the absolute supremum equal to the source's one-sided supremum. On these two events,

$$r_1 = 10\Phi(r_1) + \frac{11x}{m} \le 20\widehat{\Phi}(r_1) + \frac{31x}{m} = \widehat{\psi}(r_1),$$

and therefore $r_1 \le \widehat{r}$. For every $v \ge \sqrt{\widehat{r}}$, the sub-root property implies

$$\mathfrak{R}_{m, P_X}(v; \mathcal{S}) = \Phi(v^2) \le \frac{\psi_1(v^2)}{10} \le \frac{\sqrt{r_1}}{10} v \le \frac{\sqrt{\widehat{r}}}{10} v.$$

Since $\overline{\mathcal{F}} \subseteq \mathcal{S}$, rescaling by L yields, for $u \ge L\sqrt{\widehat{r}}$,

$$\mathfrak{R}_{m, P_X}(u; \mathcal{F}) \le \frac{\sqrt{\widehat{r}}}{10} u \le \frac{\widehat{R}}{10} u,$$

which proves (96).

For the upper comparison, (92) and $L \ge 1$ imply

$$\mathfrak{R}_{m, P_X}(v; \overline{\mathcal{F}}) \le \kappa s v, \qquad v \ge s, \quad v > 0.$$

Lemma 30 supplies a constant $C_\star$ such that, with $w = s + h_m$,

$$\Phi(r) \le C_\star w \sqrt{r} \qquad \text{whenever } \sqrt{r} \ge s \vee h_m.$$

Freeze

$$D = [120(C_\star + 1)]^2, \qquad r_0 = D w^2. \tag{98}$$

These quantities are proof devices; the statistic uses neither of them. Since $x/m = 6h_m^2 \le 6w^2$, this choice ensures

$$2r_0 \ge 20\Phi(2r_0) + \frac{26x}{m}.$$

Lemma 3.6 at $2r_0$ consequently gives

$$\{f \in \mathcal{S} : Q_m f^2 \le 2r_0\} \subseteq \{f \in \mathcal{S} : Pf^2 \le 4r_0\}$$

except with probability e^{-x}. The upper comparison in Lemma A.4, again with tuning parameter $1/2$ and range $[-1, 1]$, applied to the fixed population ball on the right gives, except with probability e^{-x},

$$\mathbb{E}_\epsilon \sup_{\substack{f\in\mathcal{S}\\ Pf^2\le 4r_0}} \left|\frac{1}{m}\sum_i \epsilon_i f(X_i)\right| \le \frac{3}{2}\Phi(4r_0) + \frac{4x}{3m}.$$

On these two events,

$$\begin{aligned}\widehat{\psi}(r_0) &\le 30\Phi(4r_0) + \frac{173x}{3m}\\ &\le \{60C_\star\sqrt{D} + 346\}w^2 \le Dw^2 = r_0.\end{aligned}$$

Thus $\widehat{r} \le r_0$ and $\widehat{R} \le L\sqrt{D}(s + h_m)$. A union bound over the four events proves the lemma. The ball inclusions and Rademacher comparisons prove the bound directly. The positive confidence term also covers arbitrarily small or zero unregularized population fixed points. □

Q.3 Sequential M1 transfers a linear certificate to the full class

Lemma 32 (Transfer with a positive radius margin)**.** *For a public class satisfying sequential M1 and the common envelope, every sample size $m \ge 1$, every $u \ge 0$, and every $\eta > 0$ satisfy*

$$\mathfrak{R}^*_{m,P_X}(u; \Delta\mathcal{G}) \le \mathfrak{R}_{m,P_X}(u + \eta; \Delta\mathcal{G}^{\mathrm{rep}}). \tag{99}$$

Consequently, if for some $R \ge 0$ and $c \ge 0$ the core has the linear certificate

$$\mathfrak{R}_{m,P_X}(v; \Delta\mathcal{G}^{\mathrm{rep}}) \le cRv, \qquad v \ge R, \quad v > 0,$$

then the same certificate holds for the full outer complexity at every $u \ge R$ with $u > 0$.

Proof. Fix $f = g - g' \in \Delta\mathcal{G}$ with $\|f\|_2 \le u$. Sequential M1 gives represented sequences converging pointwise everywhere to g and g'. Their differences f_k converge pointwise everywhere to f and, by the common envelope and dominated convergence, in $L_2(P_X)$. Thus $\|f_k\|_2 \le u + \eta$ eventually. For every covariate sample and every sign vector, the signed sample average of f_k converges to that of f. Its limit in absolute value is bounded by the core supremum at radius $u + \eta$. Taking the supremum over all such full-class f proves the pointwise inequality

$$Z_{\Delta\mathcal{G},m}(u) \le Z_{\Delta\mathcal{G}^{\mathrm{rep}},m}(u + \eta).$$

The right side is measurable and is therefore an admissible majorant in the definition of outer expectation. This proves (99). Under the linear certificate, the right side in that display is at most $cR(u + \eta)$; let $\eta \downarrow 0$. The positive margin supplies the needed domination between the two localized complexities. □

Q.4 The three-block rule and its risk

We now prove Theorem 29. All constants of Appendix P are frozen for the known regularity tuple

$$\rho' = (B_0, B_{\mathcal{G}}, v_-, v_+, 2c_{\mathrm{rad}}).$$

Let $m_{\rho'}$ be its fixed lower sample-size threshold, enlarged to at least two. If $m = \lfloor 2n/3 \rfloor < m_{\rho'}$, define the rule to be zero. There are only finitely many such n, with a bound depending only on ρ, so its absolute loss, at most three, is absorbed by $C_\rho q$. This includes $n = 1$.

For the remaining n, split the first $3m \le 2n$ observations deterministically into three disjoint blocks, each of size m. The first is the calibration block; its treatments and outcomes are not used. The

second and third supply the two input halves of the rule of Theorem 27 at sample size m. Any leftover observations are unused. In particular, $m \le n$ and $n/m \le 2$.

For each $\nu \in \{\mu, \pi\}$, apply (93)–(95) on the calibration covariates to $\mathcal{F}_\nu = \Delta\mathcal{G}_\nu^{\mathrm{rep}}$, with $L = \max\{1, 2B_{\mathcal{G}}\}$. Denote its output by $\widehat{R}_\nu$. Define

$$C_{\mathrm{cap}} = \max\{1, (60c_{\mathrm{rad}})^{-1}\}, \qquad S = \min\{1, C_{\mathrm{cap}}\widehat{R}_\mu\}, \qquad T = \min\{1, C_{\mathrm{cap}}\widehat{R}_\pi\}. \tag{100}$$

The final rule is the explicit estimator of Appendix P, with public instance Γ, regularity tuple ρ', sample size m, and stochastic inputs (S, T), applied to the last two blocks. Here T in (100) denotes the scalar stochastic input, not a treatment observation.

The estimator just specified is jointly Borel in its two stochastic inputs and its observations. To see this directly from Appendix P, its initial fits and pilot are Borel and do not use those inputs; each calibration tolerance is a continuous function of their squares; and its constraints, countable feasible-set decisions, countable infima, and first strict approximate minimizers are Borel jointly in inputs and observations. The constants are fixed for ρ' before either input is selected. Composition with (95) and (100) therefore gives one Borel rule on $\mathcal{Z}^{2n}$. Its definition contains none of the four declared budgets.

Fix now any (a, s, b, t) and compatible P, solely for the proof. For every fixed population-localized countable core, adding $n - m$ independent signed observations and applying conditional Jensen gives

$$\mathfrak{R}_{m,P_X}(u; \mathcal{F}_\nu) \le \frac{n}{m}\mathfrak{R}_{n,P_X}(u; \mathcal{F}_\nu). \tag{101}$$

Indeed, condition on the first m covariates and signs; the added signed summands have conditional mean zero, and the supremum of absolute linear averages is convex. Countability permits ordinary conditional expectation. Combining this with core inclusion into the full class and the original certificate gives

$$\mathfrak{R}_{m,P_X}(u; \mathcal{F}_\mu) \le 6c_{\mathrm{rad}}su \quad (u \ge s,\ u > 0),$$

and the corresponding inequality with t for $\mathcal{F}_\pi$. Thus Lemma 31 applies with $\kappa = 6c_{\mathrm{rad}}$. There is an event E measurable with respect to the calibration covariates, of probability at least $1 - 8(em)^{-6}$, on which both output certificates hold and

$$\widehat{R}_\mu \le C_\rho(s + h_m), \qquad \widehat{R}_\pi \le C_\rho(t + h_m). \tag{102}$$

One may take E to be the intersection of the four fixed-radius events constructed for each core in the proof of that lemma. This description makes its measurability explicit, without a supremum over random population radii.

On E, Lemma 32 turns each output core certificate into the same full outer certificate. If $S < 1$, then $S = C_{\mathrm{cap}}\widehat{R}_\mu \ge \widehat{R}_\mu$, so for every $u \ge S$,

$$\mathfrak{R}^*_{m,P_X}(u; \Delta\mathcal{G}_\mu) \le \frac{\widehat{R}_\mu}{10}u \le 6c_{\mathrm{rad}}Su = 3c'_{\mathrm{rad}}Su.$$

If $S = 1$, use $s \le 1$ in the original certificate and (101): for every $v \ge 1$,

$$\mathfrak{R}_{m,P_X}(v; \mathcal{F}_\mu) \le 6c_{\mathrm{rad}}sv \le 6c_{\mathrm{rad}}v.$$

The positive-margin transfer of Lemma 32, followed by letting the margin tend to zero, gives the required full outer bound $6c_{\mathrm{rad}}u$ for every $u \ge 1$. The same two cases apply to T and the propensity class. This argument uses ordinary Jensen only for countable cores; it does not require a conditional-Jensen identity for nonmeasurable full-class suprema.

It follows that, for every calibration realization in E,

$$P \in \mathcal{P}_{\rho',m,a,S,b,T}(\Gamma). \tag{103}$$

The approximation budgets in this assertion are the original a, b: neither public class, either approximation infimum, nor the law has changed. Only the sample size, the stochastic certificates, and the fixed complexity constant have changed. Also, from (102),

$$S \le C_\rho(s + h_m), \qquad T \le C_\rho(t + h_m) \quad \text{on } E.$$

Condition on the calibration covariates. Independence of the three blocks leaves the other two blocks iid with joint law P^{2m}. For realizations in E, apply Theorem 27 using (103). Its conditional expected absolute loss is bounded by

$$C_\rho \left[m^{-1/2} + \min\{a(b+T) + a^2 + S^2,\ b(a+S) + (b+T)^2\} \right].$$

Quadratic absorption gives, on this same event,

$$a(b+T) + a^2 + S^2 \le C_\rho\{a(b+t) + a^2 + s^2 + h_m^2\},$$
$$b(a+S) + (b+T)^2 \le C_\rho\{b(a+s) + (b+t)^2 + h_m^2\}.$$

For the first inequality use $ah_m \le (a^2 + h_m^2)/2$. For the second use $bh_m \le (b^2 + h_m^2)/2$ and $b^2 \le (b+t)^2$. Since m is a fixed proportion of n,

$$m^{-1/2} \le Cq, \qquad h_m^2 = \frac{\log(em)}{m} \le Cq.$$

Both the output and the target lie in $[-3, 3]$, so the contribution from E^c is at most $48(em)^{-6} \le Cq$. Averaging proves the first inequality in (89). The second follows from $a^2 + s^2 \le (a+s)^2$ and the definition of W.

All radii in the calibration rule are strictly positive because of the confidence term. Zero declared s or t enters the proof only through its original positive-radius certificate, which then vanishes; Lemma 30 and all subsequent steps still apply. All quotients retain positive radii, and the process bounds use the original model assumptions. This completes the proof of Theorem 29.