GEOMETRY-LIMITED POLICY CONTRASTS FOR DISTRIBUTION-VALUED OUTCOMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Many interventions produce a distribution for each individual rather than a scalar response. Comparing only means or Wasserstein barycenters discards heterogeneity across individual distributions. We study population matching costs between policy-specific outer laws, allowing all matching mechanisms compatible with the conditional marginals. Conditional Kantorovich duality gives sharp bounds for each policy pair. We construct split-sample lower and upper certificates by extending finite transport duals to the full quantile cone with direction-specific discretization corrections. Asymmetric convex costs encode one-sided harm, tolerance thresholds, and escalating severity. Global feasibility ensures finite-sample validity simultaneously over the policy class, even when the outer-law learner is misspecified. Evaluation uses inverse propensity weighting (IPW) for known propensities or calibrated augmented IPW (AIPW) for unknown propensities on fixed finite covariate support. Validity, however, does not imply informativeness. With fixed finite covariate and action sets, distributions on a one-dimensional outcome space, and a bounded convex displacement cost, local contact of order β_h imposes a worst-case error lower bound of order $(\log n)^{-\beta_h}$ at both endpoints, even for bi-Lipschitz individual quantiles. A global rectangular order γ_h yields an attainable rate of $(\log n)^{-\gamma_h}$. When the orders match, the empirical-dual estimators and their simultaneous honest bands attain the minimax rate; the cost W_p^p is the special case $\gamma_h = \beta_h = p \wedge 2$. Under conditional exchangeability and uniform overlap, both IPW and calibrated AIPW attain this geometric rate without a loss in rate from exact simultaneous honesty. A cost-active high-probability quantization condition suffices for a polynomial upper rate. Simulations and an electricity-template study distinguish validity from informativeness, examine outer-law misspecification, and assess simultaneous coverage after policy search.

1 INTRODUCTION

Learning whom to treat and how to allocate interventions is a central problem in data-driven decision making. Policy-learning methods formalize this goal by comparing the welfare achieved by alternative treatment rules (Kitagawa & Tetenov, 2018; Athey & Wager, 2021; Zhou et al., 2023). A single response need not capture what an intervention changes. Repeated measurements, such as activity readings from a wearable device, can instead describe an entire distribution for each person; causal inference for such outcomes has therefore moved beyond scalar responses (Lin et al., 2023; Kurisu et al., 2024). Optimal transport provides a way to compare these distributions while respecting the magnitude of changes in the underlying measurements (Panaretos & Zemel, 2019; Peyre & Cuturi, 2019), and recent work uses their Wasserstein barycenter, a geometric average distribution, to define policy welfare (Huang et al., 2026). These developments raise a further question: how should policies be compared when differences *among people's distributions* also matter?

Consider a stylized activity intervention. Under one policy, everyone has the same distribution of daily activity. Under another, half the population has that distribution shifted upward and half has it shifted downward by the same amount. The average activity is unchanged, and even the Wasserstein barycenter is identical, although the second population contains two distinct groups. Every pairing of a person from the first population with a person from the second has a positive discrepancy. Comparing means or Wasserstein barycenters thus cannot distinguish a homogeneous population from one divided between higher and lower activity. Retaining this distinction requires describing

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.

both variation within each person and variation across people. We call the population distribution of individual distributions the *outer law*.

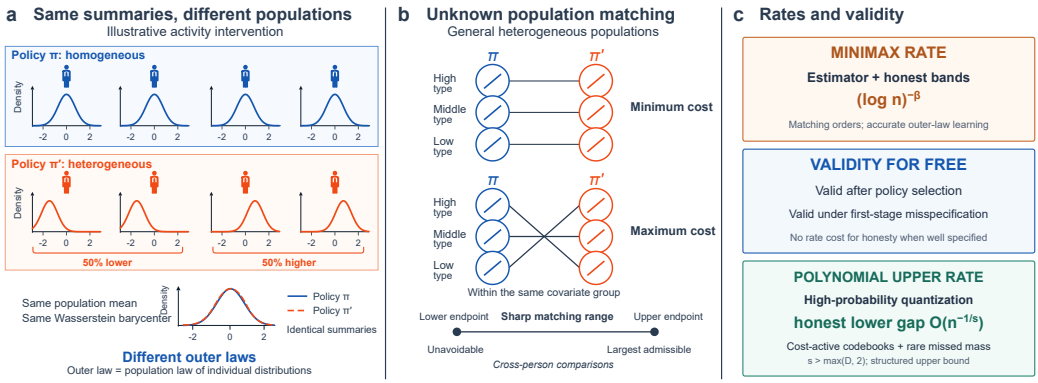


Figure 1: **Minimax-rate estimation with honest policy certificates.** (a) Policies can have the same mean and Wasserstein barycenter while inducing different outer laws. (b) Matching individuals across policy populations is unspecified, giving a range of compatible costs. (c) When $\beta_h = \gamma_h$ and the first stage learns the outer laws at the required rate, empirical-dual estimators and simultaneous honest bands attain the minimax rate $(\log n)^{-\beta_h}$ with IPW or calibrated finite-cell AIPW. Coverage persists after policy selection and outer-law misspecification; misspecification need not preserve tightness. Under high-probability quantization with $s > \max\{D, 2\}$, the honest lower certificate also has an attainable polynomial expected gap $O(n^{-1/s})$. A matching lower bound for this structured class remains open.

Comparing two outer laws introduces a substantive choice: which people should be paired across the two policy populations? Within groups with the same observed characteristics, pairing similar distributions can give a small average discrepancy, whereas pairing dissimilar ones can give a large discrepancy. We leave this population matching rule unspecified and study the full range of average costs compatible with the two outer laws. Its lower endpoint measures unavoidable discrepancy; its upper endpoint measures the largest discrepancy an admissible matching can produce. The target compares distributions across people under the two policies. The cost can also distinguish increases from decreases or tolerate small changes, so the comparison can reflect the direction and severity of changes relevant to a decision.

The statistical challenge is to make this richer comparison both reliable and precise. Optimal-transport duality turns matching problems into bounds that can be evaluated on observed outcomes, and feasible dual bounds can remain valid even when a fitted outcome model is wrong (Ji et al., 2023). For distribution-valued outcomes, however, constraints checked on finitely many training distributions must hold for every possible distribution. Moreover, valid bounds can still be too conservative to distinguish policies. The slow empirical convergence of transport between random distributions (Catalano & Lavenant, 2024) makes it essential to ask whether this imprecision is an artifact of estimation or a fundamental limit, and what population structure can overcome it.

Validity and informativeness depend on different conditions. Enforcing the dual inequalities everywhere protects coverage; the complexity of the population, as measured by the chosen cost, controls how quickly the bounds become informative. In a fixed finite design with known or unknown assignment, we show that even regular individual distributions can leave a logarithmically slow statistical barrier. Making each person's distribution smoother does not by itself simplify the diversity across people. Conversely, when that diversity admits an accurate representation by relatively few representative distributions, polynomial accuracy becomes possible under our stated approximation condition. This distinction identifies where additional structure is needed to make distribution-sensitive policy comparisons useful. Figure 1 connects the matching ambiguity to our main result: minimax-rate estimation and honest inference under the same empirical-dual construction.

Contributions. We make three contributions:

- We construct finite-sample lower and upper certificates for the sharp matching range by extending finite transport duals to all individual distributions. IPW with known assignment and finite-cell AIPW with unknown assignment give simultaneous coverage despite arbitrary outer-law misspecification.
- We establish lower and attainable upper accuracy bounds for convex comparison costs, with matching minimax rates when their local and global regularity orders agree. In the fixed finite design, the logarithmic barrier persists for regular individual distributions, while simultaneous coverage preserves the minimax rate. IPW and finite-cell AIPW attain the same geometric minimax rates.
- We give a condition on population approximation that yields polynomial certificate rates, together with explicit numerical error control. Simulations and a semi-synthetic electricity-load study examine coverage, precision, and protection after policy search.

Related work. Distribution-valued causal effects (Lin et al., 2023) and barycentric policy learning (Huang et al., 2026) target geometric summaries; scalar welfare methods control policy regret (Kitagawa & Tetenov, 2018; Athey & Wager, 2021; Zhou et al., 2023). Counterfactual utilities depend on within-individual associations (Ben-Michael et al., 2024; Cui & Han, 2026). Our cross-person matching target retains outer-law heterogeneity lost by barycentric compression.

Conditional and multi-marginal OT characterize causal bounds (Ober-Reynolds, 2023; Lin et al., 2025; 2026; Gao et al., 2025). We inherit conditional duality (Bogachev & Malofeev, 2020) and the misspecification-robust feasible-dual principle of Ji et al. (2023); the additional tasks are feasibility on the full quantile cone and simultaneous inference over policy pairs. Population transport between random measures is also established (Emami & Pass, 2025); our focus is its statistical informativeness.

Empirical OT rates depend on ground-space complexity (Weed & Bach, 2019; Hundrieser et al., 2024). Catalano & Lavenant (2024) exhibit slower-than-polynomial empirical convergence for Wasserstein-over-Wasserstein and introduce a hierarchical IPM with parametric convergence. We retain the matching-cost target and determine when its scalar endpoints admit sharp minimax rates, including honest inference, rather than replacing the discrepancy.

2 From Population Matching to a Sharp Target

We define the range of costs over admissible population matchings. Conditional duality establishes that both endpoints are attainable and expresses them through potentials whose feasibility determines the direction of the statistical bounds in Section 3.

Let $I = [I_-, I_+] \subset \mathbb{R}$ be compact, $\mathcal{Z} = \mathcal{P}(I)$ with the weak topology, $\mathcal{A} = \{1, \dots, d\}$, and Π a class of measurable deterministic policies on a standard Borel covariate space $\mathcal{X}$. Individual potential outcomes satisfy $Y(a) \in \mathcal{Z}$. We observe i.i.d. $O_i = (X_i, A_i, Y_i)$ with $Y_i = Y_i(A_i)$ and propensity $e_a(x) = \mathbb{P}(A = a \mid X = x)$. Consistency, no interference, conditional exchangeability, and overlap $e_a(x) \geq \underline{e} > 0$ identify $P_{a,x} = \mathcal{L}\{Y(a) \mid X = x\}$. The cost $c_x : \mathcal{Z}^2 \to [0, C_c]$ is jointly measurable in (x, y, y'), continuous in (y, y'), and satisfies $c_x(y, y) = 0$. The full assumptions are in Appendix A. IPW uses known propensities; the finite-design unknown-propensity procedure uses AIPW in Theorem 3.2.

For every action pair, the population matching set is

$$\mathfrak{G}_x(a, b) = \Gamma(P_{a,x}, P_{b,x}), \qquad (a, b) \in \mathcal{A}^2, \tag{1}$$

where $\Gamma(P, Q)$ is the set of couplings of P and Q. A coupling pairs distributions from two distinct individuals in the policy populations; independence is one possible matching rule, not a restriction. Identical policy populations can therefore have a positive maximal matching cost, reflecting heterogeneity among their individual distributions. For an admissible kernel $\eta_x \in \mathfrak{G}_x\{\pi(x), \pi'(x)\}$, define

$$\Theta_c(\pi, \pi'; \eta) = \mathbb{E}_X \left[\int c_X(y, y') \, d\eta_X(y, y') \right].$$

The range is generated by varying η. Define

$$\Theta_c^L(\pi, \pi') = \mathbb{E}_X \left[\inf_{\gamma \in \Gamma(P_{\pi(X),X}, P_{\pi'(X),X})} \int c_X \, d\gamma \right], \tag{2}$$

with Θ_c^U defined by replacing $\inf$ with $\sup$. Conditional Kantorovich duality establishes attainability and gives variational representations of both endpoints.

Theorem 2.1 (Sharpness and dual bounds). *Under the preceding conditions, for every fixed* (π, π'),

$$\Theta_c^L(\pi, \pi') \le \Theta_c(\pi, \pi'; \eta) \le \Theta_c^U(\pi, \pi')$$

for every admissible η*, and every value in this interval is attained. Moreover,*

$$\Theta_c^L(\pi, \pi') = \mathbb{E}_X \left[\sup_{\phi(y)+\psi(y') \le c_X(y,y')} \{ P_{\pi(X),X}\phi + P_{\pi'(X),X}\psi \} \right]. \tag{3}$$

The upper endpoint satisfies

$$\Theta_c^U(\pi, \pi') = \mathbb{E}_X \left[\inf_{u(y)+v(y') \ge c_X(y,y')} \{ P_{\pi(X),X}u + P_{\pi'(X),X}v \} \right]. \tag{4}$$

Consequently, any measurable globally subfeasible dual collection with finite population marginal integrals has population value at most Θ_c^L*, while any measurable globally superfeasible collection with finite population marginal integrals has population value at least* Θ_c^U*.*

Each policy pair has its own matching mechanism and sharp interval. Globally subfeasible potentials bound the lower endpoint from below; superfeasible potentials bound the upper endpoint from above. To use these inequalities for inference, the numerical construction must preserve feasibility on the full outcome space.

3 Finite-Sample Validity Is Geometry-Free

Globally feasible duals support two evaluation procedures: inverse propensity weighting (IPW) when assignment probabilities are known, and augmented IPW (AIPW) when they are estimated. Both use the same bounded potentials and target the same sharp endpoints. Their confidence corrections differ; neither requires accurate outer-law learning for validity. The feasible-dual principle and augmented evaluation follow Ji et al. (2023); our geometric analysis concerns their distance to the sharp endpoints.

On a training fold D_1 of size ℓ, learn measurable action-pair-indexed potentials $(\widehat{\phi}, \widehat{\psi})$ with absolute bound M and

$$\widehat{\phi}_{ab,x}(y) + \widehat{\psi}_{ab,x}(y') \le c_x(y, y') \quad \text{for all } y, y'.$$

Finite OT, conservative quantile binning, and safe extension produce such potentials (Appendix E). Fix $u = (\pi, \pi')$ and abbreviate $a = \pi(X)$, $b = \pi'(X)$, $\phi = \widehat{\phi}_{ab,X}(Y)$, $\psi = \widehat{\psi}_{ab,X}(Y)$. On an independent evaluation fold D_E of size m, the two scores are

$$S_u^{\mathrm{IPW}}(O) = \frac{\mathbb{1}\{A=a\}}{e_a(X)}\phi + \frac{\mathbb{1}\{A=b\}}{e_b(X)}\psi, \tag{5}$$

$$S_u^{\mathrm{AIPW}}(O) = \widehat{\mu}_\phi(X) + \widehat{\mu}_\psi(X) + \frac{\mathbb{1}\{A=a\}}{\widehat{e}_a(X)}\{\phi - \widehat{\mu}_\phi(X)\} + \frac{\mathbb{1}\{A=b\}}{\widehat{e}_b(X)}\{\psi - \widehat{\mu}_\psi(X)\}. \tag{6}$$

Write $\overline{S} = m^{-1}\sum_{i \in D_E} S(O_i)$ for either score. For AIPW, a third, independent nuisance fold D_N estimates e_a and the conditional means of the fitted scalar potentials, $\mu_\phi(x) = P_{a,x}\widehat{\phi}_{ab,x}$ and $\mu_\psi(x) = P_{b,x}\widehat{\psi}_{ab,x}$, with the action-pair indices suppressed. Clip $\widehat{e}_a$ to $[\underline{e}, 1]$ and the means to $[-M, M]$. Conditional on training, the IPW mean is the population dual value g_u. Conditional on both training folds, the AIPW mean is $g_u + R_u$, where

$$R_u = \mathbb{E}_X \left[\frac{\widehat{e}_a - e_a}{\widehat{e}_a}(\widehat{\mu}_\phi - \mu_\phi) + \frac{\widehat{e}_b - e_b}{\widehat{e}_b}(\widehat{\mu}_\psi - \mu_\psi) \right]. \tag{7}$$

Thus either exact nuisance component removes the evaluation bias; it does not remove the geometric gap $\Theta_c^L - g_u$.

Theorem 3.1 (IPW: finite-sample validity). *Suppose propensities are known, overlap holds, and the bounded conditional score class is pointwise measurable. Let* $\widehat{\mathfrak{R}}_m$ *be its conditional empirical*

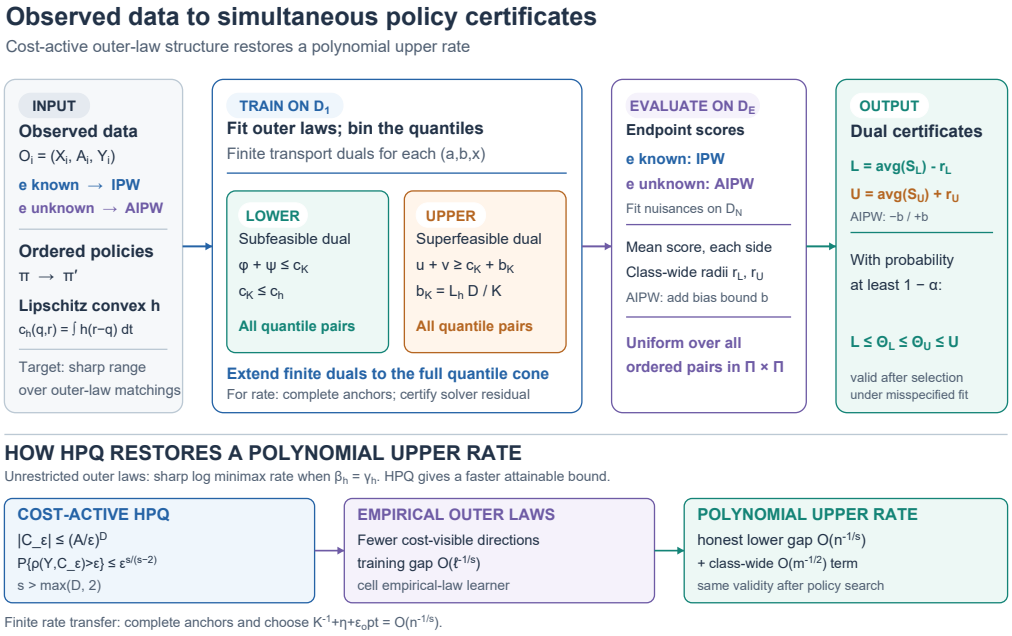


Figure 2: **Shared dual construction, two evaluation procedures.** The same globally feasible potentials are evaluated by IPW when propensities are known or by AIPW when they are estimated. AIPW uses an independent nuisance fold and the calibrated bias correction in Theorem 3.2. Both yield simultaneous certificates; cost-active population structure improves their tightness.

Rademacher complexity and $B_I = 2M/\underline{e}$. *With*

$$r_I = 2\widehat{\mathfrak{R}}_m + 3B_I\sqrt{2\log(2/\alpha)/m}, \qquad L_I(u) = \overline{S}_u^{\mathrm{IPW}} - r_I, \tag{8}$$

$\Pr\{L_I(u) \le \Theta_c^L(u),\ \forall u \in \Pi^2\} \ge 1-\alpha$. *Finite uniform policy-entropy integral gives* $r_I = O(m^{-1/2})$ *for fixed* α *(Appendix D).*

Theorem 3.2 (AIPW: finite-sample validity)**.** *Let* N_Π *be the finite number of policy restrictions to the covariate support. Suppose a* $\sigma(D_1, D_N)$*-measurable* $b \ge 0$ *satisfies* $\Pr\{\sup_u |R_u| \le b\} \ge 1-\delta$, *where* $0 < \delta < \alpha$. *With* $B_A = 2M + 4M/\underline{e}$, *set*

$$r_A = B_A\sqrt{2\log\{2N_\Pi^2/(\alpha-\delta)\}/m}, \qquad L_A(u) = \overline{S}_u^{\mathrm{AIPW}} - r_A - b. \tag{9}$$

Then $\Pr\{L_A(u) \le \Theta_c^L(u),\ \forall u \in \Pi^2\} \ge 1-\alpha$. *On fixed finite covariate support with cell masses and propensities bounded below, Appendix H supplies an observable* b *with* $\mathbb{E}b = O(\log n/n)$ *and* $\delta = O(n^{-1})$ *under a fixed-proportion three-fold split.*

Corollary 3.3 (IPW and AIPW: two-sided sharp intervals)**.** *The two validity arguments also apply to bounded costs with nonzero diagonal. Apply either construction to the complement cost* $C_c - c$ *to obtain a lower bound* $\bar{L}$ *on* $C_c - \Theta_c^U$, *and set* $U = C_c - \bar{L}$. *Allocate* $\alpha/2$ *to each endpoint and clip the endpoints to* $[0, C_c]$. *The resulting* $[L, U]$ *simultaneously contains* $[\Theta_c^L(u), \Theta_c^U(u)]$ *for every policy pair with probability at least* $1-\alpha$.

All three statements allow misspecified outer-law learning and policy selection on the evaluation fold. AIPW additionally uses a certified finite-sample nuisance envelope. Feasibility controls direction; the next section controls the gap to the sharp endpoints.

4 Minimax Informativeness for Convex Costs

Section 3 guarantees coverage even for an inaccurate first-stage learner; here we determine when a valid certificate is informative. We separate the local contact condition (11), which yields a lower bound, from the global interaction condition (12), which controls an attainable upper rate.

Theorem 4.1 combines them into a minimax law when their orders match. Theorem 4.2 shows why smooth individual distributions do not remove the outer-law barrier, while Corollary 4.3 identifies the power-cost case. We then give a structured route to faster rates through (15) and account for implementation error in (16).

Suppose $\mathrm{supp}(X) = \{x_1, \ldots, x_{K_X}\}$ has masses at least κ, conditional exchangeability holds with propensity at least $\underline{e}$, and policies induce at least two distinct action vectors. The support is fixed and known, and the overlap lower bound is available. Use IPW for known propensities or the finite-cell AIPW construction for unknown propensities; write $t = |D_N|$ in the latter case. After rescaling the outcome interval to $[0, 1]$, let $\mathcal{Q}$ be the monotone quantile cone and consider the general convex displacement cost

$$c_h(q, r) = \int_0^1 h\{r(t) - q(t)\}\, dt, \qquad h : [-1, 1] \to [0, \bar{h}], \quad h(0) = 0. \tag{10}$$

The minimax law is determined by two different features of h: a local lower contact that creates statistically indistinguishable hard instances, and a global upper interaction that controls the complexity of learned duals.

We say that h has finite-order contact $\beta_h \geq 1$ if, for some $z_h \in (-1, 1)$, $g_h \in \partial h(z_h)$, and $\underline{c}_h > 0$,

$$h(z_h + u) - h(z_h) - g_h u \geq \underline{c}_h |u|^{\beta_h} \quad \text{locally in } u. \tag{11}$$

For $0 < \gamma \leq 2$, define

$$\vartheta_\gamma(u, v) = \begin{cases} \min\{|u|^\gamma, |v|^\gamma\}, & \gamma \leq 1, \\ \min\{|u||v|^{\gamma-1}, |v||u|^{\gamma-1}\}, & \gamma > 1. \end{cases}$$

The global rectangular order is γ_h if h is $(\gamma_h \wedge 1)$-Hölder and, whenever all arguments are in $[-1, 1]$,

$$|h(z + u + v) - h(z + u) - h(z + v) + h(z)| \leq L_h \vartheta_{\gamma_h}(u, v). \tag{12}$$

The local contact condition gives a lower bound; the global rectangular condition gives an attainable upper rate. Write $T_{h,P} = \Theta^L_{c_h}$ and $T^U_{h,P} = \Theta^U_{c_h}$ under law P. Below, worst-case error is the supremum over the stated model of expected worst-policy absolute error for estimators, or positive excess length for honest bands. Appendix F gives the formal risk definitions.

Theorem 4.1 (General convex-cost minimax law)**.** *For both the known- and unknown-propensity models, fix the preceding finite design and $0 < \alpha < 1/3$. For all sufficiently large n, the following statements hold.*

(a) *If (11) holds, every simultaneous honest lower band, every simultaneous honest upper band, and every point estimator of either endpoint has worst-case error at least $c\{\log(n + 1)\}^{-\beta_h}$. For honest bands the statement also holds in probability: some law makes the worst-policy gap exceed this quantity with probability at least $1/3$.*

(b) *If (12) holds, the split-sample empirical dual estimators for the lower endpoint and for the complement cost $\bar{h} - c_h$ satisfy*

$$\sup_P \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_h - T_{h,P}| \;\vee\; \sup_P \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}^U_h - T^U_{h,P}| \\ \leq C\{\log(\ell\kappa\underline{e} + 1)\}^{-\gamma_h} + q_{\mathrm{ev}} + \frac{C\bar{h}}{\underline{e}} \sqrt{\frac{K_X \log d}{m}}. \tag{13}$$

Here $q_{\mathrm{ev}} = 0$ for IPW and $q_{\mathrm{ev}} = C \log(n)/t$ for finite-cell AIPW. The honest bands add $\log(1/\alpha)$ inside the square root; constants may depend on the fixed design and overlap.

(c) *If the two orders match, $\gamma_h = \beta_h$, then lower and upper endpoint estimation and lower and upper honest inference all have minimax rate*

$$(\log n)^{-\beta_h} \tag{14}$$

under a fixed-proportion split. Both IPW and AIPW empirical-dual estimators and their simultaneous honest bands attain this rate.

Cost-visible quantile codes and testing give the lower bounds; dual entropy gives attainability. Appendix F proves both, including a separate upper-endpoint hard code; Appendix F.2 shows why contact alone is insufficient.

Appendix H proves the AIPW transfer: unknown assignment adds a calibrated evaluation remainder but leaves the geometric training gap and the known-assignment lower submodels intact. General continuous-covariate assignment requires separate theory.

Theorem 4.2 (Individual smoothness does not reduce outer-law complexity)**.** *For $L > 0$, let*

$$\mathcal{Q}_L = \{q \in \mathcal{Q} : |q(t) - q(s)| \le L|t - s| \text{ for every } s, t \in [0, 1]\},$$

and restrict every conditional outer law to $\mathcal{P}(\mathcal{Q}_L)$. Under the local contact condition (11), the four lower bounds in Theorem 4.1 remain $c_{h,L}\{\log(n+1)\}^{-\beta_h}$ on this smaller model class. This covers lower and upper honest bands and point estimation of both endpoints. If the global order also satisfies $\gamma_h = \beta_h$, all four restricted risks are $(\log n)^{-\beta_h}$, and both IPW and AIPW empirical-dual procedures attain them.

The hard outer laws may be supported entirely on strictly increasing, uniformly bi-Lipschitz quantile curves whose induced distributions have compact supports and densities bounded above and away from zero on their supports.

The hard construction uses bi-Lipschitz quantiles, so every individual distribution is atomless and has a density bounded above and away from zero on its support. The outer law still has exponentially many directions that the cost distinguishes. A code with J blocks separates the endpoints by order $J^{-\beta_h}$, leaving the logarithmic minimax exponent unchanged.

Corollary 4.3 (Power costs are a sharp special case)**.** *For $h(z) = |z|^p$ with $p \ge 1$, both orders equal $\gamma_h = \beta_h = p \wedge 2$. Consequently, for $c_p = W_p^p$, the four minimax risks in Theorem 4.1 are all $(\log n)^{-(p\wedge 2)}$. Its smooth-law version is the corresponding special case of Theorem 4.2.*

For $1 \le p \le 2$, the least-favorable code is visible at zero displacement and has order p. For $p > 2$, zero is too flat; moving the construction to a nonzero displacement activates quadratic curvature, so the exponent saturates at two.

The same theorem gives explicit rates for directional losses (Table 1); asymmetry need not change the exponent. For the squared hinge, quadratic contact occurs above the threshold, not at its flat junction. Appendix G.1 verifies the orders.

Table 1: Directional losses with matching orders in the fixed finite design. Rates hold for IPW and finite-cell AIPW; $0 \le \tau < 1$, $\lambda_\pm > 0$, and $p \ge 1$ are fixed.

Loss $h(z)$	$\beta_h = \gamma_h$	Estimation / honest-band rate
$(z-\tau)_+$	1	$(\log n)^{-1}$
$(z-\tau)_+^2$	2	$(\log n)^{-2}$
$\lambda_+ z_+^p + \lambda_-(-z)_+^p$	$p \wedge 2$	$(\log n)^{-(p\wedge 2)}$

For both estimators, these rates require accurate outer-law learning. Global feasibility protects IPW coverage; AIPW also requires calibrated nuisance correction. Neither removes a persistent misspecification gap.

The logarithmic law is for unrestricted outer distributions. Polynomial rates return when the mass is compressible in the cost-active pseudometrics

$$\rho^c_{x,1}(y, \widetilde{y}) = \sup_z |c_x(y, z) - c_x(\widetilde{y}, z)|, \quad \rho^c_{x,2}(y, \widetilde{y}) = \sup_z |c_x(z, y) - c_x(z, \widetilde{y})|.$$

For a codebook $\mathcal{C}$, write $\rho^c_{x,k}(y, \mathcal{C}) = \inf_{z\in\mathcal{C}} \rho^c_{x,k}(y, z)$. The high-probability quantization condition requires, uniformly in action, covariate, and $k \in \{1, 2\}$,

$$|\mathcal{C}_{a,x,k,\varepsilon}| \le (A/\varepsilon)^D, \qquad P_{a,x}\{\rho^c_{x,k}(Y, \mathcal{C}_{a,x,k,\varepsilon}) > \varepsilon\} \le \varepsilon^{s/(s-2)}. \tag{15}$$

If $s > \max\{D, 2\}$, the IPW or finite-cell AIPW empirical-dual band has expected gap

$$O\left((\ell\kappa\underline{e} + 1)^{-1/s} + q_{\text{ev}} + \frac{1}{\underline{e}}\sqrt{\frac{\log(N_\Pi \vee 2) + \log(2/\alpha)}{m}}\right),$$

where N_Π counts policy restrictions to the covariate cells and q_{ev} is as above. For fixed-proportion folds, both retain $O(n^{-1/s})$, since $s > 2$. The AIPW claim follows from Corollary H.5 and the

same cellwise training bound. A matching lower bound for this structured class remains open. Exact risk definitions and all proofs are in Appendices F and I. Finite within-individual sampling creates a separate information budget; its nonidentification and correction theory are given in Appendix J.

Table 2 gives concrete HPQ models. For Lipschitz h, $\rho^c_{x,k} \le \mathrm{Lip}(h)\|q_y - q_{\tilde{y}}\|_1$; individual smoothness alone does not imply such outer-support covers.

Table 2: Sufficient HPQ models, uniformly in (a, x, k); $s > \max\{D, 2\}$. See Remarks I.5–I.6.

Outer-law model	Sufficient structure / codebook	Missed mass
At most M_0 prototypes	M_0 centers; any $D > 0$, suitable A	0
Bounded r-parameter family	Lipschitz map in active metrics; $O(\varepsilon^{-r})$ centers, $D = r$	0
Compact r-manifold	Active covers of size $O(\varepsilon^{-r})$, $D = r$	0
Rare-regime mixture	$O(\varepsilon^{-r})$ covers per regime, tail $\le C_0 e^{-c_0 J}$; retain $J = O(\log(1/\varepsilon))$, $D = r + \delta$, $\delta > 0$	$\le \varepsilon^{s/(s-2)}$

The implemented certificate also incurs computational approximation error. For a Lipschitz convex displacement cost satisfying (15), exactly observed training and evaluation curves, and the completed finite-anchor construction of Appendix E.5, the implemented lower certificate, evaluated by IPW or finite-cell AIPW, satisfies

$$\begin{aligned}&\mathbb{E} \sup_{(\pi,\pi')\in\Pi^2} \left\{\Theta^L_{c_h}(\pi,\pi') - L^{\mathrm{comp}}_\alpha(\pi,\pi')\right\}_+ \\ &\lesssim (\ell\kappa\underline{e}+1)^{-1/s} + \left(K^{-1} + \eta + \varepsilon_{\mathrm{opt}}\right) + q_{\mathrm{ev}} + \frac{1}{\underline{e}}\sqrt{\frac{\log(N_\Pi \vee 2) + \log(2/\alpha)}{m}}.\end{aligned} \tag{16}$$

Here K is the number of quantile bins, η is the covering radius of the completed anchor set, and $\varepsilon_{\mathrm{opt}}$ is a certified primal–dual optimization residual. The terms separate outer-law learning, numerical approximation, nuisance correction, and evaluation. Binning, safe extension, and certified optimization precede either score; AIPW fits means of the final potentials and calibrates their actual envelope. Upper certificates retain the direction-specific binning correction. Binning acts on the quantile curves within each covariate cell. For unrestricted outer laws with $c_h = W^p_p$, the training term is instead $\{\log(\ell\kappa\underline{e}+1)\}^{-(p\wedge 2)}$. The discretization, anchor construction, and precision choices that keep the computational terms within the corresponding statistical rate are given in Appendix E.5, particularly Corollary E.5; combining its gap bound with Proposition H.3 gives the AIPW version.

Directional comparisons use (10) with

$$h(z) = \lambda_+(z-\tau_+)^p_+ + \lambda_-(-z-\tau_-)^p_+, \qquad \lambda_\pm \ge 0, \quad \tau_\pm \in [0,1), \quad p \ge 1,$$

to encode weights, tolerances, and severity. Write $\Theta^e_h = \Theta^e_{c_h}$. Reversing policies reverses the loss: $\Theta^e_h(\pi',\pi) = \Theta^e_{\check{h}}(\pi,\pi')$, where $\check{h}(z) = h(-z)$ and $e \in \{L, U\}$. If larger outcomes are worse, $h_\tau(z) = (z-\tau)_+$ measures excess harm: $\Theta^L_{h_\tau} > 0$ implies positive average excess under every matching; $\Theta^U_{h_\tau} = 0$ implies zero excess almost surely under every matching. A positive simultaneous lower certificate supports the former after policy search. For $h_\pm(z) = (\pm z)_+$, every admissible η satisfies

$$\Theta_{h_+}(\pi,\pi';\eta) - \Theta_{h_-}(\pi,\pi';\eta) = \mu(\pi') - \mu(\pi), \tag{17}$$

where $\mu(\pi) = \mathbb{E}_X \int\{\int_0^1 q_y(t)\,dt\}\,dP_{\pi(X),X}(y)$ is the policy mean. For $h^{(p)}_{\tau,+}(z) = (z-\tau)^p_+$, $\tau \in [0,1)$, $p \ge 1$, and $\Delta = q_{Y'} - q_Y$,

$$\Theta_{h^{(p)}_{\tau,+}}(\pi,\pi';\eta) = p\int_\tau^1 (s-\tau)^{p-1}\mathbb{E}_\eta[\lambda\{t : \Delta(t) > s\}]\,ds, \tag{18}$$

where λ is Lebesgue measure and $\mathbb{E}_\eta$ averages over X and its matching kernel. Thresholds need not share an extremizing matching. Identification and validity apply; rates require the contact and interaction conditions. See Appendix G for the dictionary and proofs.

5 EXPERIMENTS

We test calibration, distributional contrasts, and nuisance overhead (Appendix K).

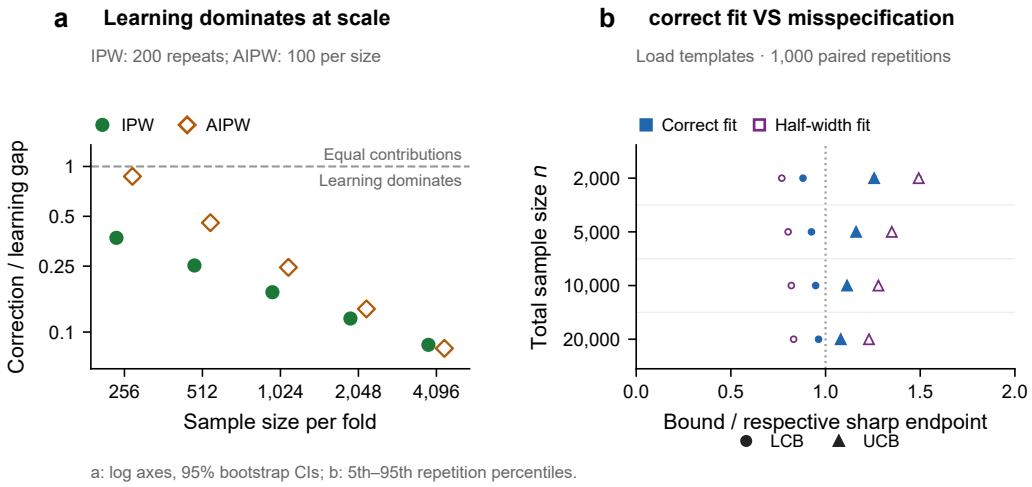


Figure 3: **Learning, calibration, and first-stage accuracy.** (a) Mean correction / mean maximal learning gap versus per-fold sample size (IPW $n/2$; AIPW $n/3$). Correction: IPW radius; AIPW nuisance bound plus radius. Bars: 95% within-codebook bootstrap intervals (200/100 repetitions per size; four codebooks, ten at the largest AIPW size). Points are horizontally offset for visibility. (b) Mean bounds normalized by their respective sharp endpoints, with 5th–95th repetition percentiles (1,000 repetitions per size). Valid lower/upper bounds lie left/right of one.

A: Learning versus calibration. For IPW with W_1, growing smooth codes have quantile slopes in $[3/16, 5/16]$. Globally feasible duals supply a simultaneous Hoeffding radius over 16 policy pairs. As per-fold size grows from 256 to 4,096, its mean divided by the mean maximal population learning gap falls from 0.371 to 0.084; all pairs cover in every repetition (Figure 3a). Relative calibration overhead decreases on this sequence of growing smooth codes (Appendix K.1.4).

B: Contrast and tightness beyond equal barycenters. Four UCI electricity-load template pairs (Trindade, 2015) give policy populations with equal barycenters but different outer heterogeneity; their sharp W_2^2 endpoints are 0.001493310 and 0.001897748. At $n = 20{,}000$, known-propensity IPW barycenters (Huang et al., 2026) and effect maps (Lin et al., 2023) have squared contrast 5.01×10^{-5} around a zero target, whereas our correct-fit mean LCB is 0.001438, positive in every repetition. Correct and misspecified fits yield mean LCB/lower and UCB/upper ratios $(0.963, 1.080)$ and $(0.831, 1.229)$; both endpoints cover in all 1,000 repetitions per size (Figure 3b; Appendices K.2.3 and K.2.2). Each endpoint is calibrated at 95%, giving at least 90% joint coverage.

C: Learning dominates nuisance error. Three-fold AIPW on growing smooth W_1 codes estimates both nuisances. Binary-remainder and variance-adaptive calibration certify all 16 policy pairs (Appendix K.1.5). At $n = 12{,}288$, mean total correction is 8.0% of the mean maximal learning gap (95% conditional bootstrap interval $[7.9\%, 8.0\%]$; Figure 3a). All pairs cover in all 100 repetitions per size. Learning dominates certified overhead at larger sizes here; small-sample calibration remains substantial.

6 DISCUSSION

Dual feasibility ensures validity; outer-law geometry governs informativeness. Our rates cover fixed finite designs and exactly observed distributions. Continuous covariates, structured lower bounds, and joint outer/within-individual minimax theory remain open.

AI USE STATEMENT

The authors used generative AI tools to assist literature discovery, prose revision, experiment implementation and debugging, and translation of selected finite-dimensional proof statements into Lean. Every citation was checked against a primary source; every reported numerical result was regenerated from the archived code and data manifest; and the Lean kernel accepted files containing zero `sorry`, `admit`, unsafe definitions, or manuscript-specific axioms. The authors reviewed all AI-assisted work and take responsibility for the final text, claims, code, and artifacts.

ETHICS STATEMENT

This work develops statistical methods and evaluates them using simulations and a semi-synthetic study based on publicly available electricity-load data. The study does not recruit human participants, conduct interventions, or collect new personal data. The semi-synthetic study uses electricity-load series as outcome templates and generates the treatment assignment and causal structure synthetically; its results should therefore be interpreted as a methodological evaluation rather than evidence about the effects of real-world policies or individuals. We use the source data in accordance with its applicable access terms and acknowledge its original creators. Because electricity-use patterns can be sensitive, applications of these methods to real individual- or household-level data should require appropriate authorization, privacy safeguards, and careful assessment of potential harms and unequal impacts. We do not identify a direct deployment context in this work.

REPRODUCIBILITY STATEMENT

Complete assumptions and proofs appear in the appendix. The experimental designs, implementation details, and artifact layout needed to reproduce the reported simulations and semi-synthetic study are documented in Appendix K.

REFERENCES

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

Eli Ben-Michael, Kosuke Imai, and Zhichao Jiang. Policy learning with asymmetric counterfactual utilities. *Journal of the American Statistical Association*, 119(548):3045–3058, 2024. doi: 10.1080/01621459.2023.2300507.

Vladimir I. Bogachev and Ilya I. Malofeev. Kantorovich problems and conditional measures depending on a parameter. *Journal of Mathematical Analysis and Applications*, 486(1):123883, 2020. doi: 10.1016/j.jmaa.2020.123883.

Marta Catalano and Hugo Lavenant. Hierarchical integral probability metrics: A distance on random probability measures with low sample complexity. *arXiv preprint arXiv:2402.00423*, 2024.

Deeparnab Chakrabarty, Prachi Goyal, and Ravishankar Krishnaswamy. The non-uniform k-center problem. *ACM Transactions on Algorithms*, 16(4):46:1–46:19, 2020. doi: 10.1145/3392720.

Siu-On Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1193–1203. SIAM, 2014.

Yifan Cui and Sukjin Han. Policy learning with distributional welfare. *Journal of the American Statistical Association*, 121(554):1324–1335, 2026. doi: 10.1080/01621459.2025.2552514.

Richard M. Dudley. The speed of mean glivenko–cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969. doi: 10.1214/aoms/1177697802.

Pedram Emami and Brendan Pass. Optimal transport with optimal transport cost: the monge–kantorovich problem on wasserstein spaces. *Calculus of Variations and Partial Differential Equations*, 64(2):43, 2025.

Fuchang Gao and Jon A. Wellner. Entropy estimate for high-dimensional monotonic functions. *Journal of Multivariate Analysis*, 98(10):2106–2117, 2007.

Zijun Gao, Shu Ge, and Jian Qian. Bridging multiple worlds: Multi-marginal optimal transport for causal partial-identification problem. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 721–729. PMLR, 2025.

David Haussler and Philip M. Long. A generalization of sauer's lemma. *Journal of Combinatorial Theory, Series A*, 71(2):219–240, 1995.

Yiyan Huang, Cheuk Hang Leung, Qi Wu, and Zhiheng Zhang. Wasserstein policy learning for distributional outcomes. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 3349–3350. PMLR, 2026.

Shayan Hundrieser, Thomas Staudt, and Axel Munk. Empirical optimal transport between different measures adapts to lower complexity. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 60(2):824–846, 2024.

Wenlong Ji, Lihua Lei, and Asher Spector. Model-agnostic covariate-assisted inference on partially identified causal effects. *arXiv preprint arXiv:2310.08115*, 2023. doi: 10.48550/arXiv.2310.08115.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Daisuke Kurisu, Yidong Zhou, Taisuke Otsu, and Hans-Georg Mueller. Geodesic causal inference. *arXiv preprint arXiv:2406.19604*, 2024.

Sirui Lin, Zijun Gao, Jose Blanchet, and Peter Glynn. Tightening causal bounds via covariate-aware optimal transport. *arXiv preprint arXiv:2502.01164*, 2025.

Sirui Lin, Zijun Gao, José Blanchet, and Peter Glynn. Causal partial identification via conditional optimal transport. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 1189–1197. PMLR, 2026.

Zhenhua Lin, Dehan Kong, and Linbo Wang. Causal inference on distribution functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2):378–398, 2023. doi: 10.1093/jrsssb/qkad008.

Pascal Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pp. 1269–1283, 1990.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.

Daniel Ober-Reynolds. Estimating functionals of the joint distribution of potential outcomes with optimal transport. *arXiv preprint arXiv:2311.09435*, 2023. doi: 10.48550/arXiv.2311.09435.

Victor M. Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6:405–431, 2019.

Gabriel Peyre and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, 2015.

Artur Trindade. Electricityloaddiagrams20112014, 2015. URL https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams201120.

Cedric Villani. *Optimal Transport: Old and New*. Springer, 2009.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019. doi: 10.3150/18-BEJ1065.

Zhengyuan Zhou, Susan Athey, and Stefan Wager. Offline multi-action policy learning: Generalization and optimization. *Operations Research*, 71(1):148–183, 2023.

Supplementary Material

CONTENTS

A Formal Setup and Assumptions **14**

B Sharp Endpoints and Conditional Duality **14**

C Estimator Definition and Quantile Notation **16**

D Finite-Sample Validity and First-Stage Tightness **18**

E Safe Quantile Computation and Finite Anchors **27**

F Geometric Minimax Rates for Policy Certificates **38**

G Directional Convex Costs and Their Rate Orders **59**

H Unknown Propensities and the Geometric Limit **61**

I Cost-Active Quantization and Polynomial Upper Rates **70**

J Finite-Bag Identification and Certificates **79**

K Experimental Designs, Numerical Results, and Reproducibility **87**

K.1 Coverage and calibration 87

K.2 Endpoint tightness and first-stage learning 100

K.3 Geometry and computational diagnostics 108

K.4 Finite bags and measurement error 117

K.5 Reproducibility and formal verification 119

Guide to the proofs. Appendix B proves the sharpness and duality claim in Theorem 2.1. Appendices D and E establish IPW validity and its finite implementation; Appendix H gives the AIPW transfer. Appendix F proves the geometric lower and upper rates, with the unknown-propensity extension again in Appendix H. The directional-cost consequences, cost-active quantization bound, and finite-bag results follow in Appendices G, I, and J.

A Formal Setup and Assumptions

Let $I = [I_-, I_+] \subset \mathbb{R}$ be compact and let $\mathcal{Z} = \mathcal{P}(I)$ be the compact Polish space of Borel probability measures on I, equipped with weak convergence. The two probability levels are central: an individual potential outcome $Y(a)$ is a probability measure in $\mathcal{Z}$, whereas

$$P_{a,x} = \mathcal{L}\{Y(a) \mid X = x\} \in \mathcal{P}(\mathcal{Z})$$

is an *outer law* describing population variation among individual measures. Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ be standard Borel, $\mathcal{A} = \{1, \dots, d\}$ a finite action set, and Π a class of measurable deterministic policies $\pi : \mathcal{X} \to \mathcal{A}$. In the baseline observation model we observe i.i.d.

$$O_i = (X_i, A_i, Y_i), \qquad Y_i = Y_i(A_i), \qquad i = 1, \dots, n, \tag{19}$$

with propensity $e_a(x) = \mathbb{P}(A = a \mid X = x)$. Section J replaces exact observation of Y_i by a finite bag of measurements from it.

Assumption A.1 (Bounded diagonal-zero cost)**.** *The analyst fixes a cost $c : \mathcal{X} \times \mathcal{Z} \times \mathcal{Z} \to [0, C_c]$ that is jointly measurable, continuous in (y, y') for every x, and satisfies $c_x(y, y) = 0$. We write $c_x(y, y') = c(x, y, y')$.*

Assumption A.2 (Consistency and no interference)**.** *If $A = a$, then $Y = Y(a)$ almost surely, and one unit's action does not change another unit's potential distribution-valued outcome.*

Assumption A.3 (Conditional exchangeability of treatment)**.** *For every $a \in \mathcal{A}$, $Y(a) \perp\!\!\!\perp A \mid X$.*

Assumption A.4 (Overlap)**.** *There is $\underline{e} > 0$ such that $e_a(x) \geq \underline{e}$ for every action a and for P_X-almost every x.*

A.1 The population matching set

For each action and covariate value, Assumptions A.2–A.4 identify the policy-specific outer law $P_{a,x}$. We compare two distinct individuals drawn from the corresponding policy populations, conditional on the same covariate value x. Their pairing is described by a matching kernel, rather than by the joint potential outcomes of one individual. For every action pair, the admissible matching set is

$$\mathfrak{G}_x(a, b) = \Gamma(P_{a,x}, P_{b,x}), \qquad (a, b) \in \mathcal{A}^2, \tag{20}$$

where $\Gamma(P, Q)$ contains all couplings with marginals P and Q. The same rule applies to every pair of policy populations.

For policies $\pi, \pi' \in \Pi$ and a measurable matching kernel $\eta = (\eta_x)_x$ with $\eta_x \in \mathfrak{G}_x\{\pi(x), \pi'(x)\}$, define

$$\Theta_c(\pi, \pi'; \eta) = \mathbb{E}_X \left[\int_{\mathcal{Z}^2} c_X(y, y') \, d\eta_X(y, y') \right]. \tag{21}$$

This is the expected cost of matching individual distributions across the two policy populations. The question is how small or large this cost can be as the matching mechanism varies while preserving the outer marginals. It is not the effect of changing the action of a particular individual.

B Sharp Endpoints and Conditional Duality

Definition B.1 (Sharp population matching bounds)**.** *For $\pi, \pi' \in \Pi$, define*

$$\Theta_c^L(\pi, \pi') = \mathbb{E}_X \left[\inf_{\gamma \in \Gamma(P_{\pi(X),X}, P_{\pi'(X),X})} \int c_X(y, y') \, d\gamma(y, y') \right], \tag{22}$$

$$\Theta_c^U(\pi, \pi') = \mathbb{E}_X \left[\sup_{\gamma \in \Gamma(P_{\pi(X),X}, P_{\pi'(X),X})} \int c_X(y, y') \, d\gamma(y, y') \right]. \tag{23}$$

Theorem B.2 (The observed outer marginals determine a sharp interval)**.** *Suppose Assumptions A.1–A.4 hold. For every fixed policy pair (π, π') and every conditional joint law η in (20),*

$$\Theta_c^L(\pi, \pi') \le \Theta_c(\pi, \pi'; \eta) \le \Theta_c^U(\pi, \pi').$$

Both conditional extrema are attained, and measurable extremizing kernels exist. The interval is sharp for the fixed policy pair: every value in it is attained by a population matching kernel with the identified marginals.

The lower endpoint has the conditional Kantorovich representation

$$\Theta_c^L(\pi, \pi') = \mathbb{E}_X\left[\sup_{(\varphi,\psi)\in\mathcal{V}_{X,\pi,\pi'}} \{P_{\pi(X),X}\varphi + P_{\pi'(X),X}\psi\}\right], \tag{24}$$

where

$$\mathcal{V}_{x,\pi,\pi'} = \{(\varphi_x, \psi_x) : \varphi_x(y) + \psi_x(y') \le c_x(y, y') \text{ for all } y, y' \in \mathcal{Z}\}. \tag{25}$$

The corresponding upper representation is

$$\Theta_c^U(\pi, \pi') = \mathbb{E}_X\left[\inf_{(\varphi,\psi)\in\mathcal{V}^U_{X,\pi,\pi'}} \{P_{\pi(X),X}\varphi + P_{\pi'(X),X}\psi\}\right], \tag{26}$$

where $\mathcal{V}^U_{x,\pi,\pi'}$ reverses the inequality in (25).

Corollary B.3 (Global feasibility fixes the direction of both endpoints)**.** *Let $\nu = (\varphi_x, \psi_x)_x$ have finite population marginal integrals and be measurable and dual feasible for P_X-almost every x. Then*

$$g_{\pi,\pi'}(\nu) := \mathbb{E}_X\{P_{\pi(X),X}\varphi_X + P_{\pi'(X),X}\psi_X\} \le \Theta_c^L(\pi, \pi') \le \Theta_c(\pi, \pi'; \eta) \tag{27}$$

for every admissible η. If $\nu^U = (u_x, v_x)_x$ has finite population marginal integrals and is measurable and globally superfeasible, $u_x(y) + v_x(y') \ge c_x(y, y')$, then

$$\Theta_c(\pi, \pi'; \eta) \le \Theta_c^U(\pi, \pi') \le g^U_{\pi,\pi'}(\nu^U) := \mathbb{E}_X\{P_{\pi(X),X}u_X + P_{\pi'(X),X}v_X\} \tag{28}$$

for every admissible η.

Remark B.4 (Pairwise matching scope)**.** *Each policy pair has its own admissible matching mechanisms. Sharpness is claimed for each fixed pair; no common multi-population matching mechanism is imposed on separately optimized pairs.*

Proof of Theorem B.2. Write $P_x = P_{\pi(x),x}$, $Q_x = P_{\pi'(x),x}$, and $\Gamma_x = \Gamma(P_x, Q_x)$. Consistency and conditional exchangeability identify $P_{a,x}$ with $\mathcal{L}(Y \mid A = a, X = x)$, while overlap makes this law observable for every relevant action and P_X-almost every x.

Because $\mathcal{Z}$ is compact Polish, Γ_x is nonempty and compact in the weak topology. The map

$$C_x(\gamma) = \int_{\mathcal{Z}\times\mathcal{Z}} c_x(y, y')\, d\gamma(y, y')$$

is continuous because c_x is bounded and continuous. Hence its minimum and maximum are attained. The kernels $x \mapsto P_x, Q_x$ are Borel and the cost is jointly measurable and continuous in its last two arguments. Put $C_c = \sup_{x,y,y'} |c_x(y, y')|$. Applying the parameterized Kantorovich theorem (Bogachev & Malofeev, 2020, Theorem 4.1) to the nonnegative costs $c_x + C_c$ and $C_c - c_x$ gives measurable optimizing kernels $x \mapsto \gamma_x^L, \gamma_x^U$ for all covariate strata. The constant shifts do not alter the optimizers.

These kernels are measurable and admissible. For any admissible kernel $x \mapsto \gamma_x$, pointwise optimality gives

$$C_x(\gamma_x^L) \le C_x(\gamma_x) \le C_x(\gamma_x^U) \qquad P_X\text{-almost surely.}$$

Integration yields

$$\Theta_c^L(\pi, \pi') \le \Theta_c(\pi, \pi'; \eta) \le \Theta_c^U(\pi, \pi').$$

The two endpoints are attained by γ^L and γ^U. If the endpoints differ and $t \in [\Theta_c^L, \Theta_c^U]$, set

$$\lambda = \frac{t - \Theta_c^L}{\Theta_c^U - \Theta_c^L}, \qquad \gamma_x^t = (1 - \lambda)\gamma_x^L + \lambda\gamma_x^U.$$

The mixture has marginals P_x, Q_x and integrated cost t. Combining it with P_X constructs a joint distribution for two matched individuals from the policy populations. This proves sharpness for the fixed policy pair without imposing a common matching mechanism across pairs.

For fixed x, classical Kantorovich duality on compact Polish spaces with continuous cost (Villani, 2009, Chapter 5) gives

$$\min_{\gamma\in\Gamma_x}\int c_x\,d\gamma = \sup_{\varphi(y)+\psi(y')\le c_x(y,y')}\{P_x\varphi + Q_x\psi\}.$$

Bounded continuous potentials suffice. The primal value is measurable by the parameterized theorem, so no measurable choice of a dual optimizer is needed to integrate the equality. Integrating over P_X gives the lower endpoint's dual representation. Applying the same argument to $-c_x$ yields

$$\max_{\gamma\in\Gamma_x}\int c_x\,d\gamma = \inf_{\varphi(y)+\psi(y')\ge c_x(y,y')}\{P_x\varphi + Q_x\psi\},$$

and gives the upper endpoint's dual representation after integration over P_X. If a measurable collection (u_x, v_x) is globally superfeasible, then for every conditional coupling γ_x,

$$P_x u_x + Q_x v_x = \int\{u_x(y) + v_x(y')\}\,d\gamma_x(y,y') \ge \int c_x(y,y')\,d\gamma_x(y,y').$$

Taking the supremum over γ_x and then integrating over P_X proves that its population dual value is at least Θ_c^U. Equivalently, $\Theta_c^U = C_c - \Theta_{C_c-c}^L$ when C_c is a constant upper bound on the cost; this identity is useful for identification, while the direct superfeasible construction retains the one-sided binning correction used in computation. □

Proof of Corollary B.3. Fix an admissible η. For almost every covariate value, global feasibility implies, for every $\gamma_x \in \Gamma(P_{\pi(x),x}, P_{\pi'(x),x})$,

$$P_{\pi(x),x}\varphi_x + P_{\pi'(x),x}\psi_x = \int\{\varphi_x(y) + \psi_x(y')\}\,d\gamma_x \le \int c_x(y,y')\,d\gamma_x.$$

Taking the infimum gives the conditional weak-duality bound. Integrating over P_X gives $g_{\pi,\pi'}(\nu) \le \Theta_c^L(\pi,\pi')$; the inequality $\Theta_c^L(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)$ follows from Theorem B.2.

For a globally superfeasible collection, the same marginal calculation with the reversed pointwise inequality gives, for every conditional coupling,

$$P_{\pi(x),x}u_x + P_{\pi'(x),x}v_x \ge \int c_x(y,y')\,d\gamma_x(y,y').$$

Taking the supremum and integrating gives $g_{\pi,\pi'}^U(\nu^U) \ge \Theta_c^U(\pi,\pi')$; the inequality $\Theta_c(\pi,\pi';\eta) \le \Theta_c^U(\pi,\pi')$ follows from Theorem B.2. □

C Estimator Definition and Quantile Notation

The first fold learns globally feasible dual potentials; the independent second fold estimates their value by IPW. Propensities are known in this construction. The next appendix proves validity and separates evaluation error from the training gap.

Definition C.1 (Dual lower-bound estimator)**.** *Given data $\{(X_i, A_i, Y_i)\}_{i=1}^n$, randomly partition the sample into two disjoint folds D_1 and D_2, independently of their values, and write $\mathcal{F}_1 = \sigma(D_1)$. For fixed policies $\pi, \pi' \in \Pi$, proceed as follows.*

Step 1: estimate a feasible dual pair. *Using D_1, construct an estimator*

$$\widehat{\nu} = (\widehat{\phi}_x, \widehat{\psi}_x)_{x\in\mathcal{X}}$$

of a dual optimizer such that, for P_X-almost every x,

$$(\widehat{\phi}_x, \widehat{\psi}_x) \in \mathcal{V}_{x,\pi,\pi'}. \tag{29}$$

One generic construction has two parts. First, estimate the conditional laws $\widehat{P}_{a,x}$ of $Y(a) \mid X = x$, for each $a \in \mathcal{A}$, using any appropriate regression or machine-learning method. Second, for every covariate value x needed on the evaluation fold, solve the plug-in conditional dual problem

$$(\widehat{\phi}_x, \widehat{\psi}_x) \in \underset{(\phi_x,\psi_x)\in\mathcal{V}_{x,\pi,\pi'}}{\arg\max} \left\{ \widehat{P}_{\pi(x),x}\phi_x + \widehat{P}_{\pi'(x),x}\psi_x \right\}. \tag{30}$$

When (30) has multiple solutions, the construction must specify a jointly measurable selection. Existence and measurability of exact plug-in optimizers are conditions of the tightness results. The finite support construction in Appendix E supplies the computational alternative for validity. Only the values $(\widehat{\phi}_{X_i}, \widehat{\psi}_{X_i})$ for $i \in D_2$ are required; thus, it is unnecessary to solve (30) over all of $\mathcal{X}$.

Step 2: estimate the dual value by IPW. *Let*

$$\widetilde{\Theta}_L(\pi, \pi') := g_{\pi,\pi'}(\widehat{\nu}).$$

By Corollary B.3, $\widetilde{\Theta}_L(\pi, \pi') \le \Theta_c^L(\pi, \pi')$ for every realization of the training fold on which feasibility holds. On the independent fold D_2, define the IPW score

$$S_i(\pi, \pi') = \left\{ \frac{\mathbb{1}\{A_i = \pi(X_i)\}}{e_{\pi(X_i)}(X_i)} \widehat{\phi}_{X_i}(Y_i) + \frac{\mathbb{1}\{A_i = \pi'(X_i)\}}{e_{\pi'(X_i)}(X_i)} \widehat{\psi}_{X_i}(Y_i) \right\}, \tag{31}$$

and the estimator

$$\widehat{\Theta}_L(\pi, \pi') = \frac{1}{|D_2|} \sum_{i\in D_2} S_i(\pi, \pi'). \tag{32}$$

Conditional on D_1, the scores in (31) are i.i.d. and

$$\mathbb{E}[S_i(\pi, \pi') \mid D_1] = \widetilde{\Theta}_L(\pi, \pi').$$

Thus, $\widehat{\Theta}_L(\pi, \pi')$ is conditionally unbiased for an $\mathcal{F}_1$-measurable lower bound on the sharp population matching endpoint $\Theta_c^L(\pi, \pi')$.

Let $m = |D_2|$, and let $\widehat{\sigma}_{\pi,\pi'}$ be the sample standard deviation of $\{S_i(\pi, \pi') : i \in D_2\}$. For a policy pair fixed before observing D_2, define the pointwise Gaussian candidate

$$\widehat{L}_\alpha^{\mathrm{pt}}(\pi, \pi') = \widehat{\Theta}_L(\pi, \pi') - \Phi^{-1}(1-\alpha)\frac{\widehat{\sigma}_{\pi,\pi'}}{\sqrt{m}}, \tag{33}$$

where Φ is the standard Gaussian distribution function. This candidate is an asymptotically valid pointwise lower confidence bound only under the moment, nondegeneracy, and evaluation-sample growth conditions of Corollary D.7. Finite-sample-valid certificates instead use the Hoeffding–Rademacher radii in Theorem D.3(ii).

In one dimension, every $y \in \mathcal{P}_2(I)$ is represented by its left-continuous nondecreasing quantile function $Q_y \in \mathcal{Q} \subset L^2([0,1])$. For a finite continuous convex displacement cost h, the associated distributional cost admits the exact quantile integral representation

$$c_h(y, y') = \int_0^1 h\big(Q_{y'}(t) - Q_y(t)\big)\, dt. \tag{34}$$

When $h(z) = |z|^2$, this reduces to the classical Wasserstein–L^2 isometry

$$W_2^2(y, y') = \int_0^1 |Q_y(t) - Q_{y'}(t)|^2\, dt, \tag{35}$$

but for a general convex h the right-hand side of (34) need not define a metric or a Hilbert norm. Dual potentials on $\mathcal{P}_2(I)$ may therefore be written as functionals of the quantile curve,

$$\varphi_x(y) = \Phi_x(Q_y), \qquad \psi_x(y) = \Psi_x(Q_y),$$

and feasibility for $c_X = c_h$ becomes (Santambrogio, 2015)

$$\Phi_x(q) + \Psi_x(q') \le \int_0^1 h\big(q'(t) - q(t)\big)\, dt, \qquad q, q' \in \mathcal{Q}. \tag{36}$$

The displacement is target minus source: $h(q' - q)$. Reversing the order replaces h by $\check{h}(z) = h(-z)$.

We normalize $h(0) = 0$ and assume that zero is a global minimizer of the convex displacement cost, equivalently $0 \in \partial h(0)$. Hence both $r \mapsto h(r)$ and $r \mapsto h(-r)$ are nondecreasing for $r \ge 0$, while positive and negative displacements may have different costs.

D Finite-Sample Validity and First-Stage Tightness

The fixed-policy construction in Section C is valid for any dual-feasible pair. To obtain a certificate that is uniform over Π^2, construct on D_1 an *action-pair-indexed* collection

$$\widehat{\nu} = \left\{ (\widehat{\phi}_{ab,x}, \widehat{\psi}_{ab,x}) : (a, b, x) \in \mathcal{A}^2 \times \mathcal{X} \right\},$$

whose elements are measurable with respect to $\mathcal{F}_1 = \sigma(D_1)$. Let $m = |D_2|$.

Assumption D.1 (Simultaneous global dual feasibility)**.** *With probability one in the training fold, the following condition holds on a common set of covariate values of full P_X-measure. For every $a, b \in \mathcal{A}$ and every $(z, z') \in \mathcal{Z}^2$,*

$$\widehat{\phi}_{ab,x}(z) + \widehat{\psi}_{ab,x}(z') \le c_x(z, z'). \tag{37}$$

Thus, for every policy pair, set

$$\widehat{\phi}_x^{\pi,\pi'} = \widehat{\phi}_{\pi(x),\pi'(x),x}, \qquad \widehat{\psi}_x^{\pi,\pi'} = \widehat{\psi}_{\pi(x),\pi'(x),x}.$$

Assumption D.1 makes this pair dual-feasible simultaneously for every $(\pi, \pi') \in \Pi^2$. It does not force same-action potentials to vanish. For the maintained nonnegative, diagonal-zero cost, choosing the zero pair for identical marginals is an optional optimal lower-endpoint normalization, not a restriction on the population matching set.

For $i \in D_2$, define

$$S_i(\pi, \pi') = \left\{ \frac{\mathbb{1}\{A_i = \pi(X_i)\}}{e_{\pi(X_i)}(X_i)} \widehat{\phi}_{X_i}^{\pi,\pi'}(Y_i) + \frac{\mathbb{1}\{A_i = \pi'(X_i)\}}{e_{\pi'(X_i)}(X_i)} \widehat{\psi}_{X_i}^{\pi,\pi'}(Y_i) \right\}, \tag{38}$$

$$\widehat{\Theta}_L(\pi, \pi') = \frac{1}{m} \sum_{i \in D_2} S_i(\pi, \pi'). \tag{39}$$

Policy-class complexity. We measure the intrinsic size of the multi-action policy class by the Hamming entropy integral used by Zhou et al. (2023). For a finite sequence $\boldsymbol{x} = (x_1, \ldots, x_q)$, let

$$H_{\boldsymbol{x}}(\pi, \widetilde{\pi}) = \frac{1}{q} \sum_{j=1}^{q} \mathbb{1}\{\pi(x_j) \ne \widetilde{\pi}(x_j)\}.$$

Let $N_H(\delta, \Pi, \boldsymbol{x})$ be the smallest cardinality of a subset of Π whose δ-balls under $H_{\boldsymbol{x}}$ cover Π, and define the worst-case Hamming covering number and entropy integral by

$$N_H(\delta, \Pi) = \sup_{q \ge 1,\, \boldsymbol{x} \in \mathcal{X}^q} N_H(\delta, \Pi, \boldsymbol{x}), \qquad \mathcal{K}_{\mathrm{H}}(\Pi) = \int_0^1 \sqrt{\log N_H(u^2, \Pi)}\, du. \tag{40}$$

Thus $\mathcal{K}_{\mathrm{H}}(\Pi)$ (the quantity denoted $\kappa(\Pi)$ in Zhou et al. (2023)) measures policy-class complexity. The Rademacher complexity serves as an intermediate empirical-process quantity connecting (40) to a simultaneous confidence radius. Finiteness of $\mathcal{K}_{\mathrm{H}}(\Pi)$ is needed only for the entropy-calibrated radius below; the data-adaptive Rademacher radius remains valid without it.

Assumption D.2 (Root-sample uniform-inference regime)**.** *Consider a sequence of problems indexed by the total sample size n. Let $m_n = |D_{2,n}|$ be the evaluation-fold size, let Π_n be the ambient deterministic policy class, and let $\alpha_n \in (0,1)$ be the confidence error probability. Suppose:*

(i) $m_n \asymp n$;

(ii) *there are deterministic envelopes M_n such that*

$$\sup_{a,b,x,z} \left(|\widehat{\phi}_{ab,x}(z)| \vee |\widehat{\psi}_{ab,x}(z)| \right) \le M_n \quad \textit{almost surely}, \qquad \sup_n M_n < \infty;$$

(iii) *the overlap constants may depend on n, but*

$$e_{a,n}(x) \ge \underline{e}_n \quad \textit{for every } a \in \mathcal{A} \textit{ and almost every } x, \qquad \inf_n \underline{e}_n > 0;$$

(iv) *the policy entropy is uniformly bounded,*

$$\sup_n \mathcal{K}_{\mathrm{H}}(\Pi_n) < \infty;$$

(v) *the confidence levels satisfy*

$$\sup_n \log(1/\alpha_n) < \infty.$$

Assumption D.2 is used only to state the rate of the safety correction. The finite-sample coverage result below remains valid at each n under its stated per-sample conditions even when this rate assumption fails.

D.1 Uniform validity

Theorem D.3 (Uniform validity of split-sample IPW dual certificates)**.** *Suppose the conditions of Theorem B.2 hold, the propensity scores are known as in Section C, and Assumption D.1 holds. Assume all quantities below are measurable; for an uncountable policy class it is enough that the conditional score class is pointwise measurable.*

(i) *For every $(\pi, \pi') \in \Pi^2$, conditional on $\mathcal{F}_1$, the scores in (38) are i.i.d. and*

$$\begin{aligned} \mu_{\pi,\pi'} &:= \mathbb{E}[S_i(\pi,\pi') \mid \mathcal{F}_1] \\ &= \mathbb{E}_X \left[P_{\pi(X),X} \widehat{\phi}_X^{\pi,\pi'} + P_{\pi'(X),X} \widehat{\psi}_X^{\pi,\pi'} \,\middle|\, \mathcal{F}_1 \right] \le \Theta_c^L(\pi,\pi'). \end{aligned} \tag{41}$$

The inequalities in (41) hold simultaneously over Π^2. Moreover, for every admissible population matching law η, Theorem B.2 gives $\Theta_c^L(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)$. No consistency or rate condition is imposed on the first-stage estimates used to construct $\widehat{\nu}$.

(ii) *Suppose, in addition, that for a deterministic $M < \infty$,*

$$\sup_{a,b,x,z} \left(|\widehat{\phi}_{ab,x}(z)| \vee |\widehat{\psi}_{ab,x}(z)| \right) \le M \quad \textit{almost surely.} \tag{42}$$

Let

$$B_S = \frac{2M}{\underline{e}}, \qquad \mathcal{S}(\mathcal{F}_1) = \{O \mapsto S(\pi,\pi';O) : (\pi,\pi') \in \Pi^2\},$$

and define the conditional empirical Rademacher complexity

$$\widehat{\mathfrak{R}}_m = \mathbb{E}_\varepsilon \left[\sup_{(\pi,\pi')\in\Pi^2} \frac{1}{m} \sum_{i \in D_2} \varepsilon_i S_i(\pi,\pi') \,\middle|\, \mathcal{F}_1, D_2 \right], \tag{43}$$

where ε_i are independent Rademacher variables, independent of the data. The action-pair indexing of the fitted duals transfers Hamming covers of Π to empirical L^2-covers of the conditional score class. In particular, whenever $\mathcal{K}_{\mathrm{H}}(\Pi) < \infty$,

$$\widehat{\mathfrak{R}}_m \le \frac{48 B_S \mathcal{K}_{\mathrm{H}}(\Pi)}{\sqrt{m}} \quad \textit{almost surely.} \tag{44}$$

For $\alpha \in (0,1)$, put

$$r_{m,\alpha} = 2\widehat{\mathfrak{R}}_m + 3B_S\sqrt{\frac{2\log(2/\alpha)}{m}}, \tag{45}$$

$$\widehat{L}^{\mathrm{unif}}_{\alpha}(\pi,\pi') = \widehat{\Theta}_L(\pi,\pi') - r_{m,\alpha}. \tag{46}$$

If $\mathcal{K}_{\mathrm{H}}(\Pi) < \infty$, the entropy-calibrated radius

$$r^{\mathrm{ent}}_{m,\alpha} = \frac{96B_S\mathcal{K}_{\mathrm{H}}(\Pi)}{\sqrt{m}} + 3B_S\sqrt{\frac{2\log(2/\alpha)}{m}} \tag{47}$$

satisfies $r_{m,\alpha} \le r^{\mathrm{ent}}_{m,\alpha}$ almost surely and may replace $r_{m,\alpha}$ in (46). Then, for every sample size,

$$\mathbb{P}\Big\{\widehat{L}^{\mathrm{unif}}_{\alpha}(\pi,\pi') \le \Theta^L_c(\pi,\pi') \quad \text{for every } (\pi,\pi') \in \Pi^2\Big\} \ge 1-\alpha. \tag{48}$$

On the same event, $\widehat{L}^{\mathrm{unif}}_{\alpha}(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)$ for every policy pair and every admissible η for that pair. Consequently, (48) holds for every data-dependent pair $(\widehat{\pi},\widehat{\pi}') \in \Pi^2$, even when that pair is selected using D_2.

If Π is finite and $N_{\mathrm{pair}} = |\Pi|^2$, the simpler computable radius

$$r^{\mathrm{fin}}_{m,\alpha} = B_S\sqrt{\frac{2\log(N_{\mathrm{pair}}/\alpha)}{m}} \tag{49}$$

may replace (45).

The proof is given in Appendix D.3.

Theorem D.4 (Uniform upper certificates and a two-sided sharp interval)**.** *In addition to the conditions of Theorem D.3, suppose the training fold produces a measurable action-pair-indexed collection $(\widehat{\phi}^U_{ab,x}, \widehat{\psi}^U_{ab,x})$ satisfying, on one common full-measure set of covariate values,*

$$\widehat{\phi}^U_{ab,x}(z) + \widehat{\psi}^U_{ab,x}(z') \ge c_x(z,z') \quad \text{for every } a,b,z,z'. \tag{50}$$

Define $S^U_i(\pi,\pi')$ by (38) with the upper collection in place of the lower collection, and put

$$\mu^U_{\pi,\pi'} := \mathbb{E}\{S^U_i(\pi,\pi') \mid \mathcal{F}_1\}.$$

Then, simultaneously over Π^2,

$$\mu^U_{\pi,\pi'} \ge \Theta^U_c(\pi,\pi'). \tag{51}$$

If $|\widehat{\phi}^U| \vee |\widehat{\psi}^U| \le M_U$ almost surely, define $B^U_S = 2M_U/\underline{e}$, the conditional empirical Rademacher complexity $\widehat{\mathfrak{R}}^U_m$ of the upper score class, and

$$r^U_{m,\alpha} = 2\widehat{\mathfrak{R}}^U_m + 3B^U_S\sqrt{\frac{2\log(2/\alpha)}{m}}, \qquad \widehat{U}^{\mathrm{unif}}_{\alpha}(\pi,\pi') = \frac{1}{m}\sum_{i\in D_2} S^U_i(\pi,\pi') + r^U_{m,\alpha}. \tag{52}$$

The entropy and finite-class alternatives in Theorem D.3(ii) apply after replacing B_S by B^U_S. For every sample size,

$$\mathbb{P}\Big\{\Theta^U_c(\pi,\pi') \le \widehat{U}^{\mathrm{unif}}_{\alpha}(\pi,\pi') \text{ for every } (\pi,\pi') \in \Pi^2\Big\} \ge 1-\alpha. \tag{53}$$

Consequently,

$$\mathbb{P}\Big\{[\Theta^L_c(\pi,\pi'), \Theta^U_c(\pi,\pi')] \subseteq [\widehat{L}^{\mathrm{unif}}_{\alpha/2}(\pi,\pi'), \widehat{U}^{\mathrm{unif}}_{\alpha/2}(\pi,\pi')] \ \forall (\pi,\pi') \in \Pi^2\Big\} \ge 1-\alpha. \tag{54}$$

The common event continues to cover a policy pair selected using D_2. No first-stage consistency condition is needed for either endpoint's directional validity.

The proof is given in Appendix D.4.

Corollary D.5 (Root-sample uniform safety radius)**.** *Suppose the conditions of Theorem D.3 hold for every n, and Assumption D.2 holds. Then*

$$r_{m_n,\alpha_n} \le r^{\mathrm{ent}}_{m_n,\alpha_n} = O(n^{-1/2}).$$

Thus simultaneous post-selection validity over Π_n^2 is obtained with a root-sample safety correction. This rate statement does not imply that the worst-case constant is small or that the resulting finite-sample lower certificate is informative. If the upper collections satisfy the analogous uniformly bounded envelope and entropy conditions, then $r^{U}_{m_n,\alpha_n} = O(n^{-1/2})$ as well. Consequently, the two-sided sharp-interval band in (54) has a root-sample safety correction at both endpoints.

The proof is given in Appendix D.5.

Remark D.6 (Policy classes covered relative to Zhou et al.)**.** *The deterministic policy classes admitted by Assumption 3 of Zhou et al. (2023) are covered here. Indeed, their condition*

$$N_H(\delta,\Pi) \le C\exp\{D\delta^{-\omega}\}, \qquad 0<\omega<\tfrac{1}{2},$$

implies

$$\mathcal{K}_{\mathrm{H}}(\Pi) \le \sqrt{\log C} + \frac{\sqrt{D}}{1-\omega} < \infty.$$

Their concrete multi-action depth-L tree class Π_L, with p_X covariates and d actions, satisfies

$$\log N_H(\delta,\Pi_L) \le (2^L-1)\log p_X + 2^L\log d + (2^L-1)\log\left(\frac{L}{\delta}+1\right).$$

Consequently, fixed (L,p_X,d) gives $\mathcal{K}_{\mathrm{H}}(\Pi_L)<\infty$, so this multi-action tree class satisfies the policy-complexity requirement in Assumption D.2. Together with the split, envelope, overlap, and confidence-level requirements there, Corollary D.5 applies. More generally, an a priori deterministic constrained subclass inherits the same entropy order, up to an immaterial constant rescaling of the covering radius.

If Π is finite, then $N_H(\delta,\Pi)\le|\Pi|$ and hence $\mathcal{K}_{\mathrm{H}}(\Pi)\le\sqrt{\log|\Pi|}$. If the action set is binary and Π has finite VC dimension, the Hamming covering bound of Haussler & Long (1995) gives

$$\mathcal{K}_{\mathrm{H}}(\Pi) \le 2.5\sqrt{\mathrm{VC}(\Pi)};$$

Thus the entropy formulation also recovers the usual finite-class and binary-VC orders. This remark concerns the deterministic policy classes used in the tree analysis of Zhou et al. (2023). Randomized policies taking values in $\Delta(\mathcal{A})$ require a different metric.

Corollary D.7 (Pointwise Gaussian asymptotic lower confidence bound)**.** *Fix in advance one policy pair (π,π'), or choose it using D_1 only. Fix a data-generating law P satisfying Assumptions A.1–A.4, with known propensity scores, and consider a sequence of sample splits such that $m=m_n=|D_{2,n}|\to\infty$. For every n, suppose the fitted fixed-pair potentials are measurable and globally dual feasible; if the pair is selected using D_1, these properties must hold for the selected pair almost surely. Write*

$$\mu_n = \mathbb{E}_P(S_i \mid \mathcal{F}_1), \qquad \sigma_n^2 = \mathrm{Var}_P(S_i \mid \mathcal{F}_1),$$

where the dependence of the score on n, through its first-stage construction, is left implicit. Suppose there are deterministic constants $C_4<\infty$ and $\underline{v}>0$, independent of n and the realization of D_1, such that, for every n, almost surely in D_1,

$$\mathbb{E}_P\left[|S_i(\pi,\pi')-\mu_n|^4 \,\middle|\, \mathcal{F}_1\right] \le C_4, \qquad \sigma_n^2 \ge \underline{v}. \tag{55}$$

Then

$$\widehat{L}^{\mathrm{pt}}_{\alpha}(\pi,\pi') = \widehat{\Theta}_L(\pi,\pi') - \Phi^{-1}(1-\alpha)\frac{\widehat{\sigma}_{\pi,\pi'}}{\sqrt{m}}$$

satisfies

$$\liminf_{n\to\infty}\mathbb{P}_P\left\{\widehat{L}^{\mathrm{pt}}_{\alpha}(\pi,\pi') \le \Theta^L_c(\pi,\pi')\right\} \ge 1-\alpha.$$

Consequently, for every admissible population matching law η, the same lower limit is at least $1-\alpha$ after replacing the event by $\{\widehat{L}^{\mathrm{pt}}_{\alpha}(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)\}$.

The proof is given in Appendix D.6.

Remark D.8 (Why the critical value must change after policy search)**.** *Simultaneous dual feasibility controls the direction of first-stage bias, but it does not control the maximum evaluation-fold sampling error over Π^2. Therefore the pointwise $\Phi^{-1}(1-\alpha)$ critical value in Corollary D.7 is not a simultaneous $(1-\alpha)$-critical value. A policy selected using D_2 requires either the finite-sample uniform radius in part (ii) or another valid uniform empirical-process critical value.*

D.2 First-stage tightness

Validity does not require a consistent first stage, but tightness does. For each (a, x), let $\widehat{P}_{a,x}$ denote the law fitted on D_1. For every action pair (a, b), the following result assumes that the corresponding feasible dual pair is an exact optimizer of the conditional dual objective with $P_{a,x}$ and $P_{b,x}$ replaced by $\widehat{P}_{a,x}$ and $\widehat{P}_{b,x}$, respectively.

For $a, b \in \mathcal{A}$, use the data-adaptive dominating measures

$$\lambda_{a,x} = P_{a,x} + \widehat{P}_{a,x}, \qquad p_{a,x} = \frac{\mathrm{d}P_{a,x}}{\mathrm{d}\lambda_{a,x}}, \qquad \widehat{p}_{a,x} = \frac{\mathrm{d}\widehat{P}_{a,x}}{\mathrm{d}\lambda_{a,x}}.$$

For a policy pair, let $a_x = \pi(x)$, $b_x = \pi'(x)$, and equip

$$\mathbb{H}_{a_x,b_x,x} = L^2(\lambda_{a_x,x}) \oplus L^2(\lambda_{b_x,x})$$

with its usual direct-sum norm $\|\cdot\|_{a_x,b_x,x}$. Let $(\phi^\star_{a,b,x}, \psi^\star_{a,b,x})$ be a measurable population conditional dual optimizer for each action pair. Define

$$\mathrm{err}_P(x;\pi,\pi') := \|(p_{a_x,x} - \widehat{p}_{a_x,x}, p_{b_x,x} - \widehat{p}_{b_x,x})\|_{a_x,b_x,x}\,, \tag{56}$$

$$\mathrm{err}_\nu(x;\pi,\pi') := \left\|\left(\phi^\star_{a_x,b_x,x} - \widehat{\phi}_{a_x,b_x,x}, \psi^\star_{a_x,b_x,x} - \widehat{\psi}_{a_x,b_x,x}\right)\right\|_{a_x,b_x,x}. \tag{57}$$

Theorem D.9 (First-stage product inequality for distribution-valued outcomes)**.** *Suppose that:*

1. *the conditional primal and dual problems satisfy strong duality and admit the measurable population optimizers above;*
2. *for every action pair, the first-stage potentials are exact plug-in optimizers as in the preceding paragraph;*
3. *the errors in (56) and (57) are measurable and finite almost surely.*

Then, conditional on D_1, for every policy pair (π, π'),

$$0 \le \Theta^L_c(\pi,\pi') - \widetilde{\Theta}_L(\pi,\pi') \le \mathbb{E}_X[\mathrm{err}_P(X;\pi,\pi')\,\mathrm{err}_\nu(X;\pi,\pi') \mid D_1]\,. \tag{58}$$

The result is finite-sample and the multiplicative constant is exactly one.

The proof is given in Appendix D.7.

Corollary D.10 (Outer $L^2(P_X)$ product rate)**.** *Define*

$$\mathcal{E}_P(\pi,\pi') := \left\{\mathbb{E}_X[\mathrm{err}_P(X;\pi,\pi')^2 \mid D_1]\right\}^{1/2},$$
$$\mathcal{E}_\nu(\pi,\pi') := \left\{\mathbb{E}_X[\mathrm{err}_\nu(X;\pi,\pi')^2 \mid D_1]\right\}^{1/2}.$$

Then

$$0 \le \Theta^L_c(\pi,\pi') - \widetilde{\Theta}_L(\pi,\pi') \le \mathcal{E}_P(\pi,\pi')\mathcal{E}_\nu(\pi,\pi'). \tag{59}$$

Hence, if $\mathcal{E}_P(\pi,\pi') = o_p(r_n)$ and $\mathcal{E}_\nu(\pi,\pi') = o_p(s_n)$, then

$$\Theta^L_c(\pi,\pi') - \widetilde{\Theta}_L(\pi,\pi') = o_p(r_n s_n).$$

In particular, if both errors are $o_p(n^{-1/4})$, then the first-stage bias is $o_p(n^{-1/2})$.

The proof is given in Appendix D.8.

Remark D.11 (Approximate plug-in solution)**.** *If the exact plug-in optimality condition is replaced by*

$$\widehat{g}_x(\widehat{\nu}_x) \geq \sup_{(\phi,\psi)\in\mathcal{V}_{x,\pi,\pi'}} \widehat{g}_x(\phi,\psi) - \varepsilon_{a,b}(x),$$

for every action pair, then the proof gives

$$\begin{aligned} 0 &\leq \Theta_c^L(\pi,\pi') - \widetilde{\Theta}_L(\pi,\pi') \\ &\leq \mathbb{E}_X\left[\mathrm{err}_P(X;\pi,\pi')\,\mathrm{err}_\nu(X;\pi,\pi') + \varepsilon_{\pi(X),\pi'(X)}(X) \,\middle|\, D_1\right]. \end{aligned}$$

Remark D.12 (Version under a common dominating measure)**.** *If all $P_{a,x}$ and $\widehat{P}_{a,x}$ are dominated by a fixed sigma-finite Borel measure ρ on $\mathcal{Z}$, and the relevant density and potential differences are square-integrable with respect to ρ, one may use the simpler norm*

$$\|(f_0,f_1)\|_{L^2(\rho)\oplus L^2(\rho)} := \left\{\int_{\mathcal{Z}} f_0(y)^2\,\rho(\mathrm{d}y) + \int_{\mathcal{Z}} f_1(y)^2\,\rho(\mathrm{d}y)\right\}^{1/2}.$$

This recovers the form of Theorem 3.2 of Ji et al. (2023), with integration over the distribution space $\mathcal{Z} = \mathcal{P}_2(I)$.

D.3 Proof of Theorem D.3

Proof. Throughout the proof, condition on $\mathcal{F}_1$. The fitted dual functions are then fixed, and the evaluation observations remain i.i.d.

Step 1: the conditional IPW identity. Fix (π,π'). For any $a,b\in\mathcal{A}$, consistency, unconfoundedness, and the definition of the propensity score imply

$$\begin{aligned} &\mathbb{E}\left[\frac{\mathbb{1}\{A=a\}}{e_a(X)}\widehat{\phi}_{ab,X}(Y)\,\middle|\,X,\mathcal{F}_1\right] \\ &\quad= \frac{\mathbb{P}(A=a\mid X)}{e_a(X)}\mathbb{E}\left[\widehat{\phi}_{ab,X}\{Y(a)\}\,\middle|\,X,A=a,\mathcal{F}_1\right] \\ &\quad= P_{a,X}\widehat{\phi}_{ab,X}. \end{aligned}$$

The same calculation gives

$$\mathbb{E}\left[\frac{\mathbb{1}\{A=b\}}{e_b(X)}\widehat{\psi}_{ab,X}(Y)\,\middle|\,X,\mathcal{F}_1\right] = P_{b,X}\widehat{\psi}_{ab,X}.$$

Substitute $a=\pi(X)$, $b=\pi'(X)$, and add the two identities. Integrating over X proves the equality in (41).

Step 2: simultaneous weak duality. Fix x outside the common null set and fix (π,π'). Let $\gamma_x\in\Gamma(P_{\pi(x),x},P_{\pi'(x),x})$ be arbitrary. Integrating (37) with $(a,b)=(\pi(x),\pi'(x))$ against γ_x yields

$$\begin{aligned} &P_{\pi(x),x}\widehat{\phi}_x^{\pi,\pi'} + P_{\pi'(x),x}\widehat{\psi}_x^{\pi,\pi'} \\ &\qquad\leq \int_{\mathcal{Z}^2} c_x(z,z')\,d\gamma_x(z,z'). \end{aligned}$$

Take the infimum over γ_x and then expectation over X. This gives $\mu_{\pi,\pi'}\leq\Theta_c^L(\pi,\pi')$. Because Assumption D.1 holds on one event for all action pairs, the inequality holds on that same event for all policy pairs. For any admissible η, the further inequality $\Theta_c^L(\pi,\pi')\leq\Theta_c(\pi,\pi';\eta)$ follows from Theorem B.2. This proves part (i).

Step 3: finite-sample uniform empirical control. Under (42) and overlap,

$$|S_i(\pi,\pi')| \leq \frac{M}{\underline{e}} + \frac{M}{\underline{e}} = B_S$$

simultaneously over Π^2. We first connect the policy entropy in (40) to the conditional score class. For two policy pairs $u=(\pi,\pi')$ and $v=(\widetilde{\pi},\widetilde{\pi}')$, write s_u,s_v for their score functions and let $d_{2,m}$

be the empirical L^2 metric on the evaluation fold. Because the fitted potentials are indexed locally by the selected action pair, $s_u(O_i) = s_v(O_i)$ whenever both policies choose the same two actions at X_i. The envelope bound therefore gives

$$d_{2,m}^2(s_u, s_v) \le 4B_S^2 \left\{ H_{\boldsymbol{X}}(\pi, \widetilde{\pi}) + H_{\boldsymbol{X}}(\pi', \widetilde{\pi}') \right\}, \tag{60}$$

where $\boldsymbol{X} = (X_i)_{i \in D_2}$. Taking the Cartesian product of two u^2-Hamming covers of Π consequently yields

$$N\Big(2\sqrt{2}B_S u, \mathcal{S}(\mathcal{F}_1), d_{2,m}\Big) \le N_H(u^2, \Pi)^2, \qquad 0 < u \le 1. \tag{61}$$

Apply the empirical Dudley entropy inequality to the translated score class (translation does not change its Rademacher complexity). Using (61) and the change of variables $t = 2\sqrt{2}B_S u$, we obtain

$$\begin{aligned} \widehat{\mathfrak{R}}_m &\le \frac{12}{\sqrt{m}} \int_0^{2\sqrt{2}B_S} \sqrt{\log N(t, \mathcal{S}(\mathcal{F}_1), d_{2,m})}\, dt \\ &\le \frac{48 B_S}{\sqrt{m}} \int_0^1 \sqrt{\log N_H(u^2, \Pi)}\, du = \frac{48 B_S \mathcal{K}_{\mathrm{H}}(\Pi)}{\sqrt{m}}. \end{aligned}$$

This proves (44). Substitution into (45) also gives $r_{m,\alpha} \le r_{m,\alpha}^{\mathrm{ent}}$.

Let P_m be the empirical law of the evaluation fold and let $P_0(\,\cdot\,) = \mathbb{E}(\,\cdot \mid \mathcal{F}_1)$. Set

$$Z_m = \sup_{s \in \mathcal{S}(\mathcal{F}_1)} (P_m - P_0)s.$$

Conditional symmetrization gives

$$\mathbb{E}(Z_m \mid \mathcal{F}_1) \le 2\, \mathbb{E}\Big(\widehat{\mathfrak{R}}_m \mid \mathcal{F}_1\Big). \tag{62}$$

Changing one evaluation observation changes Z_m by at most $2B_S/m$, and it changes $\widehat{\mathfrak{R}}_m$ by at most $2B_S/m$. McDiarmid's inequality, applied once to Z_m and once to the lower tail of $\widehat{\mathfrak{R}}_m$, therefore shows, with conditional probability at least $1 - \alpha$, that

$$\begin{aligned} Z_m &\le \mathbb{E}(Z_m \mid \mathcal{F}_1) + B_S \sqrt{\frac{2\log(2/\alpha)}{m}} \\ &\le 2\widehat{\mathfrak{R}}_m + 3B_S \sqrt{\frac{2\log(2/\alpha)}{m}} = r_{m,\alpha}. \end{aligned}$$

Hence, on this event, simultaneously for all (π, π'),

$$\widehat{\Theta}_L(\pi, \pi') - \mu_{\pi,\pi'} \le r_{m,\alpha}.$$

Combining this inequality with part (i) proves (48). Since the event is simultaneous, it also contains the inequality indexed by any data-selected $(\widehat{\pi}, \widehat{\pi}')$.

For finite Π, Hoeffding's inequality gives, for each policy pair,

$$\mathbb{P}\Big\{ \widehat{\Theta}_L(\pi, \pi') - \mu_{\pi,\pi'} > t \,\Big|\, \mathcal{F}_1 \Big\} \le \exp\left(-\frac{mt^2}{2B_S^2} \right).$$

Take $t = r_{m,\alpha}^{\mathrm{fin}}$ and use a union bound over the $N_{\mathrm{pair}} = |\Pi|^2$ policy pairs. This proves the last assertion of part (ii).

□

D.4 Proof of Theorem D.4

Proof. Condition on $\mathcal{F}_1$. The conditional IPW identity from Step 1 of the preceding proof applies without change to the upper potentials, giving

$$\mu_{\pi,\pi'}^U = \mathbb{E}_X \Big[P_{\pi(X),X} \widehat{\phi}^U_{\pi(X),\pi'(X),X} + P_{\pi'(X),X} \widehat{\psi}^U_{\pi(X),\pi'(X),X} \mid \mathcal{F}_1 \Big].$$

For any conditional coupling γ_x, global superfeasibility gives

$$P_{\pi(x),x}\widehat{\phi}^{U}_{\pi(x),\pi'(x),x} + P_{\pi'(x),x}\widehat{\psi}^{U}_{\pi(x),\pi'(x),x} \geq \int c_x \, d\gamma_x.$$

Taking the supremum over γ_x and then expectation over X proves (51) simultaneously over Π^2.

Apply Step 3 of the preceding proof to the negative upper-score class. It has the same envelope, empirical covering numbers, and conditional Rademacher complexity as the upper-score class because the Rademacher signs are symmetric. Hence, with conditional probability at least $1-\alpha$,

$$\sup_{(\pi,\pi')\in\Pi^2} \left\{ \mu^{U}_{\pi,\pi'} - \frac{1}{m}\sum_{i\in D_2} S^{U}_i(\pi,\pi') \right\} \leq r^{U}_{m,\alpha}.$$

Together with (51), this proves (53); integration over the training fold removes the conditioning. Apply the lower and upper conclusions with error probability $\alpha/2$ and take a union bound to obtain (54). Uniformity over Π^2 immediately covers any evaluation-data-dependent selection from that class. □

D.5 Proof of Corollary D.5

Proof. For the problem indexed by n, write $B_{S,n} = 2M_n/\underline{e}_n$. Parts (ii) and (iii) of Assumption D.2 imply

$$\sup_n B_{S,n} \leq \frac{2\sup_n M_n}{\inf_n \underline{e}_n} < \infty.$$

Parts (iv) and (v) imply, respectively, $\sup_n \mathcal{K}_{\mathrm{H}}(\Pi_n) < \infty$ and $\sup_n \log(2/\alpha_n) < \infty$. Substituting these three bounds into (47) gives

$$r^{\mathrm{ent}}_{m_n,\alpha_n} \leq \frac{C}{\sqrt{m_n}}$$

for a finite constant C independent of n. Part (i) gives $m_n \asymp n$, hence the right-hand side is $O(n^{-1/2})$. Finally, Theorem D.3(ii) gives $r_{m_n,\alpha_n} \leq r^{\mathrm{ent}}_{m_n,\alpha_n}$ almost surely. The identical calculation with the upper envelope M^U_n proves the upper-radius assertion. □

D.6 Proof of Corollary D.7

Proof. Fix n and a realization of D_1 in the probability-one set on which (55) holds. Suppressing the index n, put

$$\xi_i = S_i(\pi,\pi') - \mu_n, \qquad \sigma^2 = \mathbb{E}_P(\xi_i^2 \mid \mathcal{F}_1), \qquad Z_m = \frac{m^{-1/2}\sum_{i\in D_2}\xi_i}{\sigma}.$$

Conditional on $\mathcal{F}_1$, the ξ_i's are i.i.d. and centered. Moreover, (55) and conditional Lyapunov's inequality imply

$$\mathbb{E}_P(|\xi_i|^3 \mid \mathcal{F}_1) \leq C_4^{3/4}, \qquad \sigma^2 \geq \underline{v}.$$

The conditional Berry–Esseen inequality therefore gives the deterministic bound

$$\sup_{t\in\mathbb{R}} |\mathbb{P}_P(Z_m \leq t \mid \mathcal{F}_1) - \Phi(t)| \leq \frac{C_{\mathrm{BE}} C_4^{3/4}}{\underline{v}^{3/2}\sqrt{m}}. \tag{63}$$

The right-hand side is independent of n and the realization of D_1. Thus conditioning on the random first-stage fold creates a triangular array whose normal-approximation error still vanishes.

We next verify consistency of the studentizing factor. Let $\bar{\xi} = m^{-1}\sum_{i\in D_2}\xi_i$ and $Q_m = m^{-1}\sum_{i\in D_2}\xi_i^2$. Conditional independence and (55) yield

$$\mathbb{E}_P\big[(Q_m - \sigma^2)^2 \mid \mathcal{F}_1\big] \leq \frac{C_4}{m}, \qquad \mathbb{E}_P(\bar{\xi}^2 \mid \mathcal{F}_1) = \frac{\sigma^2}{m} \leq \frac{C_4^{1/2}}{m}. \tag{64}$$

Since the usual sample variance satisfies

$$\widehat{\sigma}^2_{\pi,\pi'} = \frac{m}{m-1}\{Q_m - \bar{\xi}^2\},$$

Chebyshev's and Markov's inequalities, together with $\underline{v} \le \sigma^2 \le C_4^{1/2}$, show that for every fixed $\varepsilon > 0$ there is a finite constant $C_\varepsilon = C_\varepsilon(C_4, \underline{v})$ such that, for all sufficiently large m,

$$\mathbb{P}_P\left(\left|\frac{\widehat{\sigma}^2_{\pi,\pi'}}{\sigma_n^2} - 1\right| > \varepsilon\right) \le \frac{C_\varepsilon}{m}. \tag{65}$$

The displayed unconditional probability integrates the corresponding conditional bound over D_1. In particular, $R_m := \widehat{\sigma}_{\pi,\pi'}/\sigma_n \to 1$ in P-probability. Define $T_m = Z_m/R_m$ on $\{R_m > 0\}$ and set $T_m = 0$ on $\{R_m = 0\}$; the latter event has vanishing probability by (65).

For completeness, this also supplies the required triangular-array version of Slutsky's step. For any $K < \infty$ and $0 < \delta < 1/2$,

$$\mathbb{P}_P\{|T_m - Z_m| > 2\delta K\} \le \mathbb{P}_P\{|R_m - 1| > \delta\} + \mathbb{P}_P\{|Z_m| > K\}.$$

The first term vanishes by (65); the second is arbitrarily small for large K by (63). Hence $T_m - Z_m = o_P(1)$. Together with (63), Slutsky's theorem yields $T_m \rightsquigarrow N(0,1)$ under the fixed law P. If $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$, then on $\{R_m > 0\}$ the event $\{T_m \le z_{1-\alpha}\}$ is exactly $\{\widehat{L}^{\mathrm{pt}}_\alpha(\pi,\pi') \le \mu_n\}$. Since $\mathbb{P}_P(R_m = 0) \to 0$, it follows that

$$\liminf_n \mathbb{P}_P\left\{\widehat{L}^{\mathrm{pt}}_\alpha(\pi,\pi') \le \mu_n\right\} \ge 1 - \alpha.$$

Finally, conditional on $\mathcal{F}_1$, the fixed-pair IPW identity and weak-duality argument in Steps 1–2 of the proof of Theorem D.3 give $\mu_n \le \Theta^L_c(\pi,\pi')$. Theorem B.2 further gives $\Theta^L_c(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)$ for every admissible η. Therefore the event just controlled is contained in each coverage event stated in the corollary, which proves both assertions. □

D.7 Proof of Theorem D.9

Proof. Condition on D_1 and fix $x \in \mathcal{X}$. For brevity, write $a = a_x = \pi(x)$, $b = b_x = \pi'(x)$,

$$\nu^\star_x := (\phi^\star_{a,b,x}, \psi^\star_{a,b,x}), \qquad \widehat{\nu}_x := (\widehat{\phi}_{a,b,x}, \widehat{\psi}_{a,b,x}).$$

Define the true and plug-in conditional dual objectives by

$$g_x(\phi,\psi) := P_{a,x}\phi + P_{b,x}\psi,$$
$$\widehat{g}_x(\phi,\psi) := \widehat{P}_{a,x}\phi + \widehat{P}_{b,x}\psi.$$

Because $\nu^\star_x$ maximizes g_x and $\widehat{\nu}_x$ is feasible,

$$0 \le g_x(\nu^\star_x) - g_x(\widehat{\nu}_x). \tag{66}$$

Insert and subtract the plug-in objective:

$$\begin{aligned} g_x(\nu^\star_x) - g_x(\widehat{\nu}_x) = {} & \{g_x(\nu^\star_x) - \widehat{g}_x(\nu^\star_x)\} \\ & + \{\widehat{g}_x(\nu^\star_x) - \widehat{g}_x(\widehat{\nu}_x)\} \\ & + \{\widehat{g}_x(\widehat{\nu}_x) - g_x(\widehat{\nu}_x)\}. \end{aligned} \tag{67}$$

By the plug-in optimality of $\widehat{\nu}_x$,

$$\widehat{g}_x(\nu^\star_x) - \widehat{g}_x(\widehat{\nu}_x) \le 0.$$

Consequently,

$$\begin{aligned} g_x(\nu^\star_x) - g_x(\widehat{\nu}_x) &\le \{g_x(\nu^\star_x) - \widehat{g}_x(\nu^\star_x)\} + \{\widehat{g}_x(\widehat{\nu}_x) - g_x(\widehat{\nu}_x)\} \\ &= \langle (p_{a,x} - \widehat{p}_{a,x}, p_{b,x} - \widehat{p}_{b,x}), \nu^\star_x - \widehat{\nu}_x \rangle_{a,b,x} \\ &\le \|(p_{a,x} - \widehat{p}_{a,x}, p_{b,x} - \widehat{p}_{b,x})\|_{a,b,x} \, \|\nu^\star_x - \widehat{\nu}_x\|_{a,b,x} \\ &= \mathrm{err}_P(x;\pi,\pi') \, \mathrm{err}_\nu(x;\pi,\pi'). \end{aligned} \tag{68}$$

The second line uses the Radon–Nikodym representations with respect to $\lambda_{a,x}$ and $\lambda_{b,x}$, and the third line is Cauchy–Schwarz in $\mathbb{H}_{a,b,x}$.

Finally, strong duality in Theorem B.2 and the definition of $\widetilde{\Theta}_L$ in Section C give

$$\Theta^L_c(\pi,\pi') - \widetilde{\Theta}_L(\pi,\pi') = \mathbb{E}_X[g_X(\nu^\star_X) - g_X(\widehat{\nu}_X) \,|\, D_1].$$

Integrating (66) and (68) with respect to P_X proves (58). □

D.8 Proof of Corollary D.10

Proof. Apply Cauchy–Schwarz to the outer expectation on the right-hand side of (58). □

E Safe Quantile Computation and Finite Anchors

The empirical dual in (30) is infinite-dimensional because its arguments are probability distributions. A finite collection of sampled constraints is not enough: feasibility on finitely many quantile pairs does not imply feasibility on the full quantile cone. We therefore combine a conservative finite-dimensional cost with an extension whose architecture is globally dual-feasible. Optimization and numerical discretization can then affect tightness, but not validity.

We first describe the computational core for one fixed covariate stratum and two action marginals. The rate results in Sections F and I retain finitely many covariate cells and policy indexing. Their conditional implementation is described below.

E.1 A conservative finite-dimensional quantile cost

Let $D = I_+ - I_-$, and suppose that the convex displacement function h is L_h-Lipschitz on $[-D, D]$. For $K \geq 1$, define the vector of quantile-bin averages

$$B_K(q)_k := K \int_{(k-1)/K}^{k/K} q(t)\, dt, \qquad k = 1, \dots, K, \tag{69}$$

and, for $z, r \in [I_-, I_+]^K$, define

$$c_{h,K}(z, r) := \frac{1}{K} \sum_{k=1}^{K} h(r_k - z_k), \tag{70}$$

$$d_K(z, u) := \frac{1}{K} \sum_{k=1}^{K} |z_k - u_k|. \tag{71}$$

The following theorem makes the numerical approximation one-sided and gives its exact deterministic price.

Theorem E.1 (Conservative binning error)**.** *For every pair $q, r \in \mathcal{Q}$,*

$$0 \leq c_h(q, r) - c_{h,K}\{B_K(q), B_K(r)\} \leq \frac{L_h D}{K}. \tag{72}$$

Consequently, if $B_K \# P$ denotes the pushforward of P under B_K, then

$$0 \leq T_{c_h}(P_0, P_1) - T_{c_{h,K}}(B_K \# P_0, B_K \# P_1) \leq \frac{L_h D}{K}. \tag{73}$$

The first inequality is Jensen's inequality on each bin. The upper bound uses monotonicity: the total within-bin variation of a bounded quantile curve is at most D, and its integrated deviation from its bin mean is at most half that variation divided by K. The full proof is in Appendix E.

By Theorem E.1, replacing c_h by $c_{h,K}$ can only lower the target, so finite K cannot make the reported lower certificate anti-conservative. Increasing K controls only informativeness, with an explicit error of order K^{-1}.

E.2 Empirical transport and a globally feasible extension

Let $z_{0,1}, \dots, z_{0,n_0}$ and $z_{1,1}, \dots, z_{1,n_1}$ be the encoded training curves, with strictly positive probability weights p_i and w_j. Solve the finite OT dual

$$\max_{\alpha, \beta} \left\{ \sum_{i=1}^{n_0} p_i \alpha_i + \sum_{j=1}^{n_1} w_j \beta_j \right\} \quad \text{subject to} \quad \alpha_i + \beta_j \leq c_{h,K}(z_{0,i}, z_{1,j}). \tag{74}$$

This is a standard discrete optimal-transport problem and can be solved by a network simplex or a general linear-programming solver. Select an optimal pair normalized so that

$$\alpha_i = \min_{j \le n_1} \{c_{h,K}(z_{0,i}, z_{1,j}) - \beta_j\}. \tag{75}$$

For the policy-indexed IPW implementation, fix the additive gauge by setting $a_0 = \min_i \alpha_i$ and replacing (α, β) by $(\alpha - a_0, \beta + a_0)$. This preserves feasibility, the discrete transform, and the optimal objective. Under the maintained normalization $h \ge 0$, $h(0) = 0$, it gives $0 \le \alpha_i \le L_h D$. The extensions below then satisfy

$$\sup_z |\widehat{\Phi}_K(z)| \vee \sup_r |\widehat{\Psi}_K(r)| \le L_h D. \tag{76}$$

Thus one may use $M = L_h D$ in the IPW radius. An oscillation bound alone would not specify this absolute envelope. Define, on the full encoded quantile cone,

$$\widehat{\Phi}_K(z) := \max_{i \le n_0} \{\alpha_i - L_h d_K(z, z_{0,i})\}, \tag{77}$$

$$\widehat{\Psi}_K(r) := \min_{i \le n_0} \{c_{h,K}(z_{0,i}, r) - \alpha_i\}. \tag{78}$$

Theorem E.2 (Safe extension without loss of the empirical objective)**.** *The potentials in* (77)–(78) *satisfy, for every* $q, r \in \mathcal{Q}$*,*

$$\widehat{\Phi}_K\{B_K(q)\} + \widehat{\Psi}_K\{B_K(r)\} \le c_{h,K}\{B_K(q), B_K(r)\} \le c_h(q, r). \tag{79}$$

Moreover,

$$\widehat{\Phi}_K(z_{0,i}) = \alpha_i, \qquad \widehat{\Psi}_K(z_{1,j}) = \beta_j, \tag{80}$$

for every training atom. Hence the globally feasible extension attains exactly the optimal objective value in (74)*. Both potentials have oscillation at most* $L_h D$*.*

For fixed outer laws write

$$T^U_{c_h}(P_0, P_1) := \sup_{\gamma \in \Gamma(P_0, P_1)} \int c_h \, d\gamma.$$

The upper endpoint admits a mirrored construction, but it must retain the binning remainder in the conservative direction. Solve

$$\min_{a,b} \left\{ \sum_{i=1}^{n_0} p_i a_i + \sum_{j=1}^{n_1} w_j b_j \right\} \quad \text{subject to} \quad a_i + b_j \ge c_{h,K}(z_{0,i}, z_{1,j}), \tag{81}$$

and select an optimal pair normalized so that

$$a_i = \max_{j \le n_1} \{c_{h,K}(z_{0,i}, z_{1,j}) - b_j\}. \tag{82}$$

After the invariant shift making $\min_i a_i = 0$, define

$$\widehat{\Phi}^{U,0}_K(z) := \min_{i \le n_0} \{a_i + L_h d_K(z, z_{0,i})\}, \tag{83}$$

$$\widehat{\Psi}^{U,0}_K(r) := \max_{i \le n_0} \{c_{h,K}(z_{0,i}, r) - a_i\}, \tag{84}$$

and, with $b_K = L_h D / K$, put

$$\widehat{\Phi}^U_K = \widehat{\Phi}^{U,0}_K + b_K/2, \qquad \widehat{\Psi}^U_K = \widehat{\Psi}^{U,0}_K + b_K/2. \tag{85}$$

Theorem E.3 (Safe upper extension and exact deterministic correction)**.** *For every* $q, r \in \mathcal{Q}$*,*

$$\widehat{\Phi}^U_K\{B_K(q)\} + \widehat{\Psi}^U_K\{B_K(r)\} \ge c_{h,K}\{B_K(q), B_K(r)\} + b_K \ge c_h(q, r). \tag{86}$$

The uncorrected extensions interpolate the finite upper dual:

$$\widehat{\Phi}^{U,0}_K(z_{0,i}) = a_i, \qquad \widehat{\Psi}^{U,0}_K(z_{1,j}) = b_j. \tag{87}$$

Hence the corrected extensions attain the optimal objective in (81) *plus exactly* b_K*. Both potentials have oscillation at most* $L_h D$*, and the stated normalization gives*

$$\sup_z |\widehat{\Phi}^U_K(z)| \vee \sup_r |\widehat{\Psi}^U_K(r)| \le \frac{3}{2} L_h D. \tag{88}$$

The term b_K cannot in general be omitted. The discretized inequality $c_{h,K} \le c_h$ is favorable for lower certificates but has the wrong direction for an upper certificate; the upper half of Theorem E.1 supplies the required repair.

To see the mechanism, let $i^\star$ maximize (77). The second potential is no larger than the term indexed by $i^\star$, while Lipschitz continuity of $c_{h,K}$ transfers that anchor inequality to every source curve. Thus global feasibility follows from the Lipschitz inequality on the full cone. The interpolation statement shows that safety does not reduce the training empirical OT value. Appendix E proves all claims.

E.3 SPLIT-SAMPLE IMPLEMENTATION

Let $Q^{\mathrm{tr}}_{a,i}$, $i \le n_a$, and $Q^{\mathrm{ev}}_{a,i}$, $i \le m_a$, denote independent training and evaluation curves from arm a. Conditional on the training sample, define

$$\overline{T}_K = \frac{1}{m_0}\sum_{i=1}^{m_0} \widehat{\Phi}_K\{B_K(Q^{\mathrm{ev}}_{0,i})\} + \frac{1}{m_1}\sum_{i=1}^{m_1} \widehat{\Psi}_K\{B_K(Q^{\mathrm{ev}}_{1,i})\}. \tag{89}$$

The oscillation bound in Theorem E.2 and Hoeffding's inequality give the exactly valid radius

$$r_{K,\boldsymbol{m},\alpha} = L_h D\sqrt{\frac{\log(1/\alpha)}{2}\left(\frac{1}{m_0}+\frac{1}{m_1}\right)}. \tag{90}$$

Thus

$$L^{\mathrm{comp}}_{K,\alpha} := \overline{T}_K - r_{K,\boldsymbol{m},\alpha} \tag{91}$$

satisfies

$$\inf_{P_0,P_1} \Pr\{L^{\mathrm{comp}}_{K,\alpha} \le T_{c_h}(P_0,P_1)\} \ge 1-\alpha \tag{92}$$

for every finite K, n_0, n_1, m_0, m_1. The proof is included in Appendix E.

For the upper construction, define

$$\overline{T}^U_K = \frac{1}{m_0}\sum_{i=1}^{m_0} \widehat{\Phi}^U_K\{B_K(Q^{\mathrm{ev}}_{0,i})\} + \frac{1}{m_1}\sum_{i=1}^{m_1} \widehat{\Psi}^U_K\{B_K(Q^{\mathrm{ev}}_{1,i})\}, \qquad U^{\mathrm{comp}}_{K,\alpha} = \overline{T}^U_K + r_{K,\boldsymbol{m},\alpha}. \tag{93}$$

The oscillation bound, rather than the absolute level introduced by the binning correction, determines the Hoeffding radius. Therefore

$$\inf_{P_0,P_1} \Pr\{T^U_{c_h}(P_0,P_1) \le U^{\mathrm{comp}}_{K,\alpha}\} \ge 1-\alpha. \tag{94}$$

Using $L^{\mathrm{comp}}_{K,\alpha/2}$ and $U^{\mathrm{comp}}_{K,\alpha/2}$ gives simultaneous coverage of both sharp endpoints with probability at least $1-\alpha$.

Algorithm 1 Safe quantile-dual lower confidence bound

Require: Training curves $\{Q^{\mathrm{tr}}_{a,i}\}$, evaluation curves $\{Q^{\mathrm{ev}}_{a,i}\}$, convex cost h, L_h, bins K, level α.

1: Encode every curve by the bin-average map B_K in (69).
2: Form $C_{ij} = c_{h,K}(z_{0,i}, z_{1,j})$ and solve the discrete OT dual (74).
3: Replace α by its discrete c-transform (75); set $a_0 = \min_i \alpha_i$ and apply $(\alpha,\beta) \leftarrow (\alpha - a_0, \beta + a_0)$.
4: Construct $\widehat{\Phi}_K$ and $\widehat{\Psi}_K$ from (77)–(78).
5: Evaluate $\overline{T}_K$ on the independent fold using (89).
6: Compute $r_{K,\boldsymbol{m},\alpha}$ from (90).
7: **return** $L^{\mathrm{comp}}_{K,\alpha} = \overline{T}_K - r_{K,\boldsymbol{m},\alpha}$.

With stored bin vectors, forming the cost matrix costs $O(n_0 n_1 K)$. Each subsequent potential evaluation costs $O(n_0 K)$; the roles of the two arms may be reversed so that the smaller training support supplies the anchors. The OT solve is the only non-linear scaling step and uses standard finite OT software.

Increasing K improves the approximation at greater computational cost while preserving validity. To keep the binning component below a target error, choose K so that $L_h D/K$ is no larger than that error. This component permits a polylogarithmic K at the unstructured power-cost rate, whereas at a polynomial target rate a polynomially growing K may be required. This comparison controls binning alone. The complete transfer in Appendix E.5 bounds the population dual gap including the anchor-covering error. A finite completion of the anchor set attains the ideal rates under explicit resolution and solver-residual conditions; the resulting anchor budget need not be polynomial in sample size. For training-only anchors the same bound retains a nearest-anchor term, and Theorem J.8 proves consistency.

Conditional and policy-indexed implementation. On the training fold, estimate each conditional outer law $\widehat{P}_{a,x}$. For only those covariate values and action pairs queried on the evaluation fold, draw or construct finite supports from $\widehat{P}_{a,x}$, solve (74), and apply the same safe extension. The resulting pair is globally feasible pointwise in x, independently of conditional-law misspecification or optimization accuracy. It can therefore be inserted directly into the IPW scores in (31), or into their action-pair-indexed uniform version in Section D. Cross-fitting repeats the construction with the roles of the folds exchanged. All randomness and tie-breaking used to construct the conditional potential maps must be fixed independently of evaluation outcomes; the resulting maps must be jointly measurable in the training data, covariate, and outcome. In the finite-stratum implementation, solve one problem per stratum and action pair, using a fixed fallback law for an empty training cell. Same-action pairs may be handled by the same routine. A zero pair is an optional optimal normalization for their lower endpoint only.

Boundary. The method does not remove the statistical difficulty of estimating the outer laws. It separates that difficulty from numerical validity: finite K, an inaccurate conditional-law model, and an approximate OT solve can make the certificate loose, but the one-sided cost and safe extension prevent them from making it anti-conservative. Unconstrained neural or series duals may be used as warm starts, but sampled constraint checks alone cannot replace the globally feasible extension.

E.4 Proof of Theorem E.1

Proof. Write $I_k = ((k-1)/K, k/K]$, and let q_k and r_k denote the averages of q and r on I_k. Jensen's inequality gives

$$\frac{1}{K} h(r_k - q_k) \le \int_{I_k} h\{r(t) - q(t)\}\, dt.$$

Summing over k proves

$$c_{h,K}\{B_K(q), B_K(r)\} \le c_h(q, r).$$

We next control the reverse difference. If a random variable Z takes values in an interval of length a, then

$$\mathbb{E}|Z - \mathbb{E}Z| \le \frac{a}{2}. \tag{95}$$

Indeed, after translating and rescaling it is enough to consider $0 \le Z \le 1$. If $\mu = \mathbb{E}Z$, convexity of $z \mapsto |z - \mu|$ implies

$$|z - \mu| \le (1 - z)\mu + z(1 - \mu),$$

and taking expectations yields $\mathbb{E}|Z - \mu| \le 2\mu(1 - \mu) \le 1/2$.

Apply (95) to a uniform argument on each bin. For a nondecreasing q, let

$$\Delta_{q,k} := q(k/K-) - q((k-1)/K+),$$

with the evident endpoint conventions. Then

$$\int_{I_k} |q(t) - q_k|\, dt \le \frac{\Delta_{q,k}}{2K}.$$

Monotonicity and the range restriction $q(t) \in [I_-, I_+]$ imply

$$\sum_{k=1}^{K} \Delta_{q,k} \le D,$$

because the within-bin increments are disjoint parts of the total variation; jumps at bin boundaries can only reduce their sum. Hence

$$\int_0^1 |q(t) - q_K(t)|\, dt \le \frac{D}{2K}, \tag{96}$$

where $q_K(t) = q_k$ on I_k. The same bound holds for r.

Since h is L_h-Lipschitz on $[-D, D]$,

$$c_h(q, r) - c_{h,K}\{B_K(q), B_K(r)\}$$

$$\leq L_h \int_0^1 |\{r(t) - q(t)\} - \{r_K(t) - q_K(t)\}|\, dt$$
$$\leq L_h \left(\|r - r_K\|_{L^1} + \|q - q_K\|_{L^1}\right) \leq \frac{L_h D}{K}.$$

Together with Jensen's inequality, this proves (72).

For the endpoint claim, every coupling $\gamma \in \Gamma(P_0, P_1)$ pushes forward under (B_K, B_K) to a coupling of $B_K\#P_0$ and $B_K\#P_1$. Conversely, by disintegrating P_0 and P_1 over their bin encodings, every coupling of the pushforward laws can be lifted to a coupling of P_0 and P_1. Therefore

$$T_{c_{h,K}}(B_K\#P_0, B_K\#P_1) = \inf_{\gamma \in \Gamma(P_0,P_1)} \int c_{h,K}\{B_K(q), B_K(r)\}\, d\gamma(q, r).$$

Integrating the uniform pointwise bound (72) and taking infima proves (73). □

E.5 Total error of a finite-anchor trainer

We now account for the extension error as well as the binning error. Throughout this subsection the cost is the displacement cost c_h of Section E, with $h \geq 0$, $h(0) = 0$, and Lipschitz constant L_h. Write $b_K = L_h D/K$. For one cell and ordered action pair, let P_0, P_1 be the latent laws, P°_{a,n_a} their latent empirical laws, and $\widetilde{P}_{a,n_a}$ the empirical laws of the observed training curves $\widetilde{q}_{a,i}$. Exact curves are the special case $\widetilde{q}_{a,i} = q_{a,i}$. Set

$$e_a = \frac{1}{n_a} \sum_{i=1}^{n_a} \|\widetilde{q}_{a,i} - q_{a,i}\|_1.$$

All empirical weights here are $1/n_a$; arbitrary fitted-law weights require their own statistical error bound.

A finite and explicitly completable anchor set. Start with a feasible discrete dual (α, β) whose objective is at least the finite primal optimum minus $\varepsilon_{\text{opt}} \geq 0$. Thus ε_{opt} is a certified primal–dual residual, not a solver stopping tolerance without a feasibility guarantee. Put

$$A(z) = \min_{j \leq n_1} \{c_{h,K}(z, B_K\widetilde{q}_{1,j}) - \beta_j\}.$$

Choose a finite set $\mathcal{U}$ in the encoded monotone cone containing every source training vector, assign $\alpha_u = A(u)$ for $u \in \mathcal{U}$, and use

$$\Phi_{\mathcal{U}}(z) = \max_{u \in \mathcal{U}} \{\alpha_u - L_h d_K(z, u)\}, \qquad \Psi_{\mathcal{U}}(r) = \min_{u \in \mathcal{U}} \{c_{h,K}(u, r) - \alpha_u\}. \tag{97}$$

Normalize by subtracting $\min_{u \in \mathcal{U}} \alpha_u$ from all α_u and adding it to all β_j. The proof of Theorem E.2 still gives global feasibility, the absolute envelope $M = L_h D$, and Lipschitz constant L_h for both potentials. Added anchors carry no empirical mass and require no new OT solve. They only require evaluating $A(u)$. Since A is L_h-Lipschitz, $\Phi_{\mathcal{U}}(u) = \alpha_u$; also $\Psi_{\mathcal{U}}(B_K\widetilde{q}_{1,j}) \geq \beta_j$. The empirical objective therefore remains within ε_{opt} of the finite optimum. With no additional anchors and an exact solve this is precisely Algorithm 1.

Let $\mathcal{F}_h$ be the class of all c_h-concave functions normalized by $\min f = 0$, and let $\mathcal{F}_h^{c_h}$ be their transforms. These two classes have envelopes $L_h D$ in absolute value. Define

$$U_a = \max_{\mathcal{H} \in \{\mathcal{F}_h, \mathcal{F}_h^{c_h}\}} \sup_{f \in \mathcal{H}} |(P^\circ_{a,n_a} - P_a) f|, \qquad R_0(\mathcal{U}) = \int d_K(B_K q, \mathcal{U})\, dP_0(q).$$

Constant shifts do not affect U_a.

Theorem E.4 (Deterministic total training-gap bound)**.** *For every realization with $n_0, n_1 \geq 1$, the trainer (97) satisfies*

$$0 \leq T_{c_h}(P_0, P_1) - P_0\Phi_{\mathcal{U}} \circ B_K - P_1\Psi_{\mathcal{U}} \circ B_K$$
$$\leq 4b_K + \varepsilon_{\text{opt}} + 2(U_0 + U_1) + 2L_h(e_0 + e_1) + 2L_h R_0(\mathcal{U}). \tag{98}$$

In particular, a deterministic d_K-net of radius η added to the source anchors makes the last term at most $2L_h\eta$.

Proof. Let $s_K z$ denote the step quantile encoded by z, and define the continuous-cost comparison pair

$$F(q) = \min_j \{c_h(q, s_K B_K \widetilde{q}_{1,j}) - \beta_j\}, \qquad G = F^{c_h}.$$

An infimum of cost sections is c_h-concave (apply the transform twice), so a constant shift puts F in $\mathcal{F}_h$ and G in its transformed class. Both functions are L_h-Lipschitz in L^1. The binning bound implies, uniformly in q,

$$A(B_K q) \le F(q) \le A(B_K q) + b_K. \tag{99}$$

For any target observation $\widetilde{q}_{1,j}$, feasibility of $A(z)+\beta_j \le c_{h,K}(z, B_K \widetilde{q}_{1,j})$ gives $G(\widetilde{q}_{1,j}) \ge \beta_j - b_K$. At each source observation, $F(\widetilde{q}_{0,i}) \ge \alpha_{B_K \widetilde{q}_{0,i}}$. Consequently, with $\widetilde{T}$ denoting continuous-cost OT between the two observed empirical laws,

$$\widetilde{P}_{0,n_0} F + \widetilde{P}_{1,n_1} G \ge \widetilde{T} - 2b_K - \varepsilon_{\mathrm{opt}}.$$

For a normalized population optimizer $(f^\star, (f^\star)^{c_h})$, subtract the last empirical objective from its population objective. The empirical optimizer comparison and the two marginal empirical-process bounds give

$$T_{c_h}(P_0, P_1) - P_0 F - P_1 G \le 2(U_0 + U_1) + 2L_h(e_0 + e_1) + 2b_K + \varepsilon_{\mathrm{opt}}.$$

Here each of the two candidate pairs costs at most $L_h e_a$ when its empirical integral is changed from latent to observed curves in arm a; no independence of the fitted functions and the training sample is used.

Finally, Lipschitz continuity of A gives $\Phi_{\mathcal{U}}(B_K q) \ge F(q) - b_K - 2L_h d_K(B_K q, \mathcal{U})$. For every anchor u, $F(s_K u) = A(u) = \alpha_u$, whence

$$G(r) \le \min_{u \in \mathcal{U}} \{c_h(s_K u, r) - \alpha_u\} \le \Psi_{\mathcal{U}}(B_K r) + b_K.$$

Integration proves the upper bound. Global feasibility proves the lower bound. All statements are invariant under the gauge shift. □

An explicit finite anchor budget. For any $0 < \eta \le D$, round each coordinate of a monotone encoded vector down to a grid of spacing at most η in $[I_-, I_+]$. The resulting monotone vectors form a d_K-net of radius η with at most

$$J_{K,\eta} = \binom{K + \lceil D/\eta \rceil}{K} \tag{100}$$

points. Adjoining this net is thus a finite implementation of the theorem. For $L_h D > 0$ and a desired numerical error t, it suffices to take $K \ge 16 L_h D/t$, $\eta \le t/(8L_h)$, and $\varepsilon_{\mathrm{opt}} \le t/2$. These choices make $4b_K + 2L_h\eta + \varepsilon_{\mathrm{opt}} \le t$. Forming the extra anchor values costs $O(J_{K,\eta} n_1 K)$, and an evaluation costs $O\{(n_0 + J_{K,\eta})K\}$. This construction attains the statistical bound with an anchor budget that can grow exponentially in $1/t$.

Corollary E.5 (Transfer of the ideal rates to completed finite anchors)**.** *Use the finite-cell empirical trainer, known propensities, and independent evaluation fold of the minimax and HPQ theorems. Let $q_\star = \kappa \underline{e} > 0$, $\ell = |D_1|$, and $N = |D_2|$. Use a zero potential pair in empty cells. In every nonempty cell use (97) with net radius η and residual at most $\varepsilon_{\mathrm{opt}}$. For exact training and evaluation curves, the resulting policy-uniform lower certificate obeys*

$$\mathbb{E} \sup_{\Pi^2} \{\Theta^L_{c_h} - L^{\mathrm{comp}}_\alpha\}_+ \le C \left\{ R(\ell q_\star) + K^{-1} + \eta + \varepsilon_{\mathrm{opt}} + \sqrt{\frac{\log(N_\Pi \vee 2) + \log(2/\alpha)}{N}} \right\}, \tag{101}$$

where N_Π counts distinct policy restrictions to the covariate cells. For $c_h = c_p$ on $[0,1]$, one may take $R(v) = \{\log(v+1)\}^{-(p \wedge 2)}$. For a Lipschitz displacement cost satisfying Assumption I.1, one may instead take $R(v) = (v+1)^{-1/s}$. Constants depend on the fixed model parameters. Thus $K^{-1} + \eta + \varepsilon_{\mathrm{opt}} = O\{R(n)\}$ and a fixed-proportion split attain the respective ideal rates. The first is minimax by Theorem F.1; the HPQ conclusion remains an upper attainment statement.

Proof. For power costs, Lemma F.12 bounds $\mathbb{E}U_a$ by $CR(n_a)$. In the HPQ case, every normalized transform is one-Lipschitz for its corresponding active pseudometric. Therefore U_a is bounded by the maximum of the two active W_1 empirical discrepancies, and Lemma I.7 bounds its expectation by $CR(n_a)$. Average (98) over the finitely many cells and bound a policy's gap by the largest action-pair gap in each cell. Binomial cell counts are at least $\ell q_\star/2$ except with probability at most $dK_X \exp(-\ell q_\star/8)$. Empty-cell gaps are bounded by $L_h D$; this exponential remainder is absorbed in R. Conditional Hoeffding bounds for the finitely many evaluation scores, with envelope $B_S = 2L_h D/\underline{e}$, and their integrated tails give the last term. Global feasibility yields the unchanged finite-sample coverage using the radius of Theorem D.3 with this envelope. □

The additional net is a substantive condition in Corollary E.5. For the original training-only anchors, (98) still holds with the actual $R_0(\mathcal{U})$. Binning and empirical interpolation alone do not bound that quantity at the ideal rate. In particular, cost-active HPQ does not by itself control an L^1 nearest-anchor distance. The consistency argument for the unchanged training-only extension does not assert the sharper ideal rate for that version. The continuous closure below changes that extension and removes the nearest-anchor term under a separate certified-precision condition.

E.6 Cost-adapted closure on the full quantile cone

The nearest-anchor term in Theorem E.4 belongs to the Lipschitz extension, rather than to empirical transport itself. We give an alternative for any nonnegative convex displacement loss h with $h(0) = 0$ and Lipschitz constant L_h. It preserves the finite-anchor target potential and improves the source potential on the entire cone. The guarantee is general; the cost of evaluating the continuous transform depends on h.

Write $\mathcal{Q}_K = \{z \in [I_-, I_+]^K : z_1 \le \cdots \le z_K\}$ and abbreviate $c = c_{h,K}$. Use the anchors and values $a_u = A(u)$ from (97), with every source training vector included in $\mathcal{U}$. Define

$$\Psi_{\mathcal{U}}(r) = \min_{u\in\mathcal{U}}\{c(u,r) - a_u\}, \qquad \Phi_{\mathcal{U}}(z) = \max_{u\in\mathcal{U}}\{a_u - L_h d_K(z,u)\},$$
$$\Phi_\star(z) = \inf_{r\in\mathcal{Q}_K}\{c(z,r) - \Psi_{\mathcal{U}}(r)\}. \tag{102}$$

The last infimum is over the full encoded cone, including unobserved curves.

Proposition E.6 (Feasible and pointwise improving closure)**.** *For every $z, r \in \mathcal{Q}_K$,*

$$\Phi_{\mathcal{U}}(z) \le \Phi_\star(z), \qquad \Phi_\star(z) + \Psi_{\mathcal{U}}(r) \le c(z,r). \tag{103}$$

If a numerical procedure supplies a measurable certified lower bound $b(z) \le \Phi_\star(z)$, then

$$\Phi_{\mathrm{cl}}(z) = \max\{\Phi_{\mathcal{U}}(z), b(z)\} \tag{104}$$

retains both inequalities. If additionally $\sup_z\{\Phi_\star(z) - b(z)\} \le \tau$, then $0 \le \Phi_\star - \Phi_{\mathrm{cl}} \le \tau$ uniformly. The empirical objective of $(\Phi_{\mathrm{cl}}, \Psi_{\mathcal{U}})$ is at least that of the original finite dual. Under the normalization $\min_{u\in\mathcal{U}} a_u = 0$, both returned potentials have absolute envelope $M = L_h D$.

Proof. The original extension is feasible, so $\Phi_{\mathcal{U}}(z) \le c(z,r) - \Psi_{\mathcal{U}}(r)$ for every r. Taking the infimum proves the first inequality in (103); the second follows from the definition. Taking the maximum of two lower bounds for $\Phi_\star$ preserves both claims. Since $a_u = A(u)$ is L_h-Lipschitz on $\mathcal{U}$, $\Phi_{\mathcal{U}}(z_{0,i}) = A(z_{0,i}) \ge \alpha_i$, while $\Psi_{\mathcal{U}}(z_{1,j}) \ge \beta_j$. Nonnegative empirical weights give the objective claim. For the envelope, the original pair is bounded by M under the stated normalization. Also $c(z,z) = 0$ gives $\Phi_\star(z) \le -\Psi_{\mathcal{U}}(z) \le M$, whereas $\Phi_{\mathrm{cl}} \ge \Phi_{\mathcal{U}} \ge -M$. □

With the gauge held fixed, nonnegative IPW weights make the uncalibrated evaluation score no smaller for every policy pair and every evaluation sample. The same holds for lower certificates if a common valid radius is used. It need not hold when an empirical Bernstein radius is recomputed: the changed score can have a different variance. This proposition concerns lower certificates; upper certificates require the opposite transform and an upper numerical bound.

Corollary E.7 (Training error with a certified continuous transform)**.** *Under the notation and empirical-weight assumptions of Theorem E.4, suppose the uniform precision condition in Proposition E.6 holds. Then*

$$0 \le T_{c_h}(P_0, P_1) - P_0\Phi_{\mathrm{cl}} \circ B_K - P_1\Psi_{\mathcal{U}} \circ B_K$$

$$\leq 4b_K + \varepsilon_{\mathrm{opt}} + 2(U_0 + U_1) + 2L_h(e_0 + e_1) + \tau. \tag{105}$$

In particular, completion by an η-net is unnecessary for this bound. Under the assumptions of Corollary E.5, the corresponding policy-uniform rate has computational contribution $K^{-1} + \varepsilon_{\mathrm{opt}} + \tau$ in place of $K^{-1} + \eta + \varepsilon_{\mathrm{opt}}$.

Proof. Let $s_K u$ be the step quantile encoded by u, and put

$$G(r) = \min_{u \in \mathcal{U}} \{c_h(s_K u, r) - a_u\}, \qquad F(q) = \inf_{r \in \mathcal{Q}} \{c_h(q, r) - G(r)\}.$$

As an infimum of cost sections, G is a target-side c_h-concave function and equals the transform of F. Constant shifts put this pair in the classes defining U_0, U_1. Binning gives

$$\Psi_{\mathcal{U}}(B_K r) \leq G(r) \leq \Psi_{\mathcal{U}}(B_K r) + b_K.$$

For the infimum defining F, the lower cost bound gives $F(q) \geq \Phi_\star(B_K q) - b_K$. Restricting that infimum to step quantiles, for which $G(s_K v) = \Psi_{\mathcal{U}}(v)$, gives $F(q) \leq \Phi_\star(B_K q) + b_K$. On the observed training curves the objective of (F, G) is therefore at least the finite dual objective minus b_K, and hence at least the continuous empirical optimum minus $2b_K + \varepsilon_{\mathrm{opt}}$. The same two empirical-process comparisons and measurement perturbations as in the proof of Theorem E.4 give

$$T_{c_h}(P_0, P_1) - P_0 F - P_1 G \leq 2b_K + \varepsilon_{\mathrm{opt}} + 2(U_0 + U_1) + 2L_h(e_0 + e_1).$$

Returning to the implemented pair costs at most $2b_K + \tau$. Weak duality gives the lower inequality. The rate transfer uses the same cell-count and simultaneous-score arguments as Corollary E.5, with the numerical rule fixed conditional on training and the unchanged absolute envelope. □

A common transform oracle. For finitely many $y_j \in \mathcal{Q}_K$ and values β_j, define

$$A(z) = \min_j \{c_{h,K}(z, y_j) - \beta_j\},$$

$$G_K(r) = \inf_{z \in \mathcal{Q}_K} \max_j \left\{ \beta_j + \frac{1}{K} \sum_{k=1}^{K} [h(r_k - z_k) - h(y_{jk} - z_k)] \right\}. \tag{106}$$

The direct pair (A, G_K) is feasible. To compute the closure (102), apply this same oracle to $\check{h}(v) = h(-v)$, with y_j replaced by $u \in \mathcal{U}$ and β_j by a_u. Its first potential is $\Psi_{\mathcal{U}}$ and its transformed potential is $\Phi_\star$. Thus both orientations use the same solver. Unlike the closure, replacing the original pair directly by (A, G_K) does not guarantee pointwise improvement.

Closed forms and the quadratic LP. For $h(v) = av_+ + b(-v)_+$, $a, b \geq 0$, the cost satisfies the directed triangle inequality. Consequently $c(z, y_j) \leq c(z, r) + c(r, y_j)$ gives $c(z, r) - A(z) \geq -A(r)$, with equality at $z = r$. Hence

$$G_K(r) = -A(r). \tag{107}$$

This covers W_1, both one-sided linear losses, and asymmetric linear losses. For symmetric W_1, the original source extension already equals $-\Psi_{\mathcal{U}}$, so the closure leaves it unchanged. For $h(v) = v^2$, cancellation gives the exact LP

$$G_K(r) = \min_{z \in \mathcal{Q}_K,\, t \in \mathbb{R}} \left\{ t : \; t \geq \beta_j + \frac{2\langle z, y_j - r \rangle + \|r\|^2 - \|y_j\|^2}{K} \quad \forall j \right\}. \tag{108}$$

For general convex h, the difference in (106) need not be convex in z. For example, $h(v) = (v - 0.1)_+^2$, $r = 0.3$, and $y = 0.8$ give $h(r - z) - h(y - z) = -(0.7 - z)^2$ on $[0.2, 0.7]$. A local minimum or a primal feasible value therefore cannot serve as a certified lower bound.

Certified global computation for the remaining convex costs. Within a box $l \leq z \leq u$, let $s_{jk} z_k + d_{jk}$ be the secant upper bound for the convex function $z_k \mapsto h(y_{jk} - z_k)$. Let $a_{kq} w + b_{kq} \leq h(w)$ be supporting affine functions. A lower relaxation of (106) is

$$\begin{aligned} \min_{z, v, t} \quad & t \\ \text{subject to} \quad & v_k \geq a_{kq}(r_k - z_k) + b_{kq} && \text{for all } k, q, \\ & t \geq \beta_j + K^{-1} \sum_k (v_k - s_{jk} z_k - d_{jk}) && \text{for all } j, \\ & z \in \mathcal{Q}_K \cap [l, u], \qquad 0 \leq v_k \leq H, \end{aligned} \tag{109}$$

where $H = \max_{[-D,D]} h$. One may also impose $\min_j \beta_j - H \le t \le \max_j \beta_j + H$. Every original z embeds with $v_k = h(r_k - z_k)$, so the LP gives a valid lower relaxation. Evaluating the original objective at its returned z gives an upper bound. Branching over boxes and retaining bounds for every unexplored or pruned region gives a global bracket. Box refinement and supporting lines at the current interval endpoints and midpoint control both approximation errors. Linear multi-hinge costs are piecewise linear; arbitrary powers and their nonnegative mixtures use the same construction. Worst-case computation can be exponential.

For a numerical LP $\min d^T x$ subject to $Mx \le b$, $l \le x \le u$, any $\lambda \le 0$ gives the weak-duality lower bound

$$b^T\lambda + \sum_i \min\{(d - M^T\lambda)_i l_i, (d - M^T\lambda)_i u_i\}. \tag{110}$$

This pays for the stationarity residual explicitly. The implementation adds a floating-point guard; it does not claim interval-arithmetic verification. Reaching a node budget does not establish the requested precision: the remaining bracket is recorded and (104) is still safe. Precision on finitely many evaluated curves alone does not establish the uniform τ assumption in Corollary E.7.

Finally, Proposition G.1(d) provides a route for other convex losses. If a convex interpolant satisfies $h \le h_J \le h + \varepsilon_h$, a feasible pair for $c_{h_J,K}$ becomes feasible for $c_{h,K}$ after subtracting ε_h from one potential. For mesh width Δ and an L_h-Lipschitz loss, $\varepsilon_h \le L_h\Delta/2$. This approximation contributes an additional $O(\varepsilon_h)$ term; the explicit multi-hinge costs need no such replacement.

E.7 Proof of Theorem E.2

Proof. The finite cost is Lipschitz in either argument:

$$|c_{h,K}(z,r) - c_{h,K}(u,r)| \le L_h d_K(z,u), \tag{111}$$

and analogously in r. By (75), α_i is the restriction to the source atoms of a pointwise minimum of functions that are L_h-Lipschitz in d_K. Consequently,

$$|\alpha_i - \alpha_\ell| \le L_h d_K(z_{0,i}, z_{0,\ell}). \tag{112}$$

For arbitrary z, r, let $i^\star$ attain the maximum in (77). Then

$$\begin{aligned}\widehat{\Phi}_K(z) + \widehat{\Psi}_K(r) &\le \alpha_{i^\star} - L_h d_K(z, z_{0,i^\star}) + c_{h,K}(z_{0,i^\star}, r) - \alpha_{i^\star} \\ &\le c_{h,K}(z,r),\end{aligned}$$

where the last inequality follows from (111). Combining this result with Theorem E.1 proves (79).

Taking the term $\ell = i$ in (77) gives $\widehat{\Phi}_K(z_{0,i}) \ge \alpha_i$. Conversely, (112) gives, for every ℓ,

$$\alpha_\ell - L_h d_K(z_{0,i}, z_{0,\ell}) \le \alpha_i.$$

Thus $\widehat{\Phi}_K(z_{0,i}) = \alpha_i$.

Dual feasibility and (75) imply

$$\widehat{\Psi}_K(z_{1,j}) = \min_i\{c_{h,K}(z_{0,i}, z_{1,j}) - \alpha_i\} \ge \beta_j.$$

If the weighted sum of these inequalities were strict, replacing every β_j by $\widehat{\Psi}_K(z_{1,j})$ would produce a feasible solution to (74) with a strictly larger objective, contradicting optimality. Because every w_j is strictly positive, each inequality is therefore an equality. This proves (80) and equality of the extended and finite dual objectives.

Finally, $\widehat{\Phi}_K$ is a pointwise maximum of functions that are L_h-Lipschitz in d_K, while $\widehat{\Psi}_K$ is a pointwise minimum of functions with the same Lipschitz constant. Hence both are L_h-Lipschitz. The encoded quantile cone has d_K-diameter at most D, proving that each potential has oscillation at most $L_h D$. For the anchored normalization, the same Lipschitz bound on the training values gives $0 \le \alpha_i \le L_h D$. Since $d_K \le D$, the maximum defining $\widehat{\Phi}_K$ lies in $[-L_h D, L_h D]$. Also, $0 \le c_{h,K} \le L_h D$; hence every term defining $\widehat{\Psi}_K$ is at least $-L_h D$, and the term at an anchor with $\alpha_i = 0$ is at most $L_h D$. This proves (76). □

E.8 Proof of Theorem E.3

Proof. The finite cost is L_h-Lipschitz in either argument under d_K. Equation (82) expresses each a_i as the restriction of a pointwise maximum of functions sharing this Lipschitz constant. Hence

$$|a_i - a_\ell| \le L_h d_K(z_{0,i}, z_{0,\ell}).$$

Let $i^\star$ attain the minimum in (83). Then

$$\begin{aligned}\widehat{\Phi}_K^{U,0}(z) + \widehat{\Psi}_K^{U,0}(r) &\ge a_{i^\star} + L_h d_K(z, z_{0,i^\star}) + c_{h,K}(z_{0,i^\star}, r) - a_{i^\star} \\ &\ge c_{h,K}(z, r).\end{aligned}$$

Adding $b_K/2$ to each potential and applying (72) proves (86).

The Lipschitz relation among the a_i and the term indexed by i give $\widehat{\Phi}_K^{U,0}(z_{0,i}) = a_i$. Upper-dual feasibility gives

$$\widehat{\Psi}_K^{U,0}(z_{1,j}) = \max_i \{c_{h,K}(z_{0,i}, z_{1,j}) - a_i\} \le b_j.$$

If any inequality were strict, replacing b_j by the displayed maximum would preserve every upper-dual constraint and strictly decrease the objective, because all w_j are positive. Optimality therefore gives equality and proves (87). The corrected objective is consequently the finite optimum plus b_K.

Pointwise minima and maxima preserve a common Lipschitz constant, so both uncorrected potentials have oscillation at most $L_h D$; adding constants does not change this. Under the normalization $\min_i a_i = 0$, the Lipschitz relation gives $0 \le a_i \le L_h D$. Thus $0 \le \widehat{\Phi}_K^{U,0} \le L_h D$, while $-L_h D \le \widehat{\Psi}_K^{U,0} \le L_h D$. Since $0 \le b_K \le L_h D$, (88) follows. □

E.9 Paired clipping and training-conditional range calibration

A globally feasible extension can have a much smaller certified range than the generic cost envelope. The following operation preserves feasibility and the fitted objective while making that range explicit. It applies to every nonnegative cost, including the asymmetric convex displacement costs in this paper. Write $[z]_a^b = \min\{b, \max\{a, z\}\}$ for $a \le b$.

Proposition E.8 (Paired clipping of a globally feasible lower dual)**.** *Let $c : \mathcal{Z}^2 \to [0, \infty)$ and let measurable real-valued potentials (ϕ, ψ) satisfy $\phi(y) + \psi(y') \le c(y, y')$ for every y, y'. For finite $a \le b$, define*

$$\phi^{\mathrm{cl}}(y) = [\phi(y)]_a^b - \frac{a+b}{2}, \qquad \psi^{\mathrm{cl}}(y') = [\psi(y')]_{-b}^{-a} + \frac{a+b}{2}. \tag{113}$$

Then, on the entire outcome space,

$$\phi^{\mathrm{cl}}(y) + \psi^{\mathrm{cl}}(y') \le c(y, y'), \qquad |\phi^{\mathrm{cl}}(y)| \vee |\psi^{\mathrm{cl}}(y')| \le \frac{b-a}{2}. \tag{114}$$

For two nonempty finite training supports $\{y_i\}$ and $\{y'_j\}$, choose

$$a = \min\big(\{\phi(y_i)\}_i \cup \{-\psi(y'_j)\}_j\big), \qquad b = \max\big(\{\phi(y_i)\}_i \cup \{-\psi(y'_j)\}_j\big). \tag{115}$$

For any nonnegative training weights summing to one in each arm, the empirical dual objective is unchanged. If the initial extension attains the empirical optimum, the clipped pair also attains it. A Lipschitz bound on either original potential is preserved by clipping and centering.

Proof. The centering constants cancel. If $\phi(y) < a$, the clipped sum is at most $a - a = 0 \le c(y, y')$. If $\psi(y') < -b$, it is at most $b - b = 0 \le c(y, y')$. In all other cases neither potential increases under clipping, so the original feasibility inequality applies. Their respective clipped ranges are $[a, b]$ and $[-b, -a]$; centering gives the displayed absolute bounds. Under (115), no training value is altered by clipping. The subsequent constant shifts cancel in the empirical objective because both arm weights sum to one. Finally, scalar clipping is a contraction, and additive constants do not affect Lipschitz constants. □

The intervals must be paired; arbitrary separate clipping need not preserve feasibility. This proposition starts from an already globally feasible extension, such as the constructions above. It does not turn

a dual that is feasible only on sampled constraints into a global certificate. Apply it to the final full-space potentials, after any necessary discretization correction. Its feasibility statement requires neither convexity nor a metric cost. Although the empirical objective is preserved, the population dual objective need not improve: clipping alone does not establish a smaller learning gap or preserve an asserted statistical rate.

Corollary E.9 (Training-conditional finite-class calibration)**.** *Let $\mathcal{J}$ be a fixed finite collection of J_0 policy pairs. Suppose the training fold produces globally feasible lower potentials and measurable finite bounds $A_j(D_1) \le B_j(D_1)$ such that their IPW scores satisfy*

$$A_j(D_1) \le S(o; j, D_1) \le B_j(D_1) \quad \textit{for every allowed observation } o, \quad j \in \mathcal{J}.$$

Assume the identification and known-propensity conditions for the IPW identity, and use an independent iid evaluation fold of size m. With $W_j = B_j - A_j$, define

$$L_j = \frac{1}{m} \sum_{i \in D_2} S_i(j) - W_j \sqrt{\frac{\log(J_0/\alpha)}{2m}}. \tag{116}$$

Then, almost surely in D_1,

$$\Pr\{L_j \le \Theta_c^L(j) \; \forall j \in \mathcal{J} \mid D_1\} \ge 1 - \alpha,$$

and the same inequality holds unconditionally.

Proof. Conditional on D_1, the scores are iid with fixed ranges. Hoeffding's inequality bounds each upper deviation of their average from its conditional expectation by α/J_0 at the displayed radius. A union bound controls every pair. Global feasibility and the conditional IPW identity put each score expectation below its sharp lower endpoint. Integrating over D_1 proves unconditional coverage. □

For example, apply Proposition E.8 separately to each action pair and covariate cell. If the resulting half-widths have a finite training-measurable uniform upper bound $M(D_1)$, overlap gives $|S_i| \le B(D_1) := 2M(D_1)/\underline{e}$. Taking $W_j = 2B(D_1)$ yields

$$r = B(D_1)\sqrt{2\log(J_0/\alpha)/m}.$$

The factor two in B allows both IPW terms to be active for a same-action comparison. A smaller range can be proved for a particular design. Bounds based only on observed evaluation-score extrema do not justify this conditional Hoeffding argument. For an infinite policy class, the finite union bound must instead be replaced by a valid uniform complexity bound.

Upper certificates. If $0 \le c \le C_c$ and $u + v \ge c$, then $(C_c/2 - u, C_c/2 - v)$ is a lower feasible pair for the nonnegative complement cost $C_c - c$. Paired clipping followed by the inverse transformation gives a globally superfeasible pair for c, preserving its training objective when the training-range rule is used. Applying the corresponding upper-tail calibration requires bounding the transformed IPW scores, including same-action terms and the added cost-envelope constants.

E.10 Proof of finite-sample coverage

Proof. Conditional on the training samples, the two learned potentials are fixed. By (79) and weak duality,

$$\mu_K := P_0\widehat{\Phi}_K \circ B_K + P_1\widehat{\Psi}_K \circ B_K \le T_{c_h}(P_0, P_1).$$

Theorem E.2 bounds the ranges of the two evaluation summands by $L_h D$. Hoeffding's inequality for the weighted sum of the two independent empirical means therefore yields

$$\Pr\left(\overline{T}_K - \mu_K > t \mid Q_0^{\mathrm{tr}}, Q_1^{\mathrm{tr}}\right) \le \exp\left\{-\frac{2t^2}{(L_h D)^2(m_0^{-1} + m_1^{-1})}\right\}.$$

Taking $t = r_{K,\boldsymbol{m},\alpha}$ makes the right-hand side equal to α. Thus, with conditional probability at least $1 - \alpha$,

$$L_{K,\alpha}^{\mathrm{comp}} = \overline{T}_K - r_{K,\boldsymbol{m},\alpha} \le \mu_K \le T_{c_h}(P_0, P_1).$$

Integrating over the training samples proves (92).

For the upper construction, conditional on the training sample, global superfeasibility and upper weak duality give

$$\mu_K^U := P_0\widehat{\Phi}_K^U \circ B_K + P_1\widehat{\Psi}_K^U \circ B_K \geq T_{c_h}^U(P_0, P_1).$$

Theorem E.3 gives the same $L_h D$ oscillation bound for each evaluation summand. Applying Hoeffding's inequality to the opposite deviation yields

$$\Pr\left(\mu_K^U - \overline{T}_K^U > t \mid Q_0^{\mathrm{tr}}, Q_1^{\mathrm{tr}}\right) \leq \exp\left\{-\frac{2t^2}{(L_h D)^2(m_0^{-1} + m_1^{-1})}\right\}.$$

At $t = r_{K,\boldsymbol{m},\alpha}$, with conditional probability at least $1-\alpha$,

$$T_{c_h}^U(P_0, P_1) \leq \mu_K^U \leq \overline{T}_K^U + r_{K,\boldsymbol{m},\alpha} = U_{K,\alpha}^{\mathrm{comp}}.$$

Integrating over the training sample proves (94). The two-sided statement follows by applying the lower and upper conclusions at level $\alpha/2$ and taking a union bound. □

F Geometric Minimax Rates for Policy Certificates

We now quantify how closely honest policy-indexed certificates can approach the two sharp endpoints from Section B. Unlike a two-sample comparison between two fixed actions, the inferential objects are simultaneous lower and upper confidence bands over the policy pairs in Π^2. We isolate the outcome-side difficulty while retaining the covariates, actions, and deterministic policies of Section A.

Let

$$\mathrm{supp}(X) = \mathcal{X}_{K_X} = \{x_1, \ldots, x_{K_X}\}, \qquad p_j := P_X\{x_j\} \geq \kappa > 0,$$

where $K_X < \infty$ is fixed. We work in a randomized-design experiment with known assignment probabilities

$$e_a(x_j) = P(A = a \mid X = x_j), \qquad \sum_{a=1}^{d} e_a(x_j) = 1, \qquad e_a(x_j) \geq \underline{e} > 0. \tag{RD}$$

The array $\{e_a(x_j) : a \leq d, j \leq K_X\}$ is part of the statistical experiment, is held fixed across the model class below, and is not estimated from the outcome sample. After an affine rescaling of the compact outcome interval I, every distribution-valued outcome is represented by its quantile function in

$$\mathcal{Q} = \{q \in L^1[0,1] : q \text{ has a nondecreasing representative taking values in } [0,1]\}.$$

For $1 \leq p < \infty$, define

$$d_p(q, r) = \|q - r\|_{L^p[0,1]}, \qquad c_p(q, r) = \int_0^1 |q(t) - r(t)|^p \, dt.$$

Thus this section specializes the cost in Section A to the covariate-independent power displacement cost c_p.

For an action pair (a, b) at x, put

$$T_{p,x}(a, b) := \inf_{\gamma \in \Gamma(P_{a,x}, P_{b,x})} \int_{\mathcal{Q}^2} c_p(q, r) \, d\gamma(q, r). \tag{117}$$

The policy-indexed sharp endpoint is

$$T_p(\pi, \pi') := \sum_{j=1}^{K_X} p_j T_{p,x_j}\big(\pi(x_j), \pi'(x_j)\big) = \Theta_{c_p}^L(\pi, \pi'). \tag{118}$$

Consequently, every lower confidence bound for $T_p(\pi, \pi')$ is also a lower certificate for the partially identified population matching target $\Theta_{c_p}(\pi, \pi'; \eta)$, for every admissible matching kernel η.

The sharp upper endpoint is

$$T_{p,x}^U(a, b) := \sup_{\gamma \in \Gamma(P_{a,x}, P_{b,x})} \int_{\mathcal{Q}^2} c_p(q, r) \, d\gamma(q, r), \tag{119}$$

$$T_p^U(\pi,\pi') := \sum_{j=1}^{K_X} p_j T_{p,x_j}^U\big(\pi(x_j),\pi'(x_j)\big) = \Theta_{c_p}^U(\pi,\pi'). \tag{120}$$

Every upper confidence bound for T_p^U is therefore an upper certificate for every admissible population matching kernel.

Let

$$\Pi_{K_X} := \{(\pi(x_1),\dots,\pi(x_{K_X})) : \pi \in \Pi\}, \qquad N_\Pi := |\Pi_{K_X}| \le d^{K_X}.$$

Only these action vectors are statistically relevant; policies that differ outside $\mathcal{X}_{K_X}$ induce the same endpoint and score. Let $\mathfrak{P}_{K_X,\kappa,\underline{e}}$ be the collection of observed-data laws satisfying Assumptions A.2–A.4, the finite-support conditions above, and the common known randomization mechanism (RD), with otherwise arbitrary conditional outer laws $P_{a,x} \in \mathcal{P}(\mathcal{Q})$.

Policy-uniform learned-dual certificate. Split the observations into a training fold D_1 of size ℓ and an evaluation fold D_2 of size m, with both sizes fixed positive fractions of n. On D_1, form the empirical conditional law $\widehat{P}_{a,x}$ in each action–covariate cell; an empty cell is assigned an arbitrary fixed fallback law. For every $x \in \mathcal{X}_{K_X}$ and $a,b \in \mathcal{A}$, choose an empirical dual optimizer

$$(\widehat{f}_{ab,x}, \widehat{f}_{ab,x}^{c_p}) \in \operatorname*{arg\,max}_{f(q)+g(r)\le c_p(q,r)} \{\widehat{P}_{a,x} f + \widehat{P}_{b,x} g\}, \tag{121}$$

normalized so that

$$0 \le \widehat{f}_{ab,x} \le 1, \qquad -1 \le \widehat{f}_{ab,x}^{c_p} \le 1.$$

For $a = b$, $c_p \ge 0$ and $c_p(q,q) = 0$, so the sharp lower endpoint between the identical empirical marginals is zero. In this specialized minimax lower-band construction, we select the zero dual pair for those cells. This normalization attains the lower endpoint; the upper endpoint for identical marginals can still be positive. The resulting collection is globally dual-feasible for every action pair, including when a fallback law was used.

For $i \in D_2$, define

$$\begin{aligned} S_i(\pi,\pi') = {} & \frac{\mathbb{1}\{A_i = \pi(X_i)\}}{e_{\pi(X_i)}(X_i)} \widehat{f}_{\pi(X_i),\pi'(X_i),X_i}(Q_{Y_i}) \\ & + \frac{\mathbb{1}\{A_i = \pi'(X_i)\}}{e_{\pi'(X_i)}(X_i)} \widehat{f}^{c_p}_{\pi(X_i),\pi'(X_i),X_i}(Q_{Y_i}), \end{aligned} \tag{122}$$

$$\widehat{T}_p(\pi,\pi') = \frac{1}{m} \sum_{i \in D_2} S_i(\pi,\pi'). \tag{123}$$

With

$$r_{m,\alpha}^\Pi := \frac{1}{\underline{e}} \sqrt{\frac{2\log(N_\Pi^2/\alpha)}{m}}, \qquad L_{n,p,\alpha}^\Pi(\pi,\pi') := \widehat{T}_p(\pi,\pi') - r_{m,\alpha}^\Pi, \tag{124}$$

the entire collection $\{L_{n,p,\alpha}^\Pi(\pi,\pi') : (\pi,\pi') \in \Pi^2\}$ is one simultaneous confidence band.

A possibly randomized collection $\boldsymbol{L}_n = \{L_n(\pi,\pi') : (\pi,\pi') \in \Pi^2\}$ is *uniformly honest* at level $1-\alpha$ if

$$\inf_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} P\big\{L_n(\pi,\pi') \le T_{p,P}(\pi,\pi') \text{ for every } (\pi,\pi') \in \Pi^2\big\} \ge 1-\alpha.$$

Write $\mathcal{H}_{n,\alpha}^{(p)}(K_X,\Pi)$ for all such collections and define their minimax worst-policy excess length by

$$R_{n,p}^{\mathrm{hon}}(K_X,\Pi) := \inf_{\boldsymbol{L}_n \in \mathcal{H}_{n,\alpha}^{(p)}(K_X,\Pi)} \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2} \{T_{p,P}(\pi,\pi') - L_n(\pi,\pi')\}_+\right]. \tag{125}$$

Let $\beta_p := p \wedge 2$.

Theorem F.1 (Policy-uniform minimax honest rate). *Fix $K_X, d < \infty$, $1 \le p < \infty$, $0 < \alpha < 1/3$, $0 < \kappa \le 1/K_X$, and $0 < \underline{e} \le 1/d$, and fix a known assignment array satisfying* (RD). *Suppose*

Π_{K_X} *contains at least two distinct action vectors. Then there are constants* $c, C \in (0,\infty)$ *and* $n_0 < \infty$*, depending only on* $(p, \alpha, K_X, d, \kappa, \underline{e})$*, such that for every* $n \geq n_0$*,*

$$\inf_{\boldsymbol{L}_n \in \mathcal{H}^{(p)}_{n,\alpha}(K_X,\Pi)} \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} P\left\{ \begin{array}{c} \sup_{(\pi,\pi')\in\Pi^2} \left[T_{p,P}(\pi,\pi') - L_n(\pi,\pi')\right] \\ \geq \dfrac{c}{(\log(n+1))^{\beta_p}} \end{array} \right\} \geq \frac{1}{3}, \tag{126}$$

$$c(\log(n+1))^{-\beta_p} \leq R^{\mathrm{hon}}_{n,p}(K_X,\Pi) \leq C(\log(n+1))^{-\beta_p}. \tag{127}$$

Moreover, the learned-dual band in (124) *is exactly honest for every finite* n*, and its excess length obeys*

$$\sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \left[\sup_{(\pi,\pi')\in\Pi^2} \{T_{p,P}(\pi,\pi') - L^{\Pi}_{n,p,\alpha}(\pi,\pi')\}_+ \right]$$

$$\leq C\{\log(\ell\kappa\underline{e}+1)\}^{-\beta_p} + \frac{C}{\underline{e}}\sqrt{\frac{K_X \log d + \log(1/\alpha)}{m}}. \tag{128}$$

F.1 The matching upper-endpoint minimax law

The lower-endpoint hard family cannot simply be reused for the upper endpoint: a maximal coupling can avoid a low-cost diagonal. The correct mirror image is a code whose matched labels have the largest cost and whose mismatched labels lose a controlled amount. For the attainable upper bound, put $\bar{c}_p := 1 - c_p$. Then, for every pair of outer laws,

$$T^U_p(P,Q) = 1 - T_{\bar{c}_p}(P,Q), \tag{129}$$

where $T_{\bar{c}_p}$ is the lower transport value for the bounded complement cost. Its rectangular interaction is the negative of that of c_p, so the normalized complement-cost dual class has the same entropy rate.

On the training fold, choose normalized empirical lower-dual optimizers $(\widehat{h}_{ab,x}, \widehat{h}^{\bar{c}_p}_{ab,x})$ for $\bar{c}_p$ in every action–covariate cell. On the evaluation fold, define

$$H_i(\pi,\pi') = \frac{\mathbb{1}\{A_i = \pi(X_i)\}}{e_{\pi(X_i)}(X_i)} \widehat{h}_{\pi(X_i),\pi'(X_i),X_i}(Q_{Y_i}) + \frac{\mathbb{1}\{A_i = \pi'(X_i)\}}{e_{\pi'(X_i)}(X_i)} \widehat{h}^{\bar{c}_p}_{\pi(X_i),\pi'(X_i),X_i}(Q_{Y_i}), \tag{130}$$

$$\widehat{T}^U_p(\pi,\pi') := 1 - \frac{1}{m}\sum_{i\in D_2} H_i(\pi,\pi'), \qquad U^{\Pi}_{n,p,\alpha}(\pi,\pi') := \widehat{T}^U_p(\pi,\pi') + r^{\Pi}_{m,\alpha}. \tag{131}$$

A collection $\boldsymbol{U}_n$ is *uniformly honest* for the upper endpoint if

$$\inf_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} P\big\{U_n(\pi,\pi') \geq T^U_{p,P}(\pi,\pi') \text{ for every } (\pi,\pi') \in \Pi^2\big\} \geq 1-\alpha.$$

Let $\mathcal{H}^{(p),U}_{n,\alpha}(K_X,\Pi)$ contain all such collections and define

$$R^{\mathrm{hon},U}_{n,p}(K_X,\Pi) := \inf_{\boldsymbol{U}_n \in \mathcal{H}^{(p),U}_{n,\alpha}(K_X,\Pi)} \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \left[\sup_{(\pi,\pi')\in\Pi^2} \{U_n(\pi,\pi') - T^U_{p,P}(\pi,\pi')\}_+ \right]. \tag{132}$$

For point estimation, let

$$\mathcal{R}^{\mathrm{est},U}_{n,p}(K_X,\Pi) := \inf_{\widetilde{T}^U_n} \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \left[\sup_{(\pi,\pi')\in\Pi^2} |\widetilde{T}^U_n(\pi,\pi') - T^U_{p,P}(\pi,\pi')| \right]. \tag{133}$$

Theorem F.2 (Policy-uniform minimax upper-endpoint rate)**.** *Under the conditions of Theorem F.1, there are constants* $c, C \in (0,\infty)$ *and* $n_0 < \infty$*, depending only on* $(p, \alpha, K_X, d, \kappa, \underline{e})$*, such that, for every* $n \geq n_0$*,*

$$\inf_{\boldsymbol{U}_n \in \mathcal{H}^{(p),U}_{n,\alpha}(K_X,\Pi)}$$

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} P\Big\{ \sup_{(\pi,\pi')\in\Pi^2} [U_n(\pi,\pi') - T^U_{p,P}(\pi,\pi')] \geq \frac{c}{\{\log(n+1)\}^{\beta_p}}\Big\} \geq \frac{1}{3}, \tag{134}$$

$$c \leq \{\log(n+1)\}^{\beta_p} \times R^{\mathrm{hon},U}_{n,p}(K_X,\Pi) \leq C, \tag{135}$$

$$c \leq \{\log(n+1)\}^{\beta_p} \times \mathcal{R}^{\mathrm{est},U}_{n,p}(K_X,\Pi) \leq C. \tag{136}$$

Moreover, $\boldsymbol{U}^{\Pi}_{n,p,\alpha}$ in (131) is exactly honest for every finite n, and

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2}\{U^{\Pi}_{n,p,\alpha}(\pi,\pi') - T^U_{p,P}(\pi,\pi')\}_+\right] \leq C\{\log(\ell\kappa\underline{e}+1)\}^{-\beta_p} + \frac{C}{\underline{e}}\sqrt{\frac{K_X\log d + \log(1/\alpha)}{m}}, \tag{137}$$

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}^U_p - T^U_{p,P}| \leq C\{\log(\ell\kappa\underline{e}+1)\}^{-\beta_p} + \frac{C}{\underline{e}}\sqrt{\frac{K_X\log d}{m}}. \tag{138}$$

The same conclusions hold when the model class is restricted to the uniformly Lipschitz quantile outcomes in Corollary F.11, with constants additionally allowed to depend on the fixed Lipschitz bound.

The upper and lower endpoints therefore have the same unrestricted minimax order, but for different coupling-geometric reasons. The lower-endpoint code makes label mismatch expensive; the upper-endpoint code makes label mismatch reduce the largest attainable cost. This distinction is essential: reversing the dual inequality alone proves validity, not the upper minimax lower bound.

Returning to the lower endpoint, the same construction is minimax optimal as a point estimator, not only as a lower confidence certificate. For an arbitrary estimator collection $\widetilde{\boldsymbol{T}}_n = \{\widetilde{T}_n(\pi,\pi') : (\pi,\pi')\in\Pi^2\}$, define

$$\mathcal{R}^{\mathrm{est}}_{n,p}(K_X,\Pi) := \inf_{\widetilde{\boldsymbol{T}}_n} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2} |\widetilde{T}_n(\pi,\pi') - T_{p,P}(\pi,\pi')|\right]. \tag{139}$$

The infimum ranges over all measurable, possibly randomized estimators based on the observed sample.

Theorem F.3 (Minimax optimality of the learned-dual estimator)**.** *Under the conditions of Theorem F.1, there are constants $c, C \in (0,\infty)$ such that, for all sufficiently large n,*

$$c(\log(n+1))^{-\beta_p} \leq \mathcal{R}^{\mathrm{est}}_{n,p}(K_X,\Pi) \leq \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2} |\widehat{T}_p(\pi,\pi') - T_{p,P}(\pi,\pi')|\right] \leq C\{\log(\ell\kappa\underline{e}+1)\}^{-\beta_p} + \frac{C}{\underline{e}}\sqrt{\frac{K_X\log d}{m}}. \tag{140}$$

Consequently, for a fixed-proportion split,

$$\mathcal{R}^{\mathrm{est}}_{n,p}(K_X,\Pi) \asymp \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_p - T_{p,P}| \asymp (\log n)^{-(p\wedge 2)}.$$

Thus no estimator, whether dual-based, plug-in, regularized, or randomized, can improve the worst-case convergence order uniformly over $\mathfrak{P}_{K_X,\kappa,\underline{e}}$.

Corollary F.4 (Rate-optimal feasible-dual certification)**.** *Under the conditions of Theorem F.1, with a fixed-proportion split,*

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_p - T_{p,P}| \asymp (\log n)^{-\beta_p}, \tag{141}$$

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{\varepsilon}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_p^U - T_{p,P}^U| \asymp (\log n)^{-\beta_p}. \tag{142}$$

The exactly honest bands obtained by adding or subtracting the simultaneous evaluation radius have the same order. Since their calibration term is root-m, hence $o\{(\log n)^{-\beta_p}\}$, exact simultaneous honesty and post-selection protection preserve the minimax rate of unconstrained point estimation.

Validity requires fewer assumptions than rate optimality. Conditional on the training fold, it uses only bounded global subfeasibility or superfeasibility and therefore remains true for an arbitrarily misspecified first-stage outer-law learner. Rate optimality additionally requires the learned-dual approximation gap to have the order in (128) and (137), as it does for the empirical conditional-law learner used here. Thus first-stage misspecification is free for validity, but not necessarily for tightness.

The matched endpoint theorems separate the sources of uncertainty. Learning the conditional outer laws and their dual optimizers contributes the logarithmic term. Simultaneous evaluation over the policy class contributes only the second, root-sample term because at most d^{K_X} policy action vectors occur on the support of X. For fixed K_X and d, the outcome-side logarithmic term is therefore minimax dominant. The result continues to cover a policy or policy pair selected using D_2, since the coverage event holds before selection.

The exponent has the same contact interpretation as in the two-action submodel. For $1 \le p \le 2$, a zero-displacement staircase code gives a separation of order $(\log n)^{-p}$. For $p > 2$, the cost is too flat at zero to be least favorable; a fixed nonzero displacement activates its quadratic curvature and yields order $(\log n)^{-2}$. The lower bound embeds either construction in one covariate cell and one action pair selected by two policies. Thus neither additional actions nor policy indexing remove the infinite-dimensional outcome-side obstruction.

F.2 A ONE-SIDED OBSTRUCTION FOR GENERAL CONVEX DISPLACEMENT COSTS

The preceding matched result uses the special geometry of the power cost. For a general convex displacement cost, convexity alone does not determine a rate. A lower bound nevertheless follows from a finite-order contact that is visible to the quantile code. We use the sign convention of (34): for $q, r \in \mathcal{Q}$,

$$c_h(q,r) := \int_0^1 h\{r(t) - q(t)\}\, dt.$$

Definition F.5 (Finite-order convex contact)**.** *Let $h : [-1,1] \to [0,\bar{h}]$, $0 < \bar{h} < \infty$, be finite, continuous, and convex, with $h(0) = 0$ and $0 \in \partial h(0)$. We say that h satisfies (C_{β_h}) if there are*

$$z_h \in (-1,1), \qquad g_h \in \partial h(z_h), \qquad \beta_h \in [1,\infty),$$

and constants $0 < a_h \le 1 - |z_h|$ and $\underline{c}_h > 0$ such that

$$h(z_h + u) - h(z_h) - g_h u \ge \underline{c}_h |u|^{\beta_h} \qquad \text{for every } |u| \le a_h. \tag{143}$$

The notation $\underline{c}_h$ distinguishes the contact constant from the cost functional c_h.

The contact condition is local and one-sided. Attainability is controlled by how rapidly the cost's rectangular interaction can change over the entire displacement interval. The following condition records the corresponding global upper order. For $0 < \gamma \le 2$, write

$$\vartheta_\gamma(u,v) := \begin{cases} \min\{|u|^\gamma, |v|^\gamma\}, & 0 < \gamma \le 1, \\ \min\{|u||v|^{\gamma-1}, |v||u|^{\gamma-1}\}, & 1 < \gamma \le 2. \end{cases}$$

Definition F.6 (Uniform rectangular order)**.** *We say that h satisfies (U_{γ_h}) if there is $L_h < \infty$ such that, with $s_h := \gamma_h \wedge 1$,*

$$|h(z) - h(z')| \le L_h |z - z'|^{s_h}, \tag{144}$$

$$|h(z+u+v) - h(z+u) - h(z+v) + h(z)| \le L_h \vartheta_{\gamma_h}(u,v) \tag{145}$$

whenever all four arguments in (145) belong to $[-1,1]$.

For $h(z) = |z|^p$, the elementary power inequalities give $(\mathrm{U}_{p\wedge 2})$. Condition (U_{γ_h}) is global, whereas (C_{β_h}) is local: the former controls the entropy of the whole dual class, whereas the latter supplies one cost-visible hard packing.

For laws $P, Q \in \mathcal{P}(\mathcal{Q})$, write

$$T_h(P,Q) := \inf_{\gamma\in\Gamma(P,Q)} \int c_h(q,r)\, d\gamma(q,r).$$

For this cost, define

$$T_{h,x}(a,b) := T_h(P_{a,x}, P_{b,x}),$$
$$T_{h,P}(\pi,\pi') := \sum_{j=1}^{K_X} p_j T_{h,x_j}\{\pi(x_j), \pi'(x_j)\}.$$

Let $\mathcal{H}^{(h)}_{n,\alpha}(K_X,\Pi)$ be the collections $\boldsymbol{L}_n$ satisfying

$$\inf_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} P\big\{L_n(\pi,\pi') \le T_{h,P}(\pi,\pi') \text{ for every } (\pi,\pi')\in\Pi^2\big\} \ge 1-\alpha.$$

Define $R^{\mathrm{hon}}_{n,h}(K_X,\Pi)$ by replacing $T_{p,P}$ and $\mathcal{H}^{(p)}_{n,\alpha}$ in (125) with $T_{h,P}$ and $\mathcal{H}^{(h)}_{n,\alpha}$, respectively. Likewise, let

$$\mathcal{R}^{\mathrm{est}}_{n,h}(K_X,\Pi) := \inf_{\widetilde{\boldsymbol{T}}_n} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2} |\widetilde{T}_n(\pi,\pi') - T_{h,P}(\pi,\pi')|\right]. \tag{146}$$

For the upper endpoint, put

$$T^U_h(P,Q) := \sup_{\gamma\in\Gamma(P,Q)} \int c_h(q,r)\, d\gamma(q,r), \qquad T^U_{h,P}(\pi,\pi') := \sum_{j=1}^{K_X} p_j T^U_h(P_{\pi(x_j),x_j}, P_{\pi'(x_j),x_j}).$$

Let $\mathcal{H}^{(h),U}_{n,\alpha}(K_X,\Pi)$ contain the collections $\boldsymbol{U}_n$ satisfying

$$\inf_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} P\big\{U_n(\pi,\pi') \ge T^U_{h,P}(\pi,\pi') \text{ for every } (\pi,\pi')\in\Pi^2\big\} \ge 1-\alpha.$$

Define

$$R^{\mathrm{hon},U}_{n,h}(K_X,\Pi) := \inf_{\boldsymbol{U}_n\in\mathcal{H}^{(h),U}_{n,\alpha}(K_X,\Pi)} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2}\{U_n - T^U_{h,P}\}_+, \tag{147}$$

$$\mathcal{R}^{\mathrm{est},U}_{n,h}(K_X,\Pi) := \inf_{\widetilde{\boldsymbol{T}}^U_n} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}^U_n - T^U_{h,P}|. \tag{148}$$

Theorem F.7 (Polylogarithmic obstruction under finite-order convex contact)**.** *Fix $K_X, d < \infty$, $0 < \alpha < 1/3$, $0 < \kappa \le 1/K_X$, and $0 < \underline{e} \le 1/d$, and fix a known assignment array satisfying* (RD)*. Suppose that Π_{K_X} contains at least two distinct action vectors and that h satisfies (C_{β_h}). Then there are $c > 0$ and $n_0 < \infty$, depending only on $(h, \alpha, K_X, d, \kappa, \underline{e})$, such that, for every $n \ge n_0$,*

$$\inf_{\boldsymbol{L}_n\in\mathcal{H}^{(h)}_{n,\alpha}(K_X,\Pi)} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} P\left\{\sup_{(\pi,\pi')\in\Pi^2} [T_{h,P}(\pi,\pi') - L_n(\pi,\pi')] \ge c\{\log(n+1)\}^{-\beta_h}\right\} \ge \frac{1}{3}, \tag{149}$$

$$R^{\mathrm{hon}}_{n,h}(K_X,\Pi) \ge c\{\log(n+1)\}^{-\beta_h}, \tag{150}$$

$$\mathcal{R}^{\mathrm{est}}_{n,h}(K_X,\Pi) \ge c\{\log(n+1)\}^{-\beta_h}. \tag{151}$$

Theorem F.8 (Upper-endpoint obstruction under finite-order convex contact)**.** *Under the conditions of Theorem F.7, there are $c > 0$ and $n_0 < \infty$, depending only on $(h, \alpha, K_X, d, \kappa, \underline{e})$, such that, for every $n \ge n_0$,*

$$\inf_{\boldsymbol{U}_n\in\mathcal{H}^{(h),U}_{n,\alpha}(K_X,\Pi)} \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} P\left\{\sup_{(\pi,\pi')\in\Pi^2} [U_n(\pi,\pi') - T^U_{h,P}(\pi,\pi')] \ge c\{\log(n+1)\}^{-\beta_h}\right\} \ge \frac{1}{3}, \tag{152}$$

$$R^{\mathrm{hon},U}_{n,h}(K_X,\Pi) \ge c\{\log(n+1)\}^{-\beta_h}, \tag{153}$$

$$\mathcal{R}^{\mathrm{est},U}_{n,h}(K_X,\Pi) \ge c\{\log(n+1)\}^{-\beta_h}. \tag{154}$$

To state the attainable result, repeat the split-sample construction in (121)–(123) with c_h in place of c_p, and denote the resulting estimator by $\widehat{T}_h$. For the upper endpoint use the complement cost $\bar{c}_h := \bar{h} - c_h$ and put $\widehat{T}_h^U$ equal to $\bar{h}$ minus the corresponding held-out lower-dual score average. The same global feasibility argument gives exactly honest bands after subtracting or adding

$$r_{m,\alpha}^{\Pi,h} := \frac{\bar{h}}{\underline{e}}\sqrt{\frac{2\log(N_\Pi^2/\alpha)}{m}}.$$

Theorem F.9 (Attainable rate under uniform rectangular order)**.** *Fix the finite-design quantities in Theorem F.7. If h satisfies (U_{γ_h}) for some $0 < \gamma_h \le 2$, then*

$$\sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_h - T_{h,P}| \ \vee \sup_{P\in\mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_h^U - T_{h,P}^U|$$
$$\le C\{\log(\ell\kappa\underline{e}+1)\}^{-\gamma_h} + \frac{C\bar{h}}{\underline{e}}\sqrt{\frac{K_X \log d}{m}}. \tag{155}$$

The exactly honest lower and upper bands have the same bound with $K_X \log d$ replaced by $K_X \log d + \log(1/\alpha)$. Consequently, for a fixed-proportion split, both learned-dual estimators and both honest bands have worst-case error $O\{(\log n)^{-\gamma_h}\}$.

If, in addition, h satisfies (C_{β_h}) with $\gamma_h = \beta_h$, then the lower and upper endpoint estimation risks and the two honest-band risks all satisfy

$$\mathcal{R}_{n,h}^{\mathrm{est}} \asymp \mathcal{R}_{n,h}^{\mathrm{est},U} \asymp R_{n,h}^{\mathrm{hon}} \asymp R_{n,h}^{\mathrm{hon},U} \asymp (\log n)^{-\beta_h}.$$

Thus the empirical-dual estimators are minimax rate-optimal under matching local lower and global upper orders.

The contact assumption alone still gives only a one-sided conclusion. The two obstruction theorems rule out uniformly faster worst-case convergence: no procedure has error

$$o\{(\log n)^{-\beta_h}\}$$

over the stated model class for either endpoint, but they do not imply (U_{β_h}). For example, for any $a \in (0,1)$,

$$h_a(z) = 1 - az - (1-z)^a, \qquad -1 \le z \le 1,$$

is bounded, nonnegative, and convex, has $h_a(0) = h_a'(0) = 0$, and satisfies (C_2) near zero, while $h_a(1) - h_a(1-\delta) = \delta^a - a\delta$. Hence its global modulus can have an arbitrarily smaller order than its interior contact. The positive contact constant also excludes degenerate affine costs; under the normalization above, an affine h would necessarily be $h \equiv 0$. Proofs are in Appendix F.

F.3 Persistence under smooth individual laws

For fixed $L > 0$, let

$$\mathcal{Q}_L := \{q \in \mathcal{Q} : |q(t) - q(s)| \le L|t-s| \text{ for all } s,t \in [0,1]\},$$

and let $\mathfrak{P}_{K_X,\kappa,\underline{e},L}$ restrict every conditional law $P_{a,x}$ to $\mathcal{P}(\mathcal{Q}_L)$. Append the subscript L to each of the four general-cost risks defined above when its model supremum is taken over this restricted class.

Theorem F.10 (Individual smoothness does not reduce outer-law complexity)**.** *Fix the finite-design quantities in Theorem F.7 and any $L > 0$. If h satisfies (C_{β_h}), then, for all sufficiently large n,*

$$R_{n,h,L}^{\mathrm{hon}}(K_X,\Pi) \wedge R_{n,h,L}^{\mathrm{hon},U}(K_X,\Pi) \wedge \mathcal{R}_{n,h,L}^{\mathrm{est}}(K_X,\Pi) \wedge \mathcal{R}_{n,h,L}^{\mathrm{est},U}(K_X,\Pi) \ge c_{h,L}\{\log(n+1)\}^{-\beta_h}. \tag{156}$$

The probability lower bounds in Theorems F.7 and F.8 also remain valid after this restriction. If, additionally, h satisfies (U_{γ_h}) with $\gamma_h = \beta_h$, then all four restricted risks are

$$R_{n,h,L}^{\mathrm{hon}}(K_X,\Pi) \asymp R_{n,h,L}^{\mathrm{hon},U}(K_X,\Pi) \asymp \mathcal{R}_{n,h,L}^{\mathrm{est}}(K_X,\Pi) \asymp \mathcal{R}_{n,h,L}^{\mathrm{est},U}(K_X,\Pi) \asymp (\log n)^{-\beta_h}, \tag{157}$$

and the empirical-dual estimators and their exactly honest bands attain this rate.

The hard conditional outer laws can be supported entirely on strictly increasing, uniformly bi-Lipschitz quantile curves. Their induced one-dimensional distributions have compact supports and densities bounded above and away from zero on those supports, with constants allowed to depend on h and L.

The bi-Lipschitz quantiles in the hard construction induce atomless individual outcomes with densities bounded above and away from zero on their supports. Their outer laws retain exponentially many directions that the cost distinguishes. The local contact still separates those directions at order $J^{-\beta_h}$, so the logarithmic obstruction survives unchanged.

Corollary F.11 (Power-cost consequence under Lipschitz quantile outcomes)**.** *Under the conditions of Theorem F.1, for every fixed $L > 0$, let $R^{\mathrm{hon}}_{n,p,L}(K_X, \Pi)$ denote the lower honest risk with model class $\mathfrak{P}_{K_X,\kappa,\underline{e},L}$. Then*

$$R^{\mathrm{hon}}_{n,p,L}(K_X, \Pi) \asymp_{p,\alpha,K_X,d,\kappa,\underline{e},L} (\log n)^{-(p\wedge 2)}.$$

The learned-dual band in (124) remains exactly honest and attains this rate uniformly over $\mathfrak{P}_{K_X,\kappa,\underline{e},L}$.

Boundary of the result. The fixed-K_X condition is substantive. If the covariate support grows, rare action–covariate cells reduce the effective training sample size and the policy correction need not remain negligible. The upper bound (128) displays the relevant quantities: the logarithmic term depends on $\ell\kappa\underline{e}$, while policy uniformity depends on $K_X \log d$. With unrestricted covariate support and no regularity in $x \mapsto P_{a,x}$, a uniformly honest endpoint certificate need not be uniformly consistent. Theorems F.1, F.2, and F.3 identify the matched power-cost minimax rate under controlled, rather than absent, covariate complexity. Theorem F.7 supplies the one-sided obstruction under local contact alone, while Theorem F.9 gives a matching rate when the global rectangular order agrees with that contact. Theorem F.10 shows that neither conclusion is improved merely by imposing first-order smoothness on each individual law. Proofs are in Appendix F.

Throughout this appendix,

$$\beta_p := p \wedge 2, \qquad c_p(q, r) = \int_0^1 |q(t) - r(t)|^p \, dt,$$

and $\mathcal{Q}$ is the set of nondecreasing maps from $[0, 1]$ to $[0, 1]$, identified up to almost-everywhere equality.

F.4 Dual entropy and the conditional training gap

Let $\mathcal{F}_p$ denote the normalized c_p-concave potentials

$$f(q) = \inf_{r\in\mathcal{Q}} \{c_p(q, r) - g(r)\}, \qquad 0 \le f \le 1.$$

Their transforms

$$f^{c_p}(r) = \inf_{q\in\mathcal{Q}} \{c_p(q, r) - f(q)\}$$

take values in $[-1, 1]$, and $f(q) + f^{c_p}(r) \le c_p(q, r)$.

Lemma F.12 (Uniform empirical process for power-cost potentials)**.** *There is a constant $C_p < \infty$ such that, for every probability law P on $\mathcal{Q}$ and every empirical law $\widehat{P}_s = s^{-1} \sum_{i=1}^s \delta_{Q_i}$ from P,*

$$\mathbb{E} \sup_{f\in\mathcal{F}_p} |(\widehat{P}_s - P)f| \le C_p\{\log(s + 1)\}^{-\beta_p}, \tag{158}$$

$$\mathbb{E} \sup_{f\in\mathcal{F}_p} |(\widehat{P}_s - P)f^{c_p}| \le C_p\{\log(s + 1)\}^{-\beta_p}. \tag{159}$$

Proof. The bounded monotone cone satisfies

$$\log N(\varepsilon, \mathcal{Q}, \|\cdot\|_{L^r}) \le C_r\varepsilon^{-1}, \qquad 1 \le r < \infty. \tag{160}$$

For $r = 1$, this follows directly by approximating each quantile on an equal-width grid by a nondecreasing staircase: a $2/k$-net has at most $\binom{2k}{k} \le 4^k$ elements. The general one-dimensional monotone entropy bound follows from Gao & Wellner (2007).

For the rectangular interaction

$$\Delta_{c_p}((q, r), (q', r')) = c_p(q', r) - c_p(q', r') + c_p(q, r') - c_p(q, r),$$

the elementary power inequalities give

$$|\Delta_{c_p}| \le \begin{cases} C_p\|q-q'\|_p\|r-r'\|_p^{p-1}, & 1<p\le 2,\\ C_p\|q-q'\|_2\|r-r'\|_2, & p\ge 2,\\ 2\min\{\|q-q'\|_1,\|r-r'\|_1\}, & p=1. \end{cases} \tag{161}$$

We spell out how these two bounds imply the entropy estimate. Fix $\eta \in (0,1)$, and run the following greedy active-section construction for

$$f(q) = \inf_{r\in\mathcal{Q}}\{c_p(q,r) - g(r)\}.$$

After sections $s_j(q) := c_p(q,r_j) - g(r_j)$, $j \le m$, have been chosen, set $f_m := \min_{j\le m} s_j$. If $\sup_q\{f_m(q) - f(q)\} > \eta$, choose q_{m+1} at which this inequality holds and an active r_{m+1} satisfying

$$s_{m+1}(q_{m+1}) = f(q_{m+1}).$$

(Using $\eta/4$-approximate minimizers gives the same argument when the infimum is not attained.) If $i<j$, then the choice of q_j and the contact equality at q_i give, respectively,

$$s_i(q_j) - s_j(q_j) > \eta, \qquad s_j(q_i) - s_i(q_i) \ge 0.$$

Adding the two displays shows that

$$\Delta_{c_p}((q_i,r_i),(q_j,r_j)) = s_i(q_j) - s_j(q_j) + s_j(q_i) - s_i(q_i) > \eta.$$

Hence the selected contact pairs form an η-packing for the rectangular interaction.

Let $h = a_p\eta^{1/\beta_p}$, with $a_p > 0$ sufficiently small, and cover each copy of $\mathcal{Q}$ in the product $\mathcal{Q}\times\mathcal{Q}$ by L^p-balls of radius h when $1<p\le 2$, by L^2-balls when $p\ge 2$, and by L^1-balls when $p=1$. By (161), two contact pairs in the same product cell have rectangular interaction at most η. Thus each product cell contains at most one selected pair. From (160), the number M_η of active sections therefore satisfies

$$M_\eta \le N(h,\mathcal{Q},\|\cdot\|_{L^{p\wedge 2}})^2 \le \exp(C_p h^{-1}) \le \exp(C_p\eta^{-1/\beta_p}).$$

Consequently the greedy construction stops after at most M_η steps, and its terminal minimum obeys $\|f_{M_\eta} - f\|_\infty \le \eta$.

It remains to count a finite discretization of such minima. Since all functions take values in $[0,1]$, an active section may be written $c_p(\,\cdot\,,r_j) + b_j$ with $b_j \in [-1,1]$. Moreover,

$$\sup_{q\in\mathcal{Q}} |c_p(q,r) - c_p(q,r')| \le p\|r-r'\|_1.$$

Choose the r_j's from an L^1-net of mesh $c_p'\eta$ and the intercepts b_j from a grid of mesh $c_p'\eta$. The resulting minimum changes by at most $C_p'\eta$ in uniform norm. The logarithm of the number of discretized minima with at most M_η sections is bounded by

$$M_\eta\left\{\log N(c_p'\eta,\mathcal{Q},\|\cdot\|_1) + \log(C_p/\eta)\right\} \le C_p\eta^{-1}\exp(C_p\eta^{-1/\beta_p}),$$

where the last inequality uses (160). Running the construction with a sufficiently small constant multiple of ε in place of η, and absorbing fixed rescalings into C_p, yields

$$\log N_\infty(\varepsilon,\mathcal{F}_p) \le C_p\varepsilon^{-1}\exp(C_p\varepsilon^{-1/\beta_p}). \tag{162}$$

The class is bounded and separable. The truncated Dudley inequality gives

$$\mathbb{E}\sup_{f\in\mathcal{F}_p}|(\widehat{P}_s - P)f| \le C\inf_{\delta\in(0,1)}\left\{\delta + s^{-1/2}\int_\delta^2 \sqrt{\log N_\infty(\varepsilon,\mathcal{F}_p)}\,d\varepsilon\right\}.$$

Taking $\delta = A\{\log(s+1)\}^{-\beta_p}$ with A sufficiently large makes the entropy-integral term $o(\delta)$, proving (158). Finally, c_p-transformation is one-Lipschitz in the uniform norm, so the transformed class has the same entropy up to constants, proving (159). □

Corollary F.13 (Complement-cost empirical process). *Let $\bar{c}_p = 1 - c_p$, and let $\overline{\mathcal{F}}_p$ be the normalized $\bar{c}_p$-concave potentials, with $0 \le h \le 1$. Under the notation of Lemma F.12,*

$$\mathbb{E}\sup_{h\in\overline{\mathcal{F}}_p}|(\widehat{P}_s - P)h| + \mathbb{E}\sup_{h\in\overline{\mathcal{F}}_p}|(\widehat{P}_s - P)h^{\bar{c}_p}| \le C_p\{\log(s+1)\}^{-\beta_p}. \tag{163}$$

Proof. The rectangular interaction of $\bar{c}_p$ is the negative of the rectangular interaction of c_p, and

$$\sup_q |\bar{c}_p(q,r) - \bar{c}_p(q,r')| = \sup_q |c_p(q,r) - c_p(q,r')|.$$

Therefore the active-section packing, finite discretization, and truncated Dudley argument in Lemma F.12 apply without change. The transform is again one-Lipschitz in the uniform norm. □

Lemma F.14 (Learned-dual gap with unequal cell counts)**.** *Let $P, Q \in \mathcal{P}(\mathcal{Q})$, and let $\widehat{P}_u, \widehat{Q}_v$ be empirical laws based on marginally i.i.d. samples of sizes $u, v \geq 1$; the two empirical laws may be dependent. If $(\widehat{f}, \widehat{f}^{c_p})$ is a normalized empirical dual optimizer, then*

$$0 \leq T_p(P,Q) - P\widehat{f} - Q\widehat{f}^{c_p},$$

and

$$\sup_{P,Q} \mathbb{E}\left[T_p(P,Q) - P\widehat{f} - Q\widehat{f}^{c_p}\right] \leq C_p\left\{[\log(u+1)]^{-\beta_p} + [\log(v+1)]^{-\beta_p}\right\}. \tag{164}$$

Proof. The first inequality is weak duality. Write

$$G(f) = Pf + Qf^{c_p}, \qquad \widehat{G}(f) = \widehat{P}_u f + \widehat{Q}_v f^{c_p}.$$

Population and empirical Kantorovich duality, together with the empirical optimality of $\widehat{f}$, imply

$$\begin{aligned} T_p(P,Q) - G(\widehat{f}) &= \sup_{f \in \mathcal{F}_p} G(f) - G(\widehat{f}) \\ &\leq 2 \sup_{f \in \mathcal{F}_p} |G(f) - \widehat{G}(f)| \\ &\leq 2 \sup_{f \in \mathcal{F}_p} |(P - \widehat{P}_u)f| + 2 \sup_{f \in \mathcal{F}_p} |(Q - \widehat{Q}_v)f^{c_p}|. \end{aligned}$$

Lemma F.12 bounds each marginal empirical-process term separately, so independence between the two samples is unnecessary. This proves the result. □

Lemma F.15 (Learned complement-dual gap)**.** *Let $P, Q \in \mathcal{P}(\mathcal{Q})$, let $\widehat{P}_u, \widehat{Q}_v$ be empirical laws of sizes $u, v \geq 1$, and let $(\widehat{h}, \widehat{h}^{\bar{c}_p})$ be a normalized empirical lower-dual optimizer for $\bar{c}_p = 1 - c_p$. Then*

$$0 \leq \{1 - P\widehat{h} - Q\widehat{h}^{\bar{c}_p}\} - T_p^U(P,Q),$$

and

$$\sup_{P,Q} \mathbb{E}\left[\{1 - P\widehat{h} - Q\widehat{h}^{\bar{c}_p}\} - T_p^U(P,Q)\right] \leq C_p\left\{[\log(u+1)]^{-\beta_p} + [\log(v+1)]^{-\beta_p}\right\}. \tag{165}$$

Proof. The identity $T_p^U(P,Q) = 1 - T_{\bar{c}_p}(P,Q)$ turns the displayed gap into

$$T_{\bar{c}_p}(P,Q) - P\widehat{h} - Q\widehat{h}^{\bar{c}_p}.$$

Weak duality makes it nonnegative. The empirical optimality argument in Lemma F.14, with Corollary F.13 in place of Lemma F.12, proves (165). □

F.5 Smooth cost-visible codes

Lemma F.16 (Smooth zero-contact and curved-contact codes)**.** *For every fixed $a \in (0, 1/4]$, there are constants $c, C, \zeta > 0$ such that, for every sufficiently large even integer J, there is a code $\mathcal{C}_J$ with*

$$|\mathcal{C}_J| \geq e^{\zeta J}$$

and strictly increasing quantiles $\{q_u^{(a)} : u \in \mathcal{C}_J\} \subset \mathcal{Q}$ satisfying

$$a|t - s| \leq |q_u^{(a)}(t) - q_u^{(a)}(s)| \leq 2a|t - s|. \tag{166}$$

For $u \neq v$,

$$cJ^{-p} \leq c_p(q_u^{(a)}, q_v^{(a)}) \leq CJ^{-p}. \tag{167}$$

Consequently, for label laws λ, ρ on $\mathcal{C}_J$,

$$cJ^{-p}\,\mathrm{TV}(\lambda,\rho) \le T_p(P^0_\lambda, P^0_\rho) \le CJ^{-p}\,\mathrm{TV}(\lambda,\rho), \tag{168}$$

where $P^0_\lambda = \sum_u \lambda_u \delta_{q_u^{(a)}}$.

If $p > 1$, the code can additionally be chosen to have constant Hamming weight. For fixed $\eta = \delta = 1/4$, put

$$x_u = \eta q_u^{(a)}, \qquad y_u = \delta + \eta q_u^{(a)}.$$

Then $x_u, y_u \in \mathcal{Q}$, and

$$cJ^{-2}\,\mathrm{TV}(\lambda,\rho) \le T_p(P^X_\lambda, P^Y_\rho) - \delta^p \le CJ^{-2}\,\mathrm{TV}(\lambda,\rho), \tag{169}$$

where $P^X_\lambda = \sum_u \lambda_u \delta_{x_u}$ and $P^Y_\rho = \sum_u \rho_u \delta_{y_u}$. The represented distributions have compact supports that are common within each coded family, and their densities are bounded above and away from zero on those supports, with constants depending only on a.

Proof. Partition $[0,1]$ into J intervals I_j. Let $s_{j,0}$ and $s_{j,1}$ be the indicators of the left and right halves of I_j. For $u \in \{0,1\}^J$, define

$$s_u(t) = s_{j,u_j}(t) \quad (t \in I_j), \qquad q_u^{(a)}(t) = at + a\int_0^t s_u(z)\,dz.$$

Then $(q_u^{(a)})' \in \{a, 2a\}$ almost everywhere, which proves (166). If $H(u,v)$ denotes Hamming distance, a disjoint-block calculation gives

$$c_p(q_u^{(a)}, q_v^{(a)}) = \frac{a^p H(u,v)}{2^p(p+1)J^{p+1}}. \tag{170}$$

A Gilbert–Varshamov packing supplies an exponentially large binary code with $H(u,v) \ge J/4$, proving (167). Every coupling of two label laws has off-diagonal mass at least their total variation distance, while a maximal diagonal coupling attains that mass. This proves (168).

For the curved-contact construction, use the constant-weight slice $\{u : \sum_j u_j = J/2\}$. The same packing argument gives an exponentially large subcode with pairwise Hamming distance at least $J/4$. Constant weight implies

$$\int_0^1 (q_u^{(a)} - q_v^{(a)}) = 0, \qquad \|q_u^{(a)} - q_v^{(a)}\|_2^2 \asymp J^{-2}, \qquad \|q_u^{(a)} - q_v^{(a)}\|_\infty \lesssim J^{-1}.$$

Taylor expansion around the nonzero displacement δ yields

$$\begin{aligned} c_p(x_u, y_v) &= \int_0^1 |\delta + \eta(q_v^{(a)} - q_u^{(a)})|^p \\ &= \delta^p + \frac{p(p-1)}{2}\int_0^1 \xi_{u,v}(t)^{p-2}\eta^2(q_v^{(a)} - q_u^{(a)})^2\,dt, \end{aligned}$$

because the linear term integrates to zero. For large J, $\xi_{u,v}(t)$ stays in a fixed compact subinterval of $(0,\infty)$. Hence the off-diagonal excess is of order J^{-2}, while it is zero on the diagonal. The maximal-diagonal-coupling argument proves (169). Finally, (166) gives the asserted common density bounds. □

Lemma F.17 (Smooth codes visible to the maximal coupling)**.** *For every fixed $a \in (0, 1/4]$, there are constants $c, \zeta > 0$ such that, for every sufficiently large even integer J, there is a code $\mathcal{C}_J$ with $|\mathcal{C}_J| \ge e^{\zeta J}$ and two injective families of strictly increasing quantiles $\{x_u^0, y_u^0 : u \in \mathcal{C}_J\} \subset \mathcal{Q}$ for which*

$$c_p(x_u^0, y_u^0) = \tau^0_{p,J}, \tag{171}$$

$$c_p(x_u^0, y_v^0) \le \tau^0_{p,J} - cJ^{-p}, \qquad u \ne v. \tag{172}$$

If $p > 1$, there are also injective families $\{x_u^\delta, y_u^\delta : u \in \mathcal{C}_J\} \subset \mathcal{Q}$ such that

$$c_p(x_u^\delta, y_u^\delta) = \tau^\delta_{p,J}, \tag{173}$$

$$c_p(x_u^\delta, y_v^\delta) \le \tau_{p,J}^\delta - cJ^{-2}, \qquad u \neq v. \tag{174}$$

All four families consist of uniformly bi-Lipschitz quantiles, with constants independent of J. Consequently, for either construction $(k, \Delta_J) = (0, cJ^{-p})$ or (δ, cJ^{-2}),

$$T_p^U(P_\lambda^{X,k}, P_\rho^{Y,k}) \le \tau_{p,J}^k - \Delta_J \,\mathrm{TV}(\lambda, \rho), \tag{175}$$

$$T_p^U(P_\lambda^{X,k}, P_\lambda^{Y,k}) = \tau_{p,J}^k. \tag{176}$$

Here $P_\lambda^{X,k} = \sum_u \lambda_u \delta_{x_u^k}$ and $P_\rho^{Y,k} = \sum_u \rho_u \delta_{y_u^k}$.

Proof. Partition $[0,1]$ into intervals $I_j = [(j-1)/J, j/J]$. Let ϕ_j be the tent on I_j that is zero at both endpoints, has slope $+1$ on the left half and slope -1 on the right half, and is zero outside I_j. For $u \in \{-1,1\}^J$ with $\sum_j u_j = 0$, put

$$h_u = \sum_{j=1}^{J} u_j \phi_j.$$

Then $\int h_u = 0$, $\|h_u\|_\infty \le (2J)^{-1}$, and

$$\int_{I_j} \phi_j(t)^r \, dt = \frac{1}{2^r(r+1)J^{r+1}}, \qquad r \ge 1. \tag{177}$$

A constant-weight Gilbert–Varshamov packing gives $\mathcal{C}_J \subset \{u : \sum_j u_j = 0\}$ with $|\mathcal{C}_J| \ge e^{\zeta J}$ and $H(u,v) \ge J/4$ for distinct codewords.

Fix $b_a(t) = 1/4 + at$ and $\varepsilon = a/4$. For the zero-displacement construction, set

$$x_u^0 = b_a + \varepsilon h_u, \qquad y_u^0 = b_a - \varepsilon h_u.$$

Their derivatives lie in a fixed positive interval and their ranges lie in $[0,1]$. On a block where $u_j = v_j$, $h_u + h_v = \pm 2\phi_j$; on a block where $u_j \neq v_j$, it is zero. Hence

$$c_p(x_u^0, y_v^0) = \varepsilon^p \int |h_u + h_v|^p = \frac{\varepsilon^p}{(p+1)J^p} \left\{ 1 - \frac{H(u,v)}{J} \right\}.$$

This proves (171)– (172).

For the curved construction, fix $\delta = 1/4$ and set

$$x_u^\delta = b_a + \varepsilon h_u, \qquad y_u^\delta = b_a + \delta - \varepsilon h_u.$$

These are again uniformly bi-Lipschitz quantiles in $[0,1]$. Constant weight makes the number of $+1$-to--1 mismatches equal to the number of -1-to-$+1$ mismatches. Pairing those blocks shows that the diagonal minus cross cost contributed by each mismatch pair is

$$\int_{I_j} \{(\delta - 2\varepsilon\phi_j)^p + (\delta + 2\varepsilon\phi_j)^p - 2\delta^p\} \, dt.$$

For $p > 1$, uniform strict convexity of $z \mapsto z^p$ on a fixed compact subinterval of $(0,\infty)$ bounds this display below by $c'_{p,\delta}\varepsilon^2 \int_{I_j} \phi_j^2$. Equation (177) and $H(u,v) \ge J/4$ therefore give a total loss at least cJ^{-2}. The diagonal value is independent of u, again by constant weight. This proves (173)– (174).

Finally, every coupling of the mixture laws induces a coupling matrix of the label laws. Its off-diagonal mass is at least $\mathrm{TV}(\lambda, \rho)$. All diagonal entries equal $\tau_{p,J}^k$, and every off-diagonal entry is at most $\tau_{p,J}^k - \Delta_J$, proving (175). When the label laws agree, the diagonal coupling attains $\tau_{p,J}^k$, proving (176). □

Lemma F.18 (Cost-visible code at a finite-order convex contact)**.** *Suppose that h satisfies (C_{β_h}) in Definition F.5. There are constants $A_h, \zeta > 0$ and $J_h < \infty$ such that, for every even $J \ge J_h$, there are a code $\mathcal{C}_J$ and injective families $\{x_u : u \in \mathcal{C}_J\}, \{y_u : u \in \mathcal{C}_J\} \subset \mathcal{Q}$ with*

$$|\mathcal{C}_J| \ge e^{\zeta J}$$

and

$$c_h(x_u, y_u) = h(z_h), \tag{178}$$

$$c_h(x_u, y_v) \geq h(z_h) + A_h J^{-\beta_h}, \qquad u \neq v. \tag{179}$$

Consequently, for label laws λ, ρ on $\mathcal{C}_J$,

$$T_h(P_\lambda^X, P_\rho^Y) - h(z_h) \geq A_h J^{-\beta_h} \operatorname{TV}(\lambda, \rho), \tag{180}$$

$$T_h(P_\lambda^X, P_\lambda^Y) = h(z_h), \tag{181}$$

where $P_\lambda^X = \sum_u \lambda_u \delta_{x_u}$ and $P_\rho^Y = \sum_u \rho_u \delta_{y_u}$.

Proof. Use the constant-weight version of the code in the proof of Lemma F.16, with a fixed $a \in (0, 1/4]$. Thus $\mathcal{C}_J \subset \{u \in \{0,1\}^J : \sum_j u_j = J/2\}$, $|\mathcal{C}_J| \geq e^{\zeta J}$, and $H(u, v) \geq J/4$ for $u \neq v$. The associated quantiles

$$q_u(t) = at + a \int_0^t s_u(z)\, dz$$

are injective and take values in $[0, 3a/2]$. Direct integration on the half-blocks gives, for every $u, v \in \mathcal{C}_J$,

$$\|q_u - q_v\|_\infty \leq \frac{a}{2J}, \tag{182}$$

$$\int_0^1 (q_v - q_u) = 0, \tag{183}$$

$$\int_0^1 |q_v - q_u|^{\beta_h} = \frac{a^{\beta_h} H(u, v)}{2^{\beta_h}(\beta_h + 1) J^{\beta_h + 1}}. \tag{184}$$

The middle identity follows from constant weight: the signed areas of the 0-to-1 and 1-to-0 blocks cancel.

Write $z_h^+ = \max\{z_h, 0\}$ and $z_h^- = \max\{-z_h, 0\}$. Since $z_h \in (-1, 1)$, choose a fixed $\eta_h > 0$, depending only on h, sufficiently small that

$$x_u := z_h^- + \eta_h q_u, \qquad y_u := z_h^+ + \eta_h q_u$$

take values in $[0, 1]$. They are therefore elements of $\mathcal{Q}$, and

$$y_v - x_u = z_h + \eta_h (q_v - q_u).$$

By (182), for all sufficiently large J, the perturbation has absolute value at most a_h. Applying (143) pointwise and integrating yields

$$\begin{aligned} c_h(x_u, y_v) - h(z_h) &\geq g_h \eta_h \int_0^1 (q_v - q_u) + \underline{c}_h \eta_h^{\beta_h} \int_0^1 |q_v - q_u|^{\beta_h} \\ &= \underline{c}_h \eta_h^{\beta_h} \int_0^1 |q_v - q_u|^{\beta_h}, \end{aligned}$$

where (183) removes the supporting-line term. For $u = v$, the displacement is identically z_h, so $c_h(x_u, y_u) = h(z_h)$. For $u \neq v$, the Hamming separation and (184) give

$$c_h(x_u, y_v) - h(z_h) \geq \frac{\underline{c}_h \eta_h^{\beta_h} a^{\beta_h}}{2^{\beta_h + 2}(\beta_h + 1)} J^{-\beta_h}.$$

This proves (178)– (179).

Every coupling of the two mixture laws corresponds, by injectivity, to a coupling matrix $\gamma = (\gamma_{uv})$ of their label laws. Its off-diagonal mass satisfies

$$\sum_{u \neq v} \gamma_{uv} \geq 1 - \sum_u \min(\lambda_u, \rho_u) = \operatorname{TV}(\lambda, \rho).$$

Equations (178) and (179) therefore imply (180). When $\lambda = \rho$, the diagonal coupling attains $h(z_h)$, while the same inequalities show that no coupling has smaller cost. This proves (181). □

Lemma F.19 (Finite-order contact code for maximal coupling)**.** *Suppose that h satisfies (C_{β_h}). There are constants $A_h^U, \zeta > 0$ and $J_h^U < \infty$ such that, for every sufficiently large even J, there are a code $\mathcal{C}_J$ and injective families $\{x_u^U : u \in \mathcal{C}_J\}, \{y_u^U : u \in \mathcal{C}_J\} \subset \mathcal{Q}$ with $|\mathcal{C}_J| \geq e^{\zeta J}$ for which*

$$c_h(x_u^U, y_u^U) = \tau_{h,J}, \tag{185}$$

$$c_h(x_u^U, y_v^U) \leq \tau_{h,J} - A_h^U J^{-\beta_h}, \qquad u \neq v. \tag{186}$$

Consequently, with $P_\lambda^{X,U} = \sum_u \lambda_u \delta_{x_u^U}$ and $P_\rho^{Y,U} = \sum_u \rho_u \delta_{y_u^U}$,

$$T_h^U(P_\lambda^{X,U}, P_\rho^{Y,U}) \leq \tau_{h,J} - A_h^U J^{-\beta_h} \operatorname{TV}(\lambda, \rho), \tag{187}$$

$$T_h^U(P_\lambda^{X,U}, P_\lambda^{Y,U}) = \tau_{h,J}. \tag{188}$$

Proof. Partition $[0,1]$ into the blocks I_j used in Lemma F.17, and let ϕ_j be the corresponding unit-slope tent. A constant-weight packing gives

$$\mathcal{C}_J \subset \left\{ u \in \{-1,1\}^J : \sum_j u_j = 0 \right\}, \qquad |\mathcal{C}_J| \geq e^{\zeta J}, \qquad H(u,v) \geq J/4.$$

Put $s_u = \sum_j u_j \phi_j$, $b(t) = 1/4 + t/4$, and $\varepsilon = 1/16$. Choose $\eta_h > 0$, depending only on z_h, sufficiently small that

$$x_u^U := z_h^- + \eta_h(b + \varepsilon s_u), \qquad y_u^U := z_h^+ + \eta_h(b - \varepsilon s_u) \tag{189}$$

take values in $[0,1]$, where $z_h^+ = \max\{z_h, 0\}$ and $z_h^- = \max\{-z_h, 0\}$. Their derivatives lie in a fixed positive interval, so they are injective strictly increasing quantiles. Moreover,

$$y_v^U - x_u^U = z_h - \eta_h \varepsilon (s_u + s_v).$$

The diagonal integral is independent of u: every codeword has exactly $J/2$ positive and $J/2$ negative blocks. Denote this common value by $\tau_{h,J}$. If $u_j = v_j$, the cross-cost contribution on I_j equals the corresponding diagonal contribution. If $u_j \neq v_j$, the cross displacement is the constant z_h. Constant weight makes the number of mismatches with $u_j = 1$ equal the number with $u_j = -1$. For all sufficiently large J, the perturbations stay inside the contact neighborhood, and (143) applied with both signs gives

$$\begin{aligned} &h\{z_h + 2\eta_h \varepsilon \phi_j(t)\} + h\{z_h - 2\eta_h \varepsilon \phi_j(t)\} - 2h(z_h) \\ &\quad \geq 2\underline{c}_h (2\eta_h \varepsilon)^{\beta_h} |\phi_j(t)|^{\beta_h}. \end{aligned}$$

The supporting-line terms cancel. Using (177), $H(u,v) \geq J/4$, and pairing the two mismatch orientations yields

$$\tau_{h,J} - c_h(x_u^U, y_v^U) \geq A_h^U J^{-\beta_h}, \qquad u \neq v,$$

for a constant $A_h^U > 0$. This proves (185)– (186).

Every coupling of the mixture laws induces a coupling matrix of their label laws. Its off-diagonal mass is at least $\operatorname{TV}(\lambda, \rho)$. Since diagonal entries equal $\tau_{h,J}$ and off-diagonal entries are at most $\tau_{h,J} - A_h^U J^{-\beta_h}$, maximizing over coupling matrices gives (187). When $\lambda = \rho$, the diagonal coupling attains $\tau_{h,J}$, which proves (188). □

Lemma F.20 (Honest UCB reduction to closeness testing)**.** *Let $\Psi^U(P_\lambda, P_\rho)$ be indexed by two label laws on an alphabet of size M. Suppose that, for constants τ and $\Delta > 0$,*

$$\Psi^U(P_\lambda, P_\rho) \leq \tau - \Delta \operatorname{TV}(\lambda, \rho), \qquad \Psi^U(P_\lambda, P_\lambda) = \tau.$$

If $n \leq c_{\mathrm{test}} M^{2/3}$ and $0 < \alpha < 1/3$, every uniformly honest UCB U_n for Ψ^U satisfies

$$\sup_{\lambda,\rho} P_{\lambda,\rho}\{U_n - \Psi^U(P_\lambda, P_\rho) \geq \Delta/4\} \geq \frac{1}{3}.$$

Proof. The test $\mathbb{1}\{U_n < \tau\}$ has type-I error at most α under $\lambda = \rho$. The same constant-separation closeness-testing lower bound used in Lemma F.21 supplies an alternative with $\operatorname{TV}(\lambda, \rho) \geq 1/4$ and type-II error at least $1/3$. On the type-II event, $U_n \geq \tau$, whereas $\Psi^U(P_\lambda, P_\rho) \leq \tau - \Delta/4$. □

Lemma F.21 (Honest LCB reduction to closeness testing)**.** *Let $\Psi(P_\lambda, P_\rho)$ be indexed by two label laws on an alphabet of size M. Suppose samples reveal their labels and, for constants τ and $\Delta > 0$,*

$$\Psi(P_\lambda, P_\rho) - \tau \geq \Delta\, \mathrm{TV}(\lambda, \rho), \qquad \Psi(P_\lambda, P_\lambda) = \tau.$$

If $n \leq c_{\mathrm{test}} M^{2/3}$ and $0 < \alpha < 1/3$, every uniformly honest LCB L_n for Ψ satisfies

$$\sup_{\lambda,\rho} P_{\lambda,\rho}\{\Psi(P_\lambda, P_\rho) - L_n \geq \Delta/4\} \geq \frac{1}{3}.$$

Proof. The test $\mathbb{1}\{L_n > \tau\}$ has type-I error at most α when $\lambda = \rho$. The constant-separation two-sample closeness lower bound of Chan et al. (2014) gives a pair with $\mathrm{TV}(\lambda, \rho) \geq 1/4$ and type-II error at least $1/3$ whenever $n \leq c_{\mathrm{test}} M^{2/3}$. On the type-II event, $L_n \leq \tau$ and the endpoint gap is at least $\Delta/4$. □

F.6 Proof of Theorem F.1

Proof. *Exact simultaneous validity.* Let $\mathcal{F}_1 = \sigma(D_1)$, and put

$$\mu_{\pi,\pi'} = \mathbb{E}[S_i(\pi, \pi') \mid \mathcal{F}_1].$$

Unconfoundedness and the known propensity scores give

$$\mu_{\pi,\pi'} = \sum_{x \in \mathcal{X}_{K_X}} P_X(x) \left\{ P_{\pi(x),x} \widehat{f}_{\pi(x),\pi'(x),x} + P_{\pi'(x),x} \widehat{f}^{c_p}_{\pi(x),\pi'(x),x} \right\}.$$

Every fitted pair is globally feasible, including pairs fitted using a fallback empirical law. Conditional weak duality therefore gives

$$\mu_{\pi,\pi'} \leq T_p(\pi, \pi') \quad \text{simultaneously over } \Pi^2. \tag{190}$$

The normalized potentials satisfy $0 \leq \widehat{f} \leq 1$, $-1 \leq \widehat{f}^{c_p} \leq 1$, and, by global feasibility, $-1 \leq \widehat{f}(q) + \widehat{f}^{c_p}(q) \leq c_p(q, q) = 0$. Thus a score containing either one weighted potential or their weighted sum has the same envelope. Consequently,

$$-\underline{e}^{-1} \leq S_i(\pi, \pi') \leq \underline{e}^{-1}.$$

There are at most N_Π^2 distinct score functions. Conditional Hoeffding's inequality and a union bound show that

$$P\left\{ \sup_{\Pi^2} [\widehat{T}_p(\pi, \pi') - \mu_{\pi,\pi'}] > r^{\Pi}_{m,\alpha} \,\middle|\, \mathcal{F}_1 \right\} \leq N_\Pi^2 \exp\left\{ -\frac{m\underline{e}^2 (r^{\Pi}_{m,\alpha})^2}{2} \right\} = \alpha.$$

Together with (190), this proves exact simultaneous coverage for every finite n.

Uniform excess-length upper bound. For every action pair, define

$$\Delta_{ab,x} = T_{p,x}(a, b) - P_{a,x} \widehat{f}_{ab,x} - P_{b,x} \widehat{f}^{c_p}_{ab,x},$$

Let $N_{a,x}$ be the number of training observations in cell (a, x). Conditional on the cell counts, nonempty cells contain i.i.d. observations from their corresponding conditional outer laws. Lemma F.14 therefore applies with $u = N_{a,x}$ and $v = N_{b,x}$.

Every $N_{a,x}$ has mean at least $\ell\kappa\underline{e}$. A binomial Chernoff bound and a union bound give

$$P\left\{ \min_{a,x} N_{a,x} < \frac{1}{2} \ell\kappa\underline{e} \right\} \leq Kd \exp(-c\ell\kappa\underline{e}). \tag{191}$$

On the complement of this event, Lemma F.14 bounds every conditional gap by $C[\log(\ell\kappa\underline{e}+1)]^{-\beta_p}$ in expectation. On the exceptional event the gap is bounded by a universal constant, because $0 \leq c_p \leq 1$ and the potentials are normalized. Since K_X and d are fixed,

$$\sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{(\pi,\pi') \in \Pi^2} [T_p(\pi, \pi') - \mu_{\pi,\pi'}]$$

$$\leq \sup_P \mathbb{E}_P \sum_{x \in \mathcal{X}_{K_X}} P_X(x) \max_{a,b} \Delta_{ab,x} \leq C[\log(\ell\kappa\underline{e}+1)]^{-\beta_p}. \tag{192}$$

The maximal inequality for N_Π^2 bounded score functions gives

$$\mathbb{E}\left[\sup_{\Pi^2} |\widehat{T}_p(\pi,\pi') - \mu_{\pi,\pi'}| \,\middle|\, \mathcal{F}_1\right] \leq \frac{C}{\underline{e}}\sqrt{\frac{\log(2N_\Pi^2)}{m}}. \tag{193}$$

Combining (192), (193), $N_\Pi \leq d^{K_X}$, and the confidence radius proves (128). With a fixed-proportion split, the root-sample terms are $o((\log n)^{-\beta_p})$, proving the minimax upper bound.

Minimax lower bound. Choose two policies π_0, π_1 whose action vectors differ, and select $x_\star \in \mathcal{X}_{K_X}$ with

$$a_\star := \pi_0(x_\star) \neq b_\star := \pi_1(x_\star).$$

Fix an admissible covariate distribution and write $w = P_X(x_\star) \geq \kappa$. Make all conditional outer laws other than $P_{a_\star,x_\star}$ and $P_{b_\star,x_\star}$ equal to a common point mass. The selected policy pair then satisfies

$$T_p(\pi_0,\pi_1) = wT_p(P_{a_\star,x_\star}, P_{b_\star,x_\star}). \tag{194}$$

Consider the more informative oracle experiment that supplies n observations from each of the two laws in (194). It can simulate the original n-observation experiment by independently drawing (X_i, A_i), consuming the next oracle observation in either hard cell, and returning the common point mass elsewhere. At most n observations from either oracle sample are used. Therefore, any honest policy band in the observed-data experiment induces an honest, possibly randomized, two-sample LCB for the right-hand side of (194).

Choose an even $J_n \asymp \log(n+1)$, with a sufficiently large constant, so that

$$n \leq c_{\text{test}}|\mathcal{C}_{J_n}|^{2/3}.$$

Embedding the zero-contact code of Lemma F.16 in the two hard cells and applying Lemma F.21 gives probability separation

$$wJ_n^{-p} \gtrsim \kappa(\log n)^{-p}.$$

For $p > 1$, the curved-contact code gives the additional separation

$$wJ_n^{-2} \gtrsim \kappa(\log n)^{-2}.$$

The code embeddings are injective, so observing a quantile reveals its label as required by the testing lemma. Taking the larger separation gives $(\log n)^{-(p\wedge 2)}$, proving (126). Multiplying by its probability $1/3$ proves the risk lower bound. □

F.7 Proof of Theorem F.2

Proof. Exact simultaneous validity and the attainable rate. Let $\mathcal{F}_1 = \sigma(D_1)$ and define

$$\nu_{\pi,\pi'} := \mathbb{E}[H_i(\pi,\pi') \mid \mathcal{F}_1], \qquad \mu^U_{\pi,\pi'} := 1 - \nu_{\pi,\pi'}.$$

Unconfoundedness and the known randomization probabilities imply that $\nu_{\pi,\pi'}$ is the covariate-weighted population value of the fitted complement-cost dual pair. Conditional weak duality, followed by (129), therefore gives, simultaneously over Π^2,

$$\mu^U_{\pi,\pi'} \geq T^U_{p,P}(\pi,\pi'). \tag{195}$$

For distinct selected actions, at most one term in (130) is nonzero. When the selected actions coincide, complement-dual feasibility gives $-1 \leq \widehat{h}(q) + \widehat{h}^{\bar{c}_p}(q) \leq \bar{c}_p(q,q) = 1$. Thus every score is in $[-\underline{e}^{-1}, \underline{e}^{-1}]$. Conditional Hoeffding's inequality and a union bound over at most N_Π^2 scores show that

$$P\left\{\sup_{\Pi^2}(\mu^U_{\pi,\pi'} - \widehat{T}^U_p(\pi,\pi')) > r^\Pi_{m,\alpha} \,\middle|\, \mathcal{F}_1\right\} \leq \alpha.$$

Together with (195), this proves the asserted exact finite-sample honesty.

For a cell (a, b, x), put

$$\Delta^U_{ab,x} := 1 - P_{a,x}\widehat{h}_{ab,x} - P_{b,x}\widehat{h}^{\bar{c}_p}_{ab,x} - T^U_{p,x}(a, b).$$

Lemma F.15, the cell-count event (191), and the same averaging argument as in (192) yield

$$\sup_P \mathbb{E}_P \sup_{\Pi^2}(\mu^U_{\pi,\pi'} - T^U_{p,P}(\pi, \pi')) \le C\{\log(\ell\kappa\underline{e} + 1)\}^{-\beta_p}. \tag{196}$$

The bounded-score maximal inequality used in (193) also gives

$$\mathbb{E}\left[\sup_{\Pi^2} |\widehat{T}^U_p - \mu^U| \,\middle|\, \mathcal{F}_1\right] \le \frac{C}{\underline{e}}\sqrt{\frac{\log(2N^2_\Pi)}{m}}. \tag{197}$$

Equations (196)–(197), the confidence radius, and $N_\Pi \le d^{K_X}$ prove (137). Omitting the confidence radius and using the triangle inequality proves (138). For a fixed-proportion split, both root-sample terms are $o\{(\log n)^{-\beta_p}\}$, which proves the two minimax upper bounds.

Minimax lower bound. Choose $\pi_0, \pi_1 \in \Pi$ and $x_\star$ as in the proof of Theorem F.1, and write $a_\star = \pi_0(x_\star) \ne b_\star = \pi_1(x_\star)$ and $w = P_X\{x_\star\} \ge \kappa$. In the two selected cells embed either pair of mixture families from Lemma F.17; make every remaining conditional outer law a common point mass. Then

$$T^U_{p,P}(\pi_0, \pi_1) = wT^U_p(P^{X,k}_\lambda, P^{Y,k}_\rho).$$

As before, the more informative experiment supplying n observations from each hard law simulates the randomized-design sample. Hence every honest policy UCB induces an honest, possibly randomized, UCB for the display above.

Choose an even $J_n \asymp \log(n + 1)$, with a sufficiently large constant, so that $n \le c_{\text{test}}|\mathcal{C}_{J_n}|^{2/3}$. The zero-displacement code and Lemma F.20 give a probability separation of order $wJ_n^{-p} \gtrsim \kappa(\log n)^{-p}$. For $p > 1$, the curved code gives the additional separation $wJ_n^{-2} \gtrsim \kappa(\log n)^{-2}$. Observing a coded quantile reveals its label because the embeddings are injective. Taking the larger of the two separations proves (134); multiplying by its probability $1/3$ proves the lower bound in (135).

Point estimation. For any estimator collection $\widetilde{\boldsymbol{T}}^U_n$, let

$$\rho_n := \sup_P \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}^U_n - T^U_{p,P}|.$$

If $0 < \rho_n < \infty$, then $U^{\widetilde{T}}_n(\pi, \pi') = \widetilde{T}^U_n(\pi, \pi') + \rho_n/\alpha$ is a uniformly honest simultaneous UCB by Markov's inequality, and

$$\sup_P \mathbb{E}_P \sup_{\Pi^2}\{U^{\widetilde{T}}_n - T^U_{p,P}\}_+ \le (1 + \alpha^{-1})\rho_n.$$

The cases $\rho_n = 0$ and $\rho_n = \infty$ follow by an arbitrarily small enlargement and trivially, respectively. Taking the infimum over estimators and applying (135) proves the lower bound in (136); the upper bound was established above.

Finally, for any fixed $L > 0$, use $a = \min\{1/4, L/2\}$ in Lemma F.17. The resulting slopes lie between $3a/4$ and $5a/4$, so all hard quantiles belong to $\mathcal{Q}_L$. The attainable bounds are uniform over the larger unrestricted model class. This proves the stated restricted-model conclusions, with constants allowed to depend on L. □

F.8 Proof of Theorem F.3

Proof. Upper bound. By (190), the conditional score mean satisfies $\mu_{\pi,\pi'} \le T_p(\pi, \pi')$. Hence, pointwise in the training and evaluation samples,

$$|\widehat{T}_p(\pi, \pi') - T_p(\pi, \pi')| \le |\widehat{T}_p(\pi, \pi') - \mu_{\pi,\pi'}| + T_p(\pi, \pi') - \mu_{\pi,\pi'}.$$

Taking the supremum over Π^2, expectation, and then the supremum over the model class, equations (192) and (193) give

$$\sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widehat{T}_p - T_{p,P}| \le C[\log(\ell\kappa\underline{e} + 1)]^{-\beta_p} + \frac{C}{\underline{e}}\sqrt{\frac{\log(2N^2_\Pi)}{m}}.$$

Since $N_\Pi \le d^{K_X}$, this is (140)'s upper bound.

Lower bound. We show that the honest-LCB lower bound also constrains every point estimator. Fix any estimator collection $\widetilde{\boldsymbol{T}}_n$, and let

$$\rho_n := \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{e}}} \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}_n(\pi,\pi') - T_{p,P}(\pi,\pi')|.$$

For $\rho_n > 0$, define the collection

$$L_n^{\widetilde{T}}(\pi,\pi') = \widetilde{T}_n(\pi,\pi') - \frac{\rho_n}{\alpha}.$$

Markov's inequality gives, uniformly over the model class,

$$\begin{aligned} &P\Big\{L_n^{\widetilde{T}}(\pi,\pi') > T_{p,P}(\pi,\pi') \text{ for some } (\pi,\pi') \in \Pi^2\Big\} \\ &\quad \le P\Big\{\sup_{\Pi^2} |\widetilde{T}_n - T_{p,P}| > \frac{\rho_n}{\alpha}\Big\} \le \alpha. \end{aligned}$$

Thus $\boldsymbol{L}_n^{\widetilde{T}}$ is a uniformly honest simultaneous band. Its worst-policy excess length satisfies

$$\begin{aligned} &\sup_P \mathbb{E}_P \sup_{\Pi^2} \{T_{p,P} - L_n^{\widetilde{T}}\}_+ \\ &\quad \le \sup_P \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}_n - T_{p,P}| + \frac{\rho_n}{\alpha} = (1+\alpha^{-1})\rho_n. \end{aligned}$$

The same conclusion follows by an arbitrarily small positive enlargement when $\rho_n = 0$. Taking the infimum over all point estimators yields

$$R_{n,p}^{\mathrm{hon}}(K_X,\Pi) \le (1+\alpha^{-1})\mathcal{R}_{n,p}^{\mathrm{est}}(K_X,\Pi).$$

The lower bound in (127) therefore implies

$$\mathcal{R}_{n,p}^{\mathrm{est}}(K_X,\Pi) \ge \frac{c}{1+\alpha^{-1}} (\log(n+1))^{-\beta_p}.$$

Together with the learned-dual upper bound, this proves minimax rate optimality. □

F.9 Proof of Corollary F.4

Proof. The lower-endpoint upper bound is (140), and its matching lower bound holds for every estimator by the definition of $\mathcal{R}_{n,p}^{\mathrm{est}}$. The corresponding upper-endpoint statements are (138) and (136). These pairs prove (141)– (142). The same comparison between the probability lower bounds and the attainable excess-length bounds gives the claimed orders for the exactly honest bands. With a fixed-proportion split and fixed $K_X, d, \underline{e}, \alpha$, both simultaneous evaluation terms are $O(n^{-1/2})$, which is $o\{(\log n)^{-\beta_p}\}$.

It remains to identify what uses first-stage correctness. Conditional on D_1, inequalities (190) and (195) follow from global dual feasibility alone. The bounded-score concentration argument is likewise conditional on D_1 and requires no consistency of the fitted outer laws. Hence the simultaneous coverage statements survive arbitrary first-stage misspecification. By contrast, the bounds (192) and (196) control the population loss of the learned duals and are what deliver informativeness. They hold at the stated rate for the empirical conditional-law learner used in the construction, but global feasibility alone does not force them to vanish. This proves both the no-first-order-price statement and its stated boundary. □

F.10 Proof of Theorem F.7

Proof. Choose $\pi_0, \pi_1 \in \Pi$ whose action vectors differ, and select $x_\star \in \mathcal{X}_{K_X}$ such that

$$a_\star := \pi_0(x_\star) \ne b_\star := \pi_1(x_\star).$$

Fix an admissible covariate distribution and write $w = P_X\{x_\star\} \ge \kappa$. In the two hard cells set

$$P_{a_\star,x_\star} = P_\lambda^X, \qquad P_{b_\star,x_\star} = P_\rho^Y,$$

using the families in Lemma F.18. Make all remaining conditional outer laws equal to a common point mass. Since $h(0) = 0$, the selected policy pair then satisfies

$$T_{h,P}(\pi_0, \pi_1) = wT_h(P_\lambda^X, P_\rho^Y).$$

As in the oracle reduction for (126), consider the more informative experiment supplying n observations from each of the two hard laws. It simulates the original n-observation randomized experiment by drawing (X_i, A_i), using the next oracle observation when a hard cell is selected, and returning the common point mass otherwise. No more than n observations from either oracle sample are consumed. Thus every band in $\mathcal{H}_{n,\alpha}^{(h)}(K_X, \Pi)$ induces an honest, possibly randomized, two-sample LCB for this selected endpoint.

For an even J large enough for Lemma F.18, set

$$\tau = wh(z_h), \qquad \Delta_J = wA_h J^{-\beta_h}.$$

Equations (180)– (181) imply

$$T_{h,P}(\pi_0, \pi_1) - \tau \geq \Delta_J \, \mathrm{TV}(\lambda, \rho),$$
$$T_{h,P}(\pi_0, \pi_1) = \tau \quad \text{when } \lambda = \rho.$$

The embeddings are injective, so each observed quantile reveals its label. Lemma F.21 therefore applies.

Choose an even $J_n \asymp \log(n+1)$, with a sufficiently large constant, so that

$$n \leq c_{\mathrm{test}} |\mathcal{C}_{J_n}|^{2/3}.$$

For every uniformly honest band, the testing lemma supplies a hard pair (λ, ρ) for which

$$P\left\{T_{h,P}(\pi_0, \pi_1) - L_n(\pi_0, \pi_1) \geq \frac{wA_h}{4} J_n^{-\beta_h}\right\} \geq \frac{1}{3}.$$

Since $w \geq \kappa$ and $J_n \lesssim \log(n+1)$, this proves (149). On the displayed event the positive-part worst-policy excess is at least the same threshold. Taking expectations and absorbing the factor $1/3$ into the constant proves (150).

It remains to transfer this lower bound to point estimation. For an arbitrary estimator collection $\widetilde{\boldsymbol{T}}_n$, let

$$\rho_n = \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{\varepsilon}}} \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}_n - T_{h,P}|.$$

If $\rho_n = \infty$, the claimed lower bound is immediate. When $0 < \rho_n < \infty$, define

$$L_n^{\widetilde{T}}(\pi, \pi') = \widetilde{T}_n(\pi, \pi') - \rho_n/\alpha.$$

Markov's inequality makes this a member of $\mathcal{H}_{n,\alpha}^{(h)}(K_X, \Pi)$, and

$$R_{n,h}^{\mathrm{hon}}(K_X, \Pi) \leq (1 + \alpha^{-1})\rho_n.$$

An arbitrarily small enlargement gives the same conclusion if $\rho_n = 0$. Taking the infimum over estimators and applying (150) proves (151). No upper bound for a general h is used at any point. □

F.11 Proof of Theorem F.8

Proof. Choose $\pi_0, \pi_1, x_\star, a_\star, b_\star$, and an admissible covariate law as in the proof of Theorem F.7; write $w = P_X\{x_\star\} \geq \kappa$. In the two hard cells use

$$P_{a_\star,x_\star} = P_\lambda^{X,U}, \qquad P_{b_\star,x_\star} = P_\rho^{Y,U},$$

from Lemma F.19, and make all other conditional outer laws a common point mass. Since $h(0) = 0$,

$$T_{h,P}^U(\pi_0, \pi_1) = wT_h^U(P_\lambda^{X,U}, P_\rho^{Y,U}).$$

The more informative two-sample oracle experiment used in the preceding proof again simulates the original randomized experiment. Thus every member of $\mathcal{H}_{n,\alpha}^{(h),U}(K_X, \Pi)$ induces an honest UCB for the selected upper endpoint.

For a sufficiently large even J, set

$$\tau = w\tau_{h,J}, \qquad \Delta_J = wA_h^U J^{-\beta_h}.$$

Equations (187)– (188) give

$$T_{h,P}^U(\pi_0,\pi_1) \le \tau - \Delta_J \,\mathrm{TV}(\lambda,\rho),$$
$$T_{h,P}^U(\pi_0,\pi_1) = \tau \quad \text{when } \lambda = \rho.$$

The embeddings are injective, so the oracle observations reveal their labels. Lemma F.20 therefore applies. Taking an even $J_n \asymp \log(n+1)$ for which $n \le c_{\mathrm{test}}|\mathcal{C}_{J_n}|^{2/3}$ yields a hard pair satisfying

$$P\left\{U_n(\pi_0,\pi_1) - T_{h,P}^U(\pi_0,\pi_1) \ge \frac{wA_h^U}{4} J_n^{-\beta_h}\right\} \ge \frac{1}{3}.$$

Since $w \ge \kappa$, this is (152). Integration proves (153).

For an arbitrary upper-endpoint estimator $\widetilde{\boldsymbol{T}}_n^U$, let

$$\rho_n^U = \sup_{P \in \mathfrak{P}_{K_X,\kappa,\underline{\varepsilon}}} \mathbb{E}_P \sup_{\Pi^2} |\widetilde{T}_n^U - T_{h,P}^U|.$$

If $0 < \rho_n^U < \infty$, then

$$U_n^{\widetilde{T}}(\pi,\pi') = \widetilde{T}_n^U(\pi,\pi') + \rho_n^U/\alpha$$

is uniformly honest by Markov's inequality and has worst-case expected excess at most $(1+\alpha^{-1})\rho_n^U$. The zero and infinite cases follow by an arbitrarily small enlargement and trivially, respectively. Taking the infimum over estimators and using (153) proves (154). As for the lower endpoint, this proof uses no attainable upper rate for a general h. □

F.12 Proof of Theorem F.9

Proof. We first establish the general-cost analogue of Lemma F.12. Let $\mathcal{F}_h$ be the normalized c_h-concave potentials, chosen so that $0 \le f \le \bar{h}$ and $-\bar{h} \le f^{c_h} \le \bar{h}$. We claim that, uniformly over every law P on $\mathcal{Q}$,

$$\mathbb{E} \sup_{f\in\mathcal{F}_h} |(\widehat{P}_s - P)f| + \mathbb{E} \sup_{f\in\mathcal{F}_h} |(\widehat{P}_s - P)f^{c_h}| \le C_h\{\log(s+1)\}^{-\gamma_h}. \tag{198}$$

Put $r_h = 1$ when $\gamma_h \le 1$, and $r_h = \gamma_h$ when $\gamma_h > 1$. The monotone entropy bound (160) applies in L^{r_h}. For $a = q' - q$ and $b = r' - r$, the pointwise integrand of the rectangular interaction of c_h is a four-point increment of h with increments $-a(t)$ and $b(t)$. If $\gamma_h \le 1$,

$$\begin{aligned}|\Delta_{c_h}((q,r),(q',r'))| &\le L_h \int_0^1 \min\{|a(t)|^{\gamma_h}, |b(t)|^{\gamma_h}\}\,dt \\ &\le L_h \min\{\|a\|_1^{\gamma_h}, \|b\|_1^{\gamma_h}\},\end{aligned}$$

where Jensen's inequality gives the second line. If $1 < \gamma_h \le 2$, Hölder's inequality gives

$$\begin{aligned}|\Delta_{c_h}((q,r),(q',r'))| &\le L_h \min\left\{\int |a||b|^{\gamma_h-1}, \int |b||a|^{\gamma_h-1}\right\} \\ &\le L_h \min\left\{\|a\|_{\gamma_h}\|b\|_{\gamma_h}^{\gamma_h-1}, \|b\|_{\gamma_h}\|a\|_{\gamma_h}^{\gamma_h-1}\right\}.\end{aligned}$$

Consequently, if both coordinates of two contact pairs are within δ in L^{r_h}, their rectangular interaction is at most $L_h\delta^{\gamma_h}$.

Run the active-section construction in the proof of Lemma F.12 at accuracy η. Its selected contact pairs have rectangular interaction larger than η. Taking $\delta = a_h'\eta^{1/\gamma_h}$, with a_h' sufficiently small, shows that there are at most

$$M_\eta \le N(\delta, \mathcal{Q}, \|\cdot\|_{L^{r_h}})^2 \le \exp(C_h\eta^{-1/\gamma_h})$$

active sections. It remains only to discretize those sections. By (144), with $s_h = \gamma_h \wedge 1$,

$$\sup_{q\in\mathcal{Q}} |c_h(q,r) - c_h(q,r')| \le L_h \int_0^1 |r-r'|^{s_h} \le L_h\|r-r'\|_1^{s_h}.$$

An L^1-net of mesh $c'_h \eta^{1/s_h}$, together with an intercept grid of mesh $c'_h \eta$, therefore changes a section by at most a constant multiple of η. The same counting argument as in (162) yields

$$\log N_\infty(\varepsilon, \mathcal{F}_h) \le C_h \varepsilon^{-1/s_h} \exp(C_h \varepsilon^{-1/\gamma_h}). \tag{199}$$

The truncated Dudley inequality, with $\delta = A_h\{\log(s+1)\}^{-\gamma_h}$ and A_h sufficiently large, now proves the first term in (198). The c_h-transform is one-Lipschitz in uniform norm, so the same entropy bound proves the second.

The complement cost $\bar{c}_h = \bar{h} - c_h$ has the negative rectangular interaction, the same section modulus, and the same bounded envelope. Therefore (198) also holds for its normalized lower-dual class. Repeating the empirical-optimality calculation in Lemmas F.14 and F.15 gives, for empirical cell sizes $u, v \ge 1$,

$$\sup_{P,Q} \mathbb{E}[\mathrm{Gap}_h] \ \vee \ \sup_{P,Q} \mathbb{E}[\mathrm{Gap}_{\bar{h}}] \le C_h \left\{ [\log(u+1)]^{-\gamma_h} + [\log(v+1)]^{-\gamma_h} \right\}. \tag{200}$$

We finally lift the cellwise bound to the policy class. On the event (191), every training cell contains at least $\ell\kappa\underline{e}/2$ observations, so (200), maximized over the fixed number $K_X d^2$ of cell pairs, is bounded by $C_h\{\log(\ell\kappa\underline{e}+1)\}^{-\gamma_h}$. On the complementary event the gap is bounded by a constant depending only on $\bar{h}$, and its probability is exponentially small. Conditional on the training fold, the evaluation scores have envelope at most $C\bar{h}/\underline{e}$. The same finite-class maximal inequality as (193), using $N_\Pi \le d^{K_X}$, contributes

$$\frac{C\bar{h}}{\underline{e}} \sqrt{\frac{K_X \log d}{m}}.$$

Weak duality makes the lower-score conditional mean no larger than $T_{h,P}$. The identity $T_h^U = \bar{h} - T_{\bar{c}_h}$ gives the analogous one-sided relation for the upper estimator. The triangle argument in Theorem F.3 therefore proves (155) for both endpoints.

Conditional Hoeffding inequalities and a union bound over N_Π^2 policy pairs give the radius $r_{m,\alpha}^{\Pi,h}$, exact simultaneous honesty, and the asserted extra $\log(1/\alpha)$ term. If $\gamma_h = \beta_h$, Theorems F.7 and F.8 supply the matching lower bounds for the two honest-band risks and the two point-estimation risks. This proves all claims. □

F.13 Proof of Theorem F.10

Proof. Fix $L > 0$. We first revisit the lower-endpoint code in Lemma F.18. The base quantiles there satisfy

$$a|t-s| \le |q_u(t) - q_u(s)| \le 2a|t-s|$$

for a fixed $a > 0$. In the definitions $x_u = z_h^- + \eta q_u$ and $y_u = z_h^+ + \eta q_u$, reduce the positive scale η, if necessary, so that it satisfies both the original range and contact-neighborhood conditions and $2a\eta \le L$. Then every x_u, y_u lies in $\mathcal{Q}_L$ and is bi-Lipschitz with lower slope $a\eta > 0$. The proof of the lemma is unchanged, except that its positive separation constant is now allowed to depend on L:

$$c_h(x_u, y_v) - h(z_h) \ge A_{h,L} J^{-\beta_h}, \qquad u \ne v.$$

For the upper endpoint, revisit Lemma F.19. The unscaled curves $b + \varepsilon s_u$ and $b - \varepsilon s_u$, with $b(t) = 1/4 + t/4$ and $\varepsilon = 1/16$, have almost-everywhere slopes in $[3/16, 5/16]$. Reduce the positive scale η in (189) so that, in addition to its original constraints, $5\eta/16 \le L$. The resulting upper-code curves belong to $\mathcal{Q}_L$, have lower slope $3\eta/16 > 0$, and retain off-diagonal loss $A_{h,L}^U J^{-\beta_h}$ for some $A_{h,L}^U > 0$.

Thus both hard experiments used in the proofs of Theorems F.7 and F.8 lie in $\mathfrak{P}_{K_X,\kappa,\underline{e},L}$. The testing reductions and the Markov transfers to point estimation apply verbatim, proving (156) and the two probability statements.

If (U_{γ_h}) also holds with $\gamma_h = \beta_h$, the upper bounds in Theorem F.9 are uniform over the unrestricted model class. Restricting their model suprema gives the upper half of (157); the preceding smooth-code lower bounds give the other half. Exact honesty is inherited for the same reason.

Finally, a strictly increasing quantile with slopes in a fixed interval $[m_{h,L}, L]$, $m_{h,L} > 0$, induces an absolutely continuous law whose density on its compact support lies between $1/L$ and $1/m_{h,L}$ almost everywhere. This proves the last assertion. □

F.14 Proof of Corollary F.11

Proof. For $h(z) = |z|^p$, the local contact and uniform rectangular orders agree at $p \wedge 2$. Apply Theorem F.10 and retain its lower-endpoint honest-risk conclusion. □

G Directional Convex Costs and Their Rate Orders

With I rescaled to $[0,1]$, write q_y for the quantile function of y and $\Delta_{y,y'}(t) = q_{y'}(t) - q_y(t)$. The ordered pair (π, π') fixes the sign convention. For example, when larger outcomes are undesirable, $\Delta > 0$ records deterioration under the second policy. The analyst can therefore specify a convex loss for positive and negative displacements separately.

The following proposition collects the implications of choosing such a loss within the general cost class of (10). Write $\Theta^e_h = \Theta^e_{c_h}$ for $e \in \{L, U\}$ and $c_h(y, y') = \int_0^1 h\{\Delta_{y,y'}(t)\}\, dt$.

Proposition G.1 (Ordered convex-cost dictionary)**.** *Let $h : [-1,1] \to [0, \bar{h}]$ be continuous and convex with $h(0) = 0$, and put $\check{h}(z) = h(-z)$.*

(a) *For both endpoints,*

$$\Theta^e_h(\pi', \pi) = \Theta^e_{\check{h}}(\pi, \pi').$$

Thus only an even loss guarantees invariance to swapping for all outer laws; symmetry comes from h, not from the matching estimand.

(b) *For nonnegative weights, thresholds $\tau_j, \sigma_k \in [0,1)$, and powers $p_j, q_k \geq 1$, the multi-hinge loss*

$$h_{\mathrm{mh}}(z) = \sum_j a_j (z - \tau_j)_+^{p_j} + \sum_k b_k (-z - \sigma_k)_+^{q_k} \tag{201}$$

is an admissible convex displacement loss.

(c) *Let $h_+(z) = z_+$ and $h_-(z) = (-z)_+$. Every admissible matching kernel η satisfies the signed mean decomposition (17), where $\mu(\pi) = \mathbb{E}_X \int \{\int_0^1 q_y(t)\, dt\}\, dP_{\pi(X),X}(y)$ is the policy-population mean outcome. Hence two nonnegative convex costs exactly decompose a signed, point-identified policy contrast.*

(d) *Every admissible continuous convex h can be uniformly approximated by a finite piecewise-linear loss of the form (201) with all powers equal to one. If $\|h - h_J\|_\infty \leq \varepsilon$, then*

$$|\Theta^e_h(\pi, \pi') - \Theta^e_{h_J}(\pi, \pi')| \leq \varepsilon, \qquad e \in \{L, U\}.$$

The proof follows below.

One-sided harm and direct ranking. For $h_+(z) = z_+$, the lower endpoint is the smallest possible average positive displacement from π to π'. If larger outcomes are worse, $\Theta^L_{h_+}(\pi, \pi') > 0$ certifies that every admissible population matching incurs some deterioration. Conversely, $\Theta^U_{h_+}(\pi, \pi') = 0$ certifies zero positive-displacement loss under every matching. Reversing the policy order changes the loss to h_-. Equation (17) also shows how the two one-sided losses recover the sign of the ordinary mean comparison when they are evaluated under the same matching. Interpreting positive displacement as harm requires the analyst to specify which outcome direction is undesirable.

Tolerance, tail severity, and asymmetric tradeoffs. The loss $h^{(p)}_{\tau,+}(z) = (z - \tau)_+^p$ ignores increases up to a scientifically chosen tolerance τ, counts excess linearly when $p = 1$, and increasingly emphasizes large violations when $p > 1$. For any fixed matching, the layer-cake identity (18) applies. Here λ is Lebesgue measure on $[0,1]$, and $\mathbb{E}_\eta$ averages over X and its conditional matching kernel. The target is thus an integrated directional exceedance profile. Reporting the sharp interval over several values of τ separates moderate from severe deterioration. Each threshold is pairwise sharp; different thresholds need not share one extremizing matching.

When gains and losses both matter but have different consequences, the asymmetric power loss $h(z) = \lambda_+ z_+^p + \lambda_- (-z)_+^p$ assigns distinct marginal values to the two directions. This covers,

for example, settings in which adverse increases are more costly than comparable decreases, or undersupply is penalized differently from oversupply. Finally, a sum such as $\sum_j a_j(z-\tau_j)_+$ encodes tiered penalties whose slope increases after successive thresholds. Proposition G.1(d) shows that such finite schedules approximate every convex loss allowed by our theory. The identification and honest dual certification results above apply to each chosen loss. When its contact and interaction conditions hold, the minimax results above apply as well; no new statistical construction is needed.

Proof of Proposition G.1. Fix a covariate value x and abbreviate the two conditional outer laws by P and Q. If $\gamma \in \Gamma(P,Q)$, let $\gamma^\top$ be its image under $(y,y') \mapsto (y',y)$. Transposition is a bijection from $\Gamma(P,Q)$ to $\Gamma(Q,P)$, and

$$\int c_h(y',y)\,d\gamma(y,y') = \int_0^1 \int h\{q_y(t) - q_{y'}(t)\}\,d\gamma\,dt = \int c_{\check{h}}(y,y')\,d\gamma(y,y').$$

Taking either the infimum or the supremum over couplings, and then integrating over X, proves part (a). In particular, evenness of h makes the ordered target symmetric. Conversely, if the target were symmetric for every pair of point-mass outer laws, applying it to constant quantile functions would give $h(z) = h(-z)$ for every $z \in [-1,1]$.

For $p \geq 1$, the map $u \mapsto u_+^p$ is nonnegative, continuous, convex, and nondecreasing. Its composition with either affine map $z \mapsto z - \tau$ or $z \mapsto -z - \sigma$ is convex. Nonnegative sums preserve convexity, and nonnegative thresholds imply $h_{\mathrm{mh}}(0) = 0$. Compactness of $[-1,1]$ supplies a finite upper bound. This proves part (b).

For part (c), use the pointwise identity $z_+ - (-z)_+ = z$. Fubini's theorem and the marginal restrictions on every admissible η_X yield

$$\begin{aligned}
&\Theta_{h_+}(\pi,\pi';\eta) - \Theta_{h_-}(\pi,\pi';\eta) \\
&\quad = \mathbb{E}_X\left[\int\int_0^1 \{q_{y'}(t) - q_y(t)\}\,dt\,d\eta_X(y,y')\right] \\
&\quad = \mathbb{E}_X\left[\int\int_0^1 q_{y'}(t)\,dt\,dP_{\pi'(X),X}(y') - \int\int_0^1 q_y(t)\,dt\,dP_{\pi(X),X}(y)\right] \\
&\quad = \mu(\pi') - \mu(\pi).
\end{aligned}$$

The right-hand side depends only on the identified outer marginals and not on the matching kernel.

We also verify the interpretation in (18). For $z \in [-1,1]$, $\tau \in [0,1)$, and $p \geq 1$,

$$(z-\tau)_+^p = p\int_\tau^1 (s-\tau)^{p-1}\mathbb{1}\{z > s\}\,ds.$$

Substituting $z = \Delta_{y,y'}(t)$ and applying Tonelli's theorem over s,t,X, and the matching kernel gives the displayed identity.

It remains to prove part (d). Choose a sequence of finite partitions of $[-1,1]$ that contain zero and whose mesh tends to zero, and let h_J be the piecewise-linear interpolant of h on the Jth partition. Convexity of h implies convexity of h_J; because zero is a global minimizer, $h_J \geq 0$ and $h_J(0) = 0$. Uniform continuity of h on the compact interval implies $\|h - h_J\|_\infty \to 0$.

On $[0,1]$, the successive slopes of h_J are nonnegative and nondecreasing. Hence this restriction has the representation

$$h_J(z) = a_0 z_+ + \sum_j a_j(z-\tau_j)_+, \qquad a_j \geq 0.$$

Applying the same argument to $u \mapsto h_J(-u)$ on $[0,1]$ gives $b_0(-z)_+ + \sum_k b_k(-z-\sigma_k)_+$ on the negative half-line, with all coefficients nonnegative. Thus h_J has the multi-hinge form with unit powers.

Finally, if $\|h - h_J\|_\infty \leq \varepsilon$, then for every pair of quantile functions

$$|c_h(y,y') - c_{h_J}(y,y')| \leq \int_0^1 \varepsilon\,dt = \varepsilon.$$

Every coupling value differs by at most ε, so the same is true after taking an infimum or supremum and after averaging over X. This proves the endpoint stability claim and completes the proof. □

G.1 Contact and rectangular orders for common losses

Table 1 follows directly from the contact and rectangular conditions of Theorem 4.1. For $h(z) = (z-\tau)_+$ with $0 \le \tau < 1$, choose $z_h = \tau$ and $g_h = 1/2$. The contact remainder is $|u|/2$ locally, while Lipschitz continuity bounds the rectangular increment by $2\min\{|u|, |v|\}$. Thus both orders are one. For $h(z) = (z-\tau)_+^2$, choose $z_h = (1+\tau)/2$ and restrict $|u|$ so that $z_h + u > \tau$; the contact remainder is u^2. The derivative is globally 2-Lipschitz, so the rectangular increment is at most $2|uv|$, giving both orders two.

For $h(z) = \lambda_+ z_+^p + \lambda_-(-z)_+^p$ with fixed $\lambda_\pm > 0$, the contact remainder at zero is at least $\min(\lambda_+, \lambda_-)|u|^p$ when $1 < p \le 2$. Its derivative is $(p-1)$-Hölder, which, by integrating in either increment, bounds the rectangular difference by a constant times $\min\{|u||v|^{p-1}, |v||u|^{p-1}\}$. For $p = 1$, take $g_h = (\lambda_+ - \lambda_-)/2$ at zero; the remainder is $(\lambda_+ + \lambda_-)|u|/2$, and the loss is Lipschitz. For $p > 2$, choose a positive interior contact point, where the curvature is strictly positive; the derivative is Lipschitz on $[-1, 1]$. The matching orders are therefore $p \wedge 2$. Constants may depend on τ, $\lambda_\pm$, and p; no uniform claim is made as these parameters degenerate or vary with n. All three rates apply to both endpoints and to their honest bands under the fixed-design assumptions of the main theorem.

H Unknown Propensities and the Geometric Limit

This extension separates augmented evaluation of a learned certificate from learning the sharp endpoint. The augmentation is the usual dual-score AIPW construction; the additional conclusion is that, in a fixed finite design, unknown assignment probabilities leave the geometric minimax order unchanged. All results below concern exactly observed distribution-valued outcomes. The proof combines the geometric training-gap and testing results in Appendix F with explicit observational calibration.

H.1 Main result and its scope

Observe independent copies of $O = (X, A, Y)$, with consistency, $Y(a) \perp A \mid X$, and $Y = Y(A)$. Suppose $\mathcal{X} = \{x_1, \dots, x_K\}$, $A \in \{1, \dots, d\}$, and

$$P_X(x_j) = p_j \ge \kappa > 0, \qquad e_a(x_j) = P(A = a \mid X = x_j) \ge \varepsilon > 0.$$

Here K, d are fixed positive integers, $0 < \kappa \le 1/K$, and $0 < \varepsilon \le 1/d$. The covariate support is fixed and known; ε is an available overlap lower bound, but the assignment array e is unknown and may vary throughout the model class. The conditional outer laws $P_{a,j}$ are arbitrary probability laws on the bounded monotone quantile cone $\mathcal{Q}$. Let Π_K be the distinct action vectors induced by a deterministic policy class, and put $N = |\Pi_K| \le d^K$. Assume $N \ge 2$.

For a continuous convex $h : [-1, 1] \to [0, H]$, with $H > 0$ and $h(0) = 0$, write $c_h(q, r) = \int_0^1 h(r(t) - q(t))\, dt$. Define the lower and upper conditional transport endpoints and average them over P_X to obtain $\Theta^L(u), \Theta^U(u)$ for $u = (\pi, \pi') \in \Pi_K^2$. In particular $0 \le \Theta^L(u) \le \Theta^U(u) \le H$. Write $\mathcal{P}^{\mathrm{obs}}$ for this model with unknown e.

The local notation corresponds to the main text as follows: $K = K_X$, $N = N_\Pi$, $\varepsilon = \underline{e}$, $H = \bar{h}$, $\beta = \beta_h$, and $\gamma = \gamma_h$. The geometric assumptions are the contact and rectangular conditions. To make the extension self-contained, the contact condition (C_β) means that, for some $z_0 \in (-1, 1)$, $g_0 \in \partial h(z_0)$, $c_0 > 0$, $\beta \ge 1$, and sufficiently small $a_0 > 0$,

$$h(z_0 + v) - h(z_0) - g_0 v \ge c_0 |v|^\beta \qquad (|v| \le a_0).$$

For $0 < \gamma \le 2$, condition (U_γ) requires h to be $(\gamma \wedge 1)$-Hölder and

$$|h(z + v + w) - h(z + v) - h(z + w) + h(z)| \le C_h \vartheta_\gamma(v, w),$$

whenever the four arguments belong to $[-1, 1]$, where

$$\vartheta_\gamma(v, w) = \begin{cases} \min\{|v|^\gamma, |w|^\gamma\}, & \gamma \le 1, \\ \min\{|v||w|^{\gamma-1}, |w||v|^{\gamma-1}\}, & \gamma > 1. \end{cases}$$

For an endpoint $s \in \{L, U\}$ define its estimation risk by

$$\mathcal{R}_n^{\mathrm{est,s}} = \inf_{\widehat{T}} \sup_{P \in \mathcal{P}^{\mathrm{obs}}} E_P \sup_{u \in \Pi_K^2} |\widehat{T}(u) - \Theta_P^s(u)|.$$

A lower band is honest if $\inf_P P\{L(u) \le \Theta_P^L(u),\ \forall u\} \ge 1 - \alpha$; define an honest upper band analogously. Their risks are

$$\mathcal{R}_n^{\mathrm{hon,L}} = \inf_{L \text{ honest}} \sup_P E_P \sup_u [\Theta_P^L(u) - L(u)]_+,$$
$$\mathcal{R}_n^{\mathrm{hon,U}} = \inf_{U \text{ honest}} \sup_P E_P \sup_u [U(u) - \Theta_P^U(u)]_+.$$

Theorem H.1 (Geometric minimax order with unknown assignment probabilities)**.** *Fix $0 < \alpha < 1/3$ and the finite-design quantities above. Split the sample into three independent folds of sizes ℓ, t, m, each a fixed positive fraction of n. The construction below uses only the observed data and the available constants, not the true propensity array.*

(a) *For every sample size for which the three folds are nonempty, the constructed bands satisfy the joint finite-sample guarantee*

$$\inf_{P \in \mathcal{P}^{\mathrm{obs}}} P\{L_n(u) \le \Theta_P^L(u), \quad \Theta_P^U(u) \le U_n(u), \quad \forall u \in \Pi_K^2\} \ge 1 - \alpha. \tag{202}$$

For smaller sample sizes set $L_n = 0$ and $U_n = H$. This validity statement requires bounded feasible potentials but no accuracy of the outer-law learner.

(b) *If (U_γ) holds, empirical cellwise dual optimization on the first fold gives endpoint estimators and bands with maximal expected error at most*

$$C\left\{[\log(\ell\kappa\varepsilon + 1)]^{-\gamma} + \sqrt{\frac{\log(4N^2/\alpha)}{m}} + \frac{\log\{C_0(n+1)/\alpha\}}{t} + e^{-ct}\right\}, \tag{203}$$

where error means absolute worst-policy error for either endpoint estimator or positive worst-policy excess length for either band. Constants depend only on $h, K, d, \kappa, \varepsilon$; $C_0 \ge 1$ is fixed.

(c) *If (C_β) holds, each of the four minimax risks above is at least $c[\log(n+1)]^{-\beta}$ for every sufficiently large n, where the lower-bound constant and threshold may additionally depend on α. Consequently, when both conditions hold with $\gamma = \beta$,*

$$\mathcal{R}_n^{\mathrm{est,L}} \asymp \mathcal{R}_n^{\mathrm{est,U}} \asymp \mathcal{R}_n^{\mathrm{hon,L}} \asymp \mathcal{R}_n^{\mathrm{hon,U}} \asymp (\log n)^{-\beta}. \tag{204}$$

The minimax order is preserved; the constants and confidence corrections can differ from those for known propensities. The lower bounds are inherited from the manuscript's fixed known-assignment submodels. The new upper bound accounts for estimating assignment probabilities and for obtaining an implementable finite-sample correction. For $h(z) = |z|^p$, $p \ge 1$, $\gamma = \beta = p \wedge 2$.

H.2 Augmented evaluation and its exact remainder

The following construction and identities also make sense for general covariates. Let $D_V = D_1$ be the first fold and set

$$c_0 = c_h, \qquad c_1 = H - c_h, \qquad T_0(u) = \Theta^L(u), \quad T_1(u) = H - \Theta^U(u).$$

For $k \in \{0, 1\}$ fit a measurable action-pair-indexed collection $(\phi_{ab,x}^k, \psi_{ab,x}^k)$ on D_V, globally feasible for c_k:

$$\phi_{ab,x}^k(q) + \psi_{ab,x}^k(r) \le c_k(q, r), \qquad |\phi_{ab,x}^k| \vee |\psi_{ab,x}^k| \le M.$$

Use a deterministic $M < \infty$. For the normalized empirical duals below one can take $M = H$. All fitted functions and selections are jointly measurable. For $u = (\pi, \pi')$, abbreviate $a = \pi(x)$ and $b = \pi'(x)$, and define

$$\mu_{\phi,ab}^k(x) = P_{a,x}\phi_{ab,x}^k, \qquad \mu_{\psi,ab}^k(x) = P_{b,x}\psi_{ab,x}^k,$$
$$g_k(u) = E_X[\mu_{\phi,\pi(X)\pi'(X)}^k(X) + \mu_{\psi,\pi(X)\pi'(X)}^k(X)],$$
$$\Delta_k(u) = T_k(u) - g_k(u) \ge 0.$$

All expectations here treat D_V as fixed. On an independent nuisance fold D_N fit $\widehat{e}_a \in [\varepsilon, 1]$ and $\widehat{\mu}^k_{\phi,ab}, \widehat{\mu}^k_{\psi,ab} \in [-M, M]$. Let $\mathcal{F} = \sigma(D_V, D_N)$. On the last fold D_E, the augmented score is

$$\begin{aligned} S^k_u(O) = \widehat{\mu}^k_{\phi,ab}(X) + \widehat{\mu}^k_{\psi,ab}(X) \\ + \frac{\mathbf{1}\{A=a\}}{\widehat{e}_a(X)}\{\phi^k_{ab,X}(Y) - \widehat{\mu}^k_{\phi,ab}(X)\} \\ + \frac{\mathbf{1}\{A=b\}}{\widehat{e}_b(X)}\{\psi^k_{ab,X}(Y) - \widehat{\mu}^k_{\psi,ab}(X)\}, \qquad a = \pi(X),\ b = \pi'(X). \end{aligned} \tag{205}$$

This formula includes $a = b$; the two residuals then use the same observation. Let $\overline{S}^k_u = m^{-1}\sum_{i \in D_E} S^k_u(O_i)$ and $Z_k(u) = \overline{S}^k_u - E[S^k_u(O) \mid \mathcal{F}]$.

Lemma H.2 (Evaluation error versus the geometric gap)**.** *Conditional on $\mathcal{F}$,*

$$\begin{aligned} R_k(u) &:= E[S^k_u(O) \mid \mathcal{F}] - g_k(u) \\ &= E_X\left[\frac{\widehat{e}_a - e_a}{\widehat{e}_a}(\widehat{\mu}^k_{\phi,ab} - \mu^k_{\phi,ab}) + \frac{\widehat{e}_b - e_b}{\widehat{e}_b}(\widehat{\mu}^k_{\psi,ab} - \mu^k_{\psi,ab})\right]. \end{aligned} \tag{206}$$

Consequently the exact decompositions are

$$\overline{S}^0_u - \Theta^L(u) = Z_0(u) + R_0(u) - \Delta_0(u), \tag{207}$$

$$(H - \overline{S}^1_u) - \Theta^U(u) = -Z_1(u) - R_1(u) + \Delta_1(u). \tag{208}$$

In particular R_k vanishes when all propensities are exact or when all relevant conditional potential means are exact, regardless of Δ_k.

For general covariates, define conditional $L^2(P_X)$ errors by

$$\begin{aligned} r_e^2 &= E_X \max_a |\widehat{e}_a(X) - e_a(X)|^2, \\ r_\mu^2 &= E_X \max_{k,a,b} \left\{|\widehat{\mu}^k_{\phi,ab}(X) - \mu^k_{\phi,ab}(X)|^2 \vee |\widehat{\mu}^k_{\psi,ab}(X) - \mu^k_{\psi,ab}(X)|^2\right\}. \end{aligned}$$

Then, simultaneously over all policies,

$$\sup_{k,u} |R_k(u)| \le \frac{2}{\varepsilon} r_e r_\mu. \tag{209}$$

Proof. For the first residual, conditional exchangeability and consistency give

$$E[\mathbf{1}\{A=a\}\phi^k_{ab,X}(Y) \mid X, \mathcal{F}] = e_a(X)\mu^k_{\phi,ab}(X).$$

Thus its augmented mean minus $\mu^k_{\phi,ab}$ equals $(1 - e_a/\widehat{e}_a)(\widehat{\mu}^k_{\phi,ab} - \mu^k_{\phi,ab})$. The second residual has the identical calculation, even for $a = b$. Integrate over X and use weak duality to obtain the decompositions. Cauchy–Schwarz applied to the pointwise maxima, with $\widehat{e}_a \ge \varepsilon$, proves the uniform bound. □

H.3 A FINITE-SAMPLE TRANSFER INEQUALITY

Suppose there are N distinct policy vectors (or a finite policy class of cardinality N). The score has deterministic envelope

$$|S^k_u| \le B := 2M + 4M/\varepsilon.$$

In the following formulas b denotes a radius, not an action label. Let $b \ge 0$ be $\mathcal{F}$-measurable and suppose $P\{\sup_{k,u} |R_k(u)| \le b\} \ge 1 - \delta$. For $0 < \eta < 1$ set

$$r = B\sqrt{\frac{2\log(4N^2/\eta)}{m}}, \qquad \begin{aligned} L(u) &= \mathrm{clip}_{[0,H]}(\overline{S}^0_u - r - b), \\ U(u) &= \mathrm{clip}_{[0,H]}(H - \overline{S}^1_u + r + b). \end{aligned}$$

Proposition H.3 (Simultaneous validity and informativeness)**.** *These bands satisfy joint simultaneous coverage at least* $1-\eta-\delta$*. On the same event of probability at least* $1-\eta-\delta$*,*

$$[\Theta^L(u) - L(u)]_+ \le \Delta_0(u) + 2(r+b),$$
$$[U(u) - \Theta^U(u)]_+ \le \Delta_1(u) + 2(r+b) \qquad \textit{for every } u.$$

Let $D = \sup_{k,u} \Delta_k(u)$*,* $R_* = \sup_{k,u} |R_k(u)|$*, and* $Z_* = \sup_{k,u} |Z_k(u)|$*. Without restricting to that event, either band's positive worst-policy excess is bounded by* $D + Z_* + R_* + r + b$*. Either projected endpoint estimator* $\mathrm{clip}_{[0,H]}(\overline{S}^0)$ *or* $\mathrm{clip}_{[0,H]}(H - \overline{S}^1)$ *has absolute worst-policy error at most* $D + Z_* + R_*$*.*

Proof. Conditional on $\mathcal{F}$, the evaluation observations are independent. Hoeffding's inequality and a union bound over $2N^2$ scores give $P(Z_* > r \mid \mathcal{F}) \le 4N^2 \exp\{-mr^2/(2B^2)\} = \eta$. Intersect this event with $\{R_* \le b\}$ and apply the two decompositions. Clipping to $[0, H]$ preserves the direction of a valid inequality and cannot increase distance from a target in $[0, H]$. The same decompositions and the triangle inequality give the unconditional deterministic error bounds. Also, integrating the conditional sub-Gaussian maximum bound yields

$$E[Z_* \mid \mathcal{F}] \le CB\sqrt{\frac{\log(4N^2)}{m}}. \tag{210}$$

□

Finite-sample coverage uses a high-probability remainder envelope with an explicit failure probability. The next construction obtains it from conditional means of the fitted potentials on the nuisance fold.

H.4 An implementable construction on finite covariate support

In D_N let $N_j = \sum_{i \in D_N} \mathbf{1}\{X_i = x_j\}$ and $N_{ja} = \sum_{i \in D_N} \mathbf{1}\{X_i = x_j, A_i = a\}$. Set

$$\widehat{e}_{aj} = \max\{\varepsilon, N_{ja}/N_j\} \quad (N_j > 0), \qquad \widehat{e}_{aj} = \varepsilon \quad (N_j = 0).$$

No renormalization across actions is performed: these are separately clipped denominators. Because $e_{aj} \in [\varepsilon, 1]$, clipping does not increase estimation error. For every k, a, b, j, average $\phi^k_{ab,x_j}(Y_i)$ over the (j, a) cell to form $\widehat{\mu}^k_{\phi,ab}(x_j)$, and average $\psi^k_{ab,x_j}(Y_i)$ over the (j, b) cell to form $\widehat{\mu}^k_{\psi,ab}(x_j)$. Use zero in an empty cell. All these means belong to $[-M, M]$.

There are at most $J = 4Kd^2$ fitted potential means across both costs. For $\delta_e, \delta_\mu \in (0, 1)$ define

$$q_e(s) = \begin{cases} 1, & s = 0, \\ \min\{1, \sqrt{\log(2Kd/\delta_e)/(2s)}\}, & s > 0, \end{cases}$$
$$q_\mu(s) = \begin{cases} 2M, & s = 0, \\ \min\{2M, M\sqrt{2\log(2J/\delta_\mu)/s}\}, & s > 0. \end{cases}$$

The computable radii and remainder envelope are

$$\eta_e = \max_j q_e(N_j), \qquad \eta_\mu = \max_{j,a} q_\mu(N_{ja}), \qquad b = \frac{2}{\varepsilon}\eta_e\eta_\mu. \tag{211}$$

Lemma H.4 (Finite-cell nuisance envelopes)**.** *Uniformly over* $\mathcal{P}^{\mathrm{obs}}$ *and conditional on* D_V*, the event*

$$\max_{j,a} |\widehat{e}_{aj} - e_{aj}| \le \eta_e,$$
$$\max_{k,a,b,j} (|\widehat{\mu}^k_{\phi,ab}(x_j) - \mu^k_{\phi,ab}(x_j)| \vee |\widehat{\mu}^k_{\psi,ab}(x_j) - \mu^k_{\psi,ab}(x_j)|) \le \eta_\mu$$

has probability at least $1 - \delta_e - \delta_\mu$*. On this event* $R_* \le b$*. Moreover, for constants depending only on* $K, d, \kappa, \varepsilon, M$*,*

$$Eb \le C\frac{\sqrt{\log(2Kd/\delta_e)\log(2J/\delta_\mu)}}{t} + Ce^{-ct}. \tag{212}$$

Proof. Conditional on the covariate labels in D_N, each N_{ja} is marginally binomial with N_j trials and success probability e_{aj}. Hoeffding and a union bound over Kd probabilities give the first envelope with failure probability at most δ_e; empty cells and capped radii are covered by deterministic bounds. Conditional on D_V and all covariate and action labels in D_N, the outcomes within each cell are independent with law $P_{a,j}$. Each fixed potential takes values in $[-M, M]$. Hoeffding therefore gives failure probability at most δ_μ/J per potential mean, including the deterministic treatment of empty cells. A union bound proves the second envelope. These two events need not be independent. Apply the remainder identity to their intersection.

For the expectation bound, define the count event $\mathcal{C} = \{N_{ja} \geq t\kappa\varepsilon/2 \text{ for every } j, a\}$. Each N_{ja} is binomial with success probability at least $\kappa\varepsilon$, so a multiplicative Chernoff bound gives $P(\mathcal{C}^c) \leq Kd\exp(-t\kappa\varepsilon/8)$. On $\mathcal{C}$, also $N_j \geq t\kappa\varepsilon/2$, and

$$b \leq \frac{4M}{\kappa\varepsilon^2 t}\sqrt{\log(2Kd/\delta_e)\log(2J/\delta_\mu)}.$$

Everywhere $b \leq 4M/\varepsilon$. Integrating proves the claim. □

The same joint event permits endpoint-specific envelopes. If the training fold certifies $|\phi^k| \vee |\psi^k| \leq M_k$, replace M by M_k in q_μ for cost k and denote the resulting maximum by $\eta_{\mu,k}$. The union bound still ranges over the same J potential means and still fails with probability at most δ_μ. Then $|R_k(u)| \leq b_k := 2\eta_e\eta_{\mu,k}/\varepsilon$ for every u on that joint event. Applying the two decompositions with b_0 for the lower endpoint and b_1 for the complement-cost upper endpoint gives the same joint coverage statement. A known finite outcome support can supply a valid M_k through its full declared support; for an unrestricted outcome model, the bound must hold on the entire outcome space.

H.5 Proof of the geometric rate theorem

Proof of Theorem H.1. Take $\delta_e = \delta_\mu = \delta_n := \alpha/[8(n+1)]$, $\eta = \alpha - 2\delta_n$, and use (211) in Proposition H.3. This proves (202) for arbitrary bounded feasible potentials. Notice that $\eta \geq 7\alpha/8$ for $n \geq 1$, so the evaluation radius is still of order $\sqrt{\log(4N^2/\alpha)/m}$; allocating vanishing nuisance failure probabilities does not insert $\log n$ into that radius.

For attainability, form empirical conditional outer laws on D_V. For each cost c_k, action pair, and cell, choose normalized globally feasible empirical dual optimizers, with a measurable selection. Use a fixed fallback law in empty cells. The quantile cone is compact in L^1 and the costs are continuous. Their normalized concave potentials are uniformly bounded and equicontinuous, form a closed class, and are therefore compact in the uniform norm. Empirical objectives are jointly continuous on this class and in the empirical laws, so a measurable maximizing selection exists. Normalize the first potential to have minimum zero. Its oscillation is at most H; its cost transform takes values in $[-H, H]$. Thus $M = H$ is available for both costs without reference to the unknown outcome laws. The complement cost has the same magnitude of rectangular interaction and the same section modulus as c_h.

The geometric input is the cellwise empirical-dual gap bound (200). For nonempty cell sizes s_1, s_2, it bounds the expected gap for either cost by

$$C_h\{[\log(s_1+1)]^{-\gamma} + [\log(s_2+1)]^{-\gamma}\}.$$

This input depends on the within-cell outcome laws and sample sizes, not on knowing the probabilities that generated those sizes. For every unknown array e in the present model, each training cell has probability at least $\kappa\varepsilon$. The same count argument as in Lemma H.4, with ℓ in place of t, therefore gives, uniformly over this entire model,

$$ED \leq C_h[\log(\ell\kappa\varepsilon+1)]^{-\gamma}. \tag{213}$$

Indeed, a maximum over the fixed $2Kd^2$ cellwise gaps is bounded by their sum, and bad counts contribute only an exponentially small term because bounded potentials imply a bounded gap. This also treats repeated actions.

To obtain an expectation risk, a fixed-probability argument alone would not suffice. Here $R_* \leq 4M/\varepsilon$ deterministically, while $R_* \leq b$ except on an event of probability at most $2\delta_n$. Consequently

$$ER_* \leq Eb + (8M/\varepsilon)\delta_n = Eb + \frac{M\alpha}{\varepsilon(n+1)}.$$

Combine this inequality with (212), (210), (213), and the deterministic error bounds in Proposition H.3. The logarithms in (212) are $O(\log\{C_0(n+1)/\alpha\})$. Since t is a fixed positive fraction of n, the displayed $O(n^{-1})$ failure term is absorbed by the $\log\{C_0(n+1)/\alpha\}/t$ term. Thus a $1/n$ allocation suffices for both exact finite-sample coverage and the stated expected-risk order. This proves (203) for both point estimators and both bands. Approximate optimization can be included by adding its expected uniform objective gap, provided global feasibility and the envelope are retained.

For the lower bounds, fix any admissible array e^0 and restrict $\mathcal{P}^{\mathrm{obs}}$ to laws with $e = e^0$. Giving that array to the statistician can only make estimation easier. The manuscript's contact lower bounds for the two endpoints in Theorems F.7 and F.8 apply to this known-assignment subexperiment. Any observational band honest over $\mathcal{P}^{\mathrm{obs}}$ is also honest on that submodel. Thus their estimation and honest-band lower bounds transfer unchanged. Combining them with (203) proves (204), since $n^{-1/2} + \log(n)/n = o((\log n)^{-\beta})$ for every fixed $\beta > 0$. □

H.6 Rate transfer and nuisance correction

Corollary H.5 (Rate transfer for a given geometric gap)**.** *In Proposition H.3, suppose uniformly over a model class that $ED \le a_n$, $ER_* \le Cq_n$, and $Eb \le Cq_n$, with a certified remainder envelope having the required failure probability. Then the expected positive excess of either band is at most*

$$a_n + C\left\{B\sqrt{\frac{\log(4N^2/\eta)}{m}} + q_n\right\}.$$

The same bound applies to the absolute error of either point estimator. In particular a geometric upper rate is preserved when the other displayed terms are of no larger order. It is the leading upper-bound term when they are smaller order.

Proof. Take expectations in the deterministic inequalities of Proposition H.3 and use (210). □

Structured rates and finite computation. This transfer also justifies the two evaluation procedures in the main text's structured and computational bounds. In the fixed finite design, the HPQ empirical-dual training bound of Appendix I is conditional on cell sizes and within-cell laws. The same count argument used in (213) therefore yields $ED \lesssim (\ell\kappa\varepsilon + 1)^{-1/s}$ for unknown assignment. For the completed finite-anchor lower trainer, Theorem E.4 additionally gives $K_{\mathrm{bin}}^{-1} + \eta + \varepsilon_{\mathrm{opt}}$ in this gap, where K_{bin} denotes the number of quantile bins, with the constants and Lipschitz conditions specified there. These are training bounds, independent of the subsequent choice of evaluation score. Use the final globally feasible potentials as the scalar outcomes on the nuisance fold, with their certified absolute envelope in place of M. Lemma H.4 and Proposition H.3 then add $O(m^{-1/2} + \log(n)/t)$ for fixed confidence level and finite design. Thus both evaluation procedures retain the structured $n^{-1/s}$ upper rate when the folds are fixed proportions, $s > 2$, and the numerical terms are $O(n^{-1/s})$. A matching structured lower bound remains open. Safe upper extensions must still include their deterministic binning correction before applying the complement-cost construction; arbitrary clipping or coarsening of covariates is not covered by this transfer.

For a polynomial geometric upper rate, nuisance and evaluation errors must be compared to that polynomial rate. For a logarithmic geometric rate, any polynomially decaying certified nuisance remainder is smaller order. These comparisons give upper rates for the specified nuisance classes. In particular, the geometric lower bound is a worst-case statement; it does not provide a positive lower bound on every realized dual gap and cannot justify omitting the nuisance correction.

The conditional means above concern the scalar transformed outcomes $\phi(Y)$ and $\psi(Y)$, not the full outer law. Independent estimation of those means is why their accuracy need not inherit the geometric rate for learning the endpoint. The three-fold split makes these estimation and evaluation steps independent.

Plug-in IPW also preserves the geometric rate in the same finite design: plug-in IPW has conditional bias at most $2M\eta_e/\varepsilon$ on the propensity envelope event. Its certified first-order correction is $O(\sqrt{\log n/n})$, also smaller than the logarithmic geometric term. Augmentation reduces this additional correction to $O(\log n/n)$.

The score-level mechanism follows the observational dual-bound extension of Ji et al. (2023, Section 3.4). The finite-cell envelope transfers the geometric upper and lower bounds to unknown assignment under fixed finite support, uniform overlap, conditional exchangeability, and exactly observed distributions. Extensions beyond these conditions require additional analysis. Sharpness also requires the outer-law learner's dual gap to vanish.

H.7 Experiments for the observational extension

We examine joint calibration, double robustness of certificate evaluation, and misspecification of the outer-law learner. The latter two experiments isolate mechanisms using the original synthetic outcome model. These finite-support experiments assess calibration and robustness of the observational construction.

H.7.1 Experiment 1: Joint coverage and informativeness

This experiment evaluates both sharp endpoints and all ordered policy pairs under observed-confounder-dependent assignment. It uses the 49-type outcome universe of Appendix K.1.1, two covariate cells, three actions, nine policies, and W_2^2 cost. The population cell masses are $1/2$. The treatment probabilities are $(0.2, 0.3, 0.5)$ in the first cell and $(0.5, 0.3, 0.2)$ in the second. Only the simulator and the oracle comparator receive these arrays; the feasible estimators receive the lower overlap bound 0.2. Treatment is generated conditional on X before the outcome type is drawn, so conditional exchangeability holds by construction.

With the type features v_r defined in Appendix K.1.1, conditional outcome weights are proportional to

$$\exp\left[-\frac{1}{2}\left\{\left(\frac{v_{r1}-u_{xa,1}}{0.045}\right)^2+\left(\frac{v_{r2}-u_{xa,2}}{0.030}\right)^2\right\}\right]+10^{-6},$$

where the center pairs, indexed by $x = 0, 1$ and action $a = 0, 1, 2$, are

$$\begin{pmatrix}(0.10,0.12) & (0.42,0.20) & (0.74,0.30)\\ (0.68,0.13) & (0.12,0.30) & (0.45,0.12)\end{pmatrix}.$$

The conditional laws differ from that synthetic experiment, and all population endpoints are recomputed by finite OT. The average sharp lower endpoint over 81 pairs is 0.123946; among pairs with a positive lower endpoint its mean is 0.139440 and its maximum is 0.368940.

For each $n \in \{8192, 16384, 65536\}$, use 400 independent repetitions and split each sample into three nearly equal folds. The first fold retains up to 12 highest-frequency types per action–cell stratum, solves both finite OT duals, applies the globally feasible 2-Lipschitz extension, and then applies paired clipping. The second fold estimates each propensity by its clipped cell frequency and every conditional scalar potential mean by the corresponding cell average. The third fold evaluates the scores. For the upper endpoint, we evaluate the complete complement cost in the extension rather than add a uniform binning repair. Writing $c_1(q, r) = 1 - c_h(q, r)$ and denoting the source anchors by q_i, the finite-dual coefficients α_i define

$$\phi_1(q) = \max_i\{\alpha_i - 2\|q - q_i\|_1\}, \qquad \psi_1(r) = \min_i\{c_1(q_i, r) - \alpha_i\}.$$

For an index attaining the maximum in $\phi_1(q)$, the definition of ψ_1 and the 2-Lipschitz property of c_1 in its first argument give $\phi_1(q) + \psi_1(r) \le c_1(q, r)$ on the entire quantile cone. Paired clipping preserves this feasibility. On the declared 24-step outcome types, the complete integrals equal their finite-bin evaluations, so this extension is exactly evaluated in the simulation without the generic $2/24$ correction. We call this the *direct complete-cost* extension. It differs from the generic bin-average construction, which requires the correction when the complete cost cannot be evaluated. Experiments 2 and 3 below also use the direct complete-cost extension.

The three paired methods share the same draws and trained potentials: oracle IPW knows the assignment array; estimated IPW replaces it by cell frequencies and uses the certified first-order bias allowance $2M_k\eta_e/\varepsilon$ for endpoint cost k; estimated AIPW uses the nuisance-fold scalar means and the endpoint-specific product allowance $2\eta_e\eta_{\mu,k}/\varepsilon$. The M_k are calculated over the declared 49-type outcome support, which contains every possible observed outcome in this experiment. Both latter procedures use $\delta_e = \delta_\mu = \alpha/[8(n+1)]$ and $\eta = \alpha - 2\delta_e$ at $\alpha = 0.05$. The evaluation radius uses the

exact score range over the declared finite outcome support and a one-sided Hoeffding union bound over both endpoints and all 81 pairs. All bands are clipped to $[0, 1]$. This is a joint 95% calibration, including the two endpoints; it is not two separate 95% intervals.

Table 3: **Unknown-assignment policy certificates under strong heterogeneity.** Each row has 400 repetitions. Coverage is joint over both sharp endpoints and all 81 policy pairs; ratios divide the mean LCB or UCB over pairs and repetitions by the corresponding mean sharp endpoint over all 81 pairs, including zero lower endpoints. The oracle knows the true assignment probabilities.

n	Procedure	Joint coverage	LCB/truth	UCB/truth
8,192	Oracle IPW	400/400	0.629	2.356
8,192	Est. IPW	400/400	0.034	5.437
8,192	Est. AIPW	400/400	0.265	3.329
16,384	Oracle IPW	400/400	0.735	1.954
16,384	Est. IPW	400/400	0.142	4.176
16,384	Est. AIPW	400/400	0.491	2.424
65,536	Oracle IPW	400/400	0.864	1.475
65,536	Est. IPW	400/400	0.442	2.630
65,536	Est. AIPW	400/400	0.786	1.572

Across increasing sample sizes, estimated AIPW's LCB/truth ratios are 0.265, 0.491, and 0.786, compared with 0.034, 0.142, and 0.442 for certified estimated IPW. Its UCB/truth ratios are 3.329, 2.424, and 1.572, compared with 5.437, 4.176, and 2.630. All methods have observed joint coverage $400/400$ at each size. These are empirical frequencies under this declared stress design; their 95% Wilson intervals have lower endpoint about 99.0% when all 400 cover.

The table reports whether the LCB and UCB cover their respective sharp endpoints simultaneously for all 81 pairs in each repetition. Each ratio divides the bound averaged over pairs and repetitions by the corresponding sharp endpoint averaged over all 81 pairs. Thus pairs with zero lower truth remain in the aggregate without undefined individual ratios. The oracle is an unattainable reference under unknown assignment. The observed coverage frequency does not establish a true 100% coverage probability; the theory supplies its finite-sample lower guarantee. The direct complete-cost extension uses information available in this 24-step simulation. When only bin averages can be evaluated, the generic upper correction in Theorem E.3 remains necessary. This finite simulation does not establish a rate for arbitrary curves or growing outcome supports.

The paired rerun driver is `reproducibility/observational_experiment1_exact_upper.py`; it uses the independent final core `reproducibility/observational_exact_upper_core.py` with the direct complete-cost upper extension, sizes as above, 400 repetitions, and seed 2026092602. Per-repetition records, summary, and manifest are in `reproducibility/results/observational_experiment1_strong_exact_upper/`.

The mean interval widths for oracle IPW, estimated IPW, and estimated AIPW are respectively $(0.225, 0.694, 0.395)$ at $n = 8192$, $(0.160, 0.519, 0.250)$ at $n = 16384$, and $(0.082, 0.283, 0.104)$ at $n = 65536$. No AIPW interval equals the full $[0, 1]$ range in these repetitions. Among policy pairs with positive sharp lower endpoint, the fractions with a positive AIPW lower certificate are 0.480, 0.685, and 0.999. Thus the observed coverage is accompanied by informative certificates, although the calibration remains conservative.

H.7.2 Experiment 2: Double robustness of conditional evaluation

We return to the original synthetic outcome model of Appendix K.1.1, with the same observational assignment probabilities as Experiment 1. A single training fold of 6,000 observations learns globally feasible, paired-clipped potentials using at most 12 anchors per action marginal and stratum; these potentials are then frozen. The upper endpoint uses the direct complete-cost complement extension described above, with zero additional binning correction. Its full-space feasibility follows from the complete-cost Lipschitz extension, not from checking only the 49 observed types. We cross exact and misspecified propensities with exact and misspecified conditional potential means. Exact components use simulator expectations. Misspecified propensities are $(0.35, 0.30, 0.35)$ in both cells; misspecified means are the population limits of action-specific intercept-only fits that pool the cells.

The main contrast, fixed before evaluation, compares the policies that always choose the first and third actions. For each $m \in \{2000, 20000\}$, 2,000 independent evaluation repetitions use the same observations for all four settings. No nuisance is fitted on these evaluation data, and point estimates are not clipped.

Table 4: Frozen-certificate double robustness. All entries are in units of 10^{-3}. Bias is the exact conditional evaluation bias relative to the fixed dual endpoint; RMSE is relative to the sharp endpoint, at $m = 20000$. Parentheses are Monte Carlo standard errors for RMSE.

Propensity	Potential means	Lower bias	Upper bias	Lower RMSE	Upper RMSE
Exact	Exact	0.000	0.000	7.897 (0.036)	14.983 (0.030)
Exact	Pooled	0.000	0.000	7.950 (0.038)	14.978 (0.031)
Pooled	Exact	0.000	0.000	7.880 (0.033)	14.974 (0.028)
Pooled	Pooled	−15.275	−3.109	23.075 (0.033)	11.879 (0.028)

The fixed sharp lower and upper endpoints are 0.215668 and 0.241106; their dual gaps are 0.007748 and 0.014911 in all four settings. Across all 81 pairs, either exact nuisance component makes the absolute conditional evaluation bias smaller than 2×10^{-16}. When both components are misspecified, 80 of 81 pairs have nonzero bias at each endpoint (threshold 10^{-10}), and the maximal absolute biases are 0.029612 and 0.025670. For the main upper-endpoint contrast, the negative evaluation bias under joint misspecification partly cancels the positive dual gap, explaining the smaller RMSE. This cancellation is not improved robustness. The experiment verifies the remainder identity conditional on one frozen certificate; it is not a coverage experiment or a comparison of finite-sample nuisance learners. In particular, the nonzero dual gaps persist when evaluation is unbiased. Training and evaluation seeds are 2026092603 and 2026092604. Independent expectation and observation-level score checks agree to less than 4×10^{-16}; full results are in `reproducibility/results/observational_dr_mechanism/`. The reproduction entry point is `reproducibility/observational_dr_mechanism.py`; it explicitly selects `direct_complete_cost` in the dedicated `reproducibility/observational_fixed_core.py` module. Reproduced outputs are saved separately in the `reproduced_direct_complete_cost` subdirectory.

H.7.3 Experiment 3: Outer-law misspecification

We retain the original two-cell, three-action, 49-type synthetic outcome model, but assign actions with probabilities $(0.2, 0.3, 0.5)$ and $(0.5, 0.3, 0.2)$ in the two cells. For each sample size, 200 paired repetitions compare cellwise empirical outer laws with an arm-specific pooled fit that ignores the covariate cell. All observed training types are retained as anchors (cap 49). Globally feasible potentials are constructed for the cost and its complement, using the direct complete-cost extension described in Experiment 1. The complete integrals are exact on the declared 24-step outcome types; the generic $2/24$ upper binning correction is therefore not applied. Three equal-proportion folds learn the duals, estimate cellwise nuisances, and evaluate the scores. Only the outer-law learner is misspecified: nuisance estimation continues to condition on the original cells. Range-based Hoeffding calibration and the certified nuisance correction protect both endpoints and all 81 policy pairs jointly at level 0.95. The score ranges use the full declared finite outcome universe. This experiment uses a common potential envelope for both endpoint costs and nuisance failure probabilities $\delta_e = \delta_\mu = \alpha/[8(n+1)^2]$. Together with the original outcome model and the larger anchor cap, these choices differ from Experiment 1, so absolute interval widths should be compared within each experiment. All six AIPW configurations attained 200/200 observed joint coverage; the corresponding 95% Wilson interval is [0.981,1.000]. Pooling increases the population dual gap, which augmentation does not remove. Finite-sample calibration can dominate the difference in final widths at smaller sample sizes. For cellwise learning, estimated-IPW mean widths are 0.968, 0.580, and 0.355 across the three sizes, compared with AIPW's 0.758, 0.327, and 0.149; oracle-IPW widths are 0.322, 0.170, and 0.096. Both IPW procedures also attain observed joint coverage $200/200$ in every configuration. At $n = 4096$, the proportions of fully vacuous intervals for cellwise estimated IPW and AIPW are 0.5893 and 0.0043, respectively; both are zero at the larger sizes. Pooling can reduce the score range and hence the evaluation radius, so larger dual gaps need not produce larger final widths at every sample size. Observed coverage supplements the finite-sample argument and does not establish exact calibration. The paired oracle-IPW and estimated-IPW comparisons, per-policy records, source hashes, and reproduction protocol are saved

Table 5: Observational AIPW under outer-law misspecification. Gaps are population dual errors averaged over policies and repetitions; width is the mean calibrated outer interval width. Parentheses give Monte Carlo standard errors.

n	Outer-law fit	Lower gap	Upper gap	Width
4096	Cellwise	0.00058 (0.00001)	0.00103 (0.00002)	0.7585 (0.0015)
4096	Pooled X	0.01195 (0.00002)	0.01444 (0.00005)	0.7491 (0.0014)
16384	Cellwise	0.00011 (0.00000)	0.00022 (0.00000)	0.3271 (0.0005)
16384	Pooled X	0.01185 (0.00000)	0.01412 (0.00003)	0.3323 (0.0004)
65536	Cellwise	0.00002 (0.00000)	0.00006 (0.00000)	0.1485 (0.0002)
65536	Pooled X	0.01182 (0.00000)	0.01393 (0.00003)	0.1659 (0.0001)

with `reproducibility/results/observational_outer_misspec/final/`. The entry point `reproducibility/observational_outer_misspec.py` explicitly selects `direct_complete_cost` in `reproducibility/observational_fixed_core.py`; it retains the original sample-size grid, seeds, and 200 repetitions per size. New reproduction outputs are stored in the sibling `reproduced_direct_complete_cost` directory.

Reproduction of Experiments 2 and 3. The generic-binned option is a different construction: it subtracts $1/24$ from each complement potential and increases the upper dual value by $2/24$. It must not be substituted for the direct complete-cost option when reproducing these tables. Complete reruns with the explicit direct option reproduce Experiment 2's frozen arrays and all 32,000 evaluation records, and Experiment 3's 600 repetitions, 3,600 method–learner records, and all per-policy arrays, with zero numerical difference in the recorded environment. The checker `reproducibility/observational_reproduction_verify.py` compares the new outputs with the preserved originals; verified source snapshots, hashes, and the comparison report are retained in `output/observational_repro_repair/`.

I Cost-Active Quantization and Polynomial Upper Rates

The logarithmic rates in Section F arise when the conditional outer laws of the distribution-valued outcomes may occupy the full outcome space. We now give a distribution-dependent condition under which the same policy-indexed learned-dual certificate becomes polynomially informative. Unlike a two-sample comparison between two fixed actions, this section retains the general covariate space, finite action set, and policy pairs of Section A. The bounded continuous cost may depend on the covariate. For every x and action pair, write

$$T_{c,x}(a,b) := \inf_{\gamma\in\Gamma(P_{a,x},P_{b,x})} \int c_x(y,y')\,d\gamma(y,y'). \tag{214}$$

Then

$$\Theta_c^L(\pi,\pi') = \mathbb{E}_X[T_{c,X}\{\pi(X),\pi'(X)\}]\,.$$

The cost induces two one-sided active pseudometrics on $\mathcal{Z}$; they coincide when the cost is symmetric:

$$\rho^c_{x,1}(y,\widetilde{y}) := \sup_{z\in\mathcal{Z}} |c_x(y,z)-c_x(\widetilde{y},z)|, \qquad \rho^c_{x,2}(y,\widetilde{y}) := \sup_{z\in\mathcal{Z}} |c_x(z,y)-c_x(z,\widetilde{y})|. \tag{215}$$

Directions at zero pseudodistance are identified. We normalize the active quotient spaces to have diameter at most one; a fixed nonunit diameter only changes the constants below.

Assumption I.1 (Uniform conditional high-probability cost-active quantization)**.** *Fix $A \geq 1$, $0 < D < \infty$, $\varepsilon_\star \in (0,1]$, and*

$$s > \max\{D,2\}.$$

For every action $a \in \mathcal{A}$, for P_X-almost every covariate value x, every endpoint $k \in \{1,2\}$, and $0 < \varepsilon \leq \varepsilon_\star$, there is a finite codebook $\mathcal{C}_{a,x,k,\varepsilon}$ in the corresponding active quotient such that

$$|\mathcal{C}_{a,x,k,\varepsilon}| \leq \left(\frac{A}{\varepsilon}\right)^D, \tag{216}$$

$$P_{a,x}\{y : \rho^c_{x,k}(y, \mathcal{C}_{a,x,k,\varepsilon}) > \varepsilon\} \le \varepsilon^{s/(s-2)}, \tag{217}$$

where $\rho(y, \mathcal{C}) := \min_{z\in\mathcal{C}} \rho(y, z)$.

Assumption I.1 is a conditional excess-distortion version of the trimmed covering condition underlying the upper Wasserstein dimension of Dudley (1969); Weed & Bach (2019). It allows the geometry to vary with both the action and covariate value, but requires common constants uniformly in (a, x). At resolution $\varepsilon = N^{-1/s}$, all but $N^{-1/(s-2)}$ conditional probability can be represented by at most $A^D N^{D/s}$ points. Compact supports of finite box or doubling dimension, Lipschitz images of bounded Euclidean sets, and regular finite-dimensional manifolds are sufficient examples.

Finite-sample cost-active covering diagnostic. Assumption I.1 is a population condition and cannot be verified at every vanishing scale from a finite sample. Nevertheless, its codebook inequalities induce a direct finite-resolution diagnostic. Fix a finite-support covariate value x_j, an action a, an active endpoint $k \in \{1, 2\}$, and N construction-fold observations $Y_1, \dots, Y_N$ from that cell. Write $\rho = \rho^c_{x_j,k}$, and, for integers $b \ge 1$ and $0 \le z \le N$, define the arbitrary-center and medoid robust-center radii

$$\widehat{R}^{\mathrm{arb}}_N(b, z; \rho) := \min_{\substack{|\mathcal{C}|\le b,\\ O\subseteq[N],\ |O|\le z}} \max_{\substack{\mathcal{C}\subseteq\mathcal{Z}/\sim_\rho\\ i\in[N]\setminus O}} \rho(Y_i, \mathcal{C}), \tag{218}$$

$$\widehat{R}^{\mathrm{med}}_N(b, z; \rho) := \min_{\substack{|\mathcal{C}|\le b,\\ O\subseteq[N],\ |O|\le z}} \max_{\substack{\mathcal{C}\subseteq\{Y_1,\dots,Y_N\}\\ i\in[N]\setminus O}} \rho(Y_i, \mathcal{C}), \tag{219}$$

with $\max \varnothing = 0$. The set O contains observations omitted only from the covering diagnostic; the population matching analysis retains all observations.

For a proposed (A, D, s) and resolution ε, set

$$b_\varepsilon := \lfloor (A/\varepsilon)^D \rfloor, \qquad z_\varepsilon := \left\lfloor N\varepsilon^{s/(s-2)} \right\rfloor. \tag{220}$$

The condition $\widehat{R}^{\mathrm{arb}}_N(b_\varepsilon, z_\varepsilon; \rho) \le \varepsilon$ is exactly the empirical analogue of (216)–(217). The following result separates this ideal diagnostic from its finite-metric computation.

Proposition I.2 (Robust-center diagnostic and computational gap)**.** *For every* b, z,

$$\widehat{R}^{\mathrm{arb}}_N(b, z; \rho) \le \widehat{R}^{\mathrm{med}}_N(b, z; \rho) \le 2\widehat{R}^{\mathrm{arb}}_N(b, z; \rho). \tag{221}$$

Moreover, given an exact ρ*-distance oracle or a precomputed distance matrix, the robust* k*-center-with-outliers algorithm of Chakrabarty et al. (2020, Theorem 1.4), applied to the observed finite metric quotient, runs in time polynomial in the finite instance size, returns at most* b *centers, trims at most* z *points, and has radius* R_{alg} *satisfying*

$$\widehat{R}^{\mathrm{med}}_N(b, z; \rho) \le R_{\mathrm{alg}} \le 2\widehat{R}^{\mathrm{med}}_N(b, z; \rho) \le 4\widehat{R}^{\mathrm{arb}}_N(b, z; \rho). \tag{222}$$

Consequently, $R_{\mathrm{alg}} \le \varepsilon$ *is an explicit empirical HPQ witness, whereas* $R_{\mathrm{alg}} > 4\varepsilon$ *excludes an arbitrary-center empirical cover at radius* ε*. Intermediate values are inconclusive for that arbitrary-center question.*

Operationally, run this robust k-center approximation with budgets $(b_\varepsilon, z_\varepsilon)$ on the construction fold to obtain $\widehat{\mathcal{C}}_\varepsilon$. This first stage supplies only an empirical witness. To assess population rather than construction-sample coverage, evaluate the same codebook on an independent calibration fold. If that fold contains m observations from the same cell, then, conditional on the construction fold,

$$V_\varepsilon := \sum_{i=1}^{m} \mathbf{1}\left\{\rho(Y^{\mathrm{cal}}_i, \widehat{\mathcal{C}}_\varepsilon) > \varepsilon\right\} \sim \mathrm{Binomial}(m, q_\varepsilon), \quad q_\varepsilon := P_{a,x_j}\left\{\rho(Y, \widehat{\mathcal{C}}_\varepsilon) > \varepsilon\right\}. \tag{223}$$

Write $r_s = s/(s-2)$. The lower endpoint of a computationally meaningful resolution interval should be fixed before examining the calibration failures. One defensible design is

$$\varepsilon_{\min} := \max\{\varepsilon_{\mathrm{rate}}, \varepsilon_{\mathrm{stat}}, \varepsilon_{\mathrm{meas}}, \varepsilon_{\mathrm{comp}}\}, \tag{224}$$

where, for a prespecified upper bound $\overline{M}$ on the total number of simultaneous diagnostic queries and the smallest calibration-cell size $m_{\min}$,

$$\varepsilon_{\text{rate}} = cN_{\text{eff}}^{-1/s}, \qquad \varepsilon_{\text{stat}} = \left[1 - \left(\frac{\beta}{\overline{M}}\right)^{1/m_{\min}}\right]^{1/r_s}, \tag{225}$$

$$\varepsilon_{\text{comp}} = Ab_{\max}^{-1/D}. \tag{226}$$

Here N_{eff} is the smallest effective construction-cell size, $b_{\max}$ is the largest affordable center budget, and $\varepsilon_{\text{meas}}$ is a measurement-resolution floor (zero if none is needed). The term $\varepsilon_{\text{stat}}$ is an unavoidable best-case limit: below it, even zero calibration failures cannot meet a Bonferroni-corrected exact binomial upper bound. For large $m_{\min}$, it is approximately $\{\log(\overline{M}/\beta)/m_{\min}\}^{1/r_s}$. The rate term is a design scale; its constant c must be supplied by any theorem that uses a finite-resolution version of HPQ. If $\varepsilon_{\min} > \varepsilon_\star$, the available data and computation cannot run the proposed diagnostic on a nonempty interval.

It is unnecessary to check a continuum of resolutions. Fix $\lambda \in (0,1)$, let

$$L = \left\lceil \frac{\log(\varepsilon_\star/\varepsilon_{\min})}{\log(1/\lambda)} \right\rceil, \qquad \varepsilon_j = \min\{\varepsilon_{\min}\lambda^{-j}, \varepsilon_\star\}, \quad j = 0, \dots, L, \tag{227}$$

so that $\varepsilon_j \geq \lambda\varepsilon_{j+1}$. On the resulting prespecified finite query collection $\mathcal{J}_{\text{diag}}$, use one-sided exact binomial upper bounds U_q with error allocations $\sum_{q\in\mathcal{J}_{\text{diag}}} \beta_q \leq \beta$. At a grid resolution, the inequalities

$$|\widehat{\mathcal{C}}_{\varepsilon_j}| \leq b_{\varepsilon_j}, \qquad U_{\varepsilon_j} \leq \varepsilon_j^{r_s} \tag{228}$$

certify the population coverage of that learned codebook. The following deterministic interpolation makes these finitely many checks sufficient for the whole finite-resolution interval.

Proposition I.3 (Geometric-grid interpolation)**.** *Suppose that, for every grid point in (227), there is a codebook $\mathcal{C}_j$ such that*

$$|\mathcal{C}_j| \leq (A/\varepsilon_j)^D, \qquad P_{a,x}\{\rho(Y, \mathcal{C}_j) > \varepsilon_j\} \leq \varepsilon_j^{r_s}. \tag{229}$$

Then, for every $\varepsilon \in [\varepsilon_{\min}, \varepsilon_\star]$, there is a grid codebook satisfying

$$|\mathcal{C}| \leq \left(\frac{A/\lambda}{\varepsilon}\right)^D, \qquad P_{a,x}\{\rho(Y, \mathcal{C}) > \varepsilon\} \leq \varepsilon^{r_s}. \tag{230}$$

Consequently, if (228) holds simultaneously over the grid, then (230) holds over the entire interval with confidence at least $1-\beta$. To retain the constant A instead of A/λ, impose the tighter grid budget $|\mathcal{C}_j| \leq (\lambda A/\varepsilon_j)^D$.

Algorithm 2 summarizes the single diagnostic pipeline. The construction radius R_{alg} records whether the empirical geometry is witnessed, excluded, or in the approximation gray zone; the population certificate itself is determined by independent calibration of the returned codebook.

Algorithm 2 Finite-resolution HPQ diagnostic

Require: Construction and calibration folds by cell; candidate (A, D, s); $\varepsilon_\star, \lambda, \beta$; design limits in (224)

Compute $\varepsilon_{\min}$ and the grid $\mathcal{E} = \{\varepsilon_0, \dots, \varepsilon_L\}$
Prespecify all queries $\mathcal{J}_{\text{diag}}$ and allocations $\{\beta_q : q \in \mathcal{J}_{\text{diag}}\}$ with $\sum_q \beta_q \leq \beta$
for each action–covariate cell, active endpoint, and $\varepsilon \in \mathcal{E}$ **do**
    Set $b_\varepsilon = \lfloor (A/\varepsilon)^D \rfloor$ and $z_\varepsilon = \lfloor N\varepsilon^{r_s} \rfloor$
    Run robust k-center with outliers on the construction fold to obtain $(\widehat{\mathcal{C}}_\varepsilon, R_{\text{alg}})$
    Record: witness if $R_{\text{alg}} \leq \varepsilon$, exclusion if $R_{\text{alg}} > 4\varepsilon$, and gray zone otherwise
    Count calibration failures V_ε and compute the one-sided binomial upper bound U_ε at error β_q
    Mark the query certified if $|\widehat{\mathcal{C}}_\varepsilon| \leq b_\varepsilon$ and $U_\varepsilon \leq \varepsilon^{r_s}$
end for
if every grid query is certified **then**
    return a $(1-\beta)$-confidence finite-resolution certificate with constants $(A/\lambda, D, s)$ on $[\varepsilon_{\min}, \varepsilon_\star]$
else
    return the cell–endpoint–scale failures and construction statuses
end if

The number of robust-center calls per cell and endpoint is only

$$L+1 = O\left(1+\frac{\log(\varepsilon_\star/\varepsilon_{\min})}{\log(1/\lambda)}\right), \tag{231}$$

and a pairwise construction-distance matrix can be reused across scales. A smaller λ reduces the number of calls but enlarges the constant from A to A/λ. This procedure certifies all resolutions in the stated finite interval, but it neither verifies Assumption I.1 below $\varepsilon_{\min}$ nor proves the original all-scale theorem from data. Propositions I.2–I.3 and the simultaneous calibration claim are proved in Appendix I.2.

Policy-indexed learned-dual certificate. Split the observations into a training fold D_1 of size ℓ and an independent evaluation fold D_2 of size m. Using any conditional distribution estimator, fit $\widehat{P}_{a,x}$ on D_1. For every (a,b,x), choose a measurable plug-in dual optimizer

$$(\widehat{\phi}_{ab,x}, \widehat{\psi}_{ab,x}) \in \underset{\phi(y)+\psi(y')\le c_x(y,y')}{\arg\max} \left\{\widehat{P}_{a,x}\phi + \widehat{P}_{b,x}\psi\right\}, \tag{232}$$

with both potentials selected as c_x-transforms. Because the cost is uniformly bounded, the additive normalization may be chosen so that, for a deterministic $M<\infty$,

$$\sup_{a,b,x,y}\left(|\widehat{\phi}_{ab,x}(y)| \vee |\widehat{\psi}_{ab,x}(y)|\right) \le M. \tag{233}$$

The fitted collection is globally feasible for every action pair; neither first-stage accuracy nor any fallback used by the estimator affects this property.

For $(\pi,\pi')\in\Pi^2$, use the IPW score from (38), with the potentials in (232), and write

$$\widehat{T}_c(\pi,\pi') := \frac{1}{m}\sum_{i\in D_2} S_i(\pi,\pi'). \tag{234}$$

Let $r_{m,\alpha}$ be the policy-uniform safety radius in (45), with $B_S = 2M/\underline{e}$, and define

$$L^{\mathrm{hpq}}_{n,c,\alpha}(\pi,\pi') := \widehat{T}_c(\pi,\pi') - r_{m,\alpha}. \tag{235}$$

Thus the safety loss is exactly the one used in Section D; high-probability quantization does not introduce a second validity radius. For the explicit rate below, we additionally use the finite-covariate-support setting and notation of Section F, and take $\widehat{P}_{a,x_j}$ to be the training empirical law in cell $\{A=a, X=x_j\}$, with an arbitrary fixed fallback law for an empty cell. If $\Pi_{K_X} = \{(\pi(x_1),\dots,\pi(x_{K_X})) : \pi\in\Pi\}$ is enumerated, the finite-class radius $r^{\mathrm{fin}}_{m,\alpha}$ in (49), with N_{pair} there replaced by N_Π^2 and $N_\Pi = |\Pi_{K_X}|$, is also valid.

Theorem I.4 (Uniform validity and a polynomial upper rate under HPQ)**.** *Suppose the conditions of Theorem D.3 hold, the conditional plug-in dual optimizers in (232) exist and are selected measurably, and (233) holds. Then, for every finite (ℓ,m) and $\alpha\in(0,1)$,*

$$\inf_P \Pr_P\left\{\begin{array}{l} L^{\mathrm{hpq}}_{n,c,\alpha}(\pi,\pi') \le \Theta^L_{c,P}(\pi,\pi') \le \Theta_c(\pi,\pi';\eta), \\ \text{for every } (\pi,\pi')\in\Pi^2 \text{ and every admissible } \eta \end{array}\right\} \ge 1-\alpha. \tag{236}$$

The infimum in (236) ranges over all observed-data laws satisfying the setup and boundedness conditions; Assumption I.1 is not needed for validity.

If, in addition, Assumption I.1 and the finite-support conditions of Section F hold and the conditional laws are estimated by the cell empirical laws just described, there are finite constants C_1, C_2, depending only on $(A, D, s, \varepsilon_\star, K_X, d)$, such that both suprema below, taken over this structured finite-support class, satisfy

$$\begin{aligned}\sup_P \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2}\{\Theta^L_{c,P}(\pi,\pi') - L^{\mathrm{hpq}}_{n,c,\alpha}(\pi,\pi')\}_+\right] \\ \le C_1(\ell\kappa\underline{e}+1)^{-1/s} + C_2 B_S\sqrt{\frac{\log(N_\Pi\vee 2)+\log(2/\alpha)}{m}}.\end{aligned} \tag{237}$$

Consequently, if K_X, d are fixed and both folds are fixed positive fractions of n, then

$$\sup_P \mathbb{E}_P\left[\sup_{(\pi,\pi')\in\Pi^2}\{\Theta^L_{c,P}(\pi,\pi') - L^{\mathrm{hpq}}_{n,c,\alpha}(\pi,\pi')\}_+\right] = O_\alpha(n^{-1/s}). \tag{238}$$

Remark I.5 (Conditional finite-rank and manifold outcome models)**.** *Define*

$$\mathfrak{r}_d(N) := \begin{cases} N^{-1/2}, & d < 2, \\ N^{-1/2}\log(N+1), & d = 2, \\ N^{-1/d}, & d > 2. \end{cases} \tag{239}$$

Suppose first that, in every action–covariate cell,

$$Q_{Y(a)} = q_{a,j,Z} = \bar{q}_{a,j} + \sum_{k=1}^{d_{a,j}} Z_k \phi_{a,j,k}, \qquad X = x_j,$$

where Z ranges over a bounded subset of $\mathbb{R}^{d_{a,j}}$, and the maps into both active quotients are Lipschitz. A Euclidean net verifies Assumption I.1 with zero excess-distortion probability. Alternatively, suppose each P_{a,x_j} is supported on a compact regular $d_{a,j}$-dimensional manifold of distribution-valued outcomes whose inclusions into both active quotients are Lipschitz and have covering numbers of order $\varepsilon^{-d_{a,j}}$. In either model, if $d_\star = \max_{a,j} d_{a,j}$, a fixed-proportion split gives

$$\sup_P \mathbb{E}_P \left[\sup_{(\pi,\pi')\in\Pi^2} \{\Theta^L_{c,P}(\pi,\pi') - L^{\mathrm{hpq}}_{n,c,\alpha}(\pi,\pi')\}_+ \right] = O_\alpha\{\mathfrak{r}_{d_\star}(n)\}. \tag{240}$$

For a displacement cost $c_h(y,y') = \int_0^1 h\{Q_{y'}(u) - Q_y(u)\}\,du$, the required active-metric Lipschitz property holds whenever h is Lipschitz on the relevant displacement range. For every finite $d_\star$, this rate is $o\{(\log n)^{-(p\wedge 2)}\}$, and is therefore strictly faster than the unstructured rate in Theorem F.1 for the power costs. These are upper attainment statements; matching structured lower bounds require an additional cost-visible packing and lower-mass condition.

Remark I.6 (Trimmed finite-rank and manifold models)**.** *The preceding examples have zero excess-distortion probability. The trimming in Assumption I.1 also permits models for which every finite-resolution codebook leaves a small but nonzero amount of conditional probability uncovered. Write*

$$r_s := \frac{s}{s-2}.$$

For a first example, retain the conditional finite-rank representation above, but let its bounded coefficient law be a countable rare-cluster mixture

$$\mathcal{L}(Z_{a,j}) = \sum_{\ell\geq 1} w_{a,j,\ell} G_{a,j,\ell}.$$

Each $G_{a,j,\ell}$ may, for example, be a truncated Gaussian supported on a bounded coefficient region whose image in both active quotients has covering number at most $(B/\varepsilon)^{d_{a,j}}$, uniformly in (a,j,ℓ). Suppose the cluster weights satisfy

$$\sup_{a,j} \sum_{\ell>J} w_{a,j,\ell} \leq C_0 \exp(-c_0 J).$$

At resolution ε, cover only the first

$$J_\varepsilon := \lceil c_0^{-1} \log(C_0 \varepsilon^{-r_s}) \rceil$$

coefficient clusters. The uncovered cluster tail has probability at most ε^{r_s}, while the union codebook has size at most

$$J_\varepsilon (B/\varepsilon)^{d_{a,j}} \leq (A/\varepsilon)^{d_{a,j}+\delta}$$

for every fixed $\delta > 0$ and all sufficiently small ε. Thus, if $d_\star = \max_{a,j} d_{a,j}$ and $0 < \delta < s - d_\star$, Assumption I.1 holds with $D = d_\star + \delta$. Positive-weight omitted clusters can lie away from the retained codebook at every tested scale. This is a compact, distribution-valued analogue of the Gaussian-mixture trimming mechanism in Weed & Bach (2019, Proposition 14).

For a second example, let each conditional law be a countably stratified manifold mixture

$$P_{a,x_j} = \sum_{\ell\geq 1} w_{a,j,\ell} P_{a,j,\ell}.$$

Suppose each $P_{a,j,\ell}$ is supported on a compact regular manifold $\mathcal{M}_{a,j,\ell}$ whose covering numbers in both active quotients are bounded uniformly by $(B/\varepsilon)^{d_\star}$, while the rare-stratum weights satisfy

$$\sup_{a,j} \sum_{\ell > J} w_{a,j,\ell} \le C_0 \exp(-c_0 J), \qquad J \ge 1.$$

Take

$$J_\varepsilon := \left\lceil c_0^{-1} \log(C_0 \varepsilon^{-r_s}) \right\rceil$$

and form the union of codebooks for only the first J_ε strata. The omitted strata have total probability at most ε^{r_s}, whereas the union codebook has size

$$J_\varepsilon (B/\varepsilon)^{d_\star} \le (A/\varepsilon)^{d_\star + \delta}$$

for every fixed $\delta > 0$, all sufficiently small ε, and a suitable A. If, at arbitrarily fine resolutions, some positive-weight omitted strata lie farther than ε from the retained-stratum codebook, the omitted tail is genuine even though compactness still guarantees the existence of some finite full-support cover. Hence this model again verifies Assumption I.1 with $D = d_\star + \delta < s$, while explicitly allowing rare distribution-valued outcome regimes away from the learned codebook.

Unlike the zero-excess-distortion models in Remark I.5, these two examples invoke the general bound in Theorem I.4. With a fixed-proportion split their guaranteed attainment rate is $O_\alpha(n^{-1/s})$. If the union of all clusters or strata additionally has a polynomial full-support covering, the regular-support lemma may yield the sharper rate in (240); that extra conclusion is not implied by trimming alone.

The theorem separates validity from informativeness in the same way as Sections C–D. Simultaneous global feasibility and independent IPW evaluation give exact policy-uniform coverage, with the safety loss $r_{m,\alpha}$ determined by the policy-indexed score class. Conditional high-probability quantization controls only the training-sample dual gap and changes its unstructured logarithmic rate to a polynomial rate.

The result remains an upper theorem. High-probability quantization permits point masses and thin discrete laws and hence does not identify a minimax exponent. A fixed-width noisy tube around a low-dimensional set generally provides only a pre-asymptotic low-dimensional regime and need not satisfy Assumption I.1 at every vanishing scale. Appendix I proves Theorem I.4.

All Wasserstein distances below are computed on the relevant cost-active quotient space. Its diameter is normalized to be at most one.

I.1 High-probability quantization and empirical transport

For a law P on a metric space $(\mathcal{Y}, \rho)$, write

$$N_\varepsilon(P, \tau; \rho) := \inf_{S : P(S) \ge 1-\tau} N_\varepsilon(S; \rho).$$

Lemma I.7 (Quantization-to-transport bound)**.** *Suppose that, for constants $A \ge 1$, $D < s$, $s > 2$, and $\varepsilon_\star \in (0, 1]$,*

$$N_\varepsilon\left(P, \varepsilon^{s/(s-2)}; \rho\right) \le \left(\frac{A}{\varepsilon}\right)^D \qquad (0 < \varepsilon \le \varepsilon_\star).$$

If $\widehat{P}_N$ is the empirical law of N independent observations from P, then

$$\mathbb{E} W_1^\rho(\widehat{P}_N, P) \le C_{\mathrm{hpq}} N^{-1/s}, \tag{241}$$

where $C_{\mathrm{hpq}} < \infty$ depends only on $(A, D, s, \varepsilon_\star)$.

Proof. Set

$$\bar{\varepsilon} := \min\left\{\varepsilon_\star, A^{-D/(s-D)}\right\}.$$

For every $0 < \varepsilon \le \bar{\varepsilon}$,

$$\frac{\log N_\varepsilon\left(P, \varepsilon^{s/(s-2)}; \rho\right)}{\log(1/\varepsilon)} \le D + \frac{D \log A}{\log(1/\varepsilon)} \le s.$$

Proposition 5 of Weed & Bach (2019), specialized to $p = 1$, therefore gives

$$\mathbb{E} W_1^\rho(\widehat{P}_N, P) \le C_1(s) N^{-1/s} + C_2(s, \bar{\varepsilon}) N^{-1/2}.$$

Since $s > 2$ and $N \ge 1$, $N^{-1/2} \le N^{-1/s}$. Absorbing both terms proves (241). □

I.2 Robust-center diagnostic and calibration

The diagnostic definitions and statements are given above; we prove them here.

Proof of Proposition I.2. Because every observed outcome is an admissible arbitrary center, $\widehat{R}_N^{\mathrm{arb}} \le \widehat{R}_N^{\mathrm{med}}$. For the reverse comparison, fix $\eta > 0$ and an arbitrary-center solution with radius at most $\widehat{R}_N^{\mathrm{arb}} + \eta$. Assign each untrimmed observation to a nearest center and, from every nonempty assigned cluster, select one observed representative. There are at most b representatives and the same set of at most z observations remains trimmed. If Y_i and the selected representative $Y_{i'}$ are assigned to the same center c, the triangle inequality gives

$$\rho(Y_i, Y_{i'}) \le \rho(Y_i, c) + \rho(c, Y_{i'}) \le 2\{\widehat{R}_N^{\mathrm{arb}} + \eta\}.$$

Letting $\eta \downarrow 0$ proves (221).

Theorem 1.4 of Chakrabarty et al. (2020) gives a uni-criteria two-approximation for robust k-center with outliers on a finite metric space: it respects both the center and trimming budgets. Applying that result after identifying zero-pseudodistance observations gives the first two inequalities in (222); combining them with (221) gives the last. The witness and exclusion statements follow by feasibility and contraposition. □

Proof of Proposition I.3. Fix $\varepsilon \in [\varepsilon_{\min}, \varepsilon_\star]$. If $\varepsilon = \varepsilon_\star$, use $\mathcal{C}_L$. Otherwise choose $j < L$ such that $\varepsilon_j \le \varepsilon < \varepsilon_{j+1}$, and use $\mathcal{C} = \mathcal{C}_j$. Since $\varepsilon_j \ge \lambda \varepsilon_{j+1} > \lambda\varepsilon$,

$$|\mathcal{C}| \le \left(\frac{A}{\varepsilon_j}\right)^D \le \left(\frac{A/\lambda}{\varepsilon}\right)^D.$$

Moreover, monotonicity of the uncovered event and $r_s > 0$ give

$$P_{a,x}\{\rho(Y, \mathcal{C}) > \varepsilon\} \le P_{a,x}\{\rho(Y, \mathcal{C}_j) > \varepsilon_j\} \le \varepsilon_j^{r_s} \le \varepsilon^{r_s}.$$

This proves (230). If instead $|\mathcal{C}_j| \le (\lambda A/\varepsilon_j)^D$, the same calculation replaces A/λ by A.

On the simultaneous calibration event, every upper bound satisfies $q_{\varepsilon_j} \le U_{\varepsilon_j}$. Thus (228) implies (229) for all grid queries, and the deterministic argument above applies simultaneously. The calibration event has probability at least $1 - \beta$ by the argument below. □

We finally justify the simultaneous calibration statement. Conditional on the construction fold, each learned codebook is fixed, so (223) follows from independent sampling in its cell. Let $\mathcal{J}_{\mathrm{diag}}$ be the prespecified finite collection of diagnostic queries and let U_q be an exact one-sided binomial upper confidence bound with conditional error probability β_q, where $\sum_{q \in \mathcal{J}_{\mathrm{diag}}} \beta_q \le \beta$. A union bound, which does not require independence among the queries, yields

$$\Pr(q_q \le U_q \text{ for every } q \in \mathcal{J}_{\mathrm{diag}} \mid \text{construction fold}) \ge 1 - \beta.$$

Integrating over the construction fold preserves the bound. Hence every query satisfying (228) obeys its population tail inequality simultaneously with probability at least $1 - \beta$. The argument is restricted to the prespecified finite collection and therefore does not imply the all-scale population assumption.

Lemma I.8 (Empirical transport on a regular finite-dimensional support)**.** *Let P be supported on a subset S of a metric space $(\mathcal{Y}, \rho)$ with diameter at most one. Suppose that, for some $A \ge 1$ and $d < \infty$,*

$$N_\varepsilon(S; \rho) \le \left(\frac{A}{\varepsilon}\right)^d, \qquad 0 < \varepsilon \le 1. \tag{242}$$

Then the empirical law $\widehat{P}_N$ of N independent draws from P satisfies

$$\mathbb{E} W_1^\rho(\widehat{P}_N, P) \le C_{A,d}\, \mathfrak{r}_d(N), \tag{243}$$

where $\mathfrak{r}_d$ is defined in (239).

Proof. Dudley's multiscale partition bound gives, for every $\delta \in (0, 1]$,

$$\mathbb{E}W_1^\rho(\widehat{P}_N, P) \le C\left\{\delta + \frac{1}{\sqrt{N}}\int_\delta^1 \sqrt{N_\varepsilon(S;\rho)}\, d\varepsilon\right\}.$$

Under (242), the integral is bounded, up to a constant depending only on (A, d), by

$$\int_\delta^1 \varepsilon^{-d/2}\, d\varepsilon \lesssim_d \begin{cases} 1, & d < 2, \\ \log(1/\delta), & d = 2, \\ \delta^{1-d/2}, & d > 2. \end{cases}$$

For $d < 2$, take $\delta \downarrow 0$. For $d = 2$, take $\delta = N^{-1/2}$, and for $d > 2$, take $\delta = N^{-1/d}$. These choices give (243). □

Lemma I.9 (Random cell size)**.** *Let $N \sim \mathrm{Binomial}(\ell, q)$, where $q > 0$. Conditional on $N \ge 1$, let $\widehat{P}_N$ be the empirical law of the observations in that cell, and use an arbitrary fallback law when $N = 0$. If $W_1^\rho \le 1$ and the conclusion of Lemma I.7 holds conditionally on N, then*

$$\mathbb{E}W_1^\rho(\widehat{P}_N, P) \le C(\ell q + 1)^{-1/s}. \tag{244}$$

The same conclusion holds with $N^{-1/s}$ replaced by $\mathfrak{r}_d(N)$ under Lemma I.8.

Proof. If $\ell q \le 2$, the claim follows from $W_1^\rho \le 1$. Otherwise, a Chernoff bound gives $\Pr\{N < \ell q/2\} \le \exp(-\ell q/8)$. On the complementary event, $N^{-1/s} \le (\ell q/2)^{-1/s}$, while both the small-count event and the empty-cell fallback contribute at most an exponentially small term. This proves (244). Applying the same split to each branch of $\mathfrak{r}_d$ proves the final assertion. □

I.3 Conditional cost-active stability and the learned dual gap

Lemma I.10 (Conditional endpoint stability)**.** *Fix x. For any four outer laws P_1, P_2, Q_1, Q_2 on $\mathcal{Z}$,*

$$\left|\inf_{\gamma\in\Gamma(P_1,P_2)}\int c_x\, d\gamma - \inf_{\gamma\in\Gamma(Q_1,Q_2)}\int c_x\, d\gamma\right| \le W_1^{\rho^c_{x,1}}(P_1, Q_1) + W_1^{\rho^c_{x,2}}(P_2, Q_2). \tag{245}$$

Proof. Take arbitrary couplings between Q_1 and P_1, between P_1 and P_2, and between P_2 and Q_2. The gluing lemma realizes them jointly as $(\widetilde{Q}_1, \widetilde{P}_1, \widetilde{P}_2, \widetilde{Q}_2)$. By (215),

$$c_x(\widetilde{Q}_1, \widetilde{Q}_2) - c_x(\widetilde{P}_1, \widetilde{P}_2) \le \rho^c_{x,1}(\widetilde{Q}_1, \widetilde{P}_1) + \rho^c_{x,2}(\widetilde{Q}_2, \widetilde{P}_2).$$

Integrating and optimizing the three couplings gives one direction. Interchanging P and Q gives the other. □

Lemma I.11 (Population gap of a conditional empirical dual optimizer)**.** *For a cell (a, b, x_j), define*

$$\widetilde{T}_{c,x_j}(a, b) := P_{a,x_j}\widehat{\phi}_{ab,x_j} + P_{b,x_j}\widehat{\psi}_{ab,x_j}.$$

Then

$$\begin{aligned} 0 &\le T_{c,x_j}(a, b) - \widetilde{T}_{c,x_j}(a, b) \\ &\le 2W_1^{\rho^c_{x_j,1}}(\widehat{P}_{a,x_j}, P_{a,x_j}) + 2W_1^{\rho^c_{x_j,2}}(\widehat{P}_{b,x_j}, P_{b,x_j}). \end{aligned} \tag{246}$$

Moreover,

$$\begin{aligned} |\widehat{\phi}_{ab,x_j}(y) - \widehat{\phi}_{ab,x_j}(\widetilde{y})| &\le \rho^c_{x_j,1}(y, \widetilde{y}), \\ |\widehat{\psi}_{ab,x_j}(y) - \widehat{\psi}_{ab,x_j}(\widetilde{y})| &\le \rho^c_{x_j,2}(y, \widetilde{y}). \end{aligned} \tag{247}$$

Proof. Dual feasibility gives the first inequality. The c_{x_j}-transform property gives (247), and Kantorovich–Rubinstein duality therefore gives

$$|(\widehat{P}_{a,x_j} - P_{a,x_j})\widehat{\phi}_{ab,x_j}| \le W_1^{\rho^c_{x_j,1}}(\widehat{P}_{a,x_j}, P_{a,x_j}),$$
$$|(\widehat{P}_{b,x_j} - P_{b,x_j})\widehat{\psi}_{ab,x_j}| \le W_1^{\rho^c_{x_j,2}}(\widehat{P}_{b,x_j}, P_{b,x_j}).$$

Empirical optimality identifies the empirical dual value with $T_{c,x_j}(\widehat{P}_{a,x_j}, \widehat{P}_{b,x_j})$. Add and subtract this value. Lemma I.10 bounds the change in the primal endpoint by one copy of each empirical transport error, and the two displays above bound the change in the dual objective by a second copy. □

I.4 Proof of Theorem I.4

Proof. The conditional dual optimizers in (232) form an action-pair-indexed collection satisfying simultaneous global feasibility. The bound (233) gives the score envelope $B_S = 2M/\underline{e}$. Therefore Theorem D.3, applied with the radius $r_{m,\alpha}$ from (45), directly gives (236), simultaneously for every policy pair and every admissible population matching kernel η. This argument uses neither Assumption I.1 nor accuracy of the fitted conditional laws.

It remains to bound the excess length. Conditional on D_1, put

$$\widetilde{T}_c(\pi, \pi') := \sum_{j=1}^{K_X} p_j \widetilde{T}_{c,x_j}\{\pi(x_j), \pi'(x_j)\},$$
$$\mu_{\pi,\pi'} := \mathbb{E}\{S_i(\pi, \pi') \mid D_1\} = \widetilde{T}_c(\pi, \pi').$$

Lemma I.11 implies

$$\sup_{(\pi,\pi')\in\Pi^2} \{\Theta^L_c(\pi, \pi') - \widetilde{T}_c(\pi, \pi')\}$$
$$\le 2\sum_{j=1}^{K_X} p_j \sum_{a=1}^{d} \left\{ W_1^{\rho^c_{x_j,1}}(\widehat{P}_{a,x_j}, P_{a,x_j}) + W_1^{\rho^c_{x_j,2}}(\widehat{P}_{a,x_j}, P_{a,x_j}) \right\}. \tag{248}$$

The training count in cell (a, x_j) is binomial with success probability $p_j e_a(x_j) \ge \kappa\underline{e}$. Lemmas I.7 and I.9, followed by $\sum_j p_j = 1$, show that the expectation of (248) is at most

$$C_1(\ell\kappa\underline{e} + 1)^{-1/s}. \tag{249}$$

For the evaluation component, the definition of $L^{\mathrm{hpq}}_{n,c,\alpha}$ gives

$$\sup_{(\pi,\pi')\in\Pi^2} \{\Theta^L_c(\pi, \pi') - L^{\mathrm{hpq}}_{n,c,\alpha}(\pi, \pi')\}_+$$
$$\le \sup_{(\pi,\pi')\in\Pi^2} \{\Theta^L_c(\pi, \pi') - \widetilde{T}_c(\pi, \pi')\} + \sup_{(\pi,\pi')\in\Pi^2} \left\{ \mu_{\pi,\pi'} - \widehat{T}_c(\pi, \pi') \right\} + r_{m,\alpha}. \tag{250}$$

Conditional symmetrization bounds the expected middle term by $2\mathbb{E}(\widehat{\mathfrak{R}}_m \mid D_1)$. There are at most N^2_Π distinct evaluation score vectors, so Massart's finite-class lemma and the score envelope give

$$\widehat{\mathfrak{R}}_m \le B_S\sqrt{\frac{2\log(N^2_\Pi \vee 2)}{m}}.$$

Combining this display with

$$r_{m,\alpha} = 2\widehat{\mathfrak{R}}_m + 3B_S\sqrt{\frac{2\log(2/\alpha)}{m}}$$

and (249) proves (237), after enlarging C_2.

Under a fixed-proportion split, $\ell \asymp m \asymp n$. Because $s > 2$, the evaluation term of order $n^{-1/2}$ is no larger than the training term of order $n^{-1/s}$. This proves (238). □

I.5 Proof of Remark I.5

For the conditional finite-rank model, the assumed Lipschitz map from a bounded subset of $\mathbb{R}^{d_{a,j}}$ into each active quotient gives

$$N_\varepsilon\left(\operatorname{supp} P_{a,x_j};\rho^c_{x_j,k}\right)\le\left(\frac{A_{a,j,k}}{\varepsilon}\right)^{d_{a,j}},\qquad k\in\{1,2\}.$$

The same display follows from the assumed manifold covering bound and Lipschitz inclusion in the conditional manifold model. Hence Lemma I.8 and the random-cell argument give, uniformly over all cells and endpoints,

$$\mathbb{E}W_1^{\rho^c_{x_j,k}}(\widehat{P}_{a,x_j},P_{a,x_j})\le C\mathfrak{r}_{d_{a,j}}(\ell p_j e_a(x_j)).$$

Substitution into (248), together with the evaluation bound in (250), proves (240) under a fixed-proportion split. The active-metric assertion for a Lipschitz displacement cost follows by integrating its pointwise Lipschitz inequality. Finally, every branch of $\mathfrak{r}_d(n)$ decays faster than every negative power of $\log n$, which gives the comparison with Theorem F.1.

J Finite-Bag Identification and Certificates

Exact observation of an individual's distribution-valued outcome is an idealization. We now observe, for unit i, a bag

$$V_{ij}\mid(X_i,A_i,Y_i)\overset{\text{i.i.d.}}{\sim}Y_i,\qquad j=1,\dots,m_i,\qquad\widehat{Y}_i=\frac{1}{m_i}\sum_{j=1}^{m_i}\delta_{V_{ij}}.\tag{251}$$

The bags are conditionally independent across units. Their sizes are fixed by design, or are conditioned upon and do not depend on the measurement values after conditioning on (X_i,A_i,Y_i). This observation model has two distinct information budgets: the number of individuals controls outer-law sampling error, whereas m_i controls how accurately the latent measure Y_i is observed. Both budgets enter the error bounds below.

Proposition J.1 (A fixed bag size does not identify the outer law)**.** *Fix an integer $r\ge1$. There exist two distinct outer laws $P_+,P_-\in\mathcal{P}(\mathcal{P}(\{0,1\}))$ such that a bag of r conditionally i.i.d. measurements has exactly the same distribution under P_+ and P_-. Nevertheless, for every Wasserstein exponent $s\ge1$, there is a $q\in(0,1)$ for which*

$$\int W_s^s(y,\mathrm{Bern}(q))\,dP_+(y)\ne\int W_s^s(y,\mathrm{Bern}(q))\,dP_-(y).$$

Consequently, even with infinitely many individuals, a fixed number of within-individual measurements cannot identify all matching-robust contrasts with ground cost W_s^s.

A fixed bag identifies only finitely many mixture moments, even with infinitely many individuals. Asymptotic recovery can instead be based on a measurement model, repeated-measurement structure, or a growing within-unit budget. We pursue the last route because it leaves the latent outer law nonparametric.

Let $N=|D_2|$ and $D_I=I_+-I_-$. For $\beta\in(0,1)$, define the observable radii

$$\rho_i(\beta)=D_I\min\left\{1,\sqrt{\frac{\log(2N/\beta)}{2m_i}}\right\},\qquad i\in D_2.\tag{252}$$

Lemma J.2 (Conditional DKW control of every evaluation bag)**.** *Under (251),*

$$\mathbb{P}\Big\{W_1(\widehat{Y}_i,Y_i)\le\rho_i(\beta)\text{ for every }i\in D_2\;\Big|\;(X_i,A_i,Y_i,m_i)_{i\in D_2},\mathcal{F}_1\Big\}\ge1-\beta.\tag{253}$$

The proof combines the sharp Dvoretzky–Kiefer–Wolfowitz inequality (Massart, 1990) with the one-dimensional identity $W_1(P,Q)=\int_I|F_P-F_Q|$. No smoothness or density assumption is imposed on an individual's latent measure.

For the measurement-aware certificate, suppose the training fold returns the action-pair-indexed collection in Section D. In addition to simultaneous global feasibility and the envelope bound (42), suppose that, conditionally on $\mathcal{F}_1$,

$$|\widehat{\phi}_{ab,x}(y) - \widehat{\phi}_{ab,x}(\widetilde{y})| \le \lambda^{\phi}_{ab,x} W_1(y,\widetilde{y}), \qquad |\widehat{\psi}_{ab,x}(y) - \widehat{\psi}_{ab,x}(\widetilde{y})| \le \lambda^{\psi}_{ab,x} W_1(y,\widetilde{y}), \tag{254}$$

where the nonnegative Lipschitz bounds are measurable and uniformly bounded by a deterministic $\Lambda < \infty$.

For $a_i = \pi(X_i)$ and $b_i = \pi'(X_i)$, define

$$\widehat{S}^{\mathrm{bag}}_i(\pi,\pi') = \frac{\mathbb{1}\{A_i = a_i\}}{e_{a_i}(X_i)} \widehat{\phi}_{a_i b_i, X_i}(\widehat{Y}_i) + \frac{\mathbb{1}\{A_i = b_i\}}{e_{b_i}(X_i)} \widehat{\psi}_{a_i b_i, X_i}(\widehat{Y}_i), \tag{255}$$

$$\Delta_i(\pi,\pi';\beta) = \rho_i(\beta) \left\{ \frac{\mathbb{1}\{A_i = a_i\}}{e_{a_i}(X_i)} \lambda^{\phi}_{a_i b_i, X_i} + \frac{\mathbb{1}\{A_i = b_i\}}{e_{b_i}(X_i)} \lambda^{\psi}_{a_i b_i, X_i} \right\}, \tag{256}$$

$$\overline{\Delta}_N(\pi,\pi';\beta) = \frac{1}{N} \sum_{i \in D_2} \Delta_i(\pi,\pi';\beta). \tag{257}$$

Let

$$\overline{r}^{\mathrm{ent}}_{N,\alpha} := r^{\mathrm{ent}}_{N,\alpha/2}, \qquad \widehat{L}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') := \frac{1}{N} \sum_{i \in D_2} \widehat{S}^{\mathrm{bag}}_i(\pi,\pi') - \overline{\Delta}_N(\pi,\pi';\beta) - \overline{r}^{\mathrm{ent}}_{N,\alpha}. \tag{258}$$

The use of $\alpha/2$ supplies two-sided outer empirical control, which is needed for the tightness statement below. If Π is finite, the analogous finite-class radius $r^{\mathrm{fin}}_{N,\alpha/2}$ may replace $\overline{r}^{\mathrm{ent}}_{N,\alpha}$.

Theorem J.3 (Measurement- and policy-uniform certificate)**.** *Suppose Assumptions A.1–A.4 and D.1 hold, the propensities are known, $\mathcal{K}_{\mathrm{H}}(\Pi) < \infty$, and (251)–(254) hold. Suppose also that the envelope condition (42) holds. Assume the same measurability conditions as in Theorem D.3. Then, for every $\alpha, \beta \in (0,1)$,*

$$\mathbb{P}\left\{ \begin{array}{l} \widehat{L}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') \le \Theta^L_c(\pi,\pi') \le \Theta_c(\pi,\pi';\eta), \\ \text{for all } (\pi,\pi') \in \Pi^2 \text{ and every admissible } \eta \end{array} \right\} \ge 1 - \alpha - \beta. \tag{259}$$

The event therefore covers a policy pair selected after inspecting the evaluation bags. Neither validity nor the correction requires the first-stage outer-law estimator to be consistent.

Validity is only half of the question. Define the conditional dual suboptimality

$$\mathrm{Gap}(\widehat{\nu};\pi,\pi') := \Theta^L_c(\pi,\pi') - \mathbb{E}\{S^{\circ}_i(\pi,\pi') \mid \mathcal{F}_1\} \ge 0, \tag{260}$$

where S°_i is the latent score obtained from (255) by replacing $\widehat{Y}_i$ with Y_i.

Theorem J.4 (Finite-sample informativeness decomposition)**.** *Under the conditions of Theorem J.3, with probability at least $1 - \alpha - \beta$, simultaneously over $(\pi,\pi') \in \Pi^2$,*

$$0 \le \Theta^L_c(\pi,\pi') - \widehat{L}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') \le \mathrm{Gap}(\widehat{\nu};\pi,\pi') + 2\overline{r}^{\mathrm{ent}}_{N,\alpha} + 2\overline{\Delta}_N(\pi,\pi';\beta). \tag{261}$$

The bound separates three sources of error: dual approximation, outer-sample protection, and within-individual measurement. Global feasibility removes the first term from the validity argument, but cannot remove it from informativeness.

Theorem J.5 (Measurement-aware upper certificate)**.** *Suppose, in addition to the observation and policy conditions above, that the training fold returns a globally superfeasible collection $(\widehat{\phi}^U_{ab,x}, \widehat{\psi}^U_{ab,x})$ satisfying (50), the same absolute envelope condition, and the Lipschitz bounds in (254) with constants $\lambda^{U,\phi}_{ab,x}, \lambda^{U,\psi}_{ab,x} \le \Lambda_U$. Define $\widehat{S}^{U,\mathrm{bag}}_i$ and Δ^U_i by (255)–(256) with the upper potentials and upper Lipschitz constants, let $\overline{\Delta}^U_N = N^{-1} \sum_{i \in D_2} \Delta^U_i$, and set*

$$\widehat{U}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') = \frac{1}{N} \sum_{i \in D_2} \widehat{S}^{U,\mathrm{bag}}_i(\pi,\pi') + \overline{\Delta}^U_N(\pi,\pi';\beta) + \overline{r}^{U,\mathrm{ent}}_{N,\alpha}, \tag{262}$$

where $\overline{r}^{U,\mathrm{ent}}_{N,\alpha}$ is the upper-score analogue of $\overline{r}^{\mathrm{ent}}_{N,\alpha}$. Then, with probability at least $1 - \alpha - \beta$, simultaneously over Π^2,

$$\Theta^U_c(\pi,\pi') \le \widehat{U}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi'). \tag{263}$$

If $S_i^{U,\circ}$ denotes the corresponding latent score and

$$\mathrm{Gap}^U(\widehat{\nu}^U;\pi,\pi') := \mathbb{E}\{S_i^{U,\circ}(\pi,\pi') \mid \mathcal{F}_1\} - \Theta_c^U(\pi,\pi') \geq 0, \tag{264}$$

then on the same event

$$0 \leq \widehat{U}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') - \Theta_c^U(\pi,\pi') \leq \mathrm{Gap}^U(\widehat{\nu}^U;\pi,\pi') + 2\overline{r}^{U,\mathrm{ent}}_{N,\alpha} + 2\overline{\Delta}^U_N(\pi,\pi';\beta). \tag{265}$$

Consequently, if the upper dual gap is uniformly at most a_n^U, the upper excess length has the same two-budget rate as (266), with a_n and Λ replaced by a_n^U and Λ_U. Combining the lower and upper bag certificates, each with $(\alpha/2,\beta/2)$, gives a simultaneous $(1-\alpha-\beta)$ confidence region for the full sharp interval, uniformly over policy pairs selected after inspecting the evaluation bags.

Corollary J.6 (An honest two-budget rate)**.** *Consider a sequence with $N \asymp n$, conditions* (i)–(iv) *of Assumption D.2, $m_{\min,n} = \min_{i\in D_2} m_i$, and $\sup_{a,b,x}\{\lambda^{\phi}_{ab,x} \vee \lambda^{\psi}_{ab,x}\} \leq \Lambda$. If $\sup_{\Pi_n^2} \mathrm{Gap}(\widehat{\nu};\pi,\pi') \leq a_n$ almost surely for a deterministic sequence a_n, then on an event of probability at least $1-\alpha_n-\beta_n$,*

$$\sup_{\Pi_n^2}\{\Theta_c^L - \widehat{L}^{\mathrm{bag}}_{\alpha_n,\beta_n}\} = O\left(a_n + \frac{1+\sqrt{\log(1/\alpha_n)}}{\sqrt{n}} + \sqrt{\frac{\log(n/\beta_n)}{m_{\min,n}}}\right). \tag{266}$$

For fixed confidence levels, the displayed sufficient rate becomes $a_n + O\{n^{-1/2} + \sqrt{(\log n)/m_{\min,n}}\}$. Hence the measurement term vanishes under $m_{\min,n}/\log n \to \infty$, while a fixed bag size cannot be repaired by increasing the number of individuals alone. Under this fully simultaneous DKW correction, a sufficient condition for making the measurement term no larger than the outer $n^{-1/2}$ term is $m_{\min,n} \gtrsim n\log n$. The sharper average-size bound (271) can improve this sufficient condition for unbalanced bags. Neither bound is a measurement-level minimax claim.

Remark J.7 (The training-gap assumption is separate from coverage)**.** *The deterministic bound a_n in Corollary J.6 is an additional assumption on the trained potentials. Neither global feasibility nor the bag correction proves that it tends to zero. The finite-sample decomposition in Theorem J.4 retains the random dual gap and does not require this assumption. If instead a training procedure supplies*

$$\mathbb{P}\left\{\sup_{\Pi_n^2} \mathrm{Gap}(\widehat{\nu};\pi,\pi') \leq a_n\right\} \geq 1-\gamma_n,$$

a union bound gives the same excess-length rate on an event of probability at least $1-\alpha_n-\beta_n-\gamma_n$; coverage itself remains at least $1-\alpha_n-\beta_n$. Theorem J.8 establishes such a bound for the finite-cell empirical trainer, including finite training bags, binning, solver residual, and anchor error. Corollary J.9 supplies a concrete version of the two-budget result. With unrestricted latent outer laws, the fixed-bag nonidentification result also precludes a uniform vanishing training gap from fixed-size training bags alone.

For the principal cost $c(y,y') = W_s^s(y,y')$ on a bounded interval, $c(\,\cdot\,,y')$ and $c(y,\,\cdot\,)$ are sD_I^{s-1}-Lipschitz in W_1. Exact c-transforms inherit this constant, so (254) is compatible with the globally feasible implementation.

J.1 Proof of Proposition J.1

Proof. For $p\in[0,1]$, write $\nu_p = (1-p)\delta_0 + p\delta_1$. Let L_{r+1} denote the Legendre polynomial of degree $r+1$, normalized by $L_{r+1}(1)=1$, and set

$$h(p) = L_{r+1}(2p-1), \qquad p\in[0,1].$$

The shifted polynomial h is orthogonal in $L^2[0,1]$ to every polynomial of degree at most r, satisfies $\int_0^1 h(p)\,dp = 0$, and obeys $|h(p)|\leq 1$. For any $\varepsilon\in(0,1)$, the functions

$$f_+(p) = 1+\varepsilon h(p), \qquad f_-(p) = 1-\varepsilon h(p)$$

are therefore distinct probability densities. Let F_+ and F_- be their laws and define the outer laws $P_+ = (p\mapsto\nu_p)_{\#}F_+$ and $P_- = (p\mapsto\nu_p)_{\#}F_-$.

Conditional on p, a bag of size r from ν_p is determined by its number K of ones. For every $k = 0, \dots, r$,

$$\mathbb{P}_+(K = k) - \mathbb{P}_-(K = k) = 2\varepsilon \binom{r}{k} \int_0^1 p^k (1-p)^{r-k} h(p)\, dp = 0,$$

because the integrand multiplying h is a polynomial of degree r. Thus the complete ordered bag, and hence its empirical measure, has the same distribution under P_+ and P_-.

It remains to exhibit a Wasserstein contrast that separates them. Define

$$H(q) = \int_0^1 |p - q| h(p)\, dp, \qquad q \in (0, 1).$$

The function H cannot vanish identically. Otherwise its second distributional derivative on $(0, 1)$ would be both zero and $2h$, which would imply $h = 0$ almost everywhere, a contradiction. Choose q with $H(q) \neq 0$. For every $s \geq 1$, the only mass transported between ν_p and ν_q crosses the unit distance, so

$$W_s^s(\nu_p, \nu_q) = |p - q|.$$

Consequently,

$$\int W_s^s(y, \nu_q)\, dP_+(y) - \int W_s^s(y, \nu_q)\, dP_-(y)$$
$$= 2\varepsilon \int_0^1 |p - q| h(p)\, dp = 2\varepsilon H(q) \neq 0.$$

Take a randomized two-action experiment with constant X, outer law P_+ or P_- under the first action, and the degenerate outer law δ_{ν_q} under the second. The observed treatment and bag law is identical in the two data-generating processes. Because the second outer marginal is degenerate, its coupling with the first is unique and the sharp lower and upper endpoints both equal the displayed integral. Their unequal values prove nonidentification of the contrast. □

J.2 Proof of Lemma J.2

Proof. Fix the conditioning variables in (253). Write F_i and $\widehat{F}_i$ for the distribution functions of Y_i and $\widehat{Y}_i$. On the real line,

$$W_1(\widehat{Y}_i, Y_i) = \int_I |\widehat{F}_i(t) - F_i(t)|\, dt \leq D_I \|\widehat{F}_i - F_i\|_\infty.$$

If the untruncated square-root term in (252) is at most one, the Dvoretzky–Kiefer–Wolfowitz inequality gives

$$\mathbb{P}\{W_1(\widehat{Y}_i, Y_i) > \rho_i(\beta) \mid Y_i, m_i\} \leq 2 \exp\left\{-2m_i \frac{\rho_i(\beta)^2}{D_I^2}\right\} = \frac{\beta}{N}.$$

If that term exceeds one, then $\rho_i(\beta) = D_I$ and the probability is zero because the W_1-diameter of $\mathcal{P}(I)$ is D_I. A union bound over $i \in D_2$ proves the result. The argument is conditional and therefore also permits heterogeneous latent measures and bag sizes. □

J.3 Proofs of Theorems J.3 and J.4

Proof of Theorem J.3. Condition on $\mathcal{F}_1$. For $a_i = \pi(X_i)$ and $b_i = \pi'(X_i)$, define the unobserved latent score

$$S_i^\circ(\pi, \pi') = \frac{\mathbb{1}\{A_i = a_i\}}{e_{a_i}(X_i)} \widehat{\phi}_{a_i b_i, X_i}(Y_i) + \frac{\mathbb{1}\{A_i = b_i\}}{e_{b_i}(X_i)} \widehat{\psi}_{a_i b_i, X_i}(Y_i).$$

The conditional IPW calculation in the proof of Theorem D.3 gives

$$\mu_{\pi,\pi'} := \mathbb{E}\{S_i^\circ(\pi, \pi') \mid \mathcal{F}_1\}$$
$$= \mathbb{E}_X \left[P_{\pi(X),X} \widehat{\phi}_{\pi(X)\pi'(X),X} + P_{\pi'(X),X} \widehat{\psi}_{\pi(X)\pi'(X),X} \,\middle|\, \mathcal{F}_1 \right] \leq \Theta_c^L(\pi, \pi'). \tag{267}$$

The inequality follows from global feasibility and conditional weak duality for every action pair. It therefore holds simultaneously over Π^2, and $\Theta_c^L(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)$ for every admissible η.

Apply the entropy-to-Rademacher argument in the proof of Theorem D.3 to the latent score class and also to its negative. A union bound over the two one-sided deviations, each at level $\alpha/2$, yields an event $\mathcal{E}_{\mathrm{out}}$ with conditional probability at least $1-\alpha$ on which

$$\sup_{(\pi,\pi')\in\Pi^2}\left|\frac{1}{N}\sum_{i\in D_2} S_i^{\circ}(\pi,\pi') - \mu_{\pi,\pi'}\right| \le \overline{r}_{N,\alpha}^{\mathrm{ent}}. \tag{268}$$

The same argument with a union bound over the finite score class proves the stated finite-class alternative.

Let $\mathcal{E}_{\mathrm{bag}}$ be the event in Lemma J.2. On this event, (254) and the triangle inequality give, simultaneously for every unit and policy pair,

$$\left|\widehat{S}_i^{\mathrm{bag}}(\pi,\pi') - S_i^{\circ}(\pi,\pi')\right| \le \Delta_i(\pi,\pi';\beta). \tag{269}$$

When $a_i = b_i = A_i$, both indicators equal one, so the triangle inequality uses the sum of the two Lipschitz bounds in (256). When $a_i \neq b_i$, at most one indicator is nonzero. Thus the displayed bound holds for every action pair, without requiring same-action potentials to be zero. Hence, on $\mathcal{E}_{\mathrm{out}} \cap \mathcal{E}_{\mathrm{bag}}$,

$$\begin{aligned}\widehat{L}_{\alpha,\beta}^{\mathrm{bag}}(\pi,\pi') &\le \frac{1}{N}\sum_{i\in D_2} S_i^{\circ}(\pi,\pi') - \overline{r}_{N,\alpha}^{\mathrm{ent}} \\ &\le \mu_{\pi,\pi'} \le \Theta_c^L(\pi,\pi') \le \Theta_c(\pi,\pi';\eta)\end{aligned}$$

simultaneously over the asserted indices. Lemma J.2, (268), and a union bound give probability at least $1-\alpha-\beta$. □

Proof of Theorem J.4. Work on the intersection of the two events constructed in the preceding proof. For each policy pair, add and subtract the latent empirical score and its conditional mean to obtain the exact decomposition

$$\begin{aligned}&\Theta_c^L(\pi,\pi') - \widehat{L}_{\alpha,\beta}^{\mathrm{bag}}(\pi,\pi') \\ &\quad= \mathrm{Gap}(\widehat{\nu};\pi,\pi') + \left\{\mu_{\pi,\pi'} - \frac{1}{N}\sum_{i\in D_2} S_i^{\circ}(\pi,\pi')\right\} \\ &\qquad+ \frac{1}{N}\sum_{i\in D_2}\{S_i^{\circ}(\pi,\pi') - \widehat{S}_i^{\mathrm{bag}}(\pi,\pi')\} + \overline{\Delta}_N(\pi,\pi';\beta) + \overline{r}_{N,\alpha}^{\mathrm{ent}}.\end{aligned}$$

The second term is at most $\overline{r}_{N,\alpha}^{\mathrm{ent}}$ by (268), and the third is at most $\overline{\Delta}_N(\pi,\pi';\beta)$ by (269). The upper bound in (261) follows. Its lower bound is precisely the validity statement already proved. □

Proof of Theorem J.5. Condition on $\mathcal{F}_1$ and let $S_i^{U,\circ}$ be the latent upper score. The conditional IPW identity and global superfeasibility give, simultaneously over Π^2,

$$\mu_{\pi,\pi'}^U := \mathbb{E}\{S_i^{U,\circ}(\pi,\pi') \mid \mathcal{F}_1\} \ge \Theta_c^U(\pi,\pi'). \tag{270}$$

Apply the two-sided empirical-process argument in (268) to the upper latent score class. With conditional probability at least $1-\alpha$,

$$\sup_{\Pi^2}\left|\frac{1}{N}\sum_{i\in D_2} S_i^{U,\circ}(\pi,\pi') - \mu_{\pi,\pi'}^U\right| \le \overline{r}_{N,\alpha}^{U,\mathrm{ent}}.$$

On the DKW event of Lemma J.2, the upper Lipschitz bounds give

$$\left|\widehat{S}_i^{U,\mathrm{bag}}(\pi,\pi') - S_i^{U,\circ}(\pi,\pi')\right| \le \Delta_i^U(\pi,\pi';\beta)$$

simultaneously over units and policy pairs. Therefore, on the intersection of these events,

$$\begin{aligned}\widehat{U}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') &\geq \frac{1}{N}\sum_{i\in D_2} S_i^{U,\circ}(\pi,\pi') + \overline{r}^{U,\mathrm{ent}}_{N,\alpha} \\ &\geq \mu^U_{\pi,\pi'} \geq \Theta^U_c(\pi,\pi'),\end{aligned}$$

which proves (263) with probability at least $1-\alpha-\beta$.

For informativeness, add and subtract the latent empirical score and its conditional mean:

$$\begin{aligned}&\widehat{U}^{\mathrm{bag}}_{\alpha,\beta}(\pi,\pi') - \Theta^U_c(\pi,\pi') \\ &\quad= \mathrm{Gap}^U(\widehat{\nu}^U;\pi,\pi') + \left\{\frac{1}{N}\sum_{i\in D_2} S_i^{U,\circ}(\pi,\pi') - \mu^U_{\pi,\pi'}\right\} \\ &\qquad+ \frac{1}{N}\sum_{i\in D_2}\{\widehat{S}_i^{U,\mathrm{bag}}(\pi,\pi') - S_i^{U,\circ}(\pi,\pi')\} + \overline{\Delta}^U_N(\pi,\pi';\beta) + \overline{r}^{U,\mathrm{ent}}_{N,\alpha}.\end{aligned}$$

The empirical-process and bag-comparison events bound the second and third terms by one radius and one measurement correction, respectively. This proves (265). The rate follows from the same overlap and minimum-bag-size calculation used in the proof of Corollary J.6. Finally apply the lower and upper results with $(\alpha/2,\beta/2)$ and take a union bound to obtain the two-sided statement. □

J.4 Proof of Corollary J.6

Proof. For any action pair, the sum of the two treatment indicators is at most two. Overlap and the uniform Lipschitz bound therefore imply

$$\begin{aligned}\sup_{(\pi,\pi')\in\Pi_n^2}\overline{\Delta}_N(\pi,\pi';\beta_n) &\leq \frac{2\Lambda}{\underline{e}_n N}\sum_{i\in D_2}\rho_i(\beta_n) \\ &\leq \frac{2\Lambda D_I}{\underline{e}_n}\sqrt{\frac{\log(2N/\beta_n)}{2m_{\min,n}}}.\end{aligned}$$

The minimum-bag-size bound is only a convenient simplification. Keeping the individual sizes gives the sharper, directly computable bound

$$\sup_{\Pi_n^2}\overline{\Delta}_N \leq \frac{2\Lambda D_I}{\underline{e}_n N}\sum_{i\in D_2}\min\left\{1,\sqrt{\frac{\log(2N/\beta_n)}{2m_i}}\right\}. \tag{271}$$

The implemented correction retains the individual radii and treatment indicators; a single small bag therefore need not determine its value. Conditions (i)–(iv) of Assumption D.2, with α_n left explicit, give

$$\overline{r}^{\mathrm{ent}}_{N,\alpha_n} = O\left\{n^{-1/2} + \sqrt{\frac{\log(1/\alpha_n)}{n}}\right\}.$$

Substitute the last two displays and the assumed bound on Gap into Theorem J.4; use $N \asymp n$ to obtain (266). The final claims follow by fixing α_n, β_n and inspecting the measurement term. □

J.5 Lipschitz regularity of Wasserstein transforms

For completeness, let $s \geq 1$ and $y, \widetilde{y}, y' \in \mathcal{P}(I)$. The one-dimensional quantile representation and the mean-value inequality give

$$\begin{aligned}|W_s^s(y,y') - W_s^s(\widetilde{y},y')| &\leq sD_I^{s-1}\int_0^1 |Q_y(t) - Q_{\widetilde{y}}(t)|\,dt \\ &= sD_I^{s-1}W_1(y,\widetilde{y}).\end{aligned}$$

The symmetric argument applies in the second coordinate. If, for example, $\phi(y) = \inf_{y'}\{W_s^s(y,y') - \psi(y')\}$, then ϕ is an infimum of functions sharing this Lipschitz constant and hence is itself sD_I^{s-1}-Lipschitz. The same reasoning applies to the other exact c-transform.

J.6 A CONCRETE FINITE-BAG TRAINER AND A VANISHING TRAINING GAP

We instantiate the random training-gap premise in Remark J.7. Use the finite covariate support of the minimax and HPQ results, with $H = dK_X$ action–covariate cells and $P(X = x, A = a) \geq q_\star := \kappa \underline{e} > 0$. There are ℓ independent training individuals. Within each nonempty cell use the empirical law of the bag quantiles, uniform individual weights, and the finite dual trainer of Appendix E.5; in an empty cell use the zero potential pair. Bag lengths are fixed by design, or satisfy the conditional sampling requirements in (251). In particular, conditioning on the action–covariate memberships leaves the latent outcomes i.i.d. within each cell; the proof does not condition on outcome-dependent lengths to claim this i.i.d. property. Assume a deterministic design lower bound $m_i \geq b_{\mathrm{tr}}$ on all training bag lengths. All auxiliary anchors and solver choices are functions of training data only.

Theorem J.8 (High-probability dual gap from finite training bags)**.** *Fix $\gamma \in (0, 1)$, and suppose*

$$\ell q_\star \geq 8 \log(4H/\gamma), \qquad t = \lfloor \ell q_\star / 2 \rfloor \geq 1.$$

For the completed-anchor trainer with common radius η, bins K, and feasible primal–dual residual at most $\varepsilon_{\mathrm{opt}}$, there is a constant C depending only on the fixed model parameters such that

$$\mathbb{P}\left\{ \sup_{\Pi^2} \mathrm{Gap}(\widehat{\nu}; \pi, \pi') \leq a^{\mathrm{tr}}_{\ell,\gamma} \right\} \geq 1 - \gamma, \tag{272}$$

where a deterministic choice is

$$a^{\mathrm{tr}}_{\ell,\gamma} = C \Bigg\{ R(t) + \sqrt{\frac{\log(8H/\gamma)}{t}} + K^{-1} + \eta + \varepsilon_{\mathrm{opt}} + D_I \min\left(1, \sqrt{\frac{\log(8\ell/\gamma)}{2b_{\mathrm{tr}}}}\right) \Bigg\}. \tag{273}$$

Here R is the logarithmic power-cost rate or the HPQ rate specified in Corollary E.5; HPQ is imposed on the latent *conditional laws, not on the noisy empirical bag laws.*

The original trainer using only source training observations as anchors also satisfies (272), with (273) replaced by

$$a^{\mathrm{orig}}_{\ell,\gamma} = C \Bigg\{ R(t) + \sqrt{\frac{\log(8H/\gamma)}{t}} + \frac{1}{\log(t+1)} + \frac{\log(8H/\gamma)}{t} + K^{-1} + \varepsilon_{\mathrm{opt}} + D_I \min\left(1, \sqrt{\frac{\log(8\ell/\gamma)}{2b_{\mathrm{tr}}}}\right) \Bigg\}. \tag{274}$$

The latter is a sufficient consistency bound, not an optimal rate claim for the uncompleted extension.

Proof. The proof uses four events and allocates failure probability $\gamma/4$ to each (the fourth is unnecessary with deterministic completion). First, the binomial lower-tail bound and a union bound give

$$\mathbb{P}\{\min_{a,x} n_{a,x} < t\} \leq H \exp(-\ell q_\star / 8) \leq \gamma/4.$$

Second, conditional on the cell memberships, consider the two latent empirical processes defining $U_{a,x}$ in Appendix E.5. Their conditional expectations are at most $CR(n_{a,x})$, by Lemma F.12 for power costs, or by Lemma I.7 and active-metric Lipschitz continuity for HPQ. Replacing one latent observation changes either supremum, or their maximum, by at most $2L_h D_I / n_{a,x}$. The bounded-differences inequality and a union bound thus imply, except with probability $\gamma/4$,

$$\max_{a,x} U_{a,x} \leq C \left\{ R(t) + \sqrt{\frac{\log(8H/\gamma)}{t}} \right\} \tag{275}$$

on the first event. This step is valid even though the potentials are subsequently fitted using the same bags: the supremum is over a fixed population function class.

Third, apply the conditional DKW argument of Lemma J.2 to the ℓ training bags. With probability at least $1-\gamma/4$,

$$\max_{i\in D_1}\|\widetilde{q}_i - q_i\|_1 \le E_{\mathrm{tr}} := D_I \min\left(1, \sqrt{\frac{\log(8\ell/\gamma)}{2b_{\mathrm{tr}}}}\right). \tag{276}$$

Thus $e_{a,x} \le E_{\mathrm{tr}}$ in every cell. Applying Theorem E.4 on these events, then taking the largest ordered action-pair gap in each cell and averaging with P_X, proves (273). No union bound over policies is needed for the training gap.

For completeness, the fourth event bounds nearest-anchor error for the original implementation. For a deterministic L^1-net of the full monotone cone of radius u, choose measurable nearest-center cells, each of diameter at most $2u$, and let their number be $J(u)$. Equation (160) gives $J(u) \le \exp(CD_I/u)$ for $0 < u \le D_I$, after changing the constant. For v latent observations in one action–covariate cell, let Z be the population mass of the partition cells containing no observation. For any union of partition cells with mass exceeding z, the probability that it contains no observation is at most e^{-vz}. There are at most $2^{J(u)}$ such unions. Hence

$$\mathbb{P}\left\{Z > \frac{J(u)\log 2 + \log(4H/\gamma)}{v}\right\} \le \frac{\gamma}{4H}.$$

If the cell containing a fresh quantile is occupied, a training quantile lies within $2u$ of it; otherwise the distance is at most D_I. The map B_K contracts L^1, and on (276) replacing a latent anchor by its bag version increases any nearest-anchor distance by at most E_{tr}. Consequently, simultaneously in all cells, with the additional failure probability at most $\gamma/4$,

$$R_{a,x}(\mathcal{U}) \le 2u + D_I\frac{J(u)\log 2 + \log(4H/\gamma)}{t} + E_{\mathrm{tr}}.$$

For sufficiently large t choose $u = C'D_I/\log(t+1)$ with C' large enough that $J(u) \le \sqrt{t}$; the remaining finite range of t is absorbed in the constant using $R_{a,x} \le D_I$. This yields $R_{a,x} \le C\{1/\log(t+1) + \log(8H/\gamma)/t + E_{\mathrm{tr}}\}$. Substitution into (98) proves (274) and completes the union bound. □

Corollary J.9 (A fully specified two-budget certificate)**.** *Combine the preceding trainer with the independent evaluation-bag certificate (258), using $M = L_h D_I$ and $\Lambda = L_h$. Under the assumptions of Theorems J.8 and J.3, with probability at least $1-\alpha-\beta-\gamma$, simultaneously over all policy pairs,*

$$0 \le \Theta^L_{c_h} - \widehat{L}^{\mathrm{bag}}_{\alpha,\beta} \le a^{\mathrm{tr}}_{\ell,\gamma} + 2\overline{r}^{\mathrm{ent}}_{N,\alpha} + \frac{4L_hD_I}{\underline{e}}\min\left(1, \sqrt{\frac{\log(2N/\beta)}{2b_{\mathrm{ev}}}}\right), \tag{277}$$

where b_{ev} is a deterministic lower bound on evaluation bag lengths. One may substitute $a^{\mathrm{orig}}_{\ell,\gamma}$ for the original trainer. Coverage alone still has probability at least $1-\alpha-\beta$; γ is only the price of the additional informativeness guarantee.

Proof. Intersect the training event (272) with the event in Theorem J.4, and use (271). A union bound gives the stated probability without requiring independence of these events. □

Explicit consistency and rate-preserving bag lengths. Suppose $\ell \asymp N \asymp n$ and the cell probabilities and overlap are bounded below. For example, take $\gamma_n = n^{-2}$. If $K_n \to \infty$, $\varepsilon_{\mathrm{opt},n} \to 0$, and $b_{\mathrm{tr},n}/\log n \to \infty$, then $a^{\mathrm{orig}}_{\ell,\gamma_n} \to 0$; the completed trainer also requires $\eta_n \to 0$. This proves a vanishing training gap for a concrete finite-bag learner, without an almost-sure deterministic premise on its realized gap. A vanishing evaluation measurement term additionally needs $b_{\mathrm{ev},n}/\log(n/\beta_n) \to \infty$ under the simultaneous DKW correction used here.

More precisely, for a target rate r_n of either $\{\log(n+1)\}^{-(p\wedge 2)}$ or $n^{-1/s}$ $(s > 2)$, sufficient conditions for the completed trainer and its certificate to preserve that rate at the displayed confidence levels are

$$K_n^{-1} + \eta_n + \varepsilon_{\mathrm{opt},n} = O(r_n), \qquad b_{\mathrm{tr},n} \gtrsim r_n^{-2}\log(n/\gamma_n),$$

$$b_{\mathrm{ev},n} \gtrsim r_n^{-2} \log(n/\beta_n), \quad \frac{1+\sqrt{\log(1/\alpha_n)}+\sqrt{\log(1/\gamma_n)}}{\sqrt{n}} = O(r_n). \tag{278}$$

For fixed α, β and $\gamma_n = n^{-2}$ these sufficient bag budgets are $(\log n)^{2(p\wedge 2)+1}$ in the unstructured power-cost case and $n^{2/s}\log n$ in the HPQ case, for both folds. They establish high-probability upper rates; they are not necessary bag lengths or measurement-level minimax lower bounds. In particular, the finite-bag result does not convert HPQ alone into a minimax claim.

K Experimental Designs, Numerical Results, and Reproducibility

The experiments are organized by the question they address. Each study reports its design, comparison, and result together; the common synthetic outcome universe is introduced before studies that reuse it.

K.1 Coverage and calibration

When do globally feasible certificates remain valid after policy search, and how much width does simultaneous calibration add?

K.1.1 Synthetic outer-law design and simultaneous coverage

Let $t_j = (j-1/2)/24$, $j = 1, \dots, 24$. The declared outcome universe consists of the 49 monotone vectors

$$q_{\ell s}(t_j) = \left[\ell + s t_j^{0.72+1.15(s-0.12)/0.20}\right]_{[0,1]},$$

where ℓ ranges over seven equally spaced values in $[0.01, 0.66]$ and s over seven equally spaced values in $[0.12, 0.32]$. These vectors are quantile functions of non-Dirac probability measures. For type r, let $v_r = (\bar{q}_r, \max q_r - \min q_r)$. Conditional outer weights are

$$w_{xar} \propto \exp\left[-\frac{1}{2}\left\{\left(\frac{v_{r1}-u_{xa,1}}{0.115}\right)^2 + \left(\frac{v_{r2}-u_{xa,2}}{0.070}\right)^2\right\}\right] + 2.5\times 10^{-4},$$

with centers

	$a=1$	$a=2$	$a=3$
$x=1$	$(0.13, 0.15)$	$(0.44, 0.23)$	$(0.78, 0.16)$
$x=2$	$(0.31, 0.28)$	$(0.14, 0.18)$	$(0.76, 0.29)$.

We draw X uniformly from the two cells, A uniformly from the three actions, and the observed type from $w_{X,A}$. The cost is the mean squared distance between 24-bin quantile vectors, which is exact for their step-function encodings.

On the training half, the safe method retains up to

$$J_n = \min\{28, \max(10, \mathrm{round}(7+\sqrt{n}/3.5))\}$$

highest-frequency anchors in each action–cell stratum, solves the finite dual, and applies (77)–(78). The pooled-X method fits one outer law per action. The unsafe method retains only four high-frequency anchors and extends each potential by its nearest anchor. Every reported violation is computed over all 49^2 ordered pairs, including points unseen in training.

For each pair, the uniform correction uses its exact conditional score range over the declared finite universe:

$$r_{m,\alpha}(\pi,\pi') = R_{\pi,\pi'}\sqrt{\frac{\log(|\Pi|^2/\alpha)}{2m}}.$$

This is the range-adaptive finite-class Hoeffding union bound. The pointwise baseline uses $1.644854\widehat{\sigma}/\sqrt{m}$ and is evaluated after search only to expose the invalid inferential shortcut.

Table 6 reports empirical coverage separately from constraint violations. The globally feasible methods cover in every repetition, consistent with Theorem D.3. Conversely, favorable coverage of an infeasible method at a particular n cannot repair its positive constraint violation. The pointwise experiment directly diagnoses multiplicity because it uses the same safe first-stage potentials and differs only in its critical value.

Table 6: **Complete policy-search simulation summary.** Coverage and the mean information ratio use 400 repetitions. The All-pairs column means simultaneous coverage of all 81 ordered policy pairs. Violation is the maximum over all repetitions and all declared outcome pairs.

n	Procedure	Selected	All pairs	Mean LCB/endpoint	Max violation
2048	Global feasible	1.000	1.000	0.299	6.9×10^{-17}
	Pooled X, feasible	1.000	1.000	0.328	6.9×10^{-17}
	Pointwise after search	0.895	0.693	0.939	6.9×10^{-17}
	Sampled constraints	0.995	0.995	0.584	0.458
4096	Global feasible	1.000	1.000	0.526	8.2×10^{-17}
	Pooled X, feasible	1.000	1.000	0.531	8.2×10^{-17}
	Pointwise after search	0.915	0.698	0.958	8.2×10^{-17}
	Sampled constraints	0.963	0.953	0.776	0.458

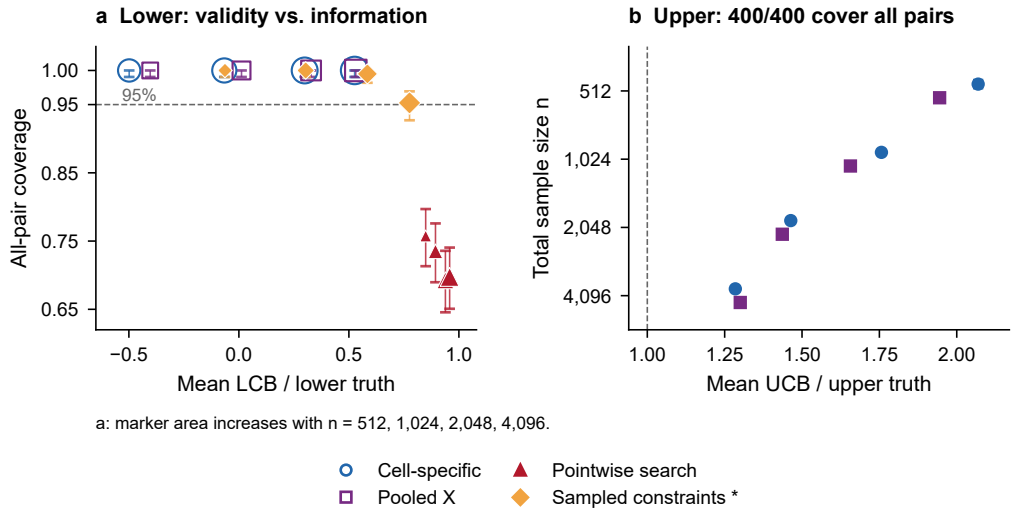


Figure 4: **Validity and information are distinct.** (a) Mean fixed-pair LCB/truth versus simultaneous coverage of all 81 pairs; vertical bars are 95% Wilson intervals from 400 repetitions. Marker area increases with sample size. The sampled-constraint method (*) has maximum violation 0.458, despite sometimes favorable coverage. (b) Upper mean UCB/truth with ± 1.96 Monte Carlo standard errors; both fits cover all pairs in 400/400 repetitions at every size. The full grid is shown; numerical results are in Tables 6 and 20.

Table 7: **Synthetic validity and informativeness.** Each row uses 400 repetitions. Selected and All pairs are coverage rates; All pairs checks all 81 ordered policy pairs simultaneously. Ratio is the mean fixed-pair LCB/truth for lower rows and UCB/truth for upper rows. Max violation checks all 49^2 declared outcome pairs; values below 10^{-15} are displayed as 0. "Ours, pooled X" changes only the first-stage fit.

Endpoint	n	Procedure	Selected	All pairs	Ratio	Max violation
Lower	2048	Ours, cell-specific	1.000	1.000	0.299	0
		Ours, pooled X	1.000	1.000	0.328	0
		Pointwise after search	0.895	0.693	0.939	0
		Sampled constraints	0.995	0.995	0.584	0.458
Lower	4096	Ours, cell-specific	1.000	1.000	0.526	0
		Ours, pooled X	1.000	1.000	0.531	0
		Pointwise after search	0.915	0.698	0.958	0
		Sampled constraints	0.963	0.953	0.776	0.458
Upper	2048	Ours, cell-specific	1.000	1.000	1.464	0
		Ours, pooled X	1.000	1.000	1.436	0
Upper	4096	Ours, cell-specific	1.000	1.000	1.284	0
		Ours, pooled X	1.000	1.000	1.300	0

K.1.2 SIMULTANEOUS VALIDITY AFTER POLICY SEARCH

To isolate the guarantee purchased by the full-class radius, fix a reference policy π_0 that takes the baseline action in each of J cells. Every nonempty subset $S \subseteq [J]$ defines a candidate π_S, so the analyst may inspect $2^J - 1$ comparisons with the same reference. The held-out contribution in cell j is the mean of m independent Rademacher scores in $[-1, 1]$. Hence $\Delta^\star(\pi_S, \pi_0) = 0$ for every S, and a positive LCB is a false certificate. The total evaluation budget is fixed at approximately 2,000 and allocated evenly across cells.

For one fixed subset, the pointwise Hoeffding radius is

$$r_{\text{pair}}(S) = \frac{2}{J}\sqrt{\frac{|S|\log(1/\alpha)}{2m}}.$$

It is valid for that prespecified comparison. Full-class calibration instead protects every cell component at level α/J and uses

$$r_{\text{class}}(S) = \frac{2|S|}{J}\sqrt{\frac{\log(J/\alpha)}{2m}}.$$

On the resulting common event, the LCB is valid for every S simultaneously, including a policy chosen after inspecting all reported bounds. Table 8 reports the fixed-pair and simultaneous coverage rates as the candidate class grows.

Table 8: **Marginal validity does not survive unrestricted policy search.** Results use 100,000 repetitions per row and one-sided $\alpha = 0.05$. "Fixed pair" is marginal coverage of one comparison; "Pointwise all" and "Full-class all" require simultaneous coverage of every candidate. "False select." is the false positive rate after selecting the largest pointwise LCB. Wilson interval endpoints are stored in the reproducibility record.

J	Candidates	Fixed pair	Pointwise all	Full-class all	False select.
4	15	0.992	0.947	0.995	0.053
6	63	0.992	0.897	0.994	0.103
8	255	0.992	0.827	0.995	0.173
10	1,023	0.993	0.744	0.996	0.256
12	4,095	0.992	0.659	0.995	0.341
16	65,535	0.993	0.483	0.995	0.517
20	1,048,575	0.993	0.306	0.996	0.694

At the largest class, full calibration produces a false selected certificate in only 0.004 of repetitions, compared with 0.694 for pointwise search. The pointwise procedure is not intrinsically invalid: its fixed-pair coverage remains about 0.993. Its failure occurs only when a collection of marginal statements is reinterpreted as one post-selection guarantee. The full-class method uses a larger radius precisely to make that stronger interpretation valid.

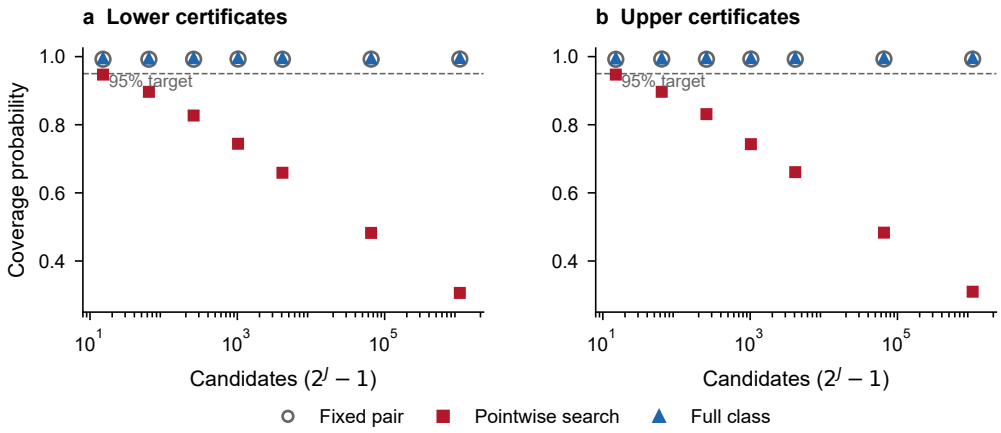


Figure 5: **Marginal coverage does not protect policy search.** Both endpoints are evaluated at all seven recorded class sizes, with 100,000 repetitions per size. Bars are 95% Wilson intervals, often smaller than the markers. The fixed-pair series measures marginal coverage; the other two series require all candidates to be covered simultaneously. The dashed line marks 0.95. Full-class calibration remains conservative while separately calibrated pointwise bounds fail after search.

Table 9: **Class-wide calibration protects data-dependent policy selection.** Coverage is over 100,000 repetitions at one-sided $\alpha = 0.05$. "Pointwise search" applies a valid fixed-pair radius to every candidate and then selects after seeing the bounds.

J	Candidates	One fixed pair	Pointwise search	Full class
4	15	0.992	0.947	0.995
8	255	0.992	0.827	0.995
12	4,095	0.992	0.659	0.995
20	1,048,575	0.993	0.306	0.996

K.1.3 Empirical Rademacher calibration over an infinite threshold class

This experiment evaluates the empirical Rademacher radius in Theorem 3.1 over an infinite policy class. It separates the cost of uniform calibration from the question of whether a prespecified contrast remains informative. Let

$$X \sim \mathrm{Unif}[0,1], \qquad \Pi = \{\pi_t : t \in [0,1]\}, \qquad \pi_t(x) = \mathbb{1}\{x \geq t\}.$$

Distinct thresholds disagree on a set of probability $|t - s| > 0$, so Π remains infinite even after identifying policies equal almost surely. Its uniform empirical Hamming covering numbers grow polynomially in inverse radius, giving $\mathcal{K}_{\mathrm{H}}(\Pi) < \infty$. The $m + 1$ action patterns on a sample are used to compute the empirical supremum over this class; they are not a prespecified finite policy grid.

Design and exact endpoints. Independently of X, draw independent $A \sim \mathrm{Bernoulli}(1/2)$ and V, with V taking values 0.38 and 0.42 with equal probability. The observed outcome is a nondegenerate distribution with quantile

$$Q_Y(u) = T + h(2u - 1), \qquad T = \tfrac{1}{2} + (2A - 1)V, \qquad h = 0.01.$$

Thus the two outer laws place equal mass on individual distributions centered at $0.08, 0.12$ and $0.88, 0.92$, respectively. The cost is W_2^2 on $\mathcal{P}([0,1])$, with trivial upper bound one. Monotone and countermonotone matching give, for every real threshold pair,

$$\Theta^L(\pi_t, \pi_s) = 0.64|t - s|, \qquad \Theta^U(\pi_t, \pi_s) = 0.0016 + 0.64|t - s|. \tag{279}$$

In particular, an equal-policy comparison has a nonzero sharp upper endpoint. We prespecify the interior pair $(t, s) = (0.1, 0.9)$, with endpoints $(0.5120, 0.5136)$, and the

whole-population pair $(0, 1)$, with endpoints $(0.6400, 0.6416)$. Total sample sizes are $n \in \{10{,}000, 20{,}000, 50{,}000, 100{,}000, 200{,}000\}$, with a 20:80 training–evaluation split, so $m = 0.8n$. Each size uses 500 independent repetitions.

Globally feasible potentials and IPW scores. The training fold estimates the difference of the two arm means of T and clips it to $[0.6, 0.95]$, giving d. An empty arm uses 0.4 for its mean of V; no training arm was empty in the reported repetitions. Put $M_L = d - d^2/2$ and define

$$f_d(z) = M_L + \begin{cases} -2dz, & z \le 1 - d, \\ z^2 - 2z + (1-d)^2, & z > 1 - d. \end{cases}$$

The bounded quadratic c-transform pair satisfies $f_d(z) + f_d(1 - z') \le (z - z')^2$ and $|f_d| \le M_L$ on $[0, 1]$. Integrating these functions along the quantiles gives globally feasible lower potentials for action pair $(0, 1)$; the reversed pair swaps them, and equal actions use zero lower potentials. For every action pair, use upper potentials

$$U(Y) = V^U(Y) = 2 \int_0^1 \{Q_Y(u) - 1/2\}^2 \, du.$$

The inequality $(z - z')^2 \le 2(z - 1/2)^2 + 2(z' - 1/2)^2$ proves global superfeasibility, and both upper potentials are bounded by $M_U = 1/2$. Consequently, the theorem's global score envelopes are $B_L = 4M_L$ and $B_U = 2$. Evaluation uses the iid IPW mean with known propensity $1/2$.

On the generated support, write the cross-action scores as

$$l_i = 4dV_i - d^2, \qquad u_i = 4(V_i^2 + h^2/3).$$

For $t \le s$, the lower score is $l_i \mathbb{1}\{t \le X_i < s\}$. The upper score equals $2u_i \mathbb{1}\{A_i = 0\}$ below t, u_i between t and s, and $2u_i \mathbb{1}\{A_i = 1\}$ at or above s. These expressions retain the equal-action upper contributions.

Computing and certifying the empirical RC. Order the evaluation observations by X, and define signed prefix sums $P_L(k) = \sum_{i \le k} \varepsilon_i l_i$ and $P_U(k) = \sum_{i \le k} \varepsilon_i u_i$, including $P_L(0) = P_U(0) = 0$. The conditional empirical complexities over all ordered threshold pairs obey the exact identities

$$\begin{aligned} \widehat{\mathfrak{R}}_L &= \frac{1}{m} \mathbb{E}_\varepsilon \max_{0 \le a \le b \le m} \{P_L(b) - P_L(a)\}, \\ \widehat{\mathfrak{R}}_U &= \frac{1}{m} \mathbb{E}_\varepsilon \{ \max_{0 \le k \le m} P_U(k) - \min_{0 \le k \le m} P_U(k) \}. \end{aligned} \tag{280}$$

The lower identity follows from the interval of policy disagreement. For the upper identity, the signed supremum is a common mean-zero term plus twice a maximum signed prefix sum; sign symmetry removes the observed action signs and converts its expectation to the expected range. The negative upper-score class has the same RC. Running prefix extrema compute each supremum in $O(m)$ operations per sign draw.

We approximate each expectation with $K = 256$ independent sign sequences. For endpoint $E \in \{L, U\}$, let $w_i^L = l_i$, $w_i^U = u_i$, and let $R_{E,\mathrm{MC}}$ denote the resulting average. Flipping one sign changes the normalized prefix statistic by at most $2|w_i^E|/m$. Bounded differences therefore supplies the conditional upper correction

$$\eta_E = \sqrt{\frac{2 \sum_i (w_i^E)^2 \log(1/\delta)}{K m^2}}, \qquad \delta = 0.0005.$$

We use $R_{E,\mathrm{safe}} = \min\{R_{E,\mathrm{MC}} + \eta_E, 4\sqrt{\sum_i (w_i^E)^2}/m\}$, where the second expression is a deterministic martingale maximal-inequality bound. With $\alpha_* = 0.0245$, the implemented radius is

$$r_E = 2R_{E,\mathrm{safe}} + 3B_E \sqrt{2 \log(2/\alpha_*)/m}. \tag{281}$$

The two endpoints together have failure probability at most $2\alpha_* + 2\delta = 0.05$, including numerical RC integration. This uses the empirical RC route in the theorem, rather than its Hamming-entropy upper bound.

Table 10: **Empirical RC calibration for an infinite threshold class.** Each row uses 500 repetitions and joint nominal coverage 0.95. Entries are mean raw bounds without truncation. Uniform coverage checks both endpoints for every real threshold pair. The true endpoints are $(0.5120, 0.5136)$ for $(0.1, 0.9)$ and $(0.6400, 0.6416)$ for $(0, 1)$.

		Interior pair $(0.1, 0.9)$		Whole pair $(0, 1)$		Uniform
n	m	LCB	UCB	LCB	UCB	coverage
10,000	8,000	0.2995	0.8671	0.4275	0.8672	1.000
20,000	16,000	0.3617	0.8012	0.4897	0.8012	1.000
50,000	40,000	0.4170	0.7425	0.5449	0.7426	1.000
100,000	80,000	0.4448	0.7131	0.5728	0.7130	1.000
200,000	160,000	0.4645	0.6922	0.5925	0.6922	1.000

Table 11: **Decomposition of the empirical RC radius.** Entries average 500 repetitions. The $2R_{E,\mathrm{safe}}$ columns include the numerical integration correction; C_E denotes the concentration term in (281). The same radius protects all pairs.

n	$2R_{L,\mathrm{safe}}$	C_L	$2R_{U,\mathrm{safe}}$	C_U
10,000	0.02145	0.19109	0.02640	0.19905
20,000	0.01516	0.13512	0.01870	0.14075
50,000	0.00960	0.08546	0.01183	0.08902
100,000	0.00680	0.06043	0.00838	0.06294
200,000	0.00480	0.04273	0.00592	0.04451

Coverage over the continuous class. Population endpoints in (279) vary between adjacent observations even when empirical action patterns are unchanged. We therefore maximize the coverage violation over both one-sided limits at each ordered observation and the endpoints $0, 1$. Between observations the target is linear in each threshold; ordered prefix extrema consequently give the supremum over all real threshold pairs, including limiting values. Small-sample exhaustive enumeration verifies both this coverage calculation and (280). Global potential inequalities and the analytic endpoints are additionally checked against quantile evaluations and a finite transport linear program.

Figure 6 separates the radius terms and compares the three calibrations.

Coverage, information, and the source of conservativeness. Table 10 reports uniform joint coverage of $500/500$ at each size; the corresponding two-sided 95% Wilson interval is approximately $[0.9924, 1]$. Across all 2,500 repetitions, the interior pair has a minimum LCB of 0.2905 and maximum UCB of 0.8762. Both prespecified pairs always have positive LCBs and UCBs below the trivial cost bound one. These observations are compatible with the uniform guarantee and indicate conservative calibration; they do not establish coverage exactly at the nominal level.

The mean uncorrected RC estimates multiplied by $\sqrt{m}$ range from 0.8023 to 0.8049 for the lower scores and from 1.0233 to 1.0276 for the upper scores, consistent with $m^{-1/2}$ decay. Table 11 shows that the concentration term dominates the total radius. At $n = 10{,}000$, the RC lower and upper radii are 0.21254 and 0.22544. Using the same global envelopes, fixed-pair Hoeffding calibration at error 0.025 per endpoint gives 0.05831 and 0.06074, respectively, but only guarantees a prespecified pair. Replacing the empirical RC by its deterministic martingale upper bound gives 0.24862 and 0.25673, with a full-class guarantee. A direct finite-class union bound cannot treat the sample-dependent $(m+1)^2$ action-pattern pairs as a fixed population policy set; other infinite-class bounds require their own complexity arguments.

The upper potential introduces a separate approximation limit. Its population value is 0.64173333 for every threshold pair, so for $\Delta = |t - s|$ its gap above the sharp upper endpoint is $0.64(1-\Delta)+4h^2/3$. The gap is 0.12813333 for the interior pair and only 0.00013333 for the whole-population pair. Thus the interior UCB does not converge to 0.5136 merely by increasing n. Uniform validity and the informativeness of these prespecified comparisons do not imply tight bounds for every policy pair. This experiment isolates calibration with a learned lower translation and a fixed upper potential; it is not an end-to-end benchmark of conditional transport learning.

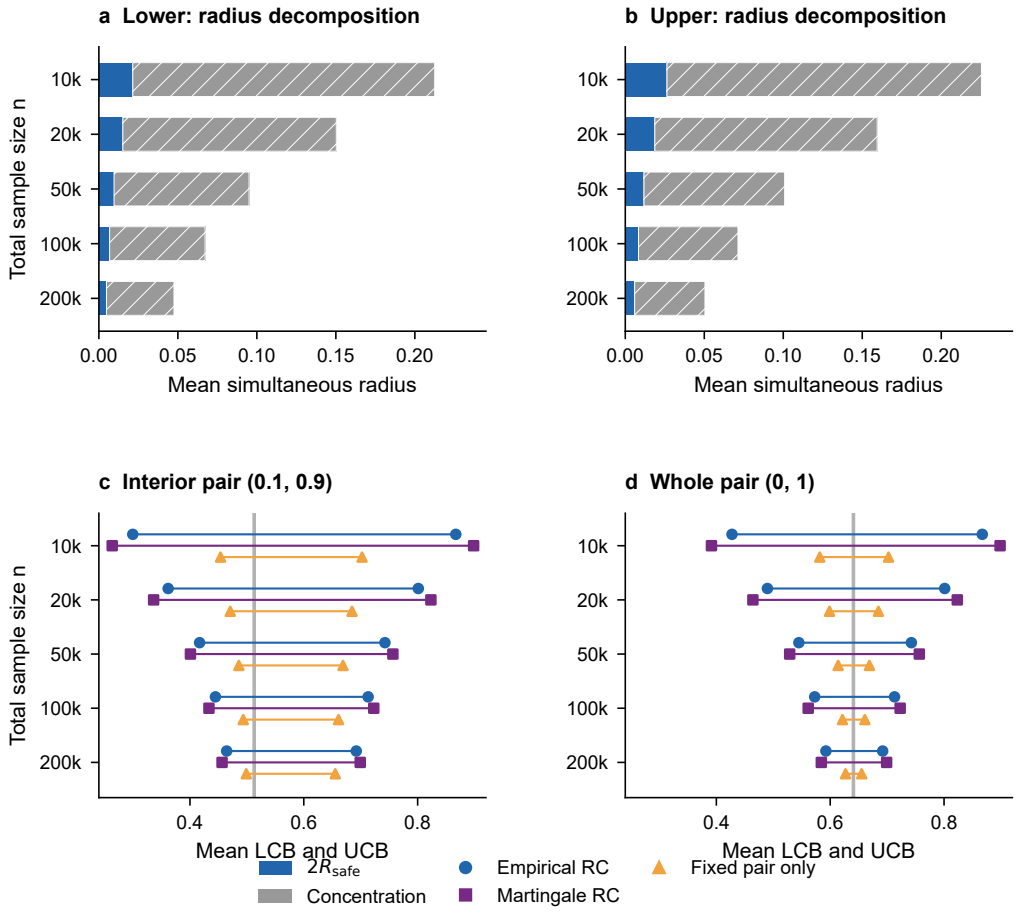


Figure 6: **Concentration dominates the empirical RC radius.** (a,b) Mean radius decomposition over 500 repetitions per size; the $2R_{\text{safe}}$ term includes the numerical integration correction. (c,d) Each segment connects the mean lower and upper bounds for a prespecified pair. Gray bands mark the analytic sharp intervals. Empirical RC and its deterministic martingale alternative protect the infinite class; fixed-pair calibration protects only the prespecified pair. All methods use the same observations and the 20:80 split.

Reproducibility. The implementation is `reproducibility/rc_threshold_study.py`, with shared dual checks in `rc_dual_checks.py`. The base seed is 2026092519; repetition $r \in \{0, \dots, 499\}$ at size n uses seed $2026092519 + 10000n + r$, with separate training, evaluation, and sign streams. The directory `reproducibility/results/rc_threshold/` contains every repetition, summaries, the numerical and continuum-supremum checks, and the execution manifest, which records the software environment and command-line parameters.

K.1.4 The relative cost of simultaneous honesty on growing smooth codes

Corollary F.4 compares minimax orders: for a fixed finite design the calibration term is lower order than the logarithmic learning rate. We examine how this separation emerges with sample size, using the training-conditional range calibration of Appendix E.9. We choose the absolute displacement loss $h(z) = |z|$, giving $c_h = W_1$, as a representative cost for this mechanism experiment. Its matching contact and interaction orders are both one, so the predicted learning scale is $(\log N)^{-1}$. On the growing smooth code family below, both the pairwise costs and population transport truth are exactly computable. This lets us measure learning and calibration errors separately without a quantile-discretization approximation. The construction retains growing outer-law complexity despite uniformly bi-Lipschitz individual quantiles. We compare radii on identical learned scores, holding the population, training procedure, and fold allocation fixed to isolate the cost of simultaneous certification.

Growing populations and fixed sample allocation. Let $k \in \{16, 32, 64, 128, 256, 512, 1024\}$ and set $\ell = m = 4k$, so the total sample size is $N = 8k$. There are two equiprobable covariate cells, two actions with known propensity $1/2$, and all four deterministic policies, giving 16 ordered policy pairs. Each fold consists of iid individuals; multinomial counts implement these draws exactly without fixing cell/action counts. Training and evaluation are independent and have equal budgets throughout.

For each k, generate four independent constant-weight binary codebooks of size $M = 8k$ and length $J = 16 \log_2 k$, with minimum Hamming distance at least $J/4$. Conditional on either covariate cell, arm zero is uniform on the first half of the codebook and arm one on its complementary half. The two cells share the population codebook but use their own training observations and fitted duals. Map each codeword to signs $u \in \{-1, 1\}^J$ and define, for $t \in [j/J, (j+1)/J]$,

$$h_u(t) = u_{j+1} \min\{t - j/J, (j+1)/J - t\}, \qquad q_u(t) = \frac{1}{4} + \frac{t}{4} + \frac{1}{16} h_u(t).$$

All quantile slopes lie in $[3/16, 5/16]$. Exact integration gives

$$W_1(q_u, q_v) = a_k \frac{d_{\mathrm{Ham}}(u, v)}{J}, \qquad a_k = \frac{1}{32J}. \tag{282}$$

Population assignment computes the sharp conditional endpoint on the two uniform half-codebooks. It is used only for truth and population-gap audits. The empirical transport fit and extension use only observed training atoms and their conditional empirical weights, without completing the anchors using unobserved population atoms. Each codebook has 50 independent repetitions, for 200 per sample size and 1,400 in total. The initial grid ended at $k = 512$; $k = 1024$ is a separately recorded extension with the same procedure and repetition budget.

Global range calibration and paired ablations. In each cell, solve the empirical transport dual on positive-mass training atoms, obtaining $\alpha_i + \beta_j \le W_1(y_i, y'_j)$, and define

$$f_0(y) = \min_j \{W_1(y, y'_j) - \beta_j\}, \qquad f(y) = [f_0(y)]_a^b - \frac{a+b}{2}, \tag{283}$$

where a, b are the extrema of f_0 on the observed training supports. The function f_0 is globally 1-Lipschitz; clipping preserves this property, so $(f, -f)$ is globally feasible and $|f| \le (b-a)/2$. Empirical optimality gives $f_0(y_i) = \alpha_i$ and $f_0(y'_j) = -\beta_j$: the positive-mass optimal dual is tight under both finite transforms, and the target transform is 1-Lipschitz on its support. Thus clipping leaves the empirical objective unchanged, as in Proposition E.8. Empty cell/arm fits use the zero pair. Write f_x for the fitted function in cell x and $s_x = b_x - a_x$. A cross-action score is

$$2\mathbb{1}\{A_i = 0\} f_{X_i}(Y_i) - 2\mathbb{1}\{A_i = 1\} f_{X_i}(Y_i).$$

Table 12: **Relative calibration cost with equal fold sizes.** Each row has 200 repetitions across four codebooks. Learning gap and radius columns are multiplied by 10^5. All 16 ordered pairs cover simultaneously in 200/200 repetitions at each size. The radius uses a globally certified training range.

N	$10^5\bar{g}$	$10^5\bar{r}_{\mathrm{span}}$	$\bar{r}_{\mathrm{span}}/\bar{g}$	$\bar{h}/\bar{e}$
128	13.494	10.874	0.806	1.802
256	10.731	5.871	0.547	1.547
512	9.026	3.344	0.371	1.371
1,024	7.752	1.956	0.252	1.252
2,048	6.781	1.180	0.174	1.174
4,096	6.032	0.728	0.121	1.121
8,192	5.434	0.455	0.084	1.084

Reversing the action order swaps the potentials and gives the same score; same-action lower comparisons use zero potentials. Thus every policy-pair score lies in $[-s, s]$ on the *entire* outcome space, where $s = \max_x s_x$. Corollary E.9 therefore justifies the training-conditional radius

$$r_{\mathrm{span}} = s\sqrt{2\log(16/\alpha)/m}, \qquad \alpha = 0.05. \tag{284}$$

No evaluation observations or known population support are used to certify this range. The two paired ablations use the identical learned potentials and evaluation data, replacing only the radius by

$$r_{\mathrm{coarse}} = 2\sqrt{2\log(16/\alpha)/m}, \qquad r_{\mathrm{point}} = s\sqrt{2\log(1/\alpha)/m}.$$

The pointwise radius only guarantees a prespecified pair. There are four distinct disagreement masks among the 16 ordered pairs; we retain the 16-pair union factor rather than using this redundancy to reduce the radius. This experiment does not study increasing policy complexity.

Learning, evaluation, and certificate errors. For repetition b, let T_j be the true lower endpoint, $\mu_{bj} = \mathbb{E}(S_i(j) \mid D_{1,b})$, and $\widehat{T}_{bj}$ the evaluation mean. We record

$$\begin{aligned} g_b &= \max_j (T_j - \mu_{bj}), & v_b &= \max_j |\widehat{T}_{bj} - \mu_{bj}|, \\ e_b &= \max_j |\widehat{T}_{bj} - T_j|, & h_b &= \max_j \{T_j - (\widehat{T}_{bj} - r_b)\}_+. \end{aligned} \tag{285}$$

The maxima are taken within each repetition, before averaging. Tables report $\bar{r}/\bar{g}$ and $\bar{h}/\bar{e}$, rather than averages of per-repetition ratios. The learning gap is computed from the true population expectation of the fitted dual. For a common radius, the positive point-estimation shortfall $e_b^- := \max_j (T_j - \widehat{T}_{bj})_+$ satisfies $|h_b - e_b^-| \le r_b$. We nevertheless report the absolute point risk e_b, which matches the point-estimation loss in the theory.

Observed relative cost and uncertainty. Table 12 shows $\bar{r}_{\mathrm{span}}/\bar{g}$ falling from 0.806 to 0.0837, while the honest-to-point error ratio falls from 1.802 to 1.084. The learning-gap ratio $\bar{g}/a_k$ stays between 0.275 and 0.278. A descriptive regression of $\log \bar{g}$ on $\log J$ has slope -0.989; the slope of $\log(\bar{r}_{\mathrm{span}}/\bar{g})$ against $\log N$ is -0.545. On this finite grid, learning error follows the contact scale and relative calibration cost decreases.

The evaluation-noise ratio $\bar{v}/\bar{g}$ decreases from 0.150 to 0.020. For the fixed all-zero versus all-one contrast, mean LCB/truth increases from -0.366 to 0.166; thus a small *relative calibration cost* need not mean a tight certificate in absolute or relative-to-truth terms. At $N = 8192$, the 95% paired bootstrap intervals are $[0.08355, 0.08387]$ for $\bar{r}_{\mathrm{span}}/\bar{g}$ and $[1.08341, 1.08397]$ for $\bar{h}/\bar{e}$. At $N = 128$ they are $[0.80045, 0.81130]$ and $[1.78456, 1.81970]$, respectively. These intervals use 2,000 within-codebook stratified resamples and condition on the four codebooks. They do not account for all possible code designs or asymptotic extrapolation. All intermediate intervals and Monte Carlo standard errors are retained in the archive.

Figure 7 compares the relative costs with their conditional bootstrap intervals.

Table 13: **Radius ablations on identical scores.** The first three columns are ratios of mean radius to mean population learning gap. The last column uses calibration for one prespecified pair. All three procedures covered all pairs in every reported repetition. The positive learning gap contributes to this conservative coverage.

N	Global range	Coarse envelope	Pointwise	$\bar{h}_{\text{point}}/\bar{e}$
128	0.806	6292.9	0.581	1.578
256	0.547	5595.3	0.394	1.394
512	0.371	4703.8	0.267	1.267
1,024	0.252	3872.6	0.182	1.182
2,048	0.174	3130.8	0.125	1.125
4,096	0.121	2488.5	0.087	1.087
8,192	0.084	1953.4	0.060	1.060

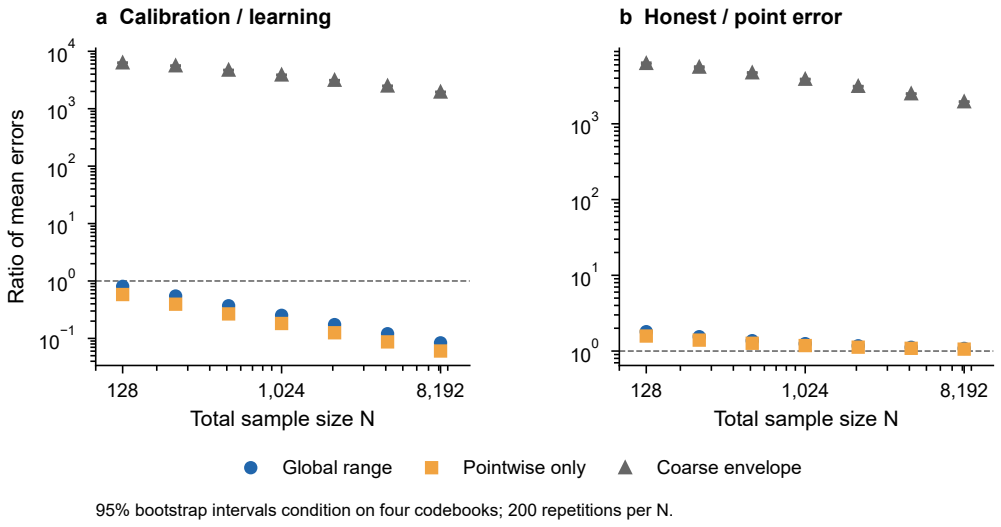


Figure 7: **Range calibration reduces the relative cost of simultaneous honesty.** (a) Ratio of mean radius to mean population learning gap. (b) Ratio of mean honest excess to mean absolute point error. All three calibrations use identical scores; the pointwise alternative does not supply a simultaneous guarantee. Error bars are existing 95% bootstrap intervals from 2,000 within-codebook stratified resamples, conditional on four codebooks (50 repetitions each). Both axes are logarithmic and all seven sizes, including the $N = 8192$ extension, are shown. The coarse-envelope radius remains orders of magnitude larger.

What the experiment establishes. The observed all-pair coverage is 200/200 at each size; the Wilson 95% interval is approximately $[0.9812, 1]$. Selected-pair coverage is also 200/200. The population learning bias helps make coverage conservative; uniform honesty itself follows from the conditional range proof, not from these observations. The result demonstrates decreasing calibration overhead for a globally valid empirical metric-dual implementation on growing smooth codes with a fixed fold ratio.

The global range s shrinks with the contact amplitude, giving a radius of approximately $a_k/\sqrt{N}$. In contrast, the fixed-envelope radius still exceeds the learning gap by orders of magnitude on this sample grid (Table 13). The minimax result concerns rate orders; its leading constants and asymptotic risk ratios are not determined by this experiment. The paired comparison isolates the contribution of calibration to the lower-endpoint error: as the sample grows, learning accounts for an increasing fraction of the total certificate gap.

Implementation and reproducibility. The driver is `reproducibility/honesty_price_study.py`, with base seed 2026092581 and separate deterministic streams for codebooks and

repetitions. The exact network-simplex solver operates on normalized Hamming costs to avoid the small physical amplitudes affecting numerical tolerances. The initial generic LP pilot was replaced for runtime; all final configurations were rerun with network simplex. Smoke runs are excluded. Independent checks in `reproducibility/verify_honesty_price.py` compare the empirical objective with a separately formulated single-potential LP, check actual off-support extensions, and integrate the signed-tent cost directly. The independent objective discrepancy is 6.8×10^{-16}; the maximum off-support Lipschitz residual in that audit is 2.8×10^{-17}. These are floating-point implementation checks, not an interval-arithmetic proof. Repetition records, conditional standard errors, bootstrap intervals, source snapshots, and hashes are in `reproducibility/results/honesty_price/`; the largest-size run also has its original records in `reproducibility/results/honesty_price_extension/`.

K.1.5 AIPW: LEARNING VERSUS CERTIFIED NUISANCE OVERHEAD

Design and estimands. We extend the growing smooth-code construction of Appendix K.1.4 to unknown assignment. For $k \in \{16, 32, 64, 128, 256, 512\}$, generate four independent constant-weight codebooks with $8k$ atoms and $J = 16 \log_2 k$ coordinates, using the same quantile map and exact W_1 costs. In each of two equiprobable covariate cells, the conditional outer laws are uniform on the two disjoint halves of the codebook. Treatment probabilities are 0.35 and 0.65 in the two cells. The estimator receives the overlap bound $\varepsilon = 0.2$, but estimates both propensities and potential means from data. The total sample size is $n = 12k$; independent training, nuisance, and evaluation folds each contain $4k$ iid observations. Each codebook has 25 independent repetitions, giving 100 per size and 600 initial repetitions. The $k = 1024$ ($n = 12{,}288$) experiment uses ten independent codebooks and ten repetitions each. This adds 100 repetitions, for 700 in total, using the same seed rule, estimator, and the reported calibration formula, fixed before running the $k = 1024$ experiment.

Empirical transport on observed training atoms and the clipped metric extension produce globally feasible potentials $(f_x, -f_x)$ with a training-certified span s_x and $|f_x| \le s_x/2$. Same-action lower comparisons use zero potentials; reversing a cross-action comparison gives the same score. All four deterministic policies and 16 ordered policy pairs are retained. The nuisance fold estimates $\widehat{e}_{ax}$ by conditional action frequencies, clipped to $[0.2, 0.8]$, and $\widehat{\mu}_{ax}$ by the conditional sample means of the fitted potential for arm a. Empty cells use propensity $1/2$ and empty cell–arm means use zero. This experiment concerns lower endpoints.

For each repetition, write $\Delta_u = \Theta^L(u) - g_u$, $G = \max_u \Delta_u$, and $R_* = \max_u |R_u|$. The exact decomposition is $\overline{S}_u - \Theta^L(u) = -\Delta_u + R_u + Z_u$. Population assignment and expectations compute the sharp truths, G, and R_* only for retrospective simulation diagnostics. They do not enter fitting, nuisance estimation, or certification. The codebook grows with sample size, so this is a sequence of populations rather than a learning curve for one fixed population.

Binary remainder correction. Set $\alpha = 0.05$ and retain $\delta_e = \delta_\mu = \delta = \alpha/[8(n+1)]$. Let N_x and N_{xa} be nuisance-fold counts. Write $p_x = e_{0x}$, $\widehat{p}_x = \widehat{e}_{0x}$ and $d_{ax} = \widehat{\mu}_{ax} - \mu_{ax}$. Since both true and fitted propensities sum to one, the cell remainder factors exactly as

$$(\widehat{p}_x - p_x)D_x, \qquad D_x = \frac{d_{0x}}{\widehat{e}_{0x}} - \frac{d_{1x}}{\widehat{e}_{1x}}.$$

For each cell, form an exact two-sided binomial interval $[L_x, U_x]$ for p_x with failure probability $\delta/2$ by inverting the two binomial tails, each at $\delta/4$. Equivalently, for $t = N_{x0}$ and $N = N_x$, its endpoints are beta quantiles $B^{-1}_{t,N-t+1}(\delta/4)$ and $B^{-1}_{t+1,N-t}(1-\delta/4)$; the lower endpoint is zero if $t = 0$, and the upper endpoint is one if $t = N$. Empty cells use $[0, 1]$. Set $q_e(x) = \max\{|\widehat{p}_x - L_x|, |\widehat{p}_x - U_x|\}$. The two intervals simultaneously contain the true propensities with probability at least $1 - \delta$.

Conditional on training and all nuisance-fold covariate/action labels, the two sample means in D_x use independent outcomes and fixed weights. Weighted Hoeffding concentration therefore gives

$$q_D(x) = \min\left\{ s_x \sum_{a=0}^{1} \frac{1}{\widehat{e}_{ax}}, \quad s_x \sqrt{\frac{\log(4/\delta)}{2} \sum_{a=0}^{1} \frac{1}{N_{xa}\widehat{e}^2_{ax}}} \right\}.$$

If either arm count is zero, use only the first, deterministic bound. Each cell's contrast bound fails with probability at most $\delta/2$, so $|D_x| \le q_D(x)$ holds in both cells with probability at least $1 - \delta$.

Table 14: **AIPW learning and certified overhead.** Ratios of means, with maxima over policy pairs taken within each repetition. Each size has 100 repetitions. All 16 pairs cover in every repetition. The binary remainder correction and variance-adaptive radius retain the stated failure allocation.

n	$\overline{R}_*/\overline{G}$	$\overline{b}_{\mathrm{bin}}/\overline{G}$	$\overline{r}/\overline{G}$	$(\overline{b}_{\mathrm{bin}}+\overline{r})/\overline{G}$
192	0.01712	2.151	1.592	3.743
384	0.00917	0.959	0.787	1.745
768	0.00361	0.459	0.412	0.871
1,536	0.00177	0.228	0.229	0.457
3,072	0.00090	0.114	0.132	0.246
6,144	0.00047	0.057	0.081	0.138
12,288	0.00028	0.029	0.051	0.080

Combining the two events gives

$$b_{\mathrm{bin}} = \max_x q_e(x)q_D(x), \qquad \Pr\{\max_u |R_u| \le b_{\mathrm{bin}}\} \ge 1-2\delta. \tag{286}$$

Indeed, R_u averages the cell remainders over disagreement cells, so their maximum absolute bound suffices without known covariate probabilities. The propensity and contrast events need not be independent. This sharper factorization uses two complementary treatment arms.

Variance-adaptive simultaneous calibration. Write $\widehat{m}_x = \sum_a \widehat{\mu}_{ax}$. For an active cell, the full-space score bounds are

$$\ell_{xa} = \widehat{m}_x + \frac{-s_x/2 - \widehat{\mu}_{ax}}{\widehat{e}_{ax}}, \qquad h_{xa} = \widehat{m}_x + \frac{s_x/2 - \widehat{\mu}_{ax}}{\widehat{e}_{ax}}.$$

There are three nonzero disagreement masks $v \in \{(0,1),(1,0),(1,1)\}$. Their full-space widths are $W_v = \max_{x,a} v_x h_{xa} - \min_{x,a} v_x \ell_{xa}$. The zero mask needs no calibration. On the evaluation fold, compute each score's unbiased sample variance $\widehat{V}_v$. Let $\eta = \alpha - 2\delta$, $m = 4k$, $t = \log(6/\eta)$, and

$$a_v = W_v\sqrt{\frac{t}{2(m-1)}}, \qquad \widehat{\sigma}_v^+ = \sqrt{\widehat{V}_v + a_v^2} + a_v.$$

The variance lower-tail inequality in the proof of Maurer & Pontil (2009, Theorem 10, inequality (5)), rescaled to width W_v, implies $\sigma_v \le \widehat{\sigma}_v^+$ except on an event of probability $\eta/6$. This follows by solving $\sigma_v^2 - W_v\sigma_v\sqrt{2t/(m-1)} \le \widehat{V}_v$ as a quadratic, without relaxing the square root. Their Theorem 3, applied to the negative score, controls its upper mean deviation with another failure probability $\eta/6$. Union over these two events for all three scores yields

$$r_v = \min\left\{W_v, \sqrt{\frac{2t}{m}}\widehat{\sigma}_v^+ + \frac{W_v t}{3m}\right\}, \qquad r = \max_v r_v, \qquad L(u) = \overline{S}_u - b_{\mathrm{bin}} - r. \tag{287}$$

The width cap is deterministic, hence valid without a further failure allocation. Conditional on both fitting folds, $\overline{S}_u \le g_u + R_u + r$ simultaneously with probability at least $1-\eta$. Intersecting with the nuisance event proves $L(u) \le g_u \le \Theta^L(u)$ for all 16 pairs with probability at least $1-\alpha$. The radius uses evaluation sample variances; its validity comes from these concentration events, without a normal approximation.

Results and interpretation. At the extension size $n = 12{,}288$, $\overline{G} = 5.44706 \times 10^{-5}$ and total correction is 0.07967 of the mean gap (95% conditional interval $[0.07931, 0.08002]$). The nuisance-bound and evaluation-radius ratios are 0.02891 and 0.05076, respectively. At $n = 6144$, $\overline{G} = 6.05665 \times 10^{-5}$ and the total correction ratio is 0.13791, with a 95% conditional bootstrap interval $[0.13709, 0.13868]$. The actual nuisance remainder ratio is 0.000467; the certified correction ratio is 0.05714. These quantities distinguish small realized bias from the cost of its certification. Learning exceeds total certified overhead in mean from $n = 768$ on this grid, whereas the small-sample correction remains substantial. This separation is observed on the reported sample grid and uses a range certificate that shrinks with the fitted potential span. At $n = 6144$, 99/100 individual repetitions have $(b_{\mathrm{bin}} + r)/G < 0.15$; the remaining repetition is retained.

The bootstrap uses 2,000 paired resamples within each codebook and conditions on the codebooks: four at the initial sizes and ten at the extension size. It does not quantify variation over all code designs. The empirical simultaneous coverage is 100/100 at each size (Wilson 95% interval approximately $[0.963, 1]$); the coverage guarantee comes from the preceding argument rather than these empirical counts.

Calibration comparisons and reproducibility. Six calibration formulas were evaluated on the same 600 repetitions after inspection of the initial calibration. At $n = 6144$, the conservative product bound with a 16-pair union factor gives a total correction ratio of 0.49633; the local product with three-score Hoeffding calibration gives 0.32339. The standard empirical Bernstein formula applied to the evaluation radius gives 0.18179, and its application to both the evaluation radius and nuisance mean bounds gives 0.19206. The binary remainder correction with the standard Bernstein formula gives 0.15271; the reported formula with exact variance inversion gives 0.13791. Fitted potentials, point estimates, learning gaps, fold allocation, and failure allocation are identical across these comparisons. The base seed is 2026092607. The driver is `reproducibility/aipw_learning_dominance.py`; the final calibration uses `reproducibility/aipw_final_moments.py` and `reproducibility/aipw_binary_remainder.py`, with independent replay statistics in `reproducibility/results/aipw_final_moments/`. Raw repetitions, conditional intervals, component arrays, and hashes are in `reproducibility/results/aipw_learning_dominance/` and `reproducibility/results/aipw_binary_remainder/`; the other calibration comparisons remain in their separate result directories. The paired comparisons on the same repetitions are descriptive rather than independent confirmation. The $n = 12{,}288$ experiment uses the reported formula, fixed before its repetitions were generated. Its driver is `reproducibility/run_aipw_extension_parallel.py`, and its protocol, raw records, moments, and calibration results are in `reproducibility/results/aipw_extension_10books/`. Independent checks verify small-instance LP objectives, off-code feasibility, exact quantile costs, and all 9,600 policy-pair decompositions; the maximal decomposition residual is 2.8×10^{-20}.

K.2 Endpoint tightness and first-stage learning

How do fitted outer laws and certificate construction affect the information retained by valid lower and upper bounds?

K.2.1 Electricity templates and endpoint calibration

The cached UCI archive has SHA-256 prefix `f6c4d0e0df12`; the complete digest is recorded in the released manifest. For each client we discard the pre-activation prefix through its first positive record, retain subsequent finite nonnegative values, and require at least 365×24 observations. With $s = 130.9398575$, the median of eligible client-specific positive medians, raw load v is mapped to $v/(v+s)$. Exact empirical-quantile bin averages q are then transformed to $1/2 + 0.45\{q - \int q\}$. This preserves monotonicity, puts every curve in $[0.05, 0.95]$, and fixes its mean at $1/2$. The four fixed template pairs are MT_343/MT_370, MT_090/MT_369, MT_130/MT_228, and MT_131/MT_363, with

$$(C_1, C_2, C_3, C_4) = (0.00888846, 0.00935252, 0.00952655, 0.00956523).$$

The largest cell contributes 25.62% of $\sum_x C_x$. The four pairs were selected by a geometry rule after examination of an earlier uneven template set. The reported results condition on these pairs. For each template pair define $d_x = q_x^+ - q_x^-$ and $C_x = \int_0^1 d_x(t)^2\, dt > 0$. The projection used by the implementation is

$$\tau_x(Y) = \mathrm{clip}_{[0,1]}\left(\frac{\int_0^1 \{Q_Y(t) - q_x^-(t)\} d_x(t)\, dt}{C_x} \right). \tag{288}$$

Clipping is nonexpansive, so Cauchy–Schwarz gives $C_x|\tau_x(Y) - \tau_x(Y')|^2 \le \|Q_Y - Q_{Y'}\|_2^2 = W_2^2(Y, Y')$. Thus any scalar dual feasible for $C_x|t - s|^2$ lifts by composition with τ_x to a feasible dual on the entire distribution space. Also $|\tau_x(Y) - \tau_x(Y')| \le \|d_x\|_\infty W_1(Y, Y')/C_x$; composing with the scalar potentials, whose Lipschitz constants are at most $2C_x$, gives the measurement constant $2\|d_x\|_\infty$ used below.

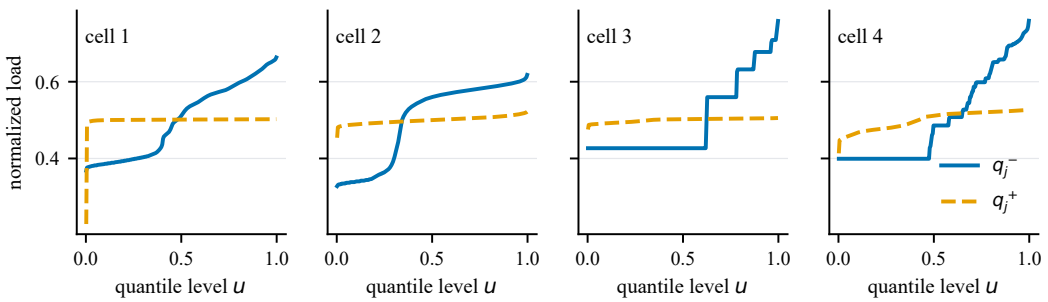


Figure 8: **Real load distributions supply heterogeneous shapes while mean location is held fixed.** Each panel is one covariate cell. Solid and dashed curves are the two transformed empirical-quantile templates; every curve has mean $1/2$, and colour plus line style distinguish the pair.

The reported experimental populations use these four template pairs. Source time dependence is part of each template, while generated units are iid conditional on the templates; they are not additional real clients.

The reported simulation compares the all-zero and all-one policies. Let $Z \sim \mathrm{Unif}[-1,1]$ and let S be an independent Rademacher variable. In every cell the coefficient laws are

$$t_0 = 0.50 + 0.05Z, \qquad t_1 = 0.50 + 0.425S + 0.025Z.$$

Both have mean $1/2$, so the individual-distribution means and Wasserstein barycenters coincide. The first outer law is concentrated near the center; the second is an equal mixture on $[0.05, 0.10]$ and $[0.90, 0.95]$. Their one-dimensional outer quantile functions differ in absolute value by 0.4 almost everywhere. The sharp all-zero versus all-one endpoint is therefore

$$\Theta^L = 0.16\,\frac{1}{4}\sum_{j=1}^{4} C_j = 0.0014933103.$$

All results use a 50–50 independent split. Within every construction cell and arm, the learner estimates only the support endpoints and fits a uniform coefficient law. It is therefore misspecified for the bimodal arm. The second potential is nevertheless the exact analytic c-transform, so the learned pair remains globally feasible. Normalization gives $0 \le f_j \le C_j$ and $-C_j \le g_j \le 0$ on the full interval.

Because $P(X = j) = 1/4$ is known, evaluation uses the four conditional arm means. Let N_{ja} be their realized counts and let $R_{f,j}, R_{g,j}$ denote the exact learned-potential ranges. Conditional on training and the counts, the fixed-pair one-sided Hoeffding radius is

$$r_{\mathrm{fixed}} = \left[\frac{\log(1/\alpha)}{2}\,\frac{1}{16}\sum_{j=1}^{4}\left\{\frac{R_{f,j}^2}{N_{j0}} + \frac{R_{g,j}^2}{N_{j1}}\right\}\right]^{1/2}.$$

For comparison, a calibration ablation splits the fixed contrast into its four cell contributions, each containing two arm means, and sums four Hoeffding radii:

$$r_{\mathrm{comp}} = \sum_{j=1}^{4}\left[\frac{\log(8/\alpha)}{2}\,\frac{1}{16}\left\{\frac{R_{f,j}^2}{N_{j0}} + \frac{R_{g,j}^2}{N_{j1}}\right\}\right]^{1/2}. \tag{289}$$

It uses the same learned dual and data as our direct fixed-pair method. It does not evaluate all 16 policies or all their ordered pairs.

The learned plug-in estimate averages about 0.001439, or 96.3% of the sharp endpoint, despite the misspecified uniform first stage. The remaining loss is therefore dominated by the explicit confidence radius. Direct fixed-pair calibration reduces the mean radius by about 62% relative to the componentwise calibration; the latter retains 0.682, 0.786, 0.838, and 0.874 of the endpoint at the four sizes. With 1,000/1,000 covered repetitions, the two-sided 95% Wilson lower endpoint is 0.996.

Figure 9 shows the repeated-sample distributions for both the heterogeneity and null contrasts.

Table 15: **A fixed heterogeneity contrast yields an informative certificate.** Each row contains 1,000 repetitions. Ratios divide the mean LCB by the analytic endpoint; both calibrations cover in every repetition. "Componentwise" is a radius ablation of our method, not another learner.

n	Ours (fixed pair)	Componentwise ablation	Ours: 5% ratio
2,000	0.856	0.682	0.850
5,000	0.895	0.786	0.892
10,000	0.915	0.838	0.913
20,000	0.929	0.874	0.927

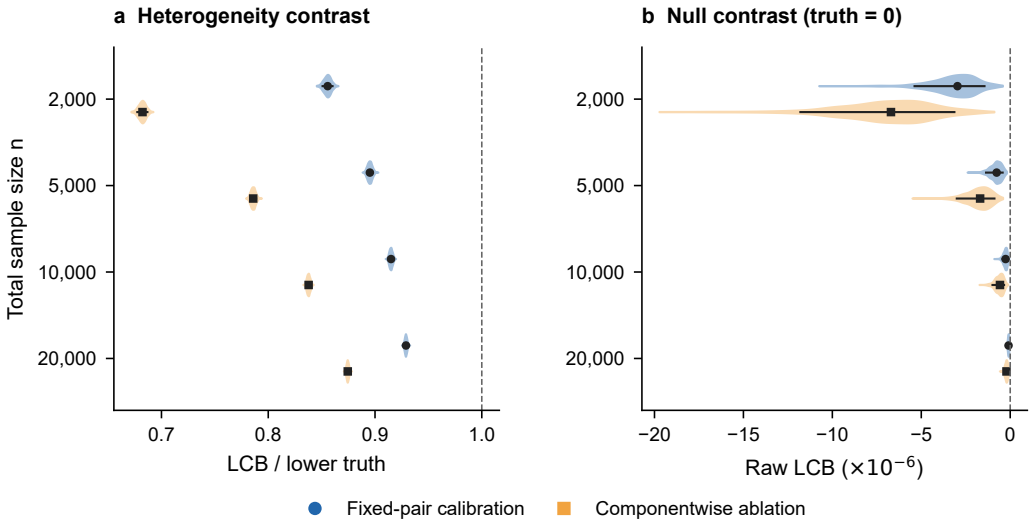


Figure 9: **Repeated-sample distributions for the focused lower certificate.** All 1,000 repetitions per design and sample size enter the violin densities; black markers and segments indicate medians and empirical 5th–95th percentiles. The two calibrations use identical learned scores and samples. (a) The heterogeneity contrast is normalized by its positive sharp endpoint. (b) The null contrast has truth zero, so raw bounds are shown in units of 10^{-6}. Densities use Scott bandwidth and stop at the observed extrema; no hypothesis test or significance filtering is applied.

Exact observation of generated Y is maintained in the focused study. The finite-bag experiments in the final block audit the measurement correction.

The release contains repetition-level CSV files, source and template hashes, command arguments, and platform metadata. The focused study uses seed 2026092101 and 1,000 repetitions per cell. The observed coverage is compatible with the theorem.

K.2.2 Paired correct and misspecified first-stage fits

Figure 3(b) reuses the fixed templates, coefficient laws, analytic endpoints, and fixed contrast of Appendix K.2.1. The four sample sizes are $n \in \{2{,}000, 5{,}000, 10{,}000, 20{,}000\}$, with 1,000 repetitions each and a 50:50 independent split. All learners receive identical draws. For zero-based sample-size index j and repetition b, the seed is $2026092101 + 100003j + b$.

First-stage comparison. In each cell, the correctly specified family estimates a uniform source interval from training extrema. Target observations are split at the prespecified coefficient threshold $1/2$; each group supplies a uniform component's endpoints, and its empirical frequency supplies its mixture weight. All parameters are estimated from the training fold. The misspecified source fit is identical. For target training extrema a, b, its target law is uniform on $[(a+b)/2 - (b-a)/4, (a+b)/2 + (b-a)/4]$. This half-width restriction understates outer heterogeneity. The factor $1/2$ was

fixed before this run, without a severity sweep. The earlier unrestricted uniform comparison remains in Appendix K.2.4 and the saved results.

Dual construction and calibration. For scalar cost $C_x(t-s)^2$, the fitted monotone and countermonotone transport maps are piecewise affine. Integrating $f'(t) = 2C_x\{t - T(t)\}$ gives piecewise quadratic potentials; the second potential is the exact infimal or supremal c-transform over $[0,1]$. The lower map is clipped to $[0,1]$ and the upper map's outer affine branches are extrapolated. The projection lift in Appendix K.2.1 and residual upper lift in Appendix K.2.4 preserve full-space feasibility. Evaluation uses cell–arm means and the fixed-pair Hoeffding calibration. Ranges are computed over the entire known template segment $t \in [0,1]$, never over the fitted or shrunken intervals. These ranges rely on the declared outcome support, whereas dual feasibility holds on the full distribution space. Each endpoint uses $\alpha = 0.05$, so a union bound guarantees at least 90% joint coverage.

Table 16: **Paired first-stage comparison.** Mean bound divided by its own sharp endpoint. Each learner has 1,000/1,000 observed lower, upper, and joint coverage at every size; the theoretical joint guarantee is at least 90%.

	Correct fit		Half-width fit	
n	LCB/truth	UCB/truth	LCB/truth	UCB/truth
2,000	0.880872	1.256481	0.768865	1.492088
5,000	0.925549	1.161206	0.802356	1.349911
10,000	0.947483	1.113653	0.819268	1.279143
20,000	0.962983	1.080150	0.831333	1.229378

At $n = 20{,}000$, the correct fit's uncalibrated mean lower and upper estimates, normalized by their respective truths, are 0.999864 and 1.000025. The remaining certificate gaps thus largely reflect calibration. Changing the fitted law changes both its population dual gap and its potential ranges; stronger misspecification need not enlarge every finite-sample UCB relative to the original uniform fit.

Reproducibility and audits. The drivers are `reproducibility/paired_first_stage_study.py` and `reproducibility/stronger_misspec_study.py`; repetition records, Monte Carlo standard errors, and manifests are retained in the corresponding `reproducibility/results/paired_first_stage/` and `reproducibility/results/stronger_misspec/` directories. Analytic transforms and ranges were independently checked against the original uniform implementation. Population dual values were integrated piecewise and verified to err in the required direction in every run. These numerical audits supplement the feasibility argument; they do not replace it. The original uniform records are preserved alongside both displayed fits.

K.2.3 Known-propensity IPW comparison with barycentric targets

We augment Experiment B with the IPW policy-evaluation estimator of Huang et al. (2026, Section 4.1) and the IPW specialization of the average causal effect map identified by Lin et al. (2023, Theorem 2). The latter comparison uses their causal target and its IPW identification formula, not their proposed doubly robust implementation. Both use the known propensity $e_a(x) = 1/2$; there is no propensity estimation or outcome regression. The policies remain the prespecified all-zero and all-one rules. Using all n observations from the same draws as the certificate experiment, define

$$\widehat{q}_a(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}\{A_i = a\}}{1/2} Q_{Y_i}(t), \qquad \widehat{\Delta}(t) = \widehat{q}_1(t) - \widehat{q}_0(t).$$

This is unnormalized Horvitz–Thompson IPW. Its positive weights preserve monotonicity of each $\widehat{q}_a$, so monotone rearrangement is the identity. With the uniform reference $\lambda(t) = t$, the causal effect map estimate is $\widehat{\Delta}$, and hence

$$\widehat{D}_{\mathrm{WPL}} := W_2^2(\widehat{\mu}_1, \widehat{\mu}_0) = \int_0^1 \widehat{\Delta}(t)^2\, dt =: \widehat{D}_{\mathrm{CEM}}.$$

The two columns therefore represent the same statistic through two literature targets, not independent algorithms. We compare the estimated barycenters themselves rather than choose a particular welfare utility or optimize over policies. At the population level, every common barycenter utility assigns these two policies the same value.

Table 17: **IPW summaries and a positive matching certificate.** All entries are multiplied by 10^3; parentheses give one Monte Carlo standard error over 1,000 paired repetitions. WPL and CEM report squared barycenter distance and squared effect-map norm, respectively, both with population value zero. The matching LCB uses the correct first-stage fit and has sharp target $10^3\Theta^L = 1.493310$. The columns describe different targets and are not a common-target accuracy ranking.

n	WPL IPW	CEM IPW	Matching LCB
2,000	0.550 (0.02490)	0.550 (0.02490)	1.315 (0.00023)
5,000	0.206 (0.00894)	0.206 (0.00894)	1.382 (0.00014)
10,000	0.106 (0.00486)	0.106 (0.00486)	1.415 (0.00010)
20,000	0.050 (0.00231)	0.050 (0.00231)	1.438 (0.00007)

The IPW estimates are not identically zero in finite samples. In this equal-barycenter design, their exact expected squared magnitude is

$$\mathbb{E}\widehat{D}_{\mathrm{WPL}} = \frac{2}{n}\sum_{a=0}^{1}\mathbb{E}\|Q_{Y(a)}\|_2^2.$$

The observed means agree with this expression to within 1.69 Monte Carlo standard errors across the four sizes and two evaluation budgets. Thus the declining discrepancy is compatible with sampling noise around a zero barycentric target. In contrast, the matching LCB is positive and below its own sharp endpoint in all 1,000 repetitions at every size. The matching certificate detects heterogeneity between the outer laws, while IPW consistently estimates their equal barycentric summaries. IPW uses the full sample without a training cost; repeating it on only the certificate's evaluation half gives the same qualitative pattern and is retained in the release.

As an implementation positive control, we add a fixed location shift $\delta = 0.04$ to each treated quantile curve while retaining the same draws. The transformed curves remain in $[0, 1]$, and the true effect map is the constant δ. At $n = 20{,}000$, the mean integrated IPW effect is 0.040384, close to 0.04. This control verifies responsiveness to a change in the barycentric and effect-map targets.

The driver `reproducibility/barycentric_ipw_benchmark.py` regenerates the original seeds and draws, checks the sufficient-statistic implementation against direct observation-level IPW, and checks the equality of the two statistics and the zero population effect. Records, means, standard errors, repetition percentiles, example curves, analytic expectations, and source hashes are in `reproducibility/results/barycentric_ipw/`.

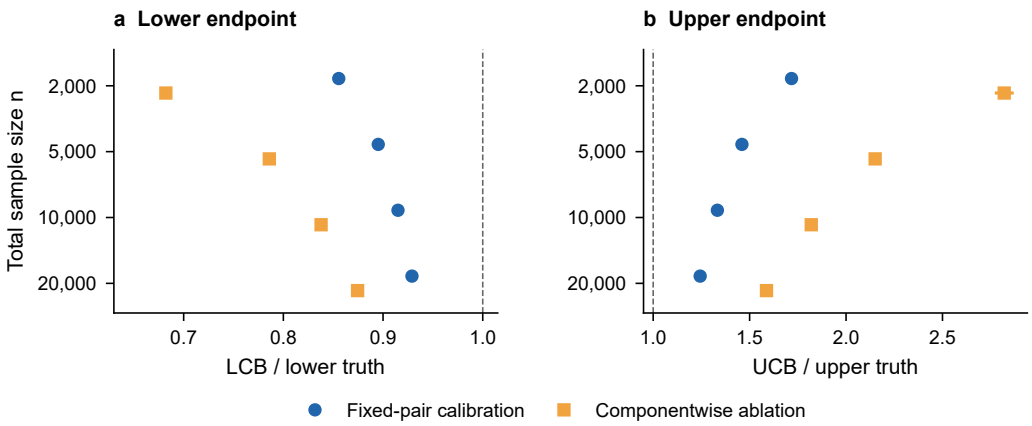


Figure 10: **Direct calibration tightens both endpoints under the same misspecified learner.** Dots show mean bound/truth; horizontal segments show the empirical 5th–95th percentiles across 1,000 paired repetitions per size, not confidence intervals for the mean. The dashed line is each endpoint's own truth. The componentwise ablation changes only the radius. Both calibrations cover every repetition; their remaining dual gaps need not vanish. Table 18 retains the numerical values.

Table 18: **Both sharp endpoints under the same misspecified learner.** Entries are mean lower LCB divided by its sharp lower endpoint and mean upper UCB divided by its sharp upper endpoint over 1,000 repetitions on the same data-generating design and evaluation samples. Observed endpoint coverage is 1.000 in every row and column. "Ours" uses direct fixed-pair calibration in each design. Each componentwise column calibrates four cell contributions, each combining two arm means, and sums their radii using $\log(8/\alpha)$; it is a calibration ablation using the same endpoint dual. Both duals are valid on the full distribution space. The misspecified learner retains nonvanishing dual gaps, so the ratios need not converge to one.

	Lower LCB / truth		Upper UCB / truth	
n	Ours (fixed pair)	Componentwise ablation	Ours (fixed pair)	Componentwise ablation
2,000	0.856	0.682	1.717	2.820
5,000	0.895	0.786	1.460	2.151
10,000	0.915	0.838	1.333	1.820
20,000	0.929	0.874	1.244	1.588

K.2.4 Upper endpoint under a misspecified uniform first stage

This experiment uses exactly the same four cells, electricity-derived template pairs, bimodal coefficient law, 50:50 split, sample sizes, repetition seeds, and fixed all-zero versus all-one contrast as the lower experiment in Appendix K.2.1. In particular, the learner fits a uniform coefficient interval in each cell and arm although the target coefficient law is bimodal. The sharp upper endpoint is $\Theta^U = (61/300)\,4^{-1}\sum_x C_x = 0.001897748$.

Let (u_x, v_x) be the scalar upper pair fitted from the same training-fold intervals as the lower learner. Its second potential is the exact scalar supremal c-transform, so $u_x(t)+v_x(s) \geq C_x(t-s)^2$ on $[0,1]^2$. The scalar pair alone does not give a full-space upper certificate. Define $P_xY = q_x^- + \tau_x(Y)(q_x^+ - q_x^-)$ using the clipped projection in (288), and extend both potentials by

$$U_x(Y) = u_x\{\tau_x(Y)\} + 2W_2(Y, P_xY), \qquad V_x(Y') = v_x\{\tau_x(Y')\} + 2W_2(Y', P_xY').$$

For distributions on $[0,1]$, changing either argument of W_2^2 changes the cost by at most twice its W_2 displacement. Consequently, $U_x(Y) + V_x(Y') \geq W_2^2(Y, Y')$ for *all* distributions Y, Y'. The residual terms vanish for every generated outcome, so the evaluated upper scores agree with the scalar

Table 19: **Upper certificate with the same misspecified uniform learner as the lower experiment.** The bimodal design uses 1,000 repetitions at each sample size. Both upper calibrations cover in every row.

n	Coverage	Fixed UCB/truth	Comp. UCB/truth	Fixed radius	Comp. radius
2,000	1.000	1.717	2.820	0.001308	0.003401
5,000	1.000	1.460	2.151	0.000819	0.002129
10,000	1.000	1.333	1.820	0.000577	0.001501
20,000	1.000	1.244	1.588	0.000407	0.001060

expressions. This validates the same misspecified first stage for an upper certificate without relying on the invalid direct projection lift.

The one-sided Hoeffding radii use the exact scalar-potential ranges on $t \in [0,1]$. The fixed-pair radius combines all eight cell-by-arm terms in one concentration bound. The componentwise ablation combines the two arm means in each cell, computes four radii with $\log(8/\alpha)$, and sums them, exactly as in the lower experiment: its formula is (289) with $R_{f,j}, R_{g,j}$ replaced by the upper potential ranges $R_{u,j}, R_{v,j}$. Both calibrations use the same learned upper dual and evaluation sample. These are valid range bounds for the declared DGP because its outcomes lie on the known template segments. If the support restriction is removed, the global residual-extended potential ranges must replace these smaller ranges. A separate random off-line quantile audit checks the implementation; the displayed Lipschitz argument supplies global feasibility. Each result below uses 1,000 repetitions at one-sided $\alpha = 0.05$.

The mean uncorrected upper score in this design remains near 0.001953, about 2.9% above the sharp endpoint as n grows. Thus coverage is robust to the misspecified outer-law fit, whereas tightness is limited by a nonvanishing dual gap. Repetition-level results and the off-line feasibility audit are in `reproducibility/results/upper_uniform_misspec/`.

K.2.5 Pooled-X first-stage misspecification at the upper endpoint

The upper mirror reuses the exact synthetic observations and split seeds from the lower experiment in Appendix K.1.1: two X cells, three actions, 49 non-Dirac step-quantile outcome types, nine policies, and all 81 ordered policy pairs. The cell-specific learner and the misspecified pooled-X learner use the same anchor cap and training weights as their lower counterparts. Pooled fitting discards the real X dependence in the outer law; the evaluation sample still follows the original cell-specific laws.

For each cell and action pair, solve the finite *upper* OT dual on the training anchors, giving coefficients a_i, b_j with $a_i + b_j \geq W_2^2(z_i, z'_j)$. Write d_1 for quantile L^1 distance. The full-space extension is

$$U(Y) = \min_i \{a_i + 2d_1(Y, z_i)\}, \qquad V(Y') = \max_i \{W_2^2(z_i, Y') - a_i\}.$$

Because W_2^2 on $\mathcal{P}([0,1])$ is 2-Lipschitz in its first argument under d_1, $U(Y) + V(Y') \geq W_2^2(Y, Y')$ for every distribution pair. The experiment evaluates these functions on all 49 types and checks all 49^2 pairwise constraints. The same finite-class Hoeffding union bound as the lower run protects all 81 policy pairs. Its score range is evaluated over the declared 49-type outcome universe, which contains the DGP support.

The fixed pair in Table 20 is the same pair used for the lower information ratio. The sharp upper endpoint for this pair is 0.297493. The misspecified pooled learner has a larger uncorrected dual value than the cell-specific fit at $n = 4096$ (0.305485 versus 0.298208), showing loss of tightness even while both retain validity. All-pair coverage is $400/400$ in every row. The occasional smaller pooled UCB at a finite n reflects differences in score ranges and does not imply better population dual approximation.

K.2.6 Upper positive control with point-mass targets

This separate upper experiment uses the same four fixed electricity-derived covariate cells and the same all-zero versus all-one contrast as the lower experiment. In cell x, let q_x be the first transformed non-Dirac load template, with mean $1/2$ and values in $[0.05, 0.95]$. For $Z \sim \mathrm{Unif}[-1,1]$, the source

Table 20: **Upper coverage under pooled-X misspecification.** Each row has 400 repetitions. "All pairs" checks simultaneous coverage of all 81 ordered pairs; UCB/truth is the mean ratio for the fixed pair.

n	First stage	All pairs	Mean estimate	UCB/truth
512	Ours, cell-specific	1.000	0.308536	2.069
	Ours, pooled X	1.000	0.317227	1.945
1024	Ours, cell-specific	1.000	0.302804	1.757
	Ours, pooled X	1.000	0.310943	1.656
2048	Ours, cell-specific	1.000	0.300411	1.464
	Ours, pooled X	1.000	0.307675	1.436
4096	Ours, cell-specific	1.000	0.298208	1.284
	Ours, pooled X	1.000	0.305485	1.300

Table 21: **Full-space upper confidence bound.** Each row has 1,000 repetitions. The target has two point-mass types. Coverage compares the fixed-pair UCB with the analytic sharp upper endpoint.

n	Upper coverage	Estimate/truth	UCB/truth	Radius
2,000	1.000	1.000	1.298	0.077679
5,000	1.000	1.000	1.188	0.049025
10,000	1.000	1.000	1.133	0.034640
20,000	1.000	1.000	1.094	0.024486

outcome has quantile $Q_0(t) = q_x(t) + hZ$, where $h = 0.04$. The target outcome is δ_a or δ_b with equal probability, independently of Z, where $(a, b) = (0.02, 0.98)$. Thus both outer populations have mean $1/2$, while their distribution-valued outcomes and heterogeneity differ. All generated distributions lie in $\mathcal{P}([0, 1])$.

The training half learns a and b as the minimum and maximum target means in each cell. Write $m_1(Y) = \int z\, dY(z)$ and $m_2(Y) = \int z^2\, dY(z)$. Given learned anchors $\widehat{a}, \widehat{b}$, define on the *full* distribution space

$$u(Y) = \max\{W_2^2(Y, \delta_{\widehat{a}}), W_2^2(Y, \delta_{\widehat{b}})\} = m_2(Y) + \max\{\widehat{a}^2 - 2\widehat{a}m_1(Y), \widehat{b}^2 - 2\widehat{b}m_1(Y)\},$$

and $v(Y') = \sup_{Y \in \mathcal{P}([0,1])}\{W_2^2(Y, Y') - u(Y)\}$. This is a globally superfeasible pair by construction. It is also computable without an infinite-dimensional optimization: if $M = \{0, (\widehat{a} + \widehat{b})/2, 1\}$, then

$$v(Y') = m_2(Y') + \max_{r \in M}\left[-2rm_1(Y') - \max\{\widehat{a}^2 - 2\widehat{a}r, \widehat{b}^2 - 2\widehat{b}r\}\right].$$

Indeed, monotone quantiles satisfy $\int Q_Y Q_{Y'} \geq m_1(Y)m_1(Y')$; the constant quantile $Q_Y \equiv r$ attains equality for each r. The remaining objective is piecewise linear in r, with possible maxima at 0, 1, or the crossing of the two affine pieces. Both potential oscillations are at most one because $0 \leq W_2^2 \leq 1$ on $\mathcal{P}([0, 1])$.

When both target anchors appear in training, $v(\delta_a) = v(\delta_b) = 0$. The upper dual meets the cost at the countermonotone outer matching: high source means pair with δ_a and low source means with δ_b. Hence its population value equals the sharp upper endpoint

$$\Theta^U = \frac{1}{4}\sum_{x=1}^{4}\int_0^1 q_x(t)^2\, dt + \frac{h^2}{3} - \frac{a+b}{2} + \frac{a^2+b^2}{2} + \frac{(b-a)h}{2} = 0.260963\ldots.$$

The experiment uses a 50–50 independent split and the same conditional cell-mean evaluation as the lower study. Conditional on training and cell counts, the fixed-pair one-sided Hoeffding radius uses the full-space range bound one for each potential. This guarantee does not assume that evaluation outcomes remain on the generated support.

The policy-search stress test is also run in the order-reversed direction. Every candidate upper endpoint is normalized to zero; a negative UCB is a false certificate. After inspecting all candidates, the pointwise procedure selects the smallest UCB. Table 22 shows that its simultaneous coverage deteriorates with class size, while the componentwise full-class event remains stable.

Table 22: **Upper confidence bounds after policy search.** Coverage is over 100,000 repetitions at one-sided $\alpha = 0.05$.

J	Candidates	One fixed pair	Pointwise search	Full class
4	15	0.992	0.947	0.994
8	255	0.992	0.831	0.995
12	4,095	0.992	0.661	0.995
20	1,048,575	0.993	0.310	0.996

K.3 Geometry and computational diagnostics

Which geometric assumptions and finite computations control endpoint rates, extension error, and numerical safety?

The additional executable study separates numerical safety, statistical learning, policy multiplicity, and active geometry. All reported results come from the full run of `complete_claim_study.py`; smoke runs are excluded. The archive contains per-repetition observations, Monte Carlo standard errors, Wilson coverage intervals, seeds, and source hashes. There are 200 repetitions per coverage configuration, 40 per hard-code configuration, and 1,000 per cost-active rate configuration.

K.3.1 Numerical budgets and general convex costs

We implement the lower and upper finite-anchor constructions in Section E. Four costs are considered: $h(z) = z^2$; $h(z) = z_+^2 + \frac{1}{4}(-z)_+^2$; $h(z) = z_+$; and $h(z) = (z - 0.15)_+^2$. On $[-1, 1]$, the Lipschitz constants used are respectively $2, 2, 1, 1.7$. For $K \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$, we compare the finite transport endpoint with the endpoint computed on the full 256-step population quantiles, check interpolation, and evaluate the extended potentials on unseen monotone curves. These evaluations audit the implementation; the global guarantee comes from the algebraic extension theorem. The lower endpoint needs no positive binning correction, whereas the upper endpoint includes L_h/K. A separate exact witness illustrates why this matters: with $K = 1$, source quantile zero and target quantile a half-zero, half-one step, quadratic cost is $1/2$ while its binned upper value is only $1/4$.

We also vary the number of source training atoms, completion of the anchor set, and a certified optimization loss $\varepsilon \in \{0, 0.01, 0.05\}$. The lower population gap is checked against $4L_h/K + \varepsilon + 2L_h \operatorname{TV}(P_0, \widehat{P}_0) + 2L_h R_0(\mathcal{U})$; the target law is exact in this deterministic audit. Here the total-variation term bounds the empirical-process term using the potential oscillation. This gives a conservative deterministic error budget.

K.3.2 All ordered policy pairs and joint endpoint coverage

There are three equally weighted covariate strata and two actions. We enumerate all eight deterministic policies and all 64 ordered policy pairs, including same-action pairs. Independent stratified training and evaluation samples have $n \in \{4096, 8192, 16392\}$ observations per action and stratum. The conditional outer mixing weights, in curve order, are

	$a = 0$	$a = 1$
$x = 1$	$(0.80, 0.10, 0.05, 0.05)$	$(0.05, 0.05, 0.85, 0.05)$
$x = 2$	$(0.05, 0.75, 0.05, 0.15)$	$(0.10, 0.05, 0.75, 0.10)$
$x = 3$	$(0.20, 0.05, 0.05, 0.70)$	$(0.05, 0.15, 0.75, 0.05)$.

Pooled-X fitting discards the covariate dependence in these outer laws. We compare empirical, pooled-X, and uniform fitted laws, with $K = 32$. The same evaluation sample checks both endpoints for every pair. To compare information with the preceding fixed-pair experiments, we also prespecify the all-zero versus all-one policy pair. Its sharp lower and upper endpoints are positive for all four costs, so mean LCB/sharp-lower and UCB/sharp-upper ratios are well defined. Joint coverage continues to check all 64 pairs, including pairs with zero lower endpoints for which a ratio would be undefined. For this fixed pair, the sharp (Θ^L, Θ^U) values are $(0.214204, 0.265427)$ for quadratic, $(0.213874, 0.250362)$ for asymmetric, $(0.397204, 0.438412)$ for hinge, and $(0.112693, 0.132870)$ for tolerance cost. The joint Hoeffding radius is $L_h\sqrt{2\log(128/0.05)/(3n)}$. This follows from the worst-case same-action oscillation $2L_h$ and independence across strata; the union bound covers

all 128 one-sided events. Known stratum weights make this a stratified version of the conditional experiment, rather than a claim about estimated propensities.

Figure 11 displays the complete cost–fit grid.

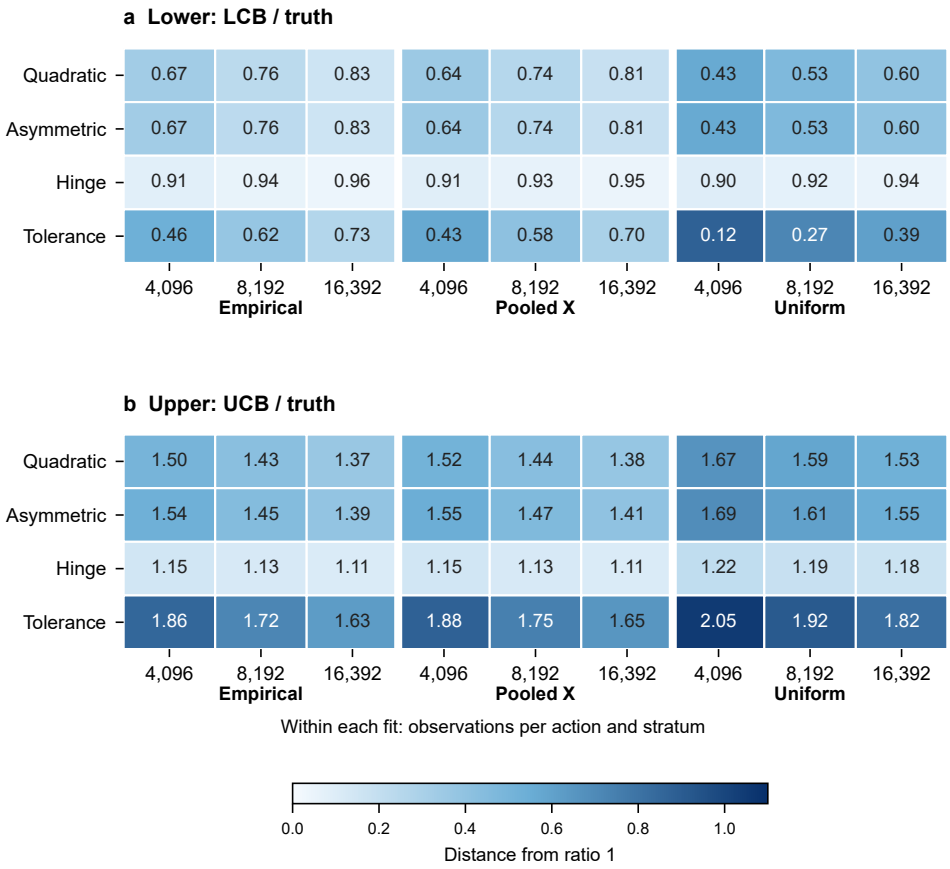


Figure 11: **Misspecification affects informativeness across convex costs.** All 36 cost–fit–sample-size configurations are shown, with 200 repetitions each. Printed values are mean fixed-pair LCB/truth or UCB/truth; a common color scale shows their distance from one, with darker cells indicating larger gaps. The fixed pair is all-zero versus all-one; n counts evaluation observations per action and stratum. Both endpoints for all 64 ordered pairs cover jointly in every recorded repetition. No row or column standardization is used.

K.3.3 Hard embeddings, smooth individual curves, and saturation

For each per-arm sample size $n \in \{16, 32, 64, 128\}$, construct $M = 8n$ constant-weight binary codewords of length $J = 16 \log_2 n$, with minimum Hamming distance at least $J/4$. Source and target label laws are uniform on complementary halves of the codebook. The code embeddings use $b(t) = 1/4 + t/4$, signed tents h_u, and $\epsilon = 1/16$: $x_u = b + \epsilon h_u$, with $y_u = b + \delta + \epsilon h_u$ for the lower construction and $y_u = b + \delta - \epsilon h_u$ for the upper construction. We use both $\delta = 0$ and $\delta = 1/4$, and $p \in \{1, 1.5, 2, 3, 4\}$, omitting the affine curved case $p = 1$. Every curve has slopes in $[3/16, 5/16]$. The exact integrated cost is an affine function of normalized Hamming distance; its amplitude is of order J^{-p} at zero contact and J^{-2} at nonzero curved contact. Finite population truths are computed by assignment, independently of the fitted empirical duals.

The empirical dual uses independent training and evaluation labels. The complete, known codebook supplies mass-free extension anchors; population label probabilities are not supplied to the trainer. Finite cost transforms on this codebook interpolate a globally feasible Lipschitz extension. This design isolates the ideal completed trainer from computational anchor error. All costs are rescaled

Table 23: Joint coverage of both endpoints over all 64 ordered policy pairs (worst across three fits), and mean fixed-pair LCB/truth (L) and UCB/truth (U). The fixed pair is prespecified as all-zero versus all-one and has positive true endpoints for every cost. Each fit has 200 repetitions; n is evaluation observations per action and stratum. Ratios measure information, while all-pair coverage measures validity.

Cost	n	All-pair cover	Emp. L	Emp. U	Pooled L	Pooled U	Uniform L	Uniform U
asymmetric	4,096	1.000	0.666	1.535	0.643	1.549	0.432	1.692
asymmetric	8,192	1.000	0.764	1.452	0.741	1.465	0.530	1.608
asymmetric	16,392	1.000	0.833	1.392	0.810	1.406	0.599	1.549
hinge	4,096	1.000	0.910	1.153	0.907	1.153	0.898	1.216
hinge	8,192	1.000	0.936	1.129	0.933	1.129	0.924	1.192
hinge	16,392	1.000	0.955	1.112	0.952	1.112	0.943	1.175
quadratic	4,096	1.000	0.666	1.505	0.642	1.517	0.431	1.668
quadratic	8,192	1.000	0.764	1.426	0.739	1.438	0.528	1.589
quadratic	16,392	1.000	0.833	1.370	0.808	1.382	0.598	1.533
tolerance	4,096	1.000	0.461	1.857	0.427	1.882	0.117	2.053
tolerance	8,192	1.000	0.619	1.723	0.584	1.748	0.274	1.919
tolerance	16,392	1.000	0.730	1.628	0.695	1.653	0.386	1.824

before optimization to avoid solver tolerances overwhelming the small contact amplitudes. Table 24 combines endpoint errors, individual smoothness, and the contact-order comparison. The support grows with sample size to retain increasing outer-law complexity. The table describes this finite sequence; the minimax guarantee follows from the testing argument in Appendix F.

As a separate observation-level stress test, an unknown randomly permuted alphabet of size $8n^2$ has a shared half-support under a null and disjoint half-supports under an alternative. We record the cross-collision test over 2,000 repetitions for sample sizes up to 4,096. Its null collision probability is at most $1/4$ by a union bound. This diagnoses one test's limited power. The larger implicit alphabet is distinct from the explicit OT codebook above.

Table 24: Combined hard-instance diagnostic. Every curve is bi-Lipschitz with slopes in $[3/16, 5/16]$. Slopes regress log mean absolute split-dual error on log code length, using four growing-support designs and 40 repetitions each; parentheses give standard errors from 1,000 stratified bootstrap resamples. These are finite-range diagnostics, not proofs of minimax optimality. All zero/curved controls are retained in the CSV.

p	Contact	$p \wedge 2$	Lower slope	Upper slope	Error/amplitude
1	zero	1	1.06 (0.05)	1.06 (0.05)	0.271–0.285
1.5	zero	1.5	1.55 (0.04)	1.56 (0.04)	0.271–0.284
2	zero	2	2.03 (0.04)	2.03 (0.04)	0.273–0.283
3	curved	2	2.04 (0.04)	2.04 (0.05)	0.275–0.282
4	curved	2	1.95 (0.05)	1.95 (0.05)	0.268–0.280

K.3.4 A DIRECT COST-ACTIVE POLYNOMIAL-RATE EXPERIMENT

For the general bounded continuous-cost result, take $c(q, r) = (\int q - \int r)^2$. Zero-mean signed-tent perturbations, with arbitrarily many nuisance bits, have exactly zero active pseudodistance. The remaining coordinate is the mean μ, and $\rho^c(\mu, \mu') \le 2|\mu - \mu'|$. A uniform mean grid therefore gives an explicit one-dimensional cover with zero tail mass. We simulate directly on this quotient, using means $\{0.2, 0.8\}$ with probabilities $(0.35, 0.65)$ and $(0.65, 0.35)$. Both split-dual endpoint errors are measured for $n = 256, 512, 2048, 8192, 32768$, with 1,000 repetitions. The archive reports the fitted polynomial slope and $\sqrt{n}$-scaled error. This demonstrates recovery under an explicit active-geometry condition for the squared-mean cost. To avoid relying only on finite atomic laws, a second experiment uses continuous mean laws $U[0.1, 0.6]$ and $U[0.3, 0.8]$, with exact endpoints 0.04 and $0.04 + 0.25/3$. At four sample sizes from 256 to 8,192, 200 repetitions evaluate the sorted empirical transport dual with a globally safe extension on an independent fold. We additionally generate monotone curves with 8, 32, 128, and 512 nuisance bits and check the cost invariance and active covering at four resolutions.

Table 25: Explicit cost-active quotient experiment: 1,000 repetitions at five sample sizes for two-atom laws, and 200 at four sizes for continuous laws. Slopes regress log mean absolute split-dual error on log sample size. The quotient admits a one-dimensional cover with zero tail mass.

Outer law	Endpoint	Polynomial slope	$\sqrt{n}$-scaled error
Two-atom	lower	0.505	0.189–0.199
Two-atom	upper	0.497	0.190–0.198
Continuous	lower	0.529	0.066–0.073
Continuous	upper	0.509	0.080–0.085

Table 26: **Local and global orders need not coincide.** Deterministic log–log slopes use the seven smallest scales, $2^{-20} \le \delta \le 2^{-14}$. The last column uses the contracted cost $h_a(0.99z)$, which excludes the singular boundary. These are geometric slopes, not fitted sample-size risk exponents.

Cost	Contact	Boundary rectangle	Endpoint width	Contracted rectangle
$h_{0.25}$	2.000	0.250	0.250	1.998
$h_{0.50}$	2.000	0.500	0.500	1.998
$h_{0.75}$	2.000	0.750	0.750	1.998
z^2	2.000	2.000	2.000	2.000

K.3.5 General convex costs with different contact and rectangular orders

The power-cost experiments above concern matching lower and upper orders. Here we separately examine the local mechanism in Theorems F.7 and F.8, and the global regularity condition in Theorem F.9, when the orders do not match. We use

$$h_a(z) = 1 - az - (1-z)^a, \qquad a \in \{0.25, 0.50, 0.75\}, \quad -1 \le z \le 1,$$

with $h(z) = z^2$ as a matching-order control. Since $h_a''(z) = a(1-a)(1-z)^{a-2} > 0$ at every interior point, the tight two-sided contact order at zero is $\beta_h = 2$; no other interior point gives a contact order below two. The contact definition also admits larger, weaker exponents. In contrast, the largest uniform rectangular order is $\gamma_h = a$. Indeed, the global a-Hölder modulus gives (U_a), whereas the boundary rectangle satisfies

$$R_a(\delta) := |h_a(1) - 2h_a(1-\delta) + h_a(1-2\delta)| = (2 - 2^a)\delta^a. \tag{290}$$

Thus no order greater than a can hold uniformly.

Local contact, global interaction, and boundary exclusion. We evaluate the symmetric contact remainder $\{h_a(\delta) + h_a(-\delta)\}/2$ and $R_a(\delta)$ for $\delta = 2^{-k}$, $k = 4, \dots, 20$. To connect the rectangular difference to the estimand, let the source outer law be uniform on the constant quantiles $\{0, \delta\}$ and the target law uniform on $\{1-\delta, 1\}$. The difference between its sharp upper and lower transport endpoints is exactly $R_a(\delta)/2$. We compute the endpoints both by transport LPs and by the two possible permutation matchings. The latter avoid loss of resolution when the two LP objectives are close. As a structural control, we repeat the rectangular calculation for $h_a((1-\varepsilon)z)$, $\varepsilon \in \{0.01, 0.1\}$, and for rectangles ending at the fixed interior point $z = 0.5$.

Table 26 separates the two mechanisms. Boundary exclusion restores a quadratic small-scale rectangle, although its constant increases as $\varepsilon \downarrow 0$. At $\delta = 2^{-20}$, using the local-curvature budget $h_a''(0)\delta^2$ for the boundary rectangle underestimates it by factors 1.49×10^{11}, 2.52×10^9, and 5.69×10^7, respectively. This diagnoses the failure of a uniform quadratic budget; it does not identify a sample-size minimax exponent.

The local difficult family retains its second-order separation. We use constant-weight codebooks with $M = 8m$, $J = 16 \log_2 m$, minimum Hamming distance $J/4$, and $m \in \{16, 32, 64, 128\}$ observations per arm in each of the independent training and evaluation samples. The signed-tent embeddings have baseline $b(t) = 1/4 + t/4$ and coefficient $1/16$, so their slopes lie in $[3/16, 5/16]$. Source and target labels are uniform on complementary halves of the codebook. If d_H denotes normalized Hamming distance, the exact cost matrices are $A_{a,J} d_H$ for the lower construction and

Table 27: **Finite-sample behavior with unmatched orders.** Boundary-exposed population, 300 repetitions; errors are mean absolute errors of globally feasible held-out estimates. Parentheses give one Monte Carlo standard error. Slack is the mean positive distance from the reported lower confidence bound to the sharp lower endpoint. All rows have joint lower/upper coverage $300/300$, with marginal 95% Wilson interval $[0.9874, 1]$.

Cost	m	Lower error	Upper error	Lower-band slack
$h_{0.25}$	1024	0.002493 (0.000103)	0.001895 (0.000077)	0.090040
$h_{0.50}$	16	0.073242 (0.002237)	0.081004 (0.001659)	0.266837
$h_{0.50}$	64	0.021717 (0.000889)	0.021762 (0.000630)	0.255607
$h_{0.50}$	256	0.005025 (0.000214)	0.004166 (0.000167)	0.122201
$h_{0.50}$	1024	0.002194 (0.000096)	0.001639 (0.000073)	0.060097
$h_{0.75}$	1024	0.001122 (0.000046)	0.000867 (0.000036)	0.030105
z^2	1024	0.003070 (0.000135)	0.002820 (0.000123)	0.120115

$A_{a,J}(1 - d_H)$ for the upper construction, where

$$A_{a,J} = \int_0^1 \frac{h_a(t/(16J)) + h_a(-t/(16J))}{2}\, dt \asymp J^{-2}.$$

The measured amplitude slope against J is -2.000 for all three costs and the quadratic control. With 100 repetitions per sample size, the lower-endpoint mean absolute errors at $m = 128$ are $(2.739 \pm 0.015) \times 10^{-9}$, $(3.652 \pm 0.020) \times 10^{-9}$, and $(2.739 \pm 0.015) \times 10^{-9}$, respectively, where $\pm$ denotes one Monte Carlo standard error. The quadratic control gives $(2.922 \pm 0.016) \times 10^{-8}$. The upper construction has identical absolute errors by the complement identity. The costs share a normalized matrix and paired draws, so they are not independent replications. Costs are rescaled before optimization; population truths are computed by assignment. Completion on the full, known finite codebook is an oracle-support diagnostic: it isolates the contact construction, but does not test an unrestricted-space attainable rate or reproduce the minimax testing-prior argument.

Globally feasible certificates under the weaker modulus. We additionally examine a fixed-population statistical experiment with one stratum, two arms, and two 32-atom outer laws. Put $t_j = (j - 1/2)/32$, $j = 1, \dots, 32$, $k = 0, \dots, 31$, and $w_k \propto \exp(-k/12)$, $v_k = w_{31-k}$. The interior scenario uses step quantiles

$$q_k(t_j) = 0.2 + 0.16k/31 + 0.1t_j, \qquad r_k(t_j) = 0.4 + 0.16k/31 + 0.1t_j.$$

The boundary-exposed scenario uses $\ell_k = 0.22(k/31)^2$ and

$$q_k(t_j) = \ell_k + 0.015t_j, \qquad r_k(t_j) = 1 - \ell_{31-k} - 0.015 + 0.015t_j.$$

Each coordinate is constant on its equal-width quantile interval, so the cost integration is exact. The latter scenario reaches displacement 0.985; exact contact with the singular boundary is tested separately by (290). We use $m \in \{16, 64, 256, 1024\}$ observations per arm per split, total sample size $4m$, and 300 repetitions per configuration.

For these non-Lipschitz costs, a value modulus valid on $[-1, 1]$ is $|h_a(z) - h_a(z')| \le L_a|z - z'|^a$, with $L_a = 1 + a2^{1-a}$. Jensen's inequality therefore bounds changes in c_{h_a} by $L_a d_1^a$. Let $H = \max h_a = 1 - a$, and shift the trained source dual coefficients α_i so that $\min_i \alpha_i = 0$. We use the globally feasible extension

$$\phi(q) = \max\{-H, \max_i[\alpha_i - L_a d_1(q, q_i)^a]\}, \qquad \psi(r) = \min_i[c_{h_a}(q_i, r) - \alpha_i].$$

Both potentials lie in $[-H, H]$. Apply the same construction to $H - c_{h_a}$ for the upper endpoint. Independent held-out means yield jointly valid one-sided bounds using $r_m = 2H\sqrt{\log(2/0.05)/m}$, clipped to $[0, H]$. For quadratic cost the *value* modulus uses exponent one and constant two, even though its rectangular order is two. We compare these extensions with empirical plug-in values and exact finite-support transforms using the same training samples. The latter oracle uses the known population support and the same conservative radius. Population endpoints are obtained from independent primal transport LPs.

Table 27 reports both endpoints for each nonmatching cost and the sample-size path for $a = 0.5$. Its lower error falls from 0.07324 to 0.002194 as m increases from 16 to 1024. The corresponding population dual gap falls from 0.07289 to 1.925×10^{-5}. Joint coverage is one in every configuration, including the interior and quadratic controls; the substantial band slack shows that this reflects conservative calibration, not precise nominal coverage. These fixed-support experiments assess implementation and finite-sample information, and are not used to fit a worst-case $(\log n)^{-\gamma_h}$ rate. Likewise, different cost scales preclude reading raw error magnitudes as a ranking of minimax difficulty.

Reproducibility and interpretation. The run uses seed 20260925 and the separate executable `convex_order_gap_study.py`; `verify_convex_order_gap.py` independently reconstructs the quantile curves and checks off-support feasibility. Direct half-block quadrature agrees with the code-cost formulas to maximum relative error 1.29×10^{-14}. An additional 278,784 off-support cost/scenario/endpoint/curve-pair checks find no positive violation; these finite checks audit the implementation, while the modulus argument supplies global feasibility. The archive under `reproducibility/results/convex_order_gap/` contains all 38,400 certificate records, 3,200 code-experiment records, codebooks, geometric scans, paired oracle comparisons, source snapshots, and manifests. Taken together, the experiments distinguish a second-order local cost separation from a weaker global interaction modulus. They do not close the gap between the $(\log n)^{-2}$ lower obstruction and the $(\log n)^{-a}$ attainable bound for this family.

K.3.6 Separating extension error from learning and simultaneous inference

This paired study examines the computational contribution in (16) for $c_h = W_2^2$. It asks whether a large certificate gap comes from estimating the outer laws or from extending the same empirical dual, and whether reducing the extension loss improves the worst policy comparison after simultaneous calibration. We compare completed anchors with the direct continuous pair (A, G_K) in (106), evaluated by the quadratic LP (108). This experiment does not evaluate the pointwise-improving closure $(\Phi_{\mathrm{cl}}, \Psi_{\mathcal{U}})$ of Proposition E.6, nor does it establish computational efficiency for all convex costs.

Population, sample budgets, and paired comparisons. There are four equiprobable covariate cells, two actions with propensity $1/2$, and all $2^4 = 16$ deterministic binary policies, giving 256 ordered policy pairs. Each cell/action outer law is uniform on 2,048 frozen, exactly observed piecewise-linear quantile curves. To construct a curve, draw a four-component Gaussian-logit mixture in $d + 1$ coordinates and apply the softmax map to obtain increments v_i. At knot $t_i = i/(d + 1)$ set

$$q(0) = 0, \qquad q(t_i) = 0.1t_i + 0.9 \sum_{j \le i} v_j, \qquad i = 1, \dots, d + 1.$$

The component templates consist of three centered Gaussian bumps and a centered alternating pattern, each with Euclidean norm 1.5; cell-specific cyclic shifts change their alignment. The logit noise has standard deviation 0.7. Mixture weights are $(0.4, 0.3, 0.2, 0.1)$ under action zero and $(0.1, 0.2, 0.3, 0.4)$ under action one, which also has an additional rank-one Gaussian perturbation of scale 0.3. Thus actions change both mixture composition and dispersion. The label d describes the latent construction, not a verified intrinsic dimension or an entropy exponent.

The primary construction uses $d = 8$ and 30 independent paired training repetitions per budget; a $d = 4$ confirmation uses 10. Individuals are sampled independently across cells, actions, and population atoms. Multinomial counts represent these draws exactly, without fixing cell/action sample sizes. The training fold uses empirical frequencies and only observed curves. Population atoms and probabilities are used separately to audit the population objective and truth. Evaluation draws are independent of training. Every method uses the same training and evaluation draws within a repetition.

The total-budget experiment starts at $n = \ell + m \in \{4096, 8192, 16384, 32768, 65536\}$, with a primary $1 : 1$ training/evaluation split; $1 : 3$ splits are also saved. A separate mechanism experiment uses $\ell \in \{4096, 8192, 16384, 32768\}$ and $m = 64\ell$ to reduce the relative inference contribution. The latter experiment has total sample size 65ℓ.

Table 28: **Extension loss and policy-uniform certificate error.** Eight-dimensional construction, 30 paired repetitions, $K = 8$, and $m = 64\ell$. $\overline{g}$ and $\overline{E}$ are defined in (291). "Continuous" uses the baseline gauge; "balanced" changes the gauge using training only. Parentheses give Monte Carlo standard errors. Balancing leaves $\overline{g}$ unchanged.

	Population gap $\overline{g}$		Uniform certificate error $\overline{E}$		
ℓ	Anchors	Continuous	Anchors	Continuous	Balanced
4096	.028522	.001880	.029043	.002457	.002135
	(.000028)	(.000009)	(.000029)	(.000018)	(.000010)
8192	.019964	.001233	.020292	.001617	.001383
	(.000026)	(.000005)	(.000028)	(.000009)	(.000007)
16384	.011229	.000743	.011443	.001003	.000841
	(.000015)	(.000003)	(.000015)	(.000007)	(.000005)
32768	.004128	.000420	.004295	.000603	.000488
	(.000018)	(.000001)	(.000018)	(.000005)	(.000003)

All methods use $K = 8$ exact quantile-bin averages and the same feasible empirical OT dual. The baseline adds a complete d_K-net of radius $\eta = 1/4$, consisting of $\binom{12}{8} = 495$ monotone grid points, to the observed source anchors. The continuous method changes only the extension. Its primary comparison uses the identical additive gauge. We also evaluate a training-only balanced gauge: subtract $\delta = (\widehat{P}_0\Phi - \widehat{P}_1\Psi)/2$ from the source potential and add it to the target potential. This leaves the population dual objective unchanged while changing the IPW score variance. The same balancing ablation is retained for the anchor method. No method is selected using its evaluation-fold certificate.

Policy-uniform metrics and calibration. Let $\mu_b(\pi, \pi')$ be the population mean of the learned score in repetition b. We first take the maximum over all 256 ordered pairs and then average over repetitions:

$$\overline{g} = \frac{1}{R}\sum_{b=1}^{R} \max_{\Pi^2}\{\Theta^L_{W_2^2} - \mu_b\},$$
$$\overline{E} = \frac{1}{R}\sum_{b=1}^{R} \max_{\Pi^2}\{\Theta^L_{W_2^2} - L_b\}_+. \tag{291}$$

The first metric excludes evaluation noise and the confidence radius; it still includes learning, binning, and extension error. The second matches the policy-uniform certificate-error quantity in the main bound. Truth is bracketed by primal and feasible-dual solutions of the population OT problem using exact integrals of the squared differences of the original piecewise-linear curves.

The tables use empirical Bernstein calibration at $\alpha = 0.05$ with an analytic score width W valid on the entire quantile cone. For each pair the radius is

$$r_B = \sqrt{\frac{2\widehat{v}\log(2 \cdot 256/\alpha)}{m}} + \frac{7W\log(2 \cdot 256/\alpha)}{3(m-1)}, \tag{292}$$

where $\widehat{v}$ is the unbiased sample variance. The union factor 256 is used even for the separately reported fixed contrast. Ranges are not estimated from the observed support. The continuous implementation additionally pays $2\tau\sqrt{2\log(2 \cdot 256/\alpha)/(m-1)}$, with $\tau = 10^{-7}$, to account for query-order-dependent numerical errors under warm starts. Indeed, the returned IPW score is below the fixed exact-transform score by at most 2τ, and their sample standard deviations differ by at most $2\tau\sqrt{m/(m-1)}$. The original envelope-based Hoeffding radius is reported as a separate comparison.

What changes when the extension is replaced. Table 28 shows a reduction of roughly 90–94% in the population gap while holding the empirical learner and gauge fixed. This isolates a substantial computational-extension loss; the old gap cannot be attributed entirely to statistical learning. The improvement persists for the worst policy comparison after simultaneous calibration. The residual gap decreases with training size, but these fixed-K curves are not a test of the asymptotic minimax rate or of the prescribed joint precision schedule.

Table 29: **Informativeness after simultaneous calibration.** Continuous transform with the training-balanced gauge, $d = 8$ and $m = 64\ell$. The radius share averages $r_{\max}/(g_{\max} + r_{\max})$ within repetitions; it is a diagnostic, not an exact decomposition of the signed final error.

ℓ	m	Fixed LCB/truth	Positive nonzero pairs	Radius share
4096	262144	.203	.601	.121
8192	524288	.480	.994	.113
16384	1048576	.685	1.000	.118
32768	2097152	.815	1.000	.134

Table 30: **Equal total sample budgets.** Mean worst-pair certificate error under the $1:1$ split, $d = 8$, 30 paired repetitions, and the same simultaneous Bernstein calibration. The last column is the mean fixed contrast LCB for the balanced continuous method.

$n = \ell + m$	Anchors	Continuous	Balanced	Fixed LCB
4096	.065544	.023120	.018286	-.015242
8192	.043609	.013128	.009821	-.007053
16384	.028147	.007467	.005352	-.002691
32768	.015917	.004373	.002910	-.000275
65536	.006952	.002691	.001674	.000943

For interpretation, the prespecified all-zero versus all-one policy contrast has true lower endpoint approximately 0.0026113. Table 29 reports its retained fraction along with a class-wide statistic: the fraction of the 240 nonzero-truth pairs with positive certificates. At the two largest training budgets, all these pairs have positive certificates in every one of the 30 repetitions. This does not imply a common lower bound on the retained fraction for every pair; near-zero endpoints make relative tightness a different question.

Budget and calibration limitations. Under equal total budgets, the uniform error also decreases markedly (Table 30), but the smallest budgets do not give informative positive certificates. At $n = 65536$, the balanced fixed contrast retains only 36.1% of its truth under the $1:1$ split, compared with 81.5% when $\ell = 32768$ and $m = 2097152$. The larger evaluation budget accounts for part of this difference. With the original envelope $B_S = 8$, the Hoeffding radius is $8\sqrt{2\log(256/0.05)/m}$. It is still 0.022832 at the largest $m = 2097152$, and the unbalanced continuous fixed-contrast LCB is -0.020627. This radius is too large to give a positive certificate at that budget. In the $d = 4$ confirmation, the balanced fixed-contrast retained fraction rises from 60.9% at $\ell = 4096$ to 90.3% at $\ell = 32768$ under $m = 64\ell$; this is supporting evidence from only 10 repetitions per budget.

Figure 12 displays every paired repetition in the primary budget comparisons, including the $d = 4$ confirmation.

Coverage, numerical checks, and reproducibility. There were no simultaneous coverage failures in the reported configurations. Thirty successful repetitions give a Wilson 95% interval with lower endpoint approximately 0.886; ten give 0.722. Thus the simulations do not themselves establish coverage of at least 95%. Different budgets, radii, and algorithms sharing a repetition are not additional independent coverage trials.

Each target potential is a weak-duality lower bound, rather than the LP primal objective. The maximum accepted transform bracket width was 7.47×10^{-8}, below $\tau = 10^{-7}$. Three strict numerical re-solves were recorded without relaxing this threshold. Independent LP checks, off-support monotone probes, and full frozen-population feasibility checks passed. The cached anchor baseline reproduces the original experiment exactly on the same repetition indices; the population gauge-invariance discrepancy is at most 1.05×10^{-16}. These floating-point checks complement the algebraic validity argument, and are not a formal interval-arithmetic proof.

The executable drivers are `certificate_total_error_study.py` and `continuous_transform_study.py`; the continuous solver is `quadratic_continuous_certificate.py`. The common seed is 202609251. Under `reproducibility/results/`

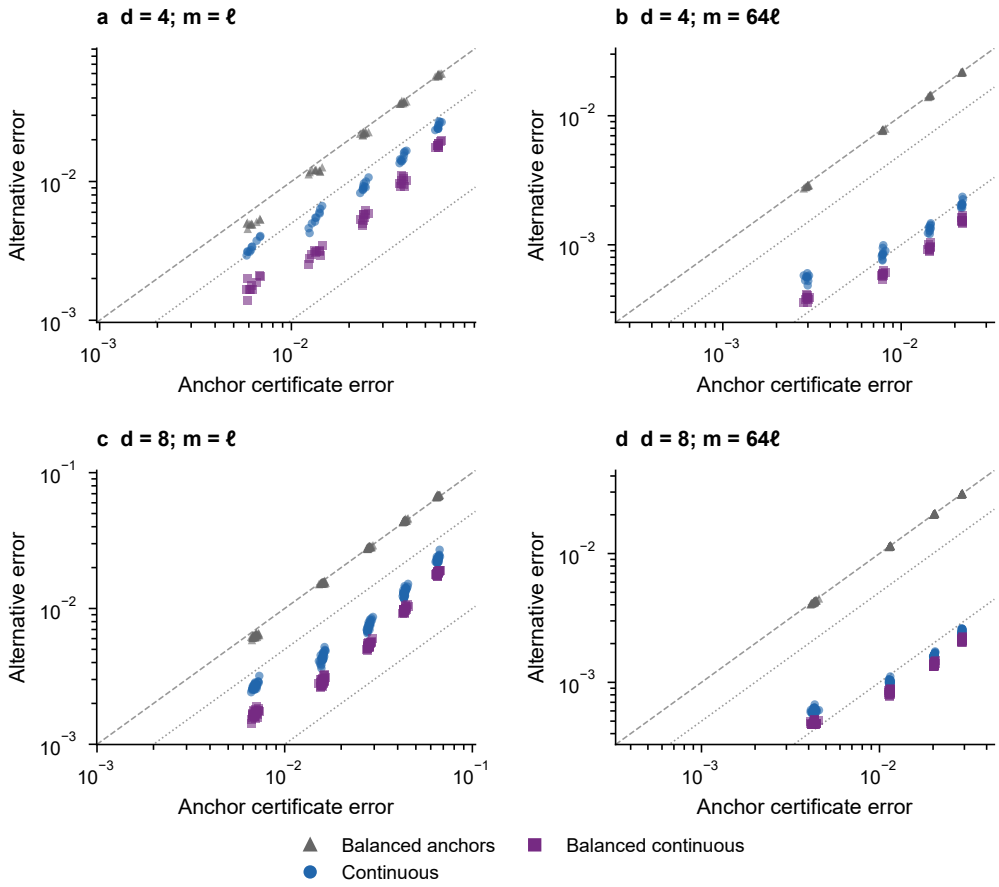


Figure 12: **Continuous extensions reduce paired certificate errors.** Each point pairs a method's worst-policy certificate error with the unbalanced anchor error on identical observations, using simultaneous Bernstein calibration. Every recorded budget in the two displayed protocols is included: 10 repetitions per budget for $d = 4$ and 30 for $d = 8$. Reference lines indicate equal error and twofold and tenfold reductions. The $m = 64\ell$ protocol spends 65ℓ observations in total, whereas $m = \ell$ spends 2ℓ; these panels must not be interpreted as equal-budget comparisons. Shared anchor records across methods are not independent trials.

`certificate_total_error/`, the directories `continuous_d8` and `continuous_d4` contain all repetitions, score arrays, manifests, and source snapshots; `continuous_report` contains paired effects and standard errors. Cached repeated curves make evaluation of this finite population inexpensive; these timings must not be extrapolated to a continuous population with a distinct LP for every evaluation curve. General convex-cost closure implementations and their separate small mechanism checks are provided in `convex_anchor_closure.py` and `convex_continuous_certificate.py`; they are not counted as additional repetitions of this W_2^2 study.

K.4 Finite bags and measurement error

What changes when an individual's distribution is observed through a finite bag rather than exactly?

K.4.1 Finite-bag design

To illustrate Proposition J.1, a treated bag contains $K \sim \mathrm{Binomial}(m, 0.35)$ ones and $m - K$ zeros. Its squared distance from δ_0 is $(K/m)^2$, so the naive large-n limit and bias are known analytically:

$$\mathbb{E}(K/m)^2 = 0.1225 + 0.2275/m, \qquad \mathrm{Bias} = 0.2275/m.$$

For each pair (n, m), 400 independent data sets are generated. The measurement-aware lower bound is

$$\frac{1}{n}\sum_{i=1}^{n}(K_i/m)^2 - 2\min\left\{1, \sqrt{\frac{\log(2n/0.025)}{2m}}\right\} - \sqrt{\frac{\log(1/0.025)}{2n}}.$$

The factor two is the certified W_1-Lipschitz constant of squared W_1 on $\mathcal{P}([0,1])$. This direct-sampling experiment is the one-stratum special case of the high-probability measurement-aware score; a union bound gives total error probability at most 0.05. It isolates the inner correction from treatment weighting.

The full two-budget surface remains available in the reproducibility artifacts. It is a diagnostic for Proposition J.1, whereas the main estimator evidence above uses exactly observed generated distributions and an informative prespecified contrast.

K.4.2 Two-budget finite-bag correction and nonidentification

For Bernoulli individual distributions ν_θ, use $W_1(\nu_\theta, \nu_{\theta'}) = |\theta - \theta'|$. The source parameters are $\{0.1, 0.7\}$ and target parameters $\{0.3, 0.9\}$, each with equal probabilities, giving true endpoints 0.2 and 0.6. Training uses independent observed bags and a prespecified two-cluster fit. Evaluation uses observed empirical Bernoulli parameters only. We cross $n \in \{256, 1024, 4096\}$ with bag sizes $m \in \{16, 256, 4096\}$ and compare omission of measurement correction, the simultaneous DKW correction, and an aggregate correction.

For this independent, equal-bag-size design, the aggregate correction has a direct justification. Conditional on the latent parameters, $\mathbb{E}|\widehat{\theta} - \theta| \le 1/(2\sqrt{m})$. Let Z be the sum of the two arm-average absolute measurement errors. Then $\mathbb{E}Z \le 1/\sqrt{m}$, and changing any of the $2nm$ independent bag observations changes Z by at most $1/(nm)$. The bounded-differences inequality gives

$$Z \le \frac{1}{\sqrt{m}} + \sqrt{\frac{\log(1/\beta)}{nm}}$$

with probability at least $1-\beta$. Since both endpoint potentials are one-Lipschitz, this one event controls both measurement errors. We allocate 0.025 to measurement and 0.025 to joint outer sampling error. This specialized correction does not replace the general heterogeneous-bag IPW formula. The naive intervals also cover in every repetition. The comparison measures the information retained by the two justified corrections. The separate construction below diagnoses the general identification limit. Finally, the Legendre construction in Proposition J.1 is checked at bag sizes $1, 2, 4, 8$: integrated bag probabilities coincide while the latent endpoints differ, verifying the nonidentification construction.

Figure 13 displays the full outer/inner budget grid.

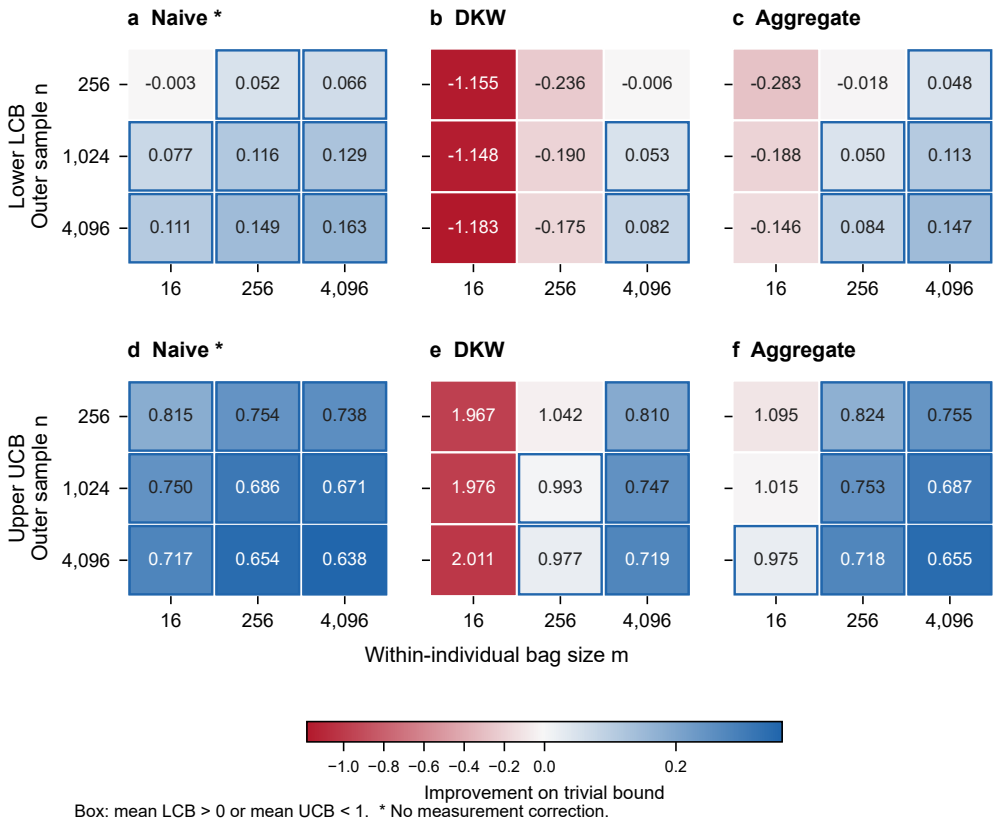


Figure 13: **Outer and within-individual budgets have different effects.** All 27 configurations use 200 repetitions each. Cell text gives the mean raw lower or upper bound, without clipping; the true endpoints are 0.2 and 0.6. Color shows improvement over the trivial lower bound zero or upper bound one. Outlined cells have positive mean lower bounds or mean upper bounds below one. Naive (*) omits measurement correction; it also covers in every recorded repetition. Aggregate and DKW are the two justified measurement corrections for this design.

Table 31: Two-endpoint finite-bag experiment, $n = m = 4096$, 200 repetitions. The true endpoints are 0.2 and 0.6. Naive omits measurement correction; aggregate uses the independently justified average-error bound for equal-size independent bags.

Correction	Joint coverage	Lower	Upper	Measurement budget
aggregate	1.000	0.147	0.655	0.016
dkw	1.000	0.082	0.719	0.081
naive	1.000	0.163	0.638	0.000

K.5 Reproducibility and formal verification

Executable record The complete source archive contains the simulation and semi-synthetic workflows, package-version manifests, and generated records. Figures are generated directly from saved CSV files and exported as vector PDF and SVG plus 300-dpi PNG. Their encodings use redundant color, marker, and line style where applicable; no smoothing alters any reported numerical value. Re-running a workflow overwrites generated results only inside `reproducibility/results/` and the corresponding revision figure stems.

Lean-checked finite foundations The accompanying `LeanProofs/` project checks selected mathematical components. With Lean 4.33.0 and mathlib 4.33.0, `lake build` kernel-checks finite-support Kantorovich weak duality, the finite-action IPW expectation identity, squared and absolute finite Cauchy–Schwarz, and the pointwise and weighted finite-bag Lipschitz corrections. The axiom audit reports only mathlib's standard `propext`, `Classical.choice`, and `Quot.sound`; the project contains no `sorry`, `admit`, `unsafe`, or custom axiom.

The theorem-by-theorem coverage record is in `LeanProofs/COVERAGE.md`. Strong conditional Kantorovich duality, the model-specific measurable selections and coverage guarantees, empirical-process arguments, and minimax lower bounds rely on the conventional proofs in the preceding appendices.