CATE Estimation or Empirical Welfare Maximization? A Statistical Analysis of Causal Decision Making under a Hard Budget

Anonymous authors
Paper under double-blind review

Abstract

Causal decision making under limited resources requires choosing whom to treat while respecting operational constraints. Two widely studied approaches to treatment allocation are (i) first estimating the conditional average treatment effect (CATE) and then using the estimated effects to determine a feasible allocation, and (ii) directly selecting an allocation by empirical welfare maximization (EWM). Their relative statistical performance, however, remains unclear when treatment costs are heterogeneous, and every deployment batch must satisfy a hard budget. To bridge this gap, we study the statistical comparison between CATE estimation and EWM for finite-batch treatment allocation under a hard budget. We develop a common framework that relates treatment effect estimation to allocation regret and characterizes the local regret of the two approaches under correct specification and model misspecification. Our analysis shows that CATE estimation can achieve lower regret by pooling information under correct specification, while misspecification can reverse the comparison in favor of EWM, with an explicit threshold separating the two cases. We extend these results to observational data using cross-fitted doubly robust scores. Finite-sample experiments confirm the predicted variance advantage and the reversal under misspecification. Source code is available at `https://anonymous.4open.science/r/STEM-B0E3`.

1 Introduction

Causal decision making concerns how treatments should be allocated using pretreatment covariates to maximize expected welfare (Kitagawa and Tetenov, 2018; Athey and Wager, 2021). In many applications, treatment is costly and available resources are limited. When a budget must be satisfied for every deployment batch, treatment decisions are coupled across units. Treating one costly unit may rule out treating several lower-cost units, so the decision maker must choose a feasible allocation jointly for the batch (Bhattacharya and Dupas, 2012; Xu et al., 2024; Cerulli, 2026).

Current causal inference studies mainly focus on two approaches to address such treatment allocation problems. The first estimates heterogeneous treatment effects and then uses the estimated effects to determine treatment decisions. A prominent example is conditional average treatment effect (CATE) estimation, which learns the expected treatment effect conditional on pretreatment covariates (Künzel et al., 2019; Kennedy, 2023), and then it can be used to determine the downstream treatment allocation problem. The second learns treatment policies by directly maximizing empirical welfare, as in empirical welfare maximization (EWM) (Kitagawa and Tetenov, 2018; Athey and Wager, 2021). Resource constraints have also been incorporated into causal treatment allocation. Existing work studies welfare maximization under budget constraints, individualized treatment rules under limited resources, treatment effect ranking under capacity constraints, and allocation with heterogeneous treatment costs (Bhattacharya and Dupas, 2012; Sun, 2026; Xu et al., 2024; Kamran et al., 2024; Arno et al., 2026; Cerulli, 2026; Haghi et al., 2026). Policy evaluation has also been developed for settings in which resources are allocated jointly across units (Mate et al., 2023).

Existing studies have shown how treatment allocations under resource constraints can be formulated, learned, and evaluated. *What remains less understood is how CATE estimation followed by allocation and empirical welfare maximization differ statistically when a hard budget couples treatment decisions*

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.

across units. Recent work begins to compare these two approaches in policy learning (Kato, 2025), but the role of hard budget constraints and heterogeneous treatment costs in this comparison remains unclear. Under such constraints, feasible allocations are compared through differences in total treatment benefit. A structured CATE model may pool information across covariate strata when estimating these differences, while observations from strata on which competing allocations agree need not enter their empirical welfare difference. As a result, the two approaches can exhibit different variance and misspecification behavior even when they use the same training sample and the same feasible set.

The different use of the same training data creates a tradeoff between variance reduction and misspecification bias. By pooling observations through a shared model, CATE estimation can improve the precision of the treatment effect contrasts that determine the allocation. If the shared model is misspecified, however, the same pooling can bias these contrasts. Whether this improves or worsens the resulting allocation therefore depends on how the gain in precision and the induced bias translate into decision quality. We measure this by *allocation regret*, defined as the expected value loss of the learned allocation relative to the optimal feasible allocation. This leads to our main investigation:

> *How does the tradeoff between variance reduction and CATE model misspecification affect the relative allocation regret of CATE estimation and empirical welfare maximization under a hard budget?*

In this paper, we study the statistical comparison between CATE estimation followed by allocation and empirical welfare maximization for treatment allocation to a finite deployment batch under a hard budget. We first identify the value of a batch allocation under standard causal assumptions and relate CATE estimation error to allocation regret, providing a common basis for comparing the two approaches. We then analyze a tractable knapsack model with heterogeneous treatment costs that makes their statistical difference explicit. CATE estimation can pool observations through a shared model, whereas EWM uses observations according to their contribution to welfare differences between feasible allocations. This distinction can reduce variance when the CATE model is correctly specified, but can introduce bias under misspecification. We characterize the resulting local allocation regret, derive a threshold at which the regret ordering reverses, and extend the analysis to observational data using cross-fitted doubly robust scores. Finite-sample experiments with oracle and estimated scores reproduce the predicted variance advantage under correct specification and the reversal under misspecification.

Our main **contributions** are summarized as follows: (1) We formulate the statistical comparison between CATE estimation followed by allocation and EWM for finite-batch treatment allocation under a hard budget, and connect CATE estimation error to allocation regret. (2) We characterize how correct specification and model misspecification affect the local allocation regret of the two approaches in a knapsack model with heterogeneous treatment costs. (3) We extend the comparison to observational data using cross-fitted doubly robust scores and verify the theoretical predictions in finite-sample experiments.

2 Related Work

Treatment-effect estimation and policy learning. Standard policy learning uses observed covariates to assign treatment to maximize policy value (Kitagawa and Tetenov, 2018; Athey and Wager, 2021). Plug-in methods estimate treatment effects before choosing actions, whereas empirical welfare maximization (EWM) directly optimizes estimated policy value (Kitagawa and Tetenov, 2018). Doubly robust policy learning combines outcome regression and propensity-score estimation (Athey and Wager, 2021; Foster and Syrgkanis, 2023). Margin conditions connect estimation near decision boundaries to policy regret (Luedtke and Chambaz, 2017). For fixed treatment capacity, treatment-effect ranking methods learn the ordering needed to prioritize a limited number of units (Kamran et al., 2024; Arno et al., 2026). Kato (2025) further shows that least-squares CATE estimation and empirical welfare maximization can coincide after reparameterization. This connection makes the empirical criterion and its parameterization central to the comparison. We therefore study how specified CATE-based and EWM criteria differ in local allocation regret under a common score class and a common hard-budget allocation rule.

Budget-constrained causal allocation. Resource constraints have long been incorporated into welfare-maximizing treatment assignment. Bhattacharya and Dupas (2012) study treatment assignment under an aggregate budget, and Athey and Wager (2021) allow policy classes to encode budget and other application constraints. Sun (2026) studies empirical welfare maximization when policy cost is itself estimated. With heterogeneous individual treatment costs, Cerulli (2026) obtains a knapsack-type policy-learning problem together with Lagrangian and ranking-based solutions. Capacity constraints also motivate end-to-end learning through dual prices (Haghi et al., 2026). For equal-cost capacity constraints, direct ranking methods target the treatment-effect ordering needed by the allocation (Kamran et al., 2024; Arno et al., 2026). These works establish how resource constraints change the feasible decision and how the resulting allocation can be learned or optimized.

Group-level policy value and the learning criterion. When a policy allocates resources jointly across several units, policy evaluation must preserve the coupling created by the allocation rule. Mate et al. (2023) study policy evaluation for algorithmic resource allocation in randomized trials. In our observational setting, a batch policy depends jointly on m covariate draws, so its value is represented by an order-m kernel built from standard doubly robust scores (Chernozhukov et al., 2018; Athey and Wager, 2021); complete U-statistics (Hoeffding, 1948) average this kernel over distinct sample tuples. Our analysis uses this representation to compare CATE estimation and EWM at the same allocation boundary. The shared treatment-effect regression can pool observations linked through a common parameter, while the EWM batch-value contrast retains observations according to where the candidate policies differ.

3 Problem Setup

Data and causal identification: We observe n independent and identically distributed observations:

$$\mathcal{D}_n = \{O_i = (X_i, A_i, Y_i)\}_{i=1}^n,$$

where $X \in \mathcal{X}$ denotes pretreatment covariates, $A \in \{0, 1\}$ is treatment, and $Y \in \mathbb{R}$ is the observed outcome. Let $Y(1)$ and $Y(0)$ denote the potential outcomes under treatment and control, respectively.

Assumption 1 (Standard causal assumptions)**.** *The observed data satisfy: (1)* ***Consistency:*** $Y = AY(1) + (1 - A)Y(0)$*; (2)* ***Conditional ignorability:*** $\{Y(0), Y(1)\} \perp A \mid X$*; (3)* ***Overlap:*** *there exists* $\eta \in (0, 1/2)$ *such that* $\eta \leq e(X) := \mathbb{P}(A = 1 \mid X) \leq 1 - \eta$ *almost surely; and (4) the stable unit treatment value assumption, excluding interference and multiple versions of treatment.*

Then, define the outcome regression functions and the CATE as follows:

$$\mu_a(x) = \mathbb{E}[Y \mid A = a, X = x], \qquad \tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x].$$

Under Assumption 1, $\tau(x) = \mu_1(x) - \mu_0(x)$. Writing $\eta_0 = (\mu_0, \mu_1, e)$, we define the following doubly robust score (Kennedy, 2023):

$$\Gamma(O; \eta_0) = \mu_1(X) - \mu_0(X) + \frac{A}{e(X)}\{Y - \mu_1(X)\} - \frac{1 - A}{1 - e(X)}\{Y - \mu_0(X)\}. \tag{1}$$

Here, this score has conditional mean $\mathbb{E}[\Gamma(O; \eta_0) \mid X] = \tau(X)$. The main local analysis first uses the oracle score $\Gamma_i = \Gamma(O_i; \eta_0)$; Theorem 4 replaces η_0 by cross-fitted nuisance estimates.

Target deployment batch: The training covariates and future deployment covariates are drawn from the same target marginal distribution P_X. A deployment batch contains m independent draws $X_{1:m} = (X_1, \ldots, X_m) \sim P_X^m$ and is independent of the training sample.

Assumption 2 (Target deployment population)**.** *The training covariate marginal equals* P_X*, and each future deployment batch is an independent draw from* P_X^m*.*

Hard-budget allocation: Each unit with covariates x has a deterministic pretreatment cost $c(x) > 0$, and the batch has a hard budget $B > 0$. For a realized batch $x_{1:m}$, define the feasible allocation set:

$$\mathcal{Z}_B(x_{1:m}) = \left\{ z \in \{0, 1\}^m : \sum_{j=1}^m c(x_j) z_j \leq B \right\}. \tag{2}$$

For a score $f : \mathcal{X} \to \mathbb{R}$, evaluate feasible allocations using the following normalized objective:

$$Q_f(z; x_{1:m}) = \frac{1}{m} \sum_{j=1}^{m} z_j f(x_j).$$

Maximizing this objective defines the common exact decoder:

$$D_B(f; x_{1:m}) \in \underset{z \in \mathcal{Z}_B(x_{1:m})}{\arg\max} \; Q_f(z; x_{1:m}), \tag{3}$$

with a fixed deterministic tie-breaking rule. A score class $\mathcal{F}$ induces the batch policy by:

$$\pi_f(x_{1:m}) = D_B(f; x_{1:m}), \qquad \Pi_{\mathcal{F}} = \{\pi_f : f \in \mathcal{F}\}.$$

Policy value and allocation regret: For a feasible batch policy $\pi : \mathcal{X}^m \to \{0,1\}^m$, define its value as:

$$V_m(\pi) = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^{m} \pi_j(X_{1:m}) \tau(X_j)\right]. \tag{4}$$

The oracle policy is $\pi_\tau = D_B(\tau; \cdot)$. We measure the allocation regret of π relative to this oracle:

$$\mathcal{R}_m(\pi) = V_m(\pi_\tau) - V_m(\pi). \tag{5}$$

Since $\Pi_{\mathcal{F}}$ is induced by $\mathcal{F}$, the following identity holds:

$$\sup_{f \in \mathcal{F}} V_m(\pi_f) = \sup_{\pi \in \Pi_{\mathcal{F}}} V_m(\pi).$$

The proof is given in Section A. Our statistical comparison therefore keeps $\mathcal{F}$, D_B, the observed sample, and compares how CATE estimation and EWM select between the same feasible batch policies.

4 Statistical Comparison of CATE Estimation and EWM

This section compares CATE-based allocation and EWM in terms of local allocation regret. We first identify the value of a finite-batch policy and record how score error can change a hard-budget allocation. These results establish the quantities that the two learning criteria seek to estimate. We then construct a four-type knapsack problem and derive an exact reduction of the two empirical criteria, which reveals the observations entering each allocation decision. Using this reduction, we compare local regret under correct CATE specification and misspecification, then extend the analysis to cross-fitted doubly robust scores.

4.1 Causal Value and Allocation Sensitivity

For a fixed feasible policy π depending only on $X_{1:m}$, define the following order-m kernel:

$$h_\pi(O_1, \ldots, O_m) = \frac{1}{m} \sum_{j=1}^{m} \pi_j(X_{1:m}) \Gamma(O_j; \eta_0). \tag{6}$$

Proposition 1 (Batch policy-value identification)**.** *Under Assumptions 1 and 2, every fixed feasible batch policy π depending only on $X_{1:m}$ satisfies the following value identity:*

$$V_m(\pi) = \mathbb{E}[h_\pi(O_1, \ldots, O_m)]. \tag{7}$$

Proposition 1 identifies the deployment value of a batch policy from the observed-data law. The coupling among the m treatment decisions appears inside the kernel through $\pi(X_{1:m})$, while each summand uses a standard doubly robust score whose conditional mean is the CATE. This representation yields the following complete ordered U-statistic:

$$\widehat{V}_m^U(\pi) = \frac{1}{(n)_m} \sum_{(i_1, \ldots, i_m) \text{ distinct}} h_\pi(O_{i_1}, \ldots, O_{i_m}), \qquad (n)_m = n(n-1)\cdots(n-m+1). \tag{8}$$

Every distinct tuple has the same law as $(O_1, \ldots, O_m)$, so $\widehat{V}_m^U(\pi)$ is unbiased for $V_m(\pi)$. The disjoint-block estimator and finite-class concentration bounds are given in Section B.

To connect score estimation to allocation quality, we next bound the decoder's sensitivity to score error. For a batch $x_{1:m}$, define the set of optimal feasible allocations:

$$\mathcal{Z}_B^*(x_{1:m}) = \operatorname*{arg\,max}_{z \in \mathcal{Z}_B(x_{1:m})} Q_\tau(z; x_{1:m})$$

The smallest positive gap between an optimal and a suboptimal allocation is defined as follows:

$$G_m(x_{1:m}) = \begin{cases} \min\limits_{z \notin \mathcal{Z}_B^*} \left\{ \max\limits_{z^* \in \mathcal{Z}_B^*} Q_\tau(z^*) - Q_\tau(z) \right\}, & \mathcal{Z}_B^* \neq \mathcal{Z}_B, \\ +\infty, & \mathcal{Z}_B^* = \mathcal{Z}_B, \end{cases} \tag{9}$$

Then, we can define the normalized Hamming diameter and its essential supremum:

$$\Delta_B(x_{1:m}) = \max_{z, z' \in \mathcal{Z}_B(x_{1:m})} \frac{\|z - z'\|_1}{m}, \qquad \bar{\Delta}_B = \operatorname{ess\,sup} \Delta_B(X_{1:m}). \tag{10}$$

Proposition 2 (Finite-batch perturbation bound)**.** *Fix a batch. Let $z_f = D_B(f; x_{1:m})$ and $z^* \in \mathcal{Z}_B^*(x_{1:m})$. If $\|f - \tau\|_\infty \leq \varepsilon$, the batch regret satisfies the following bound:*

$$0 \leq Q_\tau(z^*) - Q_\tau(z_f) \leq \varepsilon \frac{\|z^* - z_f\|_1}{m} \leq \Delta_B(x_{1:m})\varepsilon. \tag{11}$$

If z_f is suboptimal under τ, the following gap condition holds:

$$0 < G_m(x_{1:m}) \leq \Delta_B(x_{1:m})\varepsilon. \tag{12}$$

Proposition 2 shows that score error changes the batch decision only when it is large enough relative to a feasible-allocation value gap. Suppose the following margin condition holds for some $C_m < \infty$, $t_0 > 0$, and $\alpha \geq 0$:

$$\mathbb{P}(0 < G_m(X_{1:m}) \leq t) \leq C_m t^\alpha, \qquad 0 < t \leq t_0. \tag{13}$$

For $\|f - \tau\|_\infty \leq \varepsilon$ with $\bar{\Delta}_B \varepsilon \leq t_0$, this gives the following allocation-regret bound:

$$\mathcal{R}_m(\pi_f) \leq C_m \bar{\Delta}_B^{1+\alpha} \varepsilon^{1+\alpha}. \tag{14}$$

The proof and the corresponding conditional statement for a random score estimator are in Section C. With the causal target and the score-to-allocation connection established, we now compare how the two empirical criteria estimate the same allocation boundary.

4.2 Which Observations Enter the Allocation Comparison?

Consider four covariate types $X \in \{E, L_1, L_2, N\}$, with probabilities $p_E, p_1, p_2, p_N > 0$ summing to one. Deployment uses batch size $m = 3$, hard budget $B = 2$, and the following treatment costs:

$$c(E) = 2, \qquad c(L_1) = c(L_2) = c(N) = 1. \tag{15}$$

Choose $u > g > 0$ and $v \in (0, u - g)$, and define the baseline treatment effects as follows:

$$b(E) = 2u, \quad b(L_1) = u + g, \quad b(L_2) = u - g, \quad b(N) = v. \tag{16}$$

Then, to link types E and N through a common parameter, choose the following loadings:

$$\psi(E) = r > 0, \qquad \psi(L_1) = \psi(L_2) = 0, \qquad \psi(N) = \kappa \neq 0. \tag{17}$$

These functions specify the working score class:

$$\mathcal{F} = \{f_\theta = b + \theta\psi : |\theta| \leq \bar{\theta}\}. \tag{18}$$

The fixed functions b and ψ describe the local model, while the scalar θ links the treatment effects of E and N. To isolate this comparison, choose $\bar{\theta}$ so that all scores remain positive, N remains below L_2, and the decoder changes with the sign of θ only for batches containing one E, one L_1, and one L_2. The following condition is sufficient:

$$\bar{\theta} M_\psi < \min\{v,\ u - g - v,\ u - g,\ 2g,\ 2(u - v)\}, \quad M_\psi = \max\{r, |\kappa|, |r - \kappa|, |r - 2\kappa|\}. \tag{19}$$

On the critical batch, the bundle contrast satisfies the following identity:

$$f_\theta(E) - f_\theta(L_1) - f_\theta(L_2) = \theta r. \tag{20}$$

Thus π_+, induced by $\theta > 0$, treats E, while π_-, induced by $\theta < 0$, treats L_1 and L_2. The two policies agree on every other batch composition; Section D verifies this decoder property.

We first assume correct specification of the shared treatment-effect model:

$$\tau_\theta = b + \theta\psi.$$

To obtain explicit local comparisons, consider the Gaussian location experiment for the oracle scores:

$$\Gamma_i = \tau_\theta(X_i) + \varepsilon_i, \qquad \varepsilon_i \mid X_i = t \sim N(0, \sigma_t^2), \tag{21}$$

independently across observations. The Gaussian specification makes the local regret constants explicit. Section I constructs a randomized potential-outcome model that generates (21) exactly.

The first criterion fits the CATE model by unweighted squared score loss:

$$\widehat{\theta}_R \in \arg\min_{|\theta|\le\bar{\theta}} \frac{1}{n}\sum_{i=1}^{n}\{\Gamma_i - b(X_i) - \theta\psi(X_i)\}^2, \qquad \widehat{\pi}_R = \pi_{\mathrm{sign}(\widehat{\theta}_R)}. \tag{22}$$

The second criterion is that EWM maximizes the complete-U estimate of the batch value:

$$\widehat{\theta}_D \in \arg\max_{|\theta|\le\bar{\theta}} \widehat{V}_3^U(\pi_\theta), \qquad \widehat{\pi}_D = \pi_{\mathrm{sign}(\widehat{\theta}_D)}. \tag{23}$$

Proposition 3 (Exact reduction of the two empirical criteria)**.** *For the four-type model above, unweighted CATE selects its policy according to the sign of the following statistic:*

$$S_R = \sum_{i=1}^{n}\psi(X_i)\{\Gamma_i - b(X_i)\}. \tag{24}$$

Let N_t and $\bar{\Gamma}_t$ denote the sample count and score mean for type t. The complete-U value difference admits the following exact expression:

$$\widehat{V}_3^U(\pi_+) - \widehat{V}_3^U(\pi_-) = \frac{2N_EN_1N_2}{(n)_3}\left(\bar{\Gamma}_E - \bar{\Gamma}_1 - \bar{\Gamma}_2\right). \tag{25}$$

Thus, on $N_EN_1N_2 > 0$, the EWM criterion is determined by the sign of the following contrast:

$$S_D = \bar{\Gamma}_E - \bar{\Gamma}_1 - \bar{\Gamma}_2. \tag{26}$$

Proposition 3 is the key reduction for the rest of the paper. Since $\psi(E)$ and $\psi(N)$ are nonzero, type N contributes to the CATE statistic S_R. The EWM statistic S_D contains only E, L_1, L_2, because every batch involving N receives the same allocation under π_+ and π_- and therefore cancels from the empirical value difference. We next compare their local allocation regret under correct CATE specification.

4.3 Correct Shared Model: The Gain from Pooling Across Types

Consider local alternatives around the allocation boundary (Hirano and Porter, 2009),

$$\theta_{s,n} = s\frac{h}{\sqrt{n}}, \qquad s \in \{-1, +1\}, \quad h > 0. \tag{27}$$

At this scale, the two candidate policies remain statistically distinguishable with nondegenerate error probabilities. For the variance comparison, introduce the following quantities:

$$A = p_Er^2, \qquad C = p_N\kappa^2, \qquad T_D = \frac{\sigma_E^2}{p_E} + \frac{\sigma_1^2}{p_1} + \frac{\sigma_2^2}{p_2}, \tag{28}$$

where $\sigma_1^2 = \sigma_{L_1}^2$ and $\sigma_2^2 = \sigma_{L_2}^2$. The unweighted CATE and EWM criteria have the local signal-to-noise indices:

$$I_R = h\frac{A + C}{\sqrt{A\sigma_E^2 + C\sigma_N^2}}, \qquad I_D = h\frac{r}{\sqrt{T_D}}, \tag{29}$$

and the population value gap between the correct and incorrect induced policies is:

$$V_3(\pi_s) - V_3(\pi_{-s}) = \frac{2p_Ep_1p_2hr}{\sqrt{n}}. \tag{30}$$

Theorem 1 (Local regret of unweighted CATE and EWM)**.** *Under* (16)–(21) *and* (27)*, the following local regret limits hold for either* $s \in \{-1, +1\}$*:*

$$\sqrt{n}\,\mathbb{E}_{s,n}\mathcal{R}_m(\widehat{\pi}_R) \longrightarrow 2p_E p_1 p_2 hr\,\Phi(-I_R), \tag{31}$$

$$\sqrt{n}\,\mathbb{E}_{s,n}\mathcal{R}_m(\widehat{\pi}_D) \longrightarrow 2p_E p_1 p_2 hr\,\Phi(-I_D), \tag{32}$$

where Φ *is the standard normal distribution function.*

These limits make the role of type N explicit. Unweighted CATE receives both signal and noise from E and N, while EWM uses the critical contrast $E - L_1 - L_2$, with leading variance term T_D/n. Equating the two indices yields the noise-variance threshold:

$$\sigma^2_{N,*} = \frac{(A+C)^2 T_D/r^2 - A\sigma_E^2}{C}, \tag{33}$$

This threshold is positive when $C > 0$ and separates the two orderings:

$$\sigma_N^2 < \sigma^2_{N,*} \Rightarrow I_R > I_D, \qquad \sigma_N^2 > \sigma^2_{N,*} \Rightarrow I_R < I_D. \tag{34}$$

Thus, precise N observations improve unweighted CATE, while sufficiently noisy N observations can make equal weighting inefficient.

The unweighted comparison isolates the consequence of pooling all shared-model observations with the same regression weight. To characterize the best use of those observations under the correct model, let $w(x) > 0$ and define weighted CATE:

$$\widehat{\theta}_R(w) \in \arg\min_{|\theta|\le\bar{\theta}} \frac{1}{n}\sum_{i=1}^{n} w(X_i)\{\Gamma_i - b(X_i) - \theta\psi(X_i)\}^2. \tag{35}$$

The corresponding local signal-to-noise index takes the following form:

$$I_R(w) = h\frac{\mathbb{E}[w(X)\psi(X)^2]}{\sqrt{\mathbb{E}[w(X)^2\psi(X)^2\sigma(X)^2]}}. \tag{36}$$

Theorem 2 (Optimal weighting under the shared model)**.** *Under the correctly specified model* (18)–(21)*, the following inverse-variance weights maximize* (36) *on the support where* $\psi(x) \neq 0$*:*

$$w^*(x) \propto \sigma(x)^{-2}$$

The resulting optimal index is:

$$I_R^* = h\sqrt{\frac{p_E r^2}{\sigma_E^2} + \frac{p_N \kappa^2}{\sigma_N^2}}. \tag{37}$$

Moreover, the indices satisfy the following comparison:

$$I_D^2 = \frac{h^2 r^2}{T_D} < \frac{h^2 p_E r^2}{\sigma_E^2} \le (I_R^*)^2. \tag{38}$$

Thus, inverse-variance CATE has smaller local allocation regret than the raw complete-U value criterion throughout this correctly specified submodel.

Theorem 2 gives the correct-model benchmark. The shared treatment-effect relation turns type N into an additional source for learning the parameter that controls type E, and inverse-variance weighting determines how strongly those observations should enter. When N is precise, it contributes substantial information; as σ_N^2 grows, its contribution decreases continuously. The next step keeps the allocation boundary fixed and changes the accuracy of the relation that permits this pooling.

4.4 Local CATE Model Misspecification

To study local CATE misspecification, fix $\rho \ge 0$ and specify the true treatment effects:

$$\tau_{s,n,\rho}(E) = b(E) + s\frac{hr}{\sqrt{n}}, \tag{39}$$

$$\tau_{s,n,\rho}(L_1) = b(L_1), \qquad \tau_{s,n,\rho}(L_2) = b(L_2), \tag{40}$$

$$\tau_{s,n,\rho}(N) = b(N) + s\frac{h(1-\rho)\kappa}{\sqrt{n}}. \tag{41}$$

At $\rho = 0$, the shared model is correct. Increasing ρ changes the local movement of type N relative to the working model $b + \theta\psi$. Thus the decision-critical bundle contrast remains:

$$\tau_{s,n,\rho}(E) - \tau_{s,n,\rho}(L_1) - \tau_{s,n,\rho}(L_2) = s\frac{hr}{\sqrt{n}},$$

where the oracle allocation remains π_s. For all sufficiently large n, the two candidate policies also continue to agree on every batch containing N. The deviation therefore changes the relation used by CATE without changing the population contrast used by the EWM criterion.

For the weighted comparison, define the following precision-weighted quantities:

$$H_E = \frac{p_E r^2}{\sigma_E^2}, \qquad H_N = \frac{p_N \kappa^2}{\sigma_N^2}, \qquad M = H_E + H_N. \tag{42}$$

Theorem 3 (Local model deviation)**.** *Let inverse-variance CATE continue to fit the working class* (18)*, and let the EWM estimator be* (23)*. Under* (39)–(41)*, local regrets have the limits:*

$$\sqrt{n}\,\mathbb{E}_{s,n,\rho}\mathcal{R}_m(\widehat{\pi}_R(w^*)) \longrightarrow 2p_E p_1 p_2 hr\,\Phi(-I_R^*(\rho))\,, \tag{43}$$

$$\sqrt{n}\,\mathbb{E}_{s,n,\rho}\mathcal{R}_m(\widehat{\pi}_D) \longrightarrow 2p_E p_1 p_2 hr\,\Phi(-I_D). \tag{44}$$

Here, the CATE index under local deviation is defined as follows:

$$I_R^*(\rho) = h\frac{H_E + (1-\rho)H_N}{\sqrt{M}} = h\sqrt{M} - \rho h\frac{H_N}{\sqrt{M}}. \tag{45}$$

The two local regrets coincide at the following threshold:

$$\rho_* = \frac{M - r\sqrt{M/T_D}}{H_N} > 0. \tag{46}$$

For $\rho < \rho_*$, inverse-variance CATE has the smaller local regret; for $\rho > \rho_*$, the EWM criterion has the smaller local regret.

Theorem 3 completes the comparison begun by Proposition 3. CATE retains the variance reduction created by type N, while the local departure from the shared relation shifts the mean of the CATE statistic. The EWM statistic remains centered on the same bundle contrast because type N is absent from S_D. Along this local path, ρ_* marks the regret reversal: inverse-variance CATE has lower local regret below the threshold, and EWM above it.

4.5 Observational Extension with Cross-Fitted Scores

The preceding local results use the oracle score in (1). In observational data, the outcome regressions and propensity score are estimated. Split the sample into a fixed number of folds and estimate (μ_0, μ_1, e) on the complementary folds. For observation i in the held-out fold $k(i)$, compute the cross-fitted score:

$$\widehat{\Gamma}_i = \Gamma(O_i; \widehat{\eta}_{-k(i)}).$$

Theorem 4 (Cross-fitted score equivalence for the local comparison)**.** *Assume the four-type model, overlap bounded away from zero and one, finite conditional fourth moments, and a fixed number of asymptotically equal cross-fitting folds. Suppose the fold-specific nuisance estimates satisfy, uniformly over folds and along the local sequence:*

$$\|\widehat{e} - e\|_2 + \|\widehat{\mu}_1 - \mu_1\|_2 + \|\widehat{\mu}_0 - \mu_0\|_2 = o_p(1), \tag{47}$$

$$\|\widehat{e} - e\|_2\,(\|\widehat{\mu}_1 - \mu_1\|_2 + \|\widehat{\mu}_0 - \mu_0\|_2) = o_p(n^{-1/2}). \tag{48}$$

If the propensity estimates are truncated to the same overlap interval, then replacing Γ_i by $\widehat{\Gamma}_i$ in (22)*,* (23)*, and* (35) *changes the corresponding local sign statistics by $o_p(1)$ after their root-n normalizations. Consequently, the regret limits in Theorems 1 to 3 remain unchanged. Consistent type-specific variance estimates may replace the oracle inverse-variance weights.*

Theorem 4 connects the local allocation comparison to the observational causal setup. The doubly robust conditional-bias identity makes the cellwise score bias second order in nuisance errors, and cross-fitting controls the centered score difference. Under (47)–(48), the leading root-n normalized sign statistics that determine the allocation have the same limits as in the oracle-score analysis. The complete proof is given in Section K, and the finite-sample experiments are presented in Section M.4.

5 NUMERICAL EXPERIMENTS

To examine the local regret predictions of Theorems 1– 3, we use the four-type Gaussian score experiment. Set $p_E = p_1 = p_2 = 0.2$, $p_N = 0.4$, $r = \kappa = h = 1$, $n = 2500$, and use 100,000 Monte Carlo replications. The two verification experiments are as follows:

Changing the noise of type N. First take $\sigma_E^2 = \sigma_1^2 = \sigma_2^2 = 1$ and vary σ_N^2. The threshold in (33) is $\sigma_{N,*}^2 = 13$. The finite-n probabilities follow the three local limits closely; the full variance sweep is reported in Table 1. Detailed experimental setups and analyses can be found in Section J.

Changing the local model deviation. Next, set all four conditional variances to one and use (39)–(41). Here $M = 0.6$, $H_N = 0.4$, and (46) gives $\rho_* = 1$. The Monte Carlo crossing occurs at the same value: CATE has wrong-allocation probability 0.218 at $\rho = 0$ versus 0.397 for the EWM, while at $\rho = 1.5$ the probabilities are 0.500 and 0.398. The complete results are reported in Table 2. Detailed experimental setups and analyses can be found in Section J.

In addition, more experiments are provided in Section M. Specifically, in Section M.1, repetitions across five seeds assess the stability of the oracle σ_N^2 and ρ sweeps. In Section M.2, sensitivity analyses of the loadings and type proportions examine the predicted crossover thresholds under different settings. In Section M.3, local curves obtained by varying n and h examine the error probabilities and $\sqrt{n}$-scaled regret under $n^{-1/2}$ local alternatives. In Section M.4, Gaussian causal-data experiments compare oracle and cross-fitted procedures across sample sizes, providing a finite-sample check of the oracle-equivalence conclusion of Theorem 4. We repeat both σ_N^2 and ρ sweeps using estimated scores, and examine the effects of nuisance model misspecification. In Section M.5, we extend the comparison to Bernoulli outcomes, using S- and T-Learners for nuisance estimation.

Table 1: Monte Carlo wrong-allocation probabilities and their local limits. Weighted CATE uses inverse-variance weights. The EWM limit is constant because I_D does not contain σ_N^2.

σ_N^2	Unweighted CATE		EWM		Weighted CATE	
	Sim.	Limit	Sim.	Limit	Sim.	Limit
0.5	0.171	0.171	0.397	0.398	0.158	0.159
1	0.221	0.219	0.396	0.398	0.221	0.219
2	0.273	0.274	0.397	0.398	0.265	0.264
4	0.327	0.327	0.401	0.398	0.292	0.292
8	0.373	0.372	0.398	0.398	0.309	0.309
13	0.400	0.398	0.396	0.398	0.314	0.315
16	0.409	0.408	0.399	0.398	0.319	0.318
25	0.426	0.425	0.396	0.398	0.320	0.321
50	0.449	0.447	0.398	0.398	0.323	0.324

Table 2: Monte Carlo wrong-allocation probabilities under the local deviation at type N. The Limit columns report the predictions of Theorem 3. The EWM limit is constant in ρ.

ρ	Inverse-variance CATE		EWM	
	Sim.	Limit	Sim.	Limit
0.0	0.218	0.219	0.397	0.398
0.5	0.303	0.303	0.396	0.398
1.0	0.397	0.398	0.396	0.398
1.5	0.500	0.500	0.398	0.398
2.0	0.601	0.602	0.399	0.398

6 CONCLUSION

In this paper, we compare CATE-based allocation with EWM for finite batches under a hard budget. An exact reduction shows that shared CATE regression uses observations from covariate types that cancel out in EWM's empirical value contrast. Under correct specification, inverse-variance weighted CATE estimation achieves lower local allocation regret than EWM. Under local misspecification, pooling introduces bias while leaving the EWM contrast unchanged, yielding an explicit threshold where the regret ordering reverses. Cross-fitted doubly robust scores preserve these comparisons in observational data under nuisance-rate conditions. Thus, the benefit of sharing depends on the accuracy of the assumed cross-type relation.

Reproducibility Statement

The data-generating process, local alternatives, estimators, and Monte Carlo settings used in Section 5 are fully specified in the paper and in Section J. All theoretical results are stated with their assumptions, and complete proofs are provided below.

AI Use Statement

Generative AI tools were used for prose editing, LaTeX organization, algebraic checking, and code used to verify the theorem-matched Monte Carlo calculations. The scientific setup, assumptions, claims, and final mathematical verification remain the responsibility of the authors.

References

Henri Arno, Dennis Frauen, Emil Javurek, Thomas Demeester, and Stefan Feuerriegel. Rank-Learner: Orthogonal Ranking of Treatment Effects. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Susan Athey and Stefan Wager. Policy Learning with Observational Data. *Econometrica*, 89(1):133–161, 2021.

Debopam Bhattacharya and Pascaline Dupas. Inferring Welfare Maximizing Treatment Assignment under Budget Constraints. *Journal of Econometrics*, 167(1):168–196, 2012.

Giovanni Cerulli. Optimal Policy Learning under Budget and Coverage Constraints. *arXiv preprint arXiv:2605.12235*, 2026.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Mohammadsaeed Haghi, Mahdi Salmani, and Nima Kelidari. Differentiating Through Dual Prices: End-to-End Policy Learning Under Capacity Constraints. *arXiv preprint arXiv:2608.04669*, 2026.

Wassily Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

Fahad Kamran, Maggie Makar, and Jenna Wiens. Learning to Rank for Optimal Treatment Allocation Under Resource Constraints. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *PMLR*, pages 3727–3735, 2024.

Masahiro Kato. Bridging the Gap between Empirical Welfare Maximization and Conditional Average Treatment Effect Estimation in Policy Learning. *arXiv preprint arXiv:2510.26723*, 2025.

Toru Kitagawa and Aleksey Tetenov. Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*, 86(2):591–616, 2018.

Alexander Luedtke and Antoine Chambaz. Faster Rates for Policy Learning. *arXiv preprint arXiv:1704.06431*, 2017.

Aditya Mate, Bryan Wilder, Aparna Taneja, and Milind Tambe. Improved Policy Evaluation for Randomized Trials of Algorithmic Resource Allocation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *PMLR*, pages 24198–24213, 2023.

Liyang Sun. Empirical Welfare Maximization with Constraints. *Journal of Econometrics*, 253:106169, 2026.

Dylan J. Foster and Vasilis Syrgkanis. Orthogonal Statistical Learning. *The Annals of Statistics*, 51(3):879–908, 2023.

Yichun Hu, Nathan Kallus, and Xiaojie Mao. Fast Rates for Contextual Linear Optimization. *Management Science*, 68(6):4236–4245, 2022.

Edward H. Kennedy. Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023.

Qi Xu, Haoda Fu, and Annie Qu. Optimal Individualized Treatment Rule for Combination Treatments Under Budget Constraints. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(3):714–741, 2024.

Miguel A. Arcones and Evarist Giné. Limit Theorems for U-Processes. *The Annals of Probability*, 21(3):1494–1542, 1993.

Keisuke Hirano and Jack R. Porter. Asymptotics for Statistical Treatment Rules. *Econometrica*, 77(5):1683–1701, 2009.

Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins. Locally Robust Semiparametric Estimation. *Econometrica*, 90(4):1501–1535, 2022.

Stéphan Clémençon, Gábor Lugosi, and Nicolas Vayatis. Ranking and Empirical Minimization of U-statistics. *The Annals of Statistics*, 36(2):844–874, 2008.

Stephan Clémençon, Igor Colin, and Aurélien Bellet. Scaling-up Empirical Risk Minimization: Optimization of Incomplete U-statistics. *Journal of Machine Learning Research*, 17(76):1–36, 2016.

Michele Jonsson Funk, Daniel Westreich, Chris Wiesen, Til Stürmer, M. Alan Brookhart, and Marie Davidian. Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology*, 173(7):761–767, 2011.

Whitney K. Newey and James R. Robins. Cross-Fitting and Fast Remainder Rates for Semiparametric Estimation. *arXiv preprint arXiv:1801.09138*, 2018.

Deborah Nolan and David Pollard. U-Processes: Rates of Convergence. *The Annals of Statistics*, 15(2):780–799, 1987.

A Proof of the Structural Equivalence

Proposition 4 (Structural equivalence)**.** *For every score class $\mathcal{F}$,*

$$\sup_{f\in\mathcal{F}} V_m(\pi_f) = \sup_{\pi\in\Pi_{\mathcal{F}}} V_m(\pi).$$

If either supremum is attained, the corresponding maximum values are equal.

Proof. Define the score-to-policy map:

$$T : \mathcal{F} \to \Pi_{\mathcal{F}}, \qquad T(f) = \pi_f.$$

By the definition of $\Pi_{\mathcal{F}}$, T is surjective. Hence,

$$\{V_m(\pi_f) : f \in \mathcal{F}\} = \{V_m(\pi) : \pi \in \Pi_{\mathcal{F}}\}.$$

Taking the supremum over these identical sets proves the stated identity. If the common supremum is attained, the corresponding maximum values are equal. □

The parameterization is generally many-to-one. A standard equivalence relation on the score class is:

$$f \sim_B g \quad \Longleftrightarrow \quad D_B(f; X_{1:m}) = D_B(g; X_{1:m}) \quad \text{almost surely.} \tag{49}$$

The batch value depends only on the equivalence class induced by (49), while squared CATE loss generally distinguishes score functions within the same class. This difference motivates the statistical comparison developed in the main text.

B Batch Value: Identification and Estimation

B.1 Proof of the Batch Value Identification Proposition

Condition on the deployment covariates. Since the observational units are independent and π depends only on $X_{1:m}$, the conditional expectation simplifies as follows:

$$\begin{aligned}
\mathbb{E}[h_\pi(O_{1:m}) \mid X_{1:m}] &= \frac{1}{m}\sum_{j=1}^{m} \pi_j(X_{1:m})\mathbb{E}[\Gamma(O_j;\eta_0) \mid X_{1:m}] \\
&= \frac{1}{m}\sum_{j=1}^{m} \pi_j(X_{1:m})\mathbb{E}[\Gamma(O_j;\eta_0) \mid X_j] \\
&= \frac{1}{m}\sum_{j=1}^{m} \pi_j(X_{1:m})\tau(X_j).
\end{aligned}$$

Taking expectations gives (7). □

B.2 Disjoint-block estimator

Partition the first mK observations into $K = \lfloor n/m \rfloor$ non-overlapping blocks $I_1, \ldots, I_K$ of size m and define the block estimator:

$$\widehat{V}_m^{\mathrm{blk}}(\pi) = \frac{1}{K}\sum_{k=1}^{K} h_\pi(O_{I_k}). \tag{50}$$

The block kernels are independent and identically distributed with mean $V_m(\pi)$, so the estimator is unbiased for every fixed π:

$$\mathbb{E}[\widehat{V}_m^{\mathrm{blk}}(\pi)] = V_m(\pi).$$

If $|\Gamma| \le M_\Gamma$, then $|h_\pi| \le M_\Gamma$. For a finite class $|\Pi| = N$, Hoeffding's inequality (Hoeffding, 1963) for each policy and a union bound give the following uniform bound with probability at least $1-\delta$:

$$\sup_{\pi\in\Pi} |\widehat{V}_m^{\mathrm{blk}}(\pi) - V_m(\pi)| \le M_\Gamma \sqrt{\frac{2\log(2N/\delta)}{K}}. \tag{51}$$

The block estimator is therefore sufficient for identification and a basic $n^{-1/2}$ rate when m is fixed.

B.3 Complete ordered U-statistic

For (8), every ordered tuple of distinct observations has the same joint distribution as $(O_1, \ldots, O_m)$. Linearity of expectation and Proposition 1 establish unbiasedness:

$$\mathbb{E}[\widehat{V}_m^U(\pi)] = \frac{1}{(n)_m} \sum_{(i_1,\ldots,i_m)\,\text{distinct}} \mathbb{E}[h_\pi(O_{i_1}, \ldots, O_{i_m})] = V_m(\pi).$$

If the original kernel is not permutation-invariant because the coordinates of π are ordered, define its symmetrization:

$$\bar{h}_\pi(o_{1:m}) = \frac{1}{m!} \sum_{\rho \in S_m} h_\pi(o_{\rho(1)}, \ldots, o_{\rho(m)}),$$

where S_m is the set of permutations of $\{1, \ldots, m\}$. Averaging $\bar{h}_\pi$ over unordered subsets is algebraically identical to the ordered statistic in (8). Standard bounded U-statistic concentration can then be obtained by averaging over random partitions into $\lfloor n/m \rfloor$ independent blocks and applying Jensen's inequality to the exponential moment (Hoeffding, 1948; 1963; Clémençon et al., 2016). For a finite class $|\Pi| = N$, applying this bound to every $\pi \in \Pi$ and taking a union bound gives the following inequality with probability at least $1 - \delta$:

$$\sup_{\pi \in \Pi} \left| \widehat{V}_m^U(\pi) - V_m(\pi) \right| \le M_\Gamma \sqrt{\frac{2 \log(2N/\delta)}{\lfloor n/m \rfloor}}. \tag{52}$$

The displayed constant is conservative; the relevant order is $n^{-1/2}$ for fixed m.

For an infinite policy class, pointwise unbiasedness remains valid, but a uniform law of large numbers needs control of the U-process indexed by π (Nolan and Pollard, 1987). A finite VC-subgraph dimension together with an integrable envelope is one sufficient route (Arcones and Giné, 1993); an alternative is a direct bound on the corresponding U-process Rademacher complexity (Clémençon et al., 2008). No uniform convergence statement holds for an unrestricted class of all measurable batch policies.

The disjoint-block estimator discards every tuple crossing the chosen block boundaries. For the statistical comparison, we use the complete U-statistic, which averages all distinct tuples and is the welfare estimator used by EWM in Sections 4.2 and 4.3.

C Proof of the Finite-Batch Perturbation Bound

Fix a batch $x_{1:m}$ and abbreviate $Q_f(z; x_{1:m})$ by $Q_f(z)$. Let $z_f = D_B(f; x_{1:m})$ and let $z^* \in \mathcal{Z}_B^*(x_{1:m})$. Since z_f maximizes Q_f over the feasible set,

$$Q_f(z_f) \ge Q_f(z^*).$$

Therefore,

$$\begin{aligned} Q_\tau(z^*) - Q_\tau(z_f) &= Q_f(z^*) - Q_f(z_f) + \{Q_\tau - Q_f\}(z^*) - \{Q_\tau - Q_f\}(z_f) \\ &\le \frac{1}{m} \sum_{j=1}^{m} (z_j^* - z_{f,j}) \{\tau(x_j) - f(x_j)\} \\ &\le \frac{\|z^* - z_f\|_1}{m} \|f - \tau\|_\infty \\ &\le \Delta_B(x_{1:m}) \varepsilon. \end{aligned}$$

The first inequality discards the non-positive term $Q_f(z^*) - Q_f(z_f)$, and the second is Hölder's inequality. This proves (11). If z_f is suboptimal under τ, its regret is at least the smallest positive gap $G_m(x_{1:m})$, which proves (12). □

To prove (14), note that positive regret requires the following gap condition:

$$0 < G_m(X_{1:m}) \le \bar{\Delta}_B \varepsilon.$$

On this event, Proposition 2 also bounds the batch regret by $\bar{\Delta}_B \varepsilon$. Under (13), we therefore obtain the following allocation-regret bound:

$$\begin{aligned}\mathcal{R}_m(\pi_f) &\le \bar{\Delta}_B \varepsilon \, \mathbb{P}\{0 < G_m \le \bar{\Delta}_B \varepsilon\} \\ &\le C_m \bar{\Delta}_B^{1+\alpha} \varepsilon^{1+\alpha}.\end{aligned}$$

For a random score estimator $\widehat{f}$, define the training-sample event,

$$\mathcal{E} = \left\{ \bar{\Delta}_B \|\widehat{f} - \tau\|_\infty \le t_0 \right\}.$$

Conditioning on the training sample and applying the preceding argument on $\mathcal{E}$ gives the following bound:

$$\mathbb{E}\mathcal{R}_m(\pi_{\widehat{f}}) \le C_m \bar{\Delta}_B^{1+\alpha} \mathbb{E}\left[\|\widehat{f} - \tau\|_\infty^{1+\alpha} \mathbf{1}\{\mathcal{E}\} \right] + \mathbb{E}\left[\mathcal{R}_m(\pi_{\widehat{f}}) \mathbf{1}\{\mathcal{E}^c\} \right]. \tag{53}$$

The second term is the large-error tail contribution. This statement requires no bounded-regret assumption; a bounded outcome or bounded-CATE condition can further control it by a constant times $\mathbb{P}(\mathcal{E}^c)$.

D Verification of the Four-Type Decoder

This appendix verifies that (19) leaves only the (E, L_1, L_2) composition decision-sensitive. Because $m = 3$ and the budget is 2, at most one unit of type E can be treated. When the batch contains no E, every item has unit cost and all scores are positive, so the decoder selects the two largest scores. The condition $\bar{\theta} M_\psi < u - g - v$ ensures the following ordering:

$$f_\theta(L_2) - f_\theta(N) = (u - g - v) - \theta\kappa > 0.$$

Hence the ordering among L_1, L_2, N does not change with θ.

When the batch contains at least two units of type E, any feasible allocation contains at most one E. If the third item is L_1, the condition $\bar{\theta} M_\psi < u - g$ ensures the following inequality:

$$f_\theta(E) - f_\theta(L_1) = u - g + \theta r > 0$$

The corresponding differences against L_2 and N are larger under (19). Thus an E is selected for every such composition.

It remains to consider a batch with exactly one E and two unit-cost items. Since all scores are positive, the only relevant comparison is the singleton $\{E\}$ against the pair containing the two unit-cost items. The possible pairs give:

$$\begin{aligned}2f_\theta(L_1) - f_\theta(E) &= 2g - \theta r, \\ f_\theta(E) - 2f_\theta(L_2) &= 2g + \theta r, \\ f_\theta(E) - f_\theta(L_1) - f_\theta(N) &= u - g - v + \theta(r - \kappa), \\ f_\theta(E) - f_\theta(L_2) - f_\theta(N) &= u + g - v + \theta(r - \kappa), \\ f_\theta(E) - 2f_\theta(N) &= 2(u - v) + \theta(r - 2\kappa).\end{aligned}$$

Each displayed sign is fixed by (19). For the remaining pair (L_1, L_2),

$$f_\theta(E) - f_\theta(L_1) - f_\theta(L_2) = \theta r.$$

Thus, its sign changes exactly at $\theta = 0$. This proves the four decoder properties stated in Section 4.2. The positivity condition also follows from $f_\theta(N) \ge v - \bar{\theta}|\kappa| > 0$, and the remaining baseline scores are larger.

E Proof of the Exact Criterion Reduction

This section proves Proposition 3 by reducing each empirical objective to its sign statistic.

E.1 CATE CRITERION

Ignoring clipping at $\pm\bar{\theta}$ for the moment, differentiating the objective in (22) gives the unconstrained estimator:

$$\widetilde{\theta}_R = \frac{\sum_{i=1}^n \psi(X_i)\{\Gamma_i - b(X_i)\}}{\sum_{i=1}^n \psi(X_i)^2}.$$

The denominator is positive except on the exponentially small event that no observation has type E or N. Projection onto $[-\bar{\theta}, \bar{\theta}]$ preserves the sign whenever the numerator is nonzero. Thus the induced policy is determined by the sign of S_R in (24).

E.2 EWM CRITERION

The policies π_+ and π_- differ only on permutations of the (E, L_1, L_2) composition. For an ordered triple containing one observation of each of these types, the kernel difference equals one third of the score contrast, with the coordinate permutation determining which observation appears in each position. There are $3! = 6$ orderings for each unordered choice of one E, one L_1, and one L_2. Summing over these tuples gives the following value difference:

$$\begin{aligned}\widehat{V}_3^U(\pi_+) - \widehat{V}_3^U(\pi_-) &= \frac{1}{(n)_3}\frac{6}{3}\sum_{i:X_i=E}\sum_{j:X_j=L_1}\sum_{k:X_k=L_2}(\Gamma_i - \Gamma_j - \Gamma_k) \\ &= \frac{2N_E N_1 N_2}{(n)_3}(\bar{\Gamma}_E - \bar{\Gamma}_1 - \bar{\Gamma}_2),\end{aligned}$$

which is (25). Therefore S_D in (26) exactly determines the EWM empirical choice whenever all three counts are positive.

F PROOF OF THE LOCAL REGRET THEOREM

Let the data distribution under $\theta_{s,n} = sh/\sqrt{n}$ be denoted by $P_{s,n}$. For CATE,

$$S_R = \sum_{i=1}^n \psi(X_i)\{\theta_{s,n}\psi(X_i) + \varepsilon_i\}.$$

Using $\mathbb{E}[\psi(X)^2] = A + C$, we obtain the mean:

$$\mathbb{E}_{s,n}[S_R] = sh\sqrt{n}(A + C). \tag{54}$$

For the variance, the noise contribution is defined as:

$$n\mathbb{E}[\psi(X)^2\sigma(X)^2] = n(A\sigma_E^2 + C\sigma_N^2).$$

The random design contribution from $\theta_{s,n}\psi(X_i)^2$ is $O(1)$ because $\theta_{s,n}^2 = O(n^{-1})$, and its covariance with the mean-zero noise term is zero. Hence,

$$\mathrm{Var}_{s,n}(S_R) = n(A\sigma_E^2 + C\sigma_N^2) + o(n). \tag{55}$$

The triangular-array central limit theorem applied to (54)–(55) gives the limiting sign-error probability:

$$\mathbb{P}_{s,n}\{\mathrm{sign}(S_R) \neq s\} \longrightarrow \Phi(-I_R).$$

For the EWM criterion, conditional on the positive cell counts,

$$S_D = \bar{\Gamma}_E - \bar{\Gamma}_1 - \bar{\Gamma}_2.$$

The baseline contrast is zero, and only the E mean changes with θ. Thus,

$$\mathbb{E}_{s,n}[S_D] = s\frac{hr}{\sqrt{n}}. \tag{56}$$

Since $N_t/n \to p_t$ in probability for $t \in \{E, L_1, L_2\}$, the joint central limit theorem for the cell means gives the following variance expansion:

$$\mathrm{Var}_{s,n}(S_D) = \frac{1}{n}\left(\frac{\sigma_E^2}{p_E} + \frac{\sigma_1^2}{p_1} + \frac{\sigma_2^2}{p_2}\right) + o(n^{-1}) = \frac{T_D}{n} + o(n^{-1}). \tag{57}$$

The probability that any of N_E, N_1, N_2 equals zero is at most $(1-p_E)^n + (1-p_1)^n + (1-p_2)^n$, which is exponentially small. Therefore,

$$\mathbb{P}_{s,n}\{\mathrm{sign}(S_D) \neq s\} \longrightarrow \Phi(-I_D).$$

The two policies disagree only on permutations of (E, L_1, L_2), an event with probability $6p_Ep_1p_2$. On such a batch, the normalized value gap is expressed as:

$$\frac{1}{3}|\tau_{\theta_{s,n}}(E) - \tau_{\theta_{s,n}}(L_1) - \tau_{\theta_{s,n}}(L_2)| = \frac{hr}{3\sqrt{n}}.$$

Multiplying by the probability of this composition gives the population policy gap:

$$6p_Ep_1p_2\frac{hr}{3\sqrt{n}} = \frac{2p_Ep_1p_2hr}{\sqrt{n}}.$$

A sign error selects the unique inferior induced policy and incurs this gap, while a correct sign incurs zero regret. Multiplication by the two limiting sign-error probabilities proves (31)–(32). □

F.1 Explicit two-way separation

Assume $C > 0$. Solving $I_R = I_D$ for σ_N^2 gives (33). The following inequalities establish that this threshold is positive:

$$\frac{(A+C)^2T_D}{r^2} > \frac{(A+C)^2\sigma_E^2}{p_Er^2} = \frac{(A+C)^2\sigma_E^2}{A} > A\sigma_E^2.$$

Therefore $\sigma_{N,*}^2 > 0$. Since I_R is strictly decreasing in σ_N^2 when $C > 0$, (34) follows.

For completeness, the local subclasses used in the technical development can be written as:

$$\mathcal{P}_R = \{P_{s,n,\sigma_N^2} : \sigma_N^2 < \sigma_{N,*}^2\}, \qquad \mathcal{P}_D = \{P_{s,n,\sigma_N^2} : \sigma_N^2 > \sigma_{N,*}^2\}.$$

On $\mathcal{P}_R$, the canonical CATE criterion has smaller expected local regret for all sufficiently large n; on $\mathcal{P}_D$, the canonical raw complete-U EWM criterion has smaller expected local regret. These subclasses characterize the local ordering of the two specified empirical procedures.

G Proof of the Optimal Weighting Result

For the weighted criterion (35), the unconstrained estimator has the sign of the following statistic:

$$S_R(w) = \sum_{i=1}^{n} w(X_i)\psi(X_i)\{\Gamma_i - b(X_i)\}.$$

Under $\theta_{s,n} = sh/\sqrt{n}$,

$$\mathbb{E}[S_R(w)] = sh\sqrt{n}\,\mathbb{E}[w(X)\psi(X)^2],$$
$$\mathrm{Var}(S_R(w)) = n\mathbb{E}[w(X)^2\psi(X)^2\sigma(X)^2] + o(n),$$

which gives (36). By Cauchy–Schwarz,

$$\{\mathbb{E}[w\psi^2]\}^2 = \left\{\mathbb{E}\left[(w|\psi|\sigma)\frac{|\psi|}{\sigma}\right]\right\}^2$$
$$\leq \mathbb{E}[w^2\psi^2\sigma^2]\mathbb{E}\left[\frac{\psi^2}{\sigma^2}\right].$$

Equality is attained when $w(x)|\psi(x)|\sigma(x)$ is proportional to $|\psi(x)|/\sigma(x)$ on the support of ψ, which is equivalent to $w^*(x) \propto \sigma(x)^{-2}$. Substituting the four-type distribution gives the optimal index:

$$I_R^* = h\sqrt{\frac{p_Er^2}{\sigma_E^2} + \frac{p_N\kappa^2}{\sigma_N^2}}.$$

Finally, the EWM-contrast variance term obeys the following inequality:

$$T_D = \frac{\sigma_E^2}{p_E} + \frac{\sigma_1^2}{p_1} + \frac{\sigma_2^2}{p_2} > \frac{\sigma_E^2}{p_E},$$

so the indices obey the following ordering:

$$I_D^2 = \frac{h^2 r^2}{T_D} < \frac{h^2 p_E r^2}{\sigma_E^2} \leq (I_R^*)^2.$$

Since the local regret constant is proportional to $\Phi(-I)$, optimally weighted CATE has strictly smaller local regret. □

H Proof of the Local Model-Deviation Theorem

Let the inverse-variance weight be $w^*(t) = \sigma_t^{-2}$; multiplying all weights by a common positive constant leaves the estimator unchanged. Under (39)–(41), the weighted CATE sign statistic is expressed as:

$$S_R^* = \sum_{i=1}^{n} \frac{\psi(X_i)}{\sigma(X_i)^2}\{\Gamma_i - b(X_i)\}.$$

Then, write the precision-weighted quantities:

$$H_E = \frac{p_E r^2}{\sigma_E^2}, \qquad H_N = \frac{p_N \kappa^2}{\sigma_N^2}, \qquad M = H_E + H_N.$$

The local mean contribution of type E is $shr/\sqrt{n}$, while the local mean contribution of type N is $sh(1-\rho)\kappa/\sqrt{n}$. Hence,

$$\mathbb{E}_{s,n,\rho}[S_R^*] = sh\sqrt{n}\{H_E + (1-\rho)H_N\}. \tag{58}$$

The variance contributed by the score noise is expressed as:

$$\mathrm{Var}_{s,n,\rho}(S_R^*) = n\mathbb{E}\left[\frac{\psi(X)^2}{\sigma(X)^2}\right] + o(n) = nM + o(n). \tag{59}$$

The $o(n)$ term contains the variance of the $n^{-1/2}$ local mean component. The triangular-array central limit theorem therefore gives the limiting sign-error probability:

$$\mathbb{P}_{s,n,\rho}\{\mathrm{sign}(S_R^*) \neq s\} \longrightarrow \Phi\left(-h\frac{H_E + (1-\rho)H_N}{\sqrt{M}}\right) = \Phi\{-I_R^*(\rho)\}.$$

For the EWM criterion, (26) uses only E, L_1, L_2. The perturbation in (41) changes only type N, leaving the following mean and variance unchanged:

$$\mathbb{E}_{s,n,\rho}[S_D] = s\frac{hr}{\sqrt{n}}, \qquad \mathrm{Var}_{s,n,\rho}(S_D) = \frac{T_D}{n} + o(n^{-1}),$$

exactly as in Section F. Thus,

$$\mathbb{P}_{s,n,\rho}\{\mathrm{sign}(S_D) \neq s\} \longrightarrow \Phi(-I_D).$$

The critical allocation contrast remains $shr/\sqrt{n}$, and the two policies disagree only on the (E, L_1, L_2) composition. Consequently, the population value gap remains unchanged:

$$\frac{2p_E p_1 p_2 hr}{\sqrt{n}}.$$

Multiplying this gap by the two sign-error probabilities proves (43)–(44).

To find the crossing, solve $I_R^*(\rho) = I_D$:

$$h\frac{M - \rho H_N}{\sqrt{M}} = h\frac{r}{\sqrt{T_D}}.$$

Solving for ρ gives the crossing threshold:

$$\rho_* = \frac{M - r\sqrt{M/T_D}}{H_N}.$$

By (38), $\sqrt{M} > r/\sqrt{T_D}$, so the numerator is positive and $\rho_* > 0$. Since $I_R^*(\rho)$ is strictly decreasing in ρ when $H_N > 0$, the two regret orderings on either side of ρ_* follow from the monotonicity of $\Phi(-x)$. □

For every fixed ρ, the deviation of $\tau_{s,n,\rho}$ from the baseline scores is $O(n^{-1/2})$. The strict inequalities used in Section D outside the critical (E, L_1, L_2) composition are separated from zero by constants determined by u, g, v, and (19). Hence the same decoder structure holds for all sufficiently large n.

I Causal Realizability of the Oracle-Score Experiment

The Gaussian oracle-score experiment, including the local deviation in Theorem 3, can be generated by a randomized causal model. For any treatment-effect function used in the local experiment, randomize treatment independently of X:

$$A \sim \text{Bernoulli}(1/2).$$

Define the potential outcomes as follows:

$$Y(1) = \frac{1}{2}\tau(X) + \xi_1, \qquad Y(0) = -\frac{1}{2}\tau(X) + \xi_0, \tag{60}$$

where, conditional on $X = t$,

$$\xi_1, \xi_0 \overset{\text{i.i.d.}}{\sim} N(0, \sigma_t^2/4).$$

Then, the resulting outcome regressions and propensity have the following properties:

$$\mu_1(X) - \mu_0(X) = \tau(X), \qquad e(X) = 1/2.$$

Substituting these functions into (1) gives the doubly robust score:

$$\Gamma = \tau(X) + 2A\xi_1 - 2(1 - A)\xi_0.$$

Conditional on $X = t$, the residual $2A\xi_1 - 2(1 - A)\xi_0$ has the $N(0, \sigma_t^2)$ distribution because the two treatment branches have the same centered Gaussian law. Thus (21) is exactly realizable under randomized treatment, unconfoundedness, consistency, and overlap.

J Numerical Experiment Details

The simulation in Section 5 uses the following parameter settings:

$$(p_E, p_1, p_2, p_N) = (0.2, 0.2, 0.2, 0.4), \quad r = \kappa = h = 1, \quad \sigma_E^2 = \sigma_1^2 = \sigma_2^2 = 1, \quad n = 2500.$$

Type N noise-variance experiment. For each replication, the cell counts (N_E, N_1, N_2, N_N) are drawn from the multinomial distribution with total count n. Conditional on the counts and $\theta_n = 1/\sqrt{n}$, we draw the centered score sums directly:

$$\sum_{i:X_i=E} \{\Gamma_i - b(E)\} \sim N(N_E\theta_n r, N_E\sigma_E^2),$$
$$\sum_{i:X_i=L_1} \{\Gamma_i - b(L_1)\} \sim N(0, N_1\sigma_1^2),$$
$$\sum_{i:X_i=L_2} \{\Gamma_i - b(L_2)\} \sim N(0, N_2\sigma_2^2),$$
$$\sum_{i:X_i=N} \{\Gamma_i - b(N)\} \sim N(N_N\theta_n\kappa, N_N\sigma_N^2).$$

These four draws are independent conditional on the counts. The sign statistics are then evaluated exactly from (24), (26), and the inverse-variance weighted version of S_R. We use 100,000 replications for each variance value.

The theoretical entries equal $\Phi(-I_R)$, $\Phi(-I_D)$, and $\Phi(-I_R^*)$. The Monte Carlo error at 100,000 replications is below roughly 0.002 at these probabilities, consistent with the deviations shown in Table 1. Specifically, Table 1 shows that the empirical wrong-allocation probabilities closely follow the local limits. Unweighted CATE deteriorates as type N noise increases, with its comparison against EWM reversing near $\sigma^2_{N,*} = 13$. EWM remains close to 0.398, since type N cancels from its value contrast. Weighted CATE has lower error than EWM throughout the grid and improves on unweighted CATE when the variances differ; the two CATE methods coincide at $\sigma^2_N = 1$.

Local model-deviation experiment. We next set all conditional score variances to one and use (39)–(41). Conditional on the cell counts, the type N centered score sum is drawn as:

$$\sum_{i:X_i=N} \{\Gamma_i - b(N)\} \sim N\left(N_N \frac{h(1-\rho)\kappa}{\sqrt{n}}, N_N \sigma_N^2\right)$$

for $s = +1$, with the remaining cell distributions unchanged. At the stated parameter values,

$$H_E = 0.2, \qquad H_N = 0.4, \qquad M = 0.6, \qquad T_D = 15,$$

so (46) gives $\rho_* = 1$. The local CATE index simplifies as follows:

$$I_R^*(\rho) = \frac{0.6 - 0.4\rho}{\sqrt{0.6}},$$

while $I_D = 1/\sqrt{15}$. The Monte Carlo probabilities in Table 2 are obtained from 100,000 independent replications at each value of ρ. From Table 2, the CATE error increases with ρ, while EWM remains near 0.398. The methods have nearly equal error at the theoretical crossing $\rho_* = 1$; EWM performs better at the reported points above this threshold. This agrees with the local bias introduced by inaccurate pooling in Theorem 3.

K Cross-Fitted Doubly Robust Scores

We prove Theorem 4 for the finite four-type model used in the local regret analysis. Let the sample be split into a fixed number K of folds. For observation i in fold $k(i)$, estimate the nuisance functions using observations outside that fold:

$$\widehat{\eta}_{-k(i)} = (\widehat{\mu}_{0,-k(i)}, \widehat{\mu}_{1,-k(i)}, \widehat{e}_{-k(i)})$$

Define the cross-fitted score and its difference from the oracle score:

$$\widehat{\Gamma}_i = \Gamma(O_i; \widehat{\eta}_{-k(i)}), \qquad \Delta_i = \widehat{\Gamma}_i - \Gamma(O_i; \eta_0).$$

Cross-fitting makes $\widehat{\eta}_{-k(i)}$ independent of O_i conditional on the training sample used for that fold (Newey and Robins, 2018).

K.1 Conditional bias identity

For any nuisance triple $\widehat{\eta} = (\widehat{\mu}_0, \widehat{\mu}_1, \widehat{e})$ independent of an evaluation observation, direct calculation gives the conditional bias identity:

$$\mathbb{E}[\widehat{\Gamma} - \tau(X) \mid X, \widehat{\eta}] = \{\widehat{e}(X) - e(X)\} \left[\frac{\widehat{\mu}_1(X) - \mu_1(X)}{\widehat{e}(X)} + \frac{\widehat{\mu}_0(X) - \mu_0(X)}{1 - \widehat{e}(X)}\right]. \tag{61}$$

Under the common overlap bound, the absolute conditional bias is bounded by a constant times:

$$|\widehat{e}(X) - e(X)| \left(|\widehat{\mu}_1(X) - \mu_1(X)| + |\widehat{\mu}_0(X) - \mu_0(X)|\right).$$

Because X has four support points with probabilities bounded away from zero, $L_2(P_X)$ convergence controls the pointwise nuisance errors up to a fixed multiplicative constant. Therefore (48) implies, uniformly over the four types and folds,

$$\max_{t \in \{E, L_1, L_2, N\}} \left|\mathbb{E}[\Delta_i \mid X_i = t, \widehat{\eta}_{-k(i)}]\right| = o_p(n^{-1/2}). \tag{62}$$

Under finite conditional fourth moments, the overlap bound, and (47), a direct expansion also gives the following second-moment bound uniformly over folds:

$$\mathbb{E}[\Delta_i^2 \mid \widehat{\eta}_{-k(i)}] = o_p(1). \tag{63}$$

In the finite-support model, this follows by evaluating the score difference on each type and using pointwise convergence of the nuisance estimates.

K.2 Root-n equivalence of the sign statistics

Let I_k be fold k, with size n_k and $n_k/n \to 1/K$. Fix any bounded function $a(X)$. We work one held-out fold at a time, so that the nuisance estimate $\widehat{\eta}_{-k}$ is independent of the observations in I_k (Chernozhukov et al., 2022). Conditional on $\widehat{\eta}_{-k}$, decompose the score difference into centered and bias terms:

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i \in I_k} a(X_i)\Delta_i = \frac{1}{\sqrt{n}} \sum_{i \in I_k} a(X_i) \left\{\Delta_i - \mathbb{E}[\Delta_i \mid X_i, \widehat{\eta}_{-k}]\right\} \\ + \frac{1}{\sqrt{n}} \sum_{i \in I_k} a(X_i)\mathbb{E}[\Delta_i \mid X_i, \widehat{\eta}_{-k}].\end{aligned} \tag{64}$$

For the centered term, conditional independence within the held-out fold and (63) give the following variance bound:

$$\frac{n_k}{n} \|a\|_\infty^2 \max_{i \in I_k} \mathbb{E}[\Delta_i^2 \mid \widehat{\eta}_{-k}] = o_p(1).$$

By (62) and $n_k/n = O(1)$, the bias term is bounded in absolute value by the following expression:

$$\frac{n_k}{\sqrt{n}} \|a\|_\infty \max_t |\mathbb{E}[\Delta_i \mid X_i = t, \widehat{\eta}_{-k}]| = o_p(1).$$

Hence each fold contributes $o_p(1)$. Since K is fixed, summing over folds yields the desired bound:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(X_i)\Delta_i = o_p(1). \tag{65}$$

Taking $a(X) = \psi(X)$ establishes equivalence for the CATE statistic:

$$\frac{\widehat{S}_R - S_R}{\sqrt{n}} = o_p(1).$$

Taking $a(X) = w^*(X)\psi(X)$ gives the same statement for inverse-variance CATE. For the EWM statistic, fix a type t and write:

$$\bar{\widehat{\Gamma}}_t - \bar{\Gamma}_t = \frac{1}{N_t} \sum_{i=1}^{n} \mathbf{1}\{X_i = t\}\Delta_i.$$

Since $N_t/n \to p_t > 0$ in probability, (65) with $a(X) = \mathbf{1}\{X = t\}$ yields,

$$\sqrt{n}\,(\bar{\widehat{\Gamma}}_t - \bar{\Gamma}_t) = o_p(1).$$

Applying this to $t = E, L_1, L_2$ establishes equivalence for the EWM statistic:

$$\sqrt{n}\,(\widehat{S}_D - S_D) = o_p(1).$$

Slutsky's theorem now transfers the oracle local limits in Theorems 1 to 3 to the cross-fitted statistics. If the four conditional score variances are replaced by consistent estimates, the induced inverse-variance weights converge to w^* and a second application of Slutsky's theorem gives the same result. □

L Estimator-Class Inclusion

The comparison in the main text fixes the empirical criteria because the induced policy class alone does not separate CATE estimation from value-based learning. Let $\mathcal{A}_D$ be any estimator class whose outputs lie in $\Pi_{\mathcal{F}}$. If $\mathcal{A}_D$ contains the CATE-based policy $\widehat{\pi}_R$, then for every distribution class $\mathcal{P}$,

$$\inf_{\widehat{\pi}\in\mathcal{A}_D} \sup_{P\in\mathcal{P}} \mathbb{E}_P \mathcal{R}_m(\widehat{\pi}) \leq \sup_{P\in\mathcal{P}} \mathbb{E}_P \mathcal{R}_m(\widehat{\pi}_R).$$

Thus, a minimax comparison over a broad policy-estimator class inherits every estimator that the class contains. The local results in the paper instead compare explicit statistical procedures: unweighted CATE, inverse-variance CATE, and EWM. Theorems 2 and 3 then show how the working treatment-effect restriction changes the local information and local bias of these procedures.

M Additional Experiments

M.1 Oracle σ_N^2 and ρ sweeps across five seeds

Table 3: Oracle noise sweep over five seeds (2026–2030). Entries are mean wrong-allocation probabilities, with across-seed standard deviations in parentheses; local limits are reported separately.

σ_N^2	Unweighted CATE		EWM		Weighted CATE	
	Sim. mean (SD)	Limit	Sim. mean (SD)	Limit	Sim. mean (SD)	Limit
0.5	0.172 (0.0012)	0.171	0.399 (0.0025)	0.398	0.159 (0.0008)	0.159
1	0.219 (0.0009)	0.219	0.398 (0.0011)	0.398	0.219 (0.0009)	0.219
2	0.273 (0.0009)	0.274	0.399 (0.0018)	0.398	0.263 (0.0008)	0.264
4	0.328 (0.0014)	0.327	0.399 (0.0014)	0.398	0.292 (0.0014)	0.292
8	0.372 (0.0017)	0.372	0.398 (0.0028)	0.398	0.308 (0.0026)	0.309
13	0.399 (0.0019)	0.398	0.398 (0.0013)	0.398	0.314 (0.0012)	0.315
16	0.407 (0.0016)	0.408	0.398 (0.0021)	0.398	0.318 (0.0014)	0.318
25	0.426 (0.0007)	0.425	0.397 (0.0015)	0.398	0.320 (0.0007)	0.321
50	0.447 (0.0008)	0.447	0.397 (0.0017)	0.398	0.324 (0.0015)	0.324

Table 4: Oracle deviation sweep over five seeds (2026–2030). Entries are mean wrong-allocation probabilities, with across-seed standard deviations in parentheses; local limits are from Theorem 3.

ρ	Inverse-variance CATE		EWM	
	Sim. mean (SD)	Limit	Sim. mean (SD)	Limit
0.0	0.220 (0.0010)	0.219	0.399 (0.0045)	0.398
0.5	0.302 (0.0007)	0.303	0.399 (0.0029)	0.398
1.0	0.399 (0.0010)	0.398	0.399 (0.0020)	0.398
1.5	0.500 (0.0015)	0.500	0.397 (0.0021)	0.398
2.0	0.602 (0.0020)	0.602	0.398 (0.0021)	0.398

We repeat the two oracle sweeps using seeds 2026–2030 and report the mean and standard deviation of the five seed-level error estimates. The standard deviations describe across-seed variability, not standard errors of the reported means. Table 3 and Table 4 show that the five-seed averages remain close to the local limits and preserve the ordering changes near $\sigma_N^2 = 13$ and $\rho = 1$. Away from these crossings, the differences between methods are large relative to the reported across-seed variation.

M.2 Sensitivity of the theoretical crossover points

We retain $n = 2500$, $h = 1$, $s = +1$, and the oracle baseline, with 100,000 replications per setting and setting-specific seeds derived from 2026. Each setting uses its own thresholds from (33) and (46). The retained tables report comparisons at those theoretical crossings.

Loading sensitivity. We vary $\psi = (r, 0, 0, \kappa)$ over the ten reported pairs, keeping the type proportions fixed. The noise audit uses $\rho = 0$. The deviation audit uses unit score variances, so weighted and unweighted CATE coincide. The working scores and the perturbed true effects satisfy the local

Table 5: Loading sensitivity at the noise crossover in (33). Here p_* is the common local limit and $|\widehat{p}_U - \widehat{p}_E|$ is the empirical gap at $\sigma_N^2 = \sigma_{N,*}^2$. All settings satisfy (19), with $\theta M_\psi \le 0.04 < 0.3$; displayed gaps are rounded to four decimals.

Group	r	κ	$\sigma_{N,*}^2$	p_*	$\widehat{p}_U$	$\widehat{p}_E$	$\lvert\widehat{p}_U - \widehat{p}_E\rvert$
1	1	1	13.0000	0.3981	0.3980	0.3969	0.0011
2	0.5	1	30.2500	0.4486	0.4477	0.4469	0.0008
3	0.75	1	17.2292	0.4232	0.4229	0.4262	0.0032
4	1.25	1	11.4025	0.3734	0.3758	0.3749	0.0010
5	1.5	1	10.9167	0.3493	0.3524	0.3525	0.0001
6	1	0.5	11.5000	0.3981	0.3977	0.3970	0.0007
7	1	0.75	11.1528	0.3981	0.3974	0.4006	0.0032
8	1	1.25	16.0150	0.3981	0.4001	0.3996	0.0005
9	1	1.5	19.9444	0.3981	0.3983	0.3983	0.0001
10	1.5	0.5	15.6667	0.3493	0.3491	0.3511	0.0020

Table 6: Loading sensitivity at the deviation crossover in (46). Here p_* is the common local limit and W denotes weighted CATE. The final column reports the maximum over the evaluated deviations, with M_ψ calculated using the true loading $(1-\rho)\kappa$ at type N in (19); all values are below 0.3.

Group	r	κ	ρ_*	$\widehat{P}_W$	$\widehat{P}_E$	$\lvert\widehat{P}_W - \widehat{P}_E\rvert$	p_*	$\max \theta M_\psi$
1	1.00	1.00	1.0000	0.3976	0.3977	0.0001	0.3981	0.0600
2	0.50	1.00	0.9085	0.4477	0.4480	0.0003	0.4486	0.0500
3	0.75	1.00	0.9347	0.4230	0.4238	0.0008	0.4232	0.0550
4	1.25	1.00	1.1002	0.3744	0.3736	0.0008	0.3734	0.0730
5	1.50	1.00	1.2323	0.3505	0.3479	0.0026	0.3493	0.0886
6	1.00	0.50	1.5858	0.3975	0.3977	0.0002	0.3981	0.0634
7	1.00	0.75	1.1408	0.3983	0.3981	0.0002	0.3981	0.0584
8	1.00	1.25	0.9448	0.3984	0.3998	0.0014	0.3981	0.0700
9	1.00	1.50	0.9213	0.3977	0.3983	0.0006	0.3981	0.0800
10	1.50	0.50	2.6277	0.3500	0.3463	0.0037	0.3493	0.1151

geometry checks based on (19). As shown in Tables 5 and 6, the empirical error probabilities of unweighted CATE and EWM are close at the corresponding theoretical crossover points across all 10 loading settings. In the local-deviation experiment, inverse-variance and unweighted CATE coincide under unit variances. By Theorem 3, CATE has smaller local allocation regret for $\rho < \rho_*$, whereas EWM comparison has smaller local regret for $\rho > \rho_*$.

Type-proportion sensitivity. To examine whether (33) continues to predict the crossover point between the CATE and EWM regimes when the type proportions vary, as well as the ρ_* values predicted by (46) and the resulting CATE/EWM regime ranking under different type-proportion settings, we fix $r = \kappa = 1$ and use the five different type proportion vectors, recomputing both thresholds for each composition. The noise audit again uses $\rho = 0$, and the deviation audit uses unit score variances. Table 7 and Table 8 show approximately equal error probabilities at each setting-specific theoretical crossover. The recomputed ρ_* values for each setting explain the empirical crossovers. The default $\rho_* = 1$ under the original proportions cannot be directly reused.

Table 7: Type-proportion sensitivity at $\sigma_N^2 = \sigma_{N,*}^2$. The vector is ordered as (p_E, p_1, p_2, p_N); p_* is the common unweighted CATE and EWM local limit, and $|\widehat{p}_U - \widehat{p}_E|$ is their absolute empirical difference on the relative grid. P01 is the original specification; P02 is balanced, P03 is E-heavy, P04 is L-heavy, and P05 has a low E share.

Group	(p_E, p_1, p_2, p_N)	$\sigma_{N,*}^2$	p_*	$\widehat{p}_U$	$\widehat{p}_E$	$\lvert\widehat{p}_U - \widehat{p}_E\rvert$
P01	$(0.20, 0.20, 0.20, 0.40)$	13.0000	0.3981	0.3980	0.3969	0.0011
P02	$(0.25, 0.25, 0.25, 0.25)$	11.0000	0.3864	0.3882	0.3857	0.0025
P03	$(0.30, 0.20, 0.20, 0.30)$	15.0000	0.3921	0.3907	0.3921	0.0013
P04	$(0.15, 0.25, 0.25, 0.35)$	10.0476	0.3970	0.3972	0.3955	0.0018
P05	$(0.10, 0.30, 0.30, 0.30)$	8.5556	0.4032	0.4030	0.4029	0.0001

Table 8: Type-proportion sensitivity at $\rho = \rho_*$. The vector is ordered as (p_E, p_1, p_2, p_N). Here ρ_* is recomputed from (46), p_* is the common weighted CATE and EWM local limit, and the final column is their empirical gap.

Group	(p_E, p_1, p_2, p_N)	ρ_*	p_*	$\widehat{p}_W$	$\widehat{p}_E$	$\lvert\widehat{p}_W - \widehat{p}_E\rvert$
P01	$(0.20, 0.20, 0.20, 0.40)$	1.0000	0.3981	0.3976	0.3977	0.0001
P02	$(0.25, 0.25, 0.25, 0.25)$	1.1835	0.3864	0.3862	0.3873	0.0011
P03	$(0.30, 0.20, 0.20, 0.30)$	1.2929	0.3921	0.3917	0.3915	0.0001
P04	$(0.15, 0.25, 0.25, 0.35)$	0.9010	0.3970	0.3975	0.3972	0.0003
P05	$(0.10, 0.30, 0.30, 0.30)$	0.8169	0.4032	0.4039	0.4009	0.0030

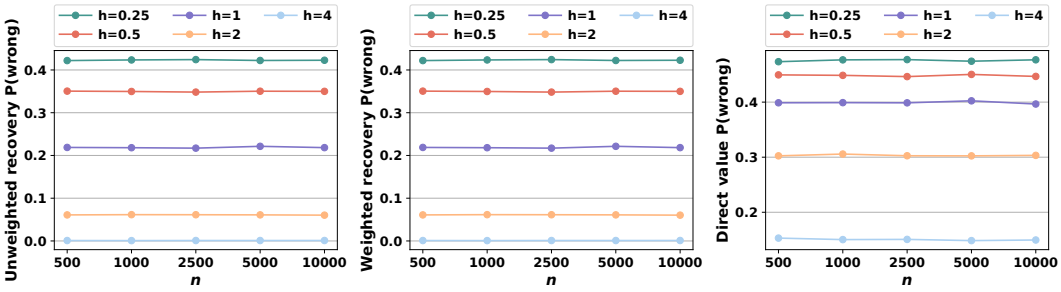


Figure 1: $n^{-1/2}$ local curve. With h fixed, the error rate remains essentially unchanged as n varies.

M.3 $n^{-1/2}$ LOCAL CURVES

Next, we examine the local error rate and $\sqrt{n}$-scaled regret under $\theta = \frac{s \cdot h}{\sqrt{n}}$, and compare CATE with EWM. Under correct sharing, we retain the oracle baseline, set all score variances to one, and use the positive local sequence $\theta = h/\sqrt{n}$. We vary $n \in \{500, 1000, 2500, 5000, 10000\}$ and $h \in \{0.25, 0.5, 1, 2, 4\}$. Each point averages five seeds. Unweighted and inverse-variance CATE coincide in this design.

For fixed h, Figure 1 shows nearly constant error probabilities across n, consistent with the nondegenerate local limits in Theorem 1 and Theorem 2. Increasing n also shrinks the signal, so these curves do not describe estimation at a fixed treatment effect. Figure 2 gives a more detailed view of the local comparison. The left panel shows the advantage of CATE under correct sharing. The middle panel shows that the $\sqrt{n}$-normalized regret remains finite and exhibits a non-monotone peak as the local signal strength changes. The right panel shows that both procedures become more accurate as h increases, with CATE maintaining the lower error rate.

M.4 GAUSSIAN CAUSAL-DATA EXPERIMENTS

These experiments generate individual-level causal observations and estimate cross-fitted doubly robust (DR) scores, extending the oracle-score comparisons. Specifically, when $e_t = 1/2$, these experiments use the causal realization in Section I, (60), which exactly generates the Gaussian oracle-score model in (21). The individual-level extensions retain the same four-type probabilities, treatment-effect specification, and potential-outcome noise parameters as the preceding oracle experiments. We set $u = 1$, $g = 0.2$, and $v = 0.3$, giving $b = (2u, u + g, u - g, v) = (2, 1.2, 0.8, 0.3)$, with $r = \kappa = 1$. The nuisance misspecification experiment instead uses the type-specific propensity vector $e = (0.2, 0.4, 0.6, 0.8)$, while the σ_N^2 and ρ sweeps use $e_t = 0.5$. In all cases, the observed outcome is generated from the selected potential outcome, $Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0)$. For the nonconstant-propensity extension, the corresponding oracle DR score variance includes the factor $\{4e_t(1 - e_t)\}^{-1}$. Following the cross-fitting construction in Section K, we use two folds and estimate the type-cell outcome regressions and propensity scores on the complementary fold. Propensity estimates are truncated to $[0.02, 0.98]$.

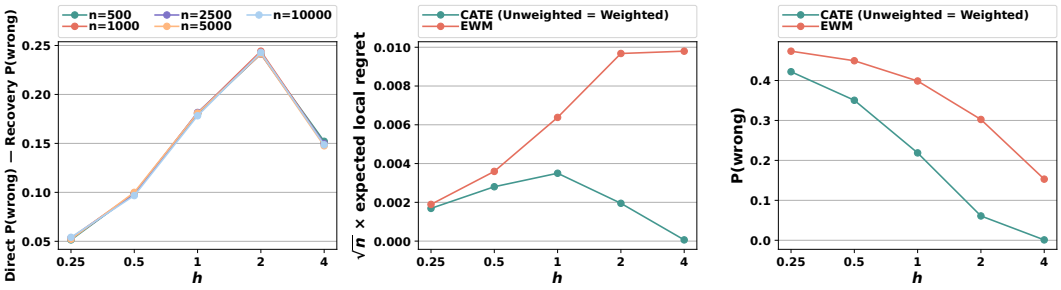


Figure 2: $n^{-1/2}$ local curves. Left: under correct sharing, CATE has a lower error rate than EWM. Middle: the $\sqrt{n}$-normalized regret remains finite and exhibits a local peak. Right: the error rate decreases as the local signal strength h increases.

M.4.1 Nuisance estimation and oracle equivalence

Under the correctly specified Gaussian cell model, we compare oracle and feasible procedures over $n \in \{500, 1000, 2500, 5000, 10000\}$, with 100,000 replications and one seed per sample size to examine the finite-sample behavior underlying Theorem 4. For weighted CATE, EE denotes estimated scores combined with estimated type-specific inverse-variance weights. Figure 3 shows that the

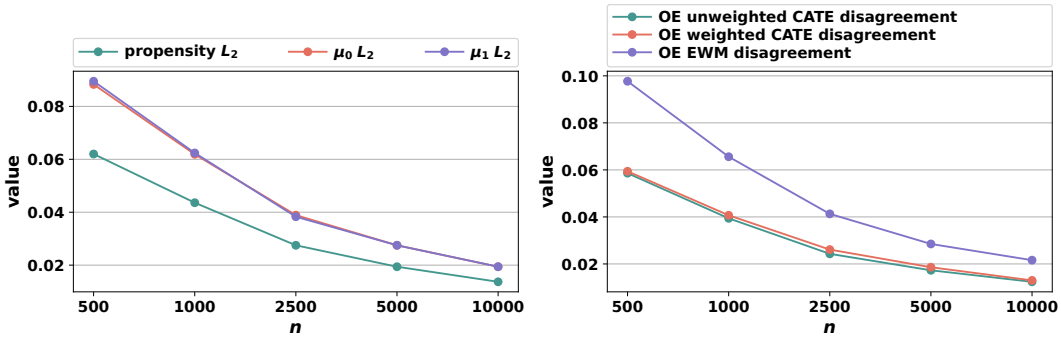


Figure 3: Nuisance L_2 errors (left) and probabilities of oracle–estimated (OE) policy disagreement (right) as n increases from 500 to 10,000.

Table 9: Wrong-allocation probabilities across sample sizes. The table compares oracle procedures with estimated procedures using estimated quantities, including estimated scores and estimated inverse-variance weights (EE).

n	Oracle			Estimated		
	Unweighted CATE	Weighted CATE	EWM	Unweighted CATE	Weighted CATE	EWM
500	0.2207	0.2207	0.3962	0.2283	0.2274	0.4019
1000	0.2188	0.2188	0.3998	0.2224	0.2226	0.4012
2500	0.2191	0.2191	0.3984	0.2201	0.2201	0.3995
5000	0.2187	0.2187	0.3992	0.2196	0.2197	0.3994
10000	0.2181	0.2181	0.3996	0.2186	0.2184	0.4004

nuisance L_2 estimation errors and oracle–estimated policy disagreement decrease as n increases from 500 to 10,000. These trends indicate improved nuisance estimation and closer agreement between the feasible and oracle decisions. Furthermore, Table 9 shows the wrong-allocation probability for the oracle and estimated cases. For unweighted CATE, the estimated–oracle gap decreases from 0.0076 at $n = 500$ to 0.0005 at $n = 10{,}000$. The corresponding gaps for weighted CATE and EWM decrease from 0.0067 to 0.0003 and from 0.0057 to 0.0008, respectively. The oracle error rates remain approximately stable, while the feasible procedures approach their oracle benchmarks. These diagnostics support the oracle-equivalence conclusion of Theorem 4. They do not by themselves establish its product-rate condition or isolate the effect of weight estimation.

Table 10: Wrong-allocation probabilities under nuisance-model misspecification.

Nuisance scenario	Unweighted CATE	Weighted CATE	EWM
Oracle	0.2680	0.2680	0.4054
Both correct	0.2720	0.2722	0.4069
Outcome correct	0.2698	0.2707	0.4069
Propensity correct	0.2739	0.2735	0.4105
Both misspecified	0.4917	0.4965	0.4240

M.4.2 NUISANCE MODEL MISSPECIFICATION

To evaluate the impact of misspecification in DR score learning, we compare correctly specified outcome and propensity nuisances, one-sided correctness, and simultaneous misspecification. Correct nuisance models use type-specific cells. Misspecified components use the corresponding constant working model. As shown in Table 10, under both-correct and one-sided correctness, all three methods remain close to the oracle. When both nuisance models are misspecified, the CATE error rate rises to approximately 0.49, indicating that double robustness does not protect against simultaneous misspecification of both nuisance models (Funk et al., 2011).

M.4.3 NOISE-VARIANCE SWEEP WITH ESTIMATED SCORES

Table 11: Wrong-allocation probabilities across type N noise variances using cross-fitted DR scores estimated from individual-level Gaussian causal data (seed 2026).

σ_N^2	Unweighted CATE	EWM	Weighted CATE
0.5	0.1752	0.4015	0.1610
1	0.2209	0.4000	0.2205
2	0.2765	0.3974	0.2658
4	0.3293	0.3981	0.2926
8	0.3701	0.3994	0.3072
13	0.3977	0.3971	0.3152
16	0.4100	0.3982	0.3185
25	0.4293	0.3996	0.3203
50	0.4472	0.3991	0.3245

Table 12: Wrong-allocation probabilities under local deviations at type N using cross-fitted DR scores and estimated inverse-variance weights (seed 2026).

ρ	Weighted CATE	Unweighted CATE	EWM
0.0	0.2207	0.2202	0.4003
0.5	0.2993	0.2995	0.3954
1.0	0.3988	0.3984	0.3971
1.5	0.5007	0.5011	0.3976
2.0	0.6010	0.6003	0.3984

After generating $Y(0), Y(1), A, Y$, we repeat the type N noise grid using DR scores estimated from individual-level Gaussian causal data and cross-fitting. As shown in Table 11, unweighted CATE deteriorates as σ_N^2 increases and approaches EWM at around 13. In contrast, weighted CATE has lower error, with a trend consistent with the oracle-score results in Table 1.

M.4.4 LOCAL DEVIATION SWEEP WITH ESTIMATED SCORES

We likewise repeat the ρ grid using cross-fitted scores and seed 2026, retaining the local perturbation at type N. As shown in Table 12, CATE error increases with ρ and crosses EWM near 1, while EWM remains near 0.398. Estimated weighting produces only small differences from unweighted CATE here. The observed reversal supports the persistence of the shared-model bias under nuisance estimation. Cross-fitting does not correct an inaccurate sharing restriction.

M.5 BERNOULLI EXTENSION

This extension keeps the four-type probabilities and the local parameterization from the preceding experiments, but replaces the Gaussian causal outcomes with binary potential outcomes. For each replication, draw an observed type $T_i \in \{E, L_1, L_2, N\}$ with probabilities $(0.2, 0.2, 0.2, 0.4)$ and generate a noisy continuous covariate as follows:

$$X_i = c_{T_i} + \eta_i, \qquad \eta_i \sim N(0, 0.9^2 I_4),$$

where the type-specific centers are defined as follows:

$$c_E = (-1.2, -1, 0.4, 0), \quad c_{L_1} = (-1, 1, 0, 0.4),$$
$$c_{L_2} = (1, -1, -0.4, 0), \qquad c_N = (1.1, 1, 0, -0.4).$$

Treatment is randomized with known propensity $e(X_i, T_i) = 0.5$. Define the control response probability as:

$$\mu_0(X_i) = \text{expit}(-2.2 + 0.25X_{i1} - 0.15X_{i2} + 0.1X_{i3}).$$

For this Bernoulli design, the type-specific treatment effect takes the following form:

$$\tau_i = b_{\text{Ber}}(T_i) + \theta\psi_{\text{Ber}}(T_i),$$

where $b_{\text{Ber}} = (0.08, 0.055, 0.025, 0.01)$ and $\psi_{\text{Ber}} = (1, 0, 0, 0.5)$. Then, the treated response probability is expressed as:

$$\mu_1(X_i, T_i) = \text{clip}\{\mu_0(X_i) + \tau_i, 10^{-5}, 1 - 10^{-5}\}.$$

Finally, conditional on (X_i, T_i),

$$Y_i(a) \sim \text{Bernoulli}(\mu_a(X_i, T_i)), \qquad Y_i = A_iY_i(1) + (1 - A_i)Y_i(0).$$

The clipping is only a numerical safeguard for valid Bernoulli probabilities and is inactive for the retained strictly local settings.

The type identifier T_i is observed by both the nuisance learners and the policy. We use two-fold cross-fitting. The S-Learner uses X_i, one-hot type indicators, treatment, and treatment-by-type interactions. The T-Learner fits separate logistic models for the two treatment arms using X_i and the type indicators (Künzel et al., 2019). The propensity $e = 0.5$ is known rather than estimated, and the fitted outcome regressions are used to construct the cross-fitted DR scores.

At $n = 2500$, S- and T-Learner nuisance estimates are combined with cross-fitting and DR scores. Each setting aggregates seeds 2026–2030 with 10,000 replications per seed, and reports the wrong-allocation probability. The original grid is $h \in \{0.1, 0.25, 0.5, 1.0\}$ with $\theta = sh/\sqrt{n}$, $s \in \{-1, +1\}$. We retain the 14 learner–sign–signal settings satisfying the actual local policy interval $-0.02 < \theta < 0.03$. In particular, $(s, h) = (-1, 1)$ lies on an excluded breakpoint, whereas $(+1, 1)$ remains inside the interval. The experimental result is shown in Table 13. From Table 13, both CATE procedures have lower wrong-allocation probabilities than EWM in every retained setting, and the S- and T-Learner results are similar. Errors decrease as h increases, while estimated inverse-variance weighting provides no consistent additional advantage. These results support the pooling benefit in this observed-type, locally valid Bernoulli design, without asserting the Gaussian local constants for binary outcomes. The reported error is a training-sample sign error. Within the certified local two-policy region, it is equivalent to a wrong-allocation decision.

Table 13: Wrong-allocation probabilities for Bernoulli outcomes with observable type identifiers. Only rows strictly inside the actual policy breakpoint interval are retained; observations at either breakpoint are excluded. The original grid is $h \in \{0.1, 0.25, 0.5, 1.0\}$ and $\text{sign} \in \{-1, +1\}$.

Nuisance model	Sign	h	Unweighted CATE	Weighted CATE	EWM
S-Learner	-1	0.10	0.4695	0.4672	0.4859
S-Learner	$+1$	0.10	0.4652	0.4669	0.4823
S-Learner	-1	0.25	0.4195	0.4175	0.4621
S-Learner	$+1$	0.25	0.4190	0.4207	0.4592
S-Learner	-1	0.50	0.3404	0.3386	0.4247
S-Learner	$+1$	0.50	0.3408	0.3426	0.4211
S-Learner	$+1$	1.00	0.2075	0.2086	0.3484
T-Learner	-1	0.10	0.4696	0.4673	0.4863
T-Learner	$+1$	0.10	0.4649	0.4668	0.4824
T-Learner	-1	0.25	0.4195	0.4176	0.4624
T-Learner	$+1$	0.25	0.4190	0.4202	0.4593
T-Learner	-1	0.50	0.3405	0.3388	0.4250
T-Learner	$+1$	0.50	0.3410	0.3423	0.4213
T-Learner	$+1$	1.00	0.2072	0.2083	0.3486