NOT ALL EXPLORATION IS INFORMATIVE: TARGETED INFERENCE IN ADAPTIVE EXPERIMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive experiments must earn deployment reward while estimating prespecified contrasts. Regret theory asks how much exploration learning requires; adaptive-inference theory recovers valid uncertainty after history-dependent assignment. We ask which exploration informs the requested directions. Equal expected counts can produce arbitrarily different inverse-propensity variance, while equal welfare and scalar information can leave different target contrasts with sharply different precision. We develop *clock-separated information flows* (CS–IF). A base policy retains its decision clock, while a logged evaluation clock allocates a welfare-priced information portfolio across target directions. Structured acquisition matrices and the predictable quadratic variation for design-based inference remain separate information states. Finite-sample matrix concentration and clock transfer connect deployment to a deterministic reference. An independent Gaussian pilot followed by a frozen evaluation design gives a pilot-to-deployment certificate: under balanced-pilot and leverage conditions, pilot-induced design and reference-information drift scale as $m^{-1/2}$. Logged propensities give exact design unbiasedness, a finite-sample interval, and, under score regularity, a studentized limit. At a fixed Gaussian reference instance, a local two-point risk lower bound and rounded fixed-design attainability identify the geometric welfare price up to a universal constant. In 500-repetition synthetic studies, targeting reduces worst-target standard error by 12.5% in the bandit model and 38.1% in the linear model under matched evaluation schedules. Across 300 train/held-out splits of a 64,000-customer randomized e-mail trial, it reduces prospective plug-in worst-contrast width by 4.3% for a 12,000-customer experiment. Information is therefore a directionally allocated resource, not a pull-count total.

1 INTRODUCTION

Adaptive experimentation links sequential decision making with statistical inference: a deployment must earn reward now and establish which interventions work before the next decision. Consider a retailer with incumbent message A, two challengers B, C, and a fixed campaign of 1500 customers. The incumbent earns 0.30 per assignment and either challenger earns 0.20, while management must still report the prespecified effects $B - A$ and $C - A$. Sending 300 customers away from A costs 30 units of welfare. How those 300 customers are randomized determines whether the campaign can answer both questions. In our reanalysis of the randomized Hillstrom e-mail trial (Hillstrom, 2008), the messages have different visit rates and residual variances (Section I.6); the same tension recurs in adaptive clinical trials and policy evaluation (Hadad et al., 2021; Ham et al., 2023).

Regret theory prices reward lost during sequential learning (Lai & Robbins, 1985; Bubeck & Cesa-Bianchi, 2012; Lattimore & Szepesvári, 2020), while precise-allocation theory resolves this price into instance-specific pull curves (Han et al., 2024). Adaptive-inference methods construct valid estimators and intervals after history-dependent assignment (Hadad et al., 2021; Bibaut et al., 2021; Ham et al., 2023). The closest joint design trades regret against worst-pairwise arm-effect error (Simchi-Levi & Wang, 2023). Prespecified, possibly structured contrasts pose a different question: which assignments repair their weakest direction? Figure 1 makes the distinction concrete: the answer depends on target geometry and the full propensity path, not exploration volume alone.

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.

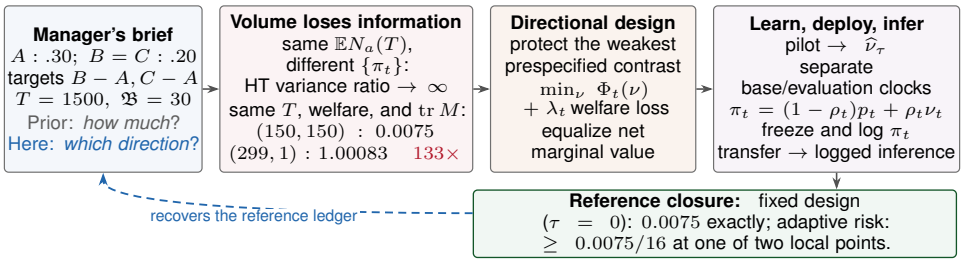


Figure 1: Target-specific information, rather than exploration volume, closes the retailer's welfare–inference loop. The two failures are formalized in Proposition 2.1; the allocation and learned deployment are certified by Theorems 3.1 and 4.1 to 4.3; logged inference follows from Theorem 4.4, Proposition F.1, and Corollaries F.1 and F.2; and the reference benchmark is completed by Theorem 4.5 and Corollary 4.1. The dashed return closes the reference ledger.

The retailer supplies both failures in one ledger. The stationary path $(0.8, 0.1, 0.1)$ and a time-varying path can have the same expected counts and welfare, while the latter's inverse-propensity variance diverges. Even the full count vector can point the wrong way: $(1200, 150, 150)$ and $(1200, 299, 1)$ have the same horizon, welfare cost, and Gaussian information trace, but their worst-target variances are 0.0075 and 1.00083, a $133\times$ loss because $C - A$ receives one observation. Target precision therefore requires a directional design, a logged propensity path, and a reference benchmark; neither a welfare budget nor scalar information determines it.

We develop *clock-separated information flows* (CS–IF) around this directional objective. A base policy proposes p_t; a randomized evaluation channel allocates its reserve by the marginal reduction in worst-target uncertainty net of welfare price. The resulting assignment propensity is logged before the outcome:

$$\pi_t = (1 - \rho_t)p_t + \rho_t\nu_t. \tag{1.1}$$

Evaluation outcomes update the information history without advancing the base policy's clock, exposing the incremental welfare bill while preserving predictable propensities. Gaussian arm-count design is recovered by $G_t(a) = e_a e_a^\top/\sigma^2$; structured acquisition uses M_T, whereas design-based inference uses the inverse-propensity score bracket. These states answer different probability questions: the campaign's 0.0075 is a Gaussian model-based quantity, while augmented inverse-propensity-weighted (AIPW) uncertainty conditions on potential outcomes and randomizes only assignments.

The proof problem is to carry a learned directional design through an unknown instance, a random decision clock, and randomized assignment. An independent pilot followed by a frozen design controls arm identification, allocation drift, welfare drift, and every requested inverse-information direction. Clock transfer explains why the nominal 1200 policy rounds fluctuate with standard deviation $\sqrt{240} \approx 15.5$; target-normalized concentration then sandwiches each reference variance. Logged propensities yield exact AIPW unbiasedness, a finite-sample interval, and a studentized limit. With $\rho_t \asymp t^{-1/3}$, the same campaign has regret $O(T^{2/3})$ and standard error $O(T^{-1/3})$. Finally, a reference-instance Gaussian bound shows that every budget-feasible adaptive policy under its local conditions pays the weak-direction price at one of two points, while rounded fixed designs attain the directional oracle.

Our contributions are:

1. **A target-specific problem and allocation rule.** Proposition 2.1 separates information direction from counts and trace; the full-horizon welfare-constrained design problem and Theorem 3.1 turn that gap into a marginal rule for choosing evaluation assignments (Sections 2 and 3).
2. **A certified path from pilot to inference.** CS–IF preserves the base trajectory, logs the full propensity, and connects a learned design to reference welfare, target geometry, exact design moments, finite-sample coverage, and studentization (Section 4).
3. **An intrinsic reference price and empirical evidence.** A local adaptive lower bound and fixed-design attainability recover $(1200, 150, 150)$ and 0.0075; synthetic and Hillstrom experiments compare target precision at matched evaluation schedules or expected welfare (Sections 4 and 5).

Section 2 defines the two information ledgers; Section 3 derives the allocation and two-clock wrapper; Section 4 proves pilot transfer, logged inference, and the local frontier; Section 5 tests each link; and Section 6 draws the implications. Appendix B gives the detailed closest-work comparison.

2 PROBLEM: WELFARE UNDER A TARGET INFORMATION REQUIREMENT

We formulate the retailer's decision as welfare-constrained information acquisition with two clocks and two statistical ledgers. The four-number campaign in Example 2.1 will recur at every step: Proposition 2.1 first shows why neither traffic volume nor scalar information can certify its two challenger effects; Section 3 then derives the allocation rule; and Section 4 follows the same ledger through learning, deployment, inference, and a local lower bound.

One campaign, two clocks. Let $\mathcal{A} = [K]$, let $\mathcal{F}_{t-1}$ contain the history before customer t, and let the base-policy distribution p_t, evaluation distribution ν_t, and evaluation probability ρ_t be predictable. Draw $E_t \sim \mathrm{Ber}(\rho_t)$, then draw the actual action $I_t \sim p_t$ when $E_t = 0$ and $I_t \sim \nu_t$ when $E_t = 1$. Hence $I_t \mid \mathcal{F}_{t-1} \sim \pi_t = (1-\rho_t)p_t + \rho_t\nu_t$, as in Equation (1.1). The filtration $\mathcal{G}_t = \mathcal{F}_{t-1} \vee \sigma(E_t, I_t)$ precedes the outcome and $\mathcal{F}_t = \mathcal{G}_t \vee \sigma(Y_t)$. Write $\mu_t(a) = \mathbb{E}\{Y_t(a) \mid \mathcal{F}_{t-1}\}$, $\mu_t^\star = \max_a \mu_t(a)$, and the predictable gap $\Delta_t(a) = \mu_t^\star - \mu_t(a) \geq 0$. With $N_a(T) = \sum_{t \leq T} \mathbf{1}\{I_t = a\}$, the conditional expected welfare loss is $\overline{R}_T = \sum_{t \leq T} \langle \pi_t, \Delta_t \rangle$; in a stationary instance, $R_T = \sum_a \Delta_a N_a(T)$.

Example 2.1 (The campaign ledger used throughout)**.** *A retailer sends incumbent A and must report the prespecified challenger effects $B - A$ and $C - A$ before roll-out. Let $\mu = (0.30, 0.20, 0.20)$ and $\Delta = (0, 0.10, 0.10)$. The directions $c_B^{\mathrm{par}} = e_B - e_A$ and $c_C^{\mathrm{par}} = e_C - e_A$ form $\mathcal{C}_{\mathrm{par}}$; in this three-arm example $\mathcal{C}_{\mathrm{arm}}$ contains the same two vectors. Setting $p_t = e_A$, $M_0 = 0$, and splitting a 20% evaluation reserve between B, C gives $\pi = (0.8, 0.1, 0.1)$. At $T = 1500$, the reference ledger is therefore* $\boxed{T = 1500,\ \mathfrak{B} = 30,\ \bar{N}^{\mathrm{ref}} = (1200, 150, 150),\ V_{\mathrm{ref}} = 0.0075}$*: the welfare bill is $0.1(150 + 150) = 30$, and under the unit-noise Gaussian acquisition model each comparison has variance $1/1200 + 1/150 = 0.0075$. Corollary 4.1 establishes that 0.0075 is also the directional frontier.*

Two complementary information ledgers. Model-based precision depends on acquisition geometry; randomization-based precision depends on the entire inverse-propensity path. These are parallel answers to different probability questions, not numerically interchangeable variance formulas.

For a structured reward model, fix $\mathcal{C}_{\mathrm{par}} = \{c_j^{\mathrm{par}} : j \leq q_{\mathrm{par}}\} \subset \mathbb{R}^d$ and $M_0 \succeq 0$; action a contributes an $\mathcal{F}_{t-1}$-measurable matrix $G_t(a) \succeq 0$. The realized and predictable acquisition matrices are $M_T = M_0 + \sum_{t \leq T} G_t(I_t)$ and $\overline{M}_T = M_0 + \sum_{t \leq T, a} \pi_t(a) G_t(a)$. In a homoskedastic Gaussian linear model with variance σ^2, $G_t(a) = \sigma^{-2} x_{t,a} x_{t,a}^\top$; the Gaussian multi-armed bandit (MAB) is the special case $x_{t,a} = e_a$. The quantities $(c_B^{\mathrm{par}})^\top M_T^{-1} c_B^{\mathrm{par}}$ and $(c_C^{\mathrm{par}})^\top M_T^{-1} c_C^{\mathrm{par}}$ expose the weaker challenger direction.

For design-based uncertainty, condition on the potential-outcome array, fix $\mathcal{C}_{\mathrm{arm}} = \{c_j^{\mathrm{arm}} : j \leq q_{\mathrm{arm}}\} \subset \mathbb{R}^K$, and take each outcome model $m_t(a)$ to be $\mathcal{F}_{t-1}$-measurable. Let $r_t(a) = Y_t(a) - m_t(a)$ and $z_t(a) = (c_{1a}^{\mathrm{arm}} r_t(a), \ldots, c_{q_{\mathrm{arm}} a}^{\mathrm{arm}} r_t(a))^\top$. Then

$$V_T = \sum_t \left\{ \sum_a \frac{z_t(a) z_t(a)^\top}{\pi_t(a)} - s_t s_t^\top \right\}, \qquad s_t = \sum_a z_t(a), \tag{2.1}$$

$$\mathbf{Q}_T = \sum_{t,a} \frac{z_t(a) z_t(a)^\top}{\pi_t(a)}, \qquad \widehat{\mathbf{Q}}_T = \sum_t \frac{z_t(I_t) z_t(I_t)^\top}{\pi_t(I_t)^2}. \tag{2.2}$$

Thus M_T records model geometry, V_T is the predictable randomization bracket, and $T^{-2}\mathbb{E}V_T$ is the exact unconditional covariance of the corresponding AIPW estimator. The matrix $\mathbf{Q}_T$ in Equation (2.2) is the allocable upper component and $\widehat{\mathbf{Q}}_T$ its observable counterpart.

Why aggregate certificates fail. Is reserving 300 customers for evaluation enough? No: the same traffic can be randomized at radically different propensities, and the same welfare can be spent in

the wrong direction. The next proposition makes both failures on the campaign itself and identifies directional information as the state needed by the design problem that follows.

Proposition 2.1 (The campaign defeats counts and scalar information)**.** *Let T be even and $0 < \varepsilon < 0.1$. (i) Compare the stationary propensity path $(0.8, 0.1, 0.1)$ with the path $(0.9-\varepsilon, \varepsilon, 0.1)$ for the first $T/2$ customers and $(0.7+\varepsilon, 0.2-\varepsilon, 0.1)$ thereafter. Both have expected counts $(0.8T, 0.1T, 0.1T)$ and the same expected welfare. If $Y_t(B) = 0.2$, the HT estimator of the B-arm mean has variance $0.36/T$ under the stationary path and*

$$\frac{0.04}{2T}\left\{\frac{1-\varepsilon}{\varepsilon} + \frac{0.8+\varepsilon}{0.2-\varepsilon}\right\}$$

under the second path; their ratio diverges as $\varepsilon \downarrow 0$. (ii) In Example 2.1, the Gaussian allocations $(1200, 150, 150)$ and $(1200, 299, 1)$ have the same size, welfare cost, and information trace. Their worst requested variances are 0.0075 and $1/1200 + 1 = 1.00083$, respectively, a factor exceeding 133.

Part (i) says that final counts erase when randomization became fragile; part (ii) says that a scalar matrix summary erases where information went. Thus the ordinary estimation–regret question, "how much exploration?", does not determine the answer here. With prespecified targets, the operative question is "which exploration informs their weakest direction?" This is the point at which our problem departs from arm-volume Pareto design (Simchi-Levi & Wang, 2023): the state variable is the target geometry or the reciprocal-propensity path.

Operational assumptions. Predictability protects the clock, positivity protects loaded targets, leverage controls geometric fluctuation, and design stability transfers pilot error. Logged propensities, normalized contributions, and first-order residuals make these conditions diagnosable; Table 1 records their roles and fallbacks. We next derive the reference allocation from the welfare and information constraints.

3 Clock-separated information flows

Proposition 2.1 leaves the manager with a constructive question: if 300 evaluation customers can be either decisive or nearly useless, what should replace "explore more" as the allocation principle? Our answer has two layers. Equation (3.0) states the scientific problem directly as target uncertainty under a welfare budget; Equations (3.1) and (3.2) give receding-horizon surrogates for its two statistical ledgers, and Theorem 3.1 converts each update into a marginal allocation rule. Algorithm 1 then implements that rule without advancing the base learner on evaluation rounds. The central path in Figure 1 thus runs from a failed volume certificate to directional design, two-clock deployment, and the transfer results of Section 4.

3.1 From the scientific constraint to a welfare-priced oracle

Why does the campaign split $(150, 150)$ follow from optimization? The primitive problem protects the weakest requested contrast under the full assignment mixture while charging every departure from the incumbent to the welfare budget. The acquisition and design-based criteria instantiate this constraint in the native units of their respective ledgers. A receding-horizon Lagrangian prices the remaining budget in Equation (3.3); Theorem 3.1 then says that every action receiving evaluation mass must deliver the same marginal information net of welfare cost. The full-horizon three-arm reference is solved directly in Corollary 4.1, where symmetry yields $150/150$; the asymmetric calculation after the theorem shows where equal splitting ceases to be optimal.

Fix a deterministic reference path and index its targets by $j \le q$. Put $r_t^{\mathrm{ref}}(a) = Y_t^{\mathrm{ref}}(a) - m_t^{\mathrm{ref}}(a)$, $w_{tja}^{\mathrm{ref}} = \{c_{ja}^{\mathrm{arm}} r_t^{\mathrm{ref}}(a)\}^2$, $s_{tj}^{\mathrm{ref}} = \sum_a c_{ja}^{\mathrm{arm}} r_t^{\mathrm{ref}}(a)$, and $M^{\mathrm{ref}}(\pi) = M_0 + \sum_{t,a} \pi_t(a) G_t^{\mathrm{ref}}(a)$. For acquisition, $\mathcal{U}_j(\pi) = (c_j^{\mathrm{par}})^\top M^{\mathrm{ref}}(\pi)^{-1} c_j^{\mathrm{par}}$, set to $+\infty$ if $M^{\mathrm{ref}}(\pi) \not\succ 0$; for design-based inference, $\mathcal{U}_j(\pi) = T^{-2} \sum_t \{\sum_a w_{tja}^{\mathrm{ref}}/\pi_t(a) - (s_{tj}^{\mathrm{ref}})^2\}$, with $0/0 = 0$ and $w/0 = +\infty$ for $w > 0$. The full-horizon master problem is

$$\min_{\nu_{1:T} \in \Delta_K^T} \max_{j \le q} \mathcal{U}_j(\pi_{1:T}^{\mathrm{ref}}) \quad \text{s.t.} \quad \sum_{t \le T} \langle \pi_t^{\mathrm{ref}}, \Delta_t^{\mathrm{ref}} \rangle \le \mathfrak{B}, \qquad \pi_t^{\mathrm{ref}} = (1-\rho_t) p_t^{\mathrm{ref}} + \rho_t \nu_t. \tag{3.0}$$

Recursively set $M_0^{\mathrm{ref}} = M_0$ and $M_{t-1}^{\mathrm{ref}} = M_0 + \sum_{s<t,a} \pi_s^{\mathrm{ref}}(a) G_s^{\mathrm{ref}}(a)$ after the preceding reference decisions. For $\pi_t^{\mathrm{ref}}(\nu) = (1-\rho_t)p_t^{\mathrm{ref}} + \rho_t \nu$, a look-ahead of length h_t yields

$$\Phi_t^{\mathrm{acq}}(\nu) = \max_{c_j^{\mathrm{par}} \in \mathcal{C}_{\mathrm{par}}} (c_j^{\mathrm{par}})^\top \left[M_{t-1}^{\mathrm{ref}} + h_t \sum_a \pi_t^{\mathrm{ref}}(\nu, a) G_t^{\mathrm{ref}}(a) \right]^{-1} c_j^{\mathrm{par}}, \tag{3.1}$$

$$\Phi_t^{\mathrm{db}}(\nu) = \max_{j \le q} \sum_a \frac{w_{tja}^{\mathrm{ref}}}{\pi_t^{\mathrm{ref}}(\nu, a)}. \tag{3.2}$$

A zero-load term in Equation (3.2) is zero at zero propensity, while a positive load requires positivity. These receding-horizon surrogates freeze the next h_t matrix contributions or retain the allocable one-step term; the design-based subtraction is allocation-invariant, and horizon scaling is absorbed into calibrated λ_t. Introducing this nonnegative multiplier gives

$$\nu_t^{\mathrm{ref}} \in \arg\min_{\nu \in \Delta_K} \left\{ \Phi_t(\nu) + \lambda_t h_t \rho_t \langle \nu, \Delta_t^{\mathrm{ref}} \rangle + \tau_t \, \mathrm{KL}(\nu \| \omega) \right\}, \tag{3.3}$$

where ω has full support. The KL term stabilizes the learned optimizer. With $\tau_t = 0$, Equation (3.3) is the unregularized per-update Lagrangian. Offline, (λ_t, h_t) are calibrated against a completed reference path satisfying the global welfare ledger in Equation (3.0); full-horizon optimality comes from solving that full-horizon master problem directly, as in Corollary 4.1. Positive τ_t supplies the sensitivity control in Theorem 4.1.

For $\tau_t = 0$, let ϕ_{tj} be the contrast-specific component of Φ_t, let α_j weight its active maximizers, and define $\mathsf{MB}_t(a; \nu) = -\sum_j \alpha_j \partial_{\nu(a)} \phi_{tj}(\nu)$. The proof gives the closed forms for both ledgers.

Theorem 3.1 (Per-update marginal-information equalization)**.** *Assume $\tau_t = 0$, $\lambda_t \ge 0$, $h_t > 0$, and $0 < \rho_t \le 1$. For the selected ledger, suppose each contrast criterion ϕ_{tj} is proper, closed, convex, and differentiable on the relative interior of their common effective domain. Assume that the epigraph program for Equation (3.3) is relatively Slater feasible and that its relevant matrix is positive definite or every loaded propensity is positive. Every optimum ν_t^{ref} of Equation (3.3) admits $\alpha_j \ge 0$ supported on active contrasts, with $\sum_j \alpha_j = 1$, and a common value ζ_t such that*

$$\mathsf{MB}_t(a; \nu_t^{\mathrm{ref}}) - \lambda_t h_t \rho_t \Delta_t^{\mathrm{ref}}(a) \begin{cases} = \zeta_t, & \nu_t^{\mathrm{ref}}(a) > 0, \\ \le \zeta_t, & \nu_t^{\mathrm{ref}}(a) = 0. \end{cases}$$

Together with feasibility and complementary slackness, these conditions are sufficient.

In the symmetric campaign, both challenger constraints bind and the rule sends 150 evaluations to each. The equality emerges from symmetry. If a fixed reserve $s = n_B + n_C$ faces $\Delta_C > \Delta_B$, then, after dropping the common A-arm term, the objective is $\max\{n_B^{-1}, n_C^{-1}\} + \lambda(\Delta_B n_B + \Delta_C n_C)$. The balanced split remains optimal only when $\lambda(\Delta_C - \Delta_B) \le 4/s^2$; above that threshold, $n_C = \{\lambda(\Delta_C - \Delta_B)\}^{-1/2} < s/2$ and $n_B = s - n_C$. Thus KKT balances directional repair against its unequal economic price.

3.2 From the oracle allocation to a valid adaptive log

An oracle allocation does not by itself define an implementable experiment. Can evaluation outcomes improve the information design without silently advancing the base learner, and can the resulting data still support inverse-propensity inference? Algorithm 1 answers both questions by freezing every score and propensity before the outcome, updating the information history on every round, and updating the base history only when $E_t = 0$. This separation is exactly what Theorem 4.2 uses to recover the base-policy count and welfare curve; the logged π_t then supports exact moments, finite-sample coverage, and studentization in Theorem 4.4, Proposition F.1, and Corollary F.1.

In Example 2.1, $p_t = e_A$, $\rho_t = 0.2$, and $\nu_t = (0, 1/2, 1/2)$ generate $\pi_t = (0.8, 0.1, 0.1)$. Evaluation responses update $\mathcal{H}_t^{\mathrm{info}}$ while the base learner remains at the same decision time. The next section asks whether pilot error, random clock length, and random assignments preserve the promised ledger rather than merely its expectation.

Algorithm 1 Clock-separated information-flow wrapper (CS–IF)

Require: base policy Base; $(\rho_t, \lambda_t, h_t, \tau_t)$; targets $\mathcal{C}_{\text{par}}$ or $\mathcal{C}_{\text{arm}}$ for the chosen ledger.
1: Initialize $\mathcal{H}_0^{\text{base}}$, $\mathcal{H}_0^{\text{info}}$ and AIPW accumulators.
2: **for** $t = 1, \dots, T$ **do**
3: Set $p_t \leftarrow \mathsf{Base}(\mathcal{H}_{t-1}^{\text{base}})$; from $\mathcal{H}_{t-1}^{\text{info}}$, compute and freeze m_t, gap estimates, and information loads.
4: Solve Equation (3.3) with plug-in inputs for ν_t; store $\pi_t = (1-\rho_t)p_t + \rho_t\nu_t$.
5: Draw $E_t \sim \text{Ber}(\rho_t)$, then $I_t \sim p_t$ if $E_t = 0$, and $I_t \sim \nu_t$ otherwise.
6: Observe Y_t; log $(E_t, I_t, Y_t, \pi_t, m_t)$ and update AIPW via Equation (4.8).
7: Update $\mathcal{H}_t^{\text{info}}$; append to $\mathcal{H}_t^{\text{base}}$ iff $E_t = 0$.
8: **end for**
9: **return** welfare/information ledgers, propensity log, and AIPW inference.

4 Theory: Learning, Transferring, and Pricing Information

The reference ledger is now explicit, but a reviewer should still ask whether it survives contact with an unknown instance and random deployment. We answer in four steps. Theorem 4.1 follows pilot error through arm selection, allocation, welfare, and target geometry; Theorems 4.2 and 4.3 then separate the count and matrix mechanisms inside that result. Next, Theorem 4.4, Proposition F.1, and Corollaries F.1 and F.2 turn the propensity log into finite- and large-sample inference. Finally, Theorem 4.5 and Corollary 4.1 show, at the reference instance and up to constants, why the weak direction must be purchased. Each step returns to the same campaign, while keeping the exact unregularized frontier separate from stabilized learned deployment.

4.1 Pilot-to-deployment guarantee

The oracle knows the reward gaps; the retailer does not. Could a small pilot error be amplified until a stable reference allocation becomes nearly one-sided? We answer in the train-once, deploy-once regime: an independent Gaussian pilot estimates the gaps, after which the plug-in design is frozen. Theorem 4.1 shows that one pilot event preserves the best arm, bounds design displacement, and propagates the same error into evaluation counts, welfare, and reference-normalized information. The assignment noise incurred after freezing is then handled separately by matrix concentration. Theorems 4.2 and 4.3 unpack these two transfer mechanisms in the next subsection.

For pilot Gram matrix V_m, feature bound L_x, and welfare coefficient κ_{w}, define $r_m(\delta_0) = \sigma\{\sqrt{d} + \sqrt{2\log(1/\delta_0)}\}/\sqrt{\lambda_{\min}(V_m)}$ and $C_m = 4\kappa_{\text{w}} L_x r_m(\delta_0)/\tau$. Here ν^{ref} is the $\tau > 0$ regularized reference design in Condition E.1; it is distinct from the $\tau = 0$ closed-form frontier. During frozen deployment let $R_{\text{eval}} = \sum_{t>m} \rho_t$, let $\varepsilon_T^{\text{eval}}$ be the maximum reference-to-plug-in evaluation count drift, and let M_T^{ref} and M_T^{plug} denote the reference and frozen plug-in acquisition matrices. In this specialization, $M_T^{\text{plug}} = \mathbb{E}(M_T \mid \mathcal{H}_{\text{pil}})$; there is no extra plug-in-to-predictable drift. Write ℓ_T^{ref} for the maximum reference-normalized one-step leverage and $\varepsilon_T^{\text{info}}$ for the normalized distance between these two matrices. Full definitions appear in Condition E.1.

Theorem 4.1 (Pilot-to-deployment certificate)**.** *For $\delta_0, \delta \in (0,1)$, under Condition E.1, if $2L_x r_m(\delta_0) < \gamma$, then, with pilot probability at least $1 - \delta_0$,*

$$\widehat{a} = a^\star, \qquad \|\widehat{\nu} - \nu^{\text{ref}}\|_1 \le C_m, \qquad \varepsilon_T^{\text{eval}} \le R_{\text{eval}} C_m, \tag{4.1}$$

$$\varepsilon_T^{\text{info}} \le R_{\text{eval}} \ell_T^{\text{ref}} C_m, \qquad |\mathbb{E}(R_{\text{dep}} \mid \mathcal{H}_{\text{pil}}) - R_{\text{dep}}^{\text{ref}}| \le \|\Delta\|_\infty R_{\text{eval}} C_m. \tag{4.2}$$

For $x = \log(2d/\delta)$, set $\eta_T = R_{\text{eval}}\ell_T^{\text{ref}} C_m + \sqrt{2\ell_T^{\text{ref}}(1 + R_{\text{eval}}\ell_T^{\text{ref}} C_m)x} + 2\ell_T^{\text{ref}} x/3$. Jointly over pilot and deployment, with probability at least $1 - \delta_0 - \delta$, if $\eta_T < 1$, then every c satisfies

$$\frac{c^\top (M_T^{\text{ref}})^{-1} c}{1 + \eta_T} \le c^\top M_T^{-1} c \le \frac{c^\top (M_T^{\text{ref}})^{-1} c}{1 - \eta_T}. \tag{4.3}$$

For the campaign, $2L_x r_m(\delta_0) < 0.10$ keeps A as the deployed incumbent. If the pilot is balanced so that $\lambda_{\min}(V_m) \ge \kappa_{\text{pil}} m$ for fixed $\kappa_{\text{pil}} > 0$, then $C_m = O(m^{-1/2})$ at fixed confidence and the plug-in design remains close to its regularized reference. The exact $(150, 150)$ allocation remains the $\tau = 0$ benchmark. The final sandwich separates learned-design error from deployment randomness.

4.2 TWO TRANSFER MECHANISMS

What exactly survives from the reference design? Counts and welfare require a time-change argument, whereas contrast precision requires a matrix argument; neither can substitute for the other. Theorem 4.2 decomposes count error into base tracking, decision-clock fluctuation, learned evaluation allocation, and evaluation randomization, then gap-weights those terms into a welfare bound. Conditional on the predictable design target, Theorem 4.3 instead normalizes matrix drift by M_T^{ref} and converts drift and leverage into a simultaneous inverse guarantee. Their composition is the deployment half of Theorem 4.1.

Let $D_T = \sum_t (1 - E_t)$, $d_T^{\mathrm{ref}} = \sum_t (1 - \rho_t)$, $v_D = \sum_t \rho_t(1 - \rho_t)$, $\bar{N}_{a,T}^{\mathrm{eval}} = \sum_t \rho_t \nu_t^{\mathrm{ref}}(a)$, and $r_T^{\mathrm{eval}} = \sum_t \rho_t$. Let $N_a^{\mathrm{base}}(n)$ be the base policy's pathwise count after n rounds on the non-evaluation subsequence, coupled to the wrapper so that $N_a(T) = N_a^{\mathrm{base}}(D_T) + \sum_t E_t \mathbf{1}\{I_t = a\}$. Let $\varepsilon_T^{\mathrm{eval}}$ bound the plug-in evaluation allocation drift.

Theorem 4.2 (Reference clock transfer)**.** *Let $\delta, \delta_{\mathrm{base}}, \delta_{\mathrm{eval}} \in (0,1)$. Assume $(\rho_t, \nu_t^{\mathrm{ref}})$ is deterministic and the implemented ν_t is $\mathcal{F}_{t-1}$-measurable. Let $n_a^\star : [0,T] \to \mathbb{R}$ be deterministic, and let $e_{a,T}, \Lambda_{a,T} \geq 0$. Put $x = \log\{2(K+1)/\delta\}$ and $u_D = \sqrt{2v_D x} + 2x/3$. If, with probability at least $1 - \delta_{\mathrm{base}}$, simultaneously for all a and integers $n \in [0,T]$ satisfying $|n - d_T^{\mathrm{ref}}| \leq u_D$,*

$$|N_a^{\mathrm{base}}(n) - n_a^\star(n)| \leq e_{a,T}, \qquad |n_a^\star(n) - n_a^\star(d_T^{\mathrm{ref}})| \leq \Lambda_{a,T} |n - d_T^{\mathrm{ref}}|, \tag{4.4}$$

and, with probability at least $1 - \delta_{\mathrm{eval}}$, $\max_a |\sum_t \rho_t \{\nu_t(a) - \nu_t^{\mathrm{ref}}(a)\}| \leq \varepsilon_T^{\mathrm{eval}}$, then, with probability at least $1 - \delta_{\mathrm{base}} - \delta_{\mathrm{eval}} - \delta$, simultaneously in a,

$$|N_a(T) - n_a^\star(d_T^{\mathrm{ref}}) - \bar{N}_{a,T}^{\mathrm{eval}}| \leq e_{a,T} + \Lambda_{a,T} u_D + \varepsilon_T^{\mathrm{eval}} + \sqrt{2 r_T^{\mathrm{eval}} x} + 2x/3. \tag{4.5}$$

The stationary pseudo-regret error around $R_T^{\mathrm{ref}} = \sum_a \Delta_a \{n_a^\star(d_T^{\mathrm{ref}}) + \bar{N}_{a,T}^{\mathrm{eval}}\}$ is at most the gap-weighted sum of the right-hand side.

For the retailer, $\mathbb{E} D_T = 1200$, while $\mathrm{Var}(D_T) = 1500(0.2)(0.8) = 240$, a standard deviation of about 15.5. Thus the base policy need not run exactly 1200 rounds and the challengers need not receive exactly 150 assignments; Theorem 4.2 states which of the four mechanisms accounts for each discrepancy.

For the geometric ledger, let $\mathcal{I}_0 \subseteq \mathcal{F}_0$ be an initial sigma-field, enlarge the working filtration by $\mathcal{I}_0$, and let $M_T^{\mathrm{ref}} \succ 0$ be $\mathcal{I}_0$-measurable. A pilot-fixed reference uses the pilot sigma-field; a deterministic reference uses the trivial field. Let deterministic $b, \ell \geq 0$ satisfy, with probability at least $1 - \delta_0$, $\|(M_T^{\mathrm{ref}})^{-1/2}(\overline{M}_T - M_T^{\mathrm{ref}})(M_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}} \leq b$, and assume almost surely that every normalized one-step contribution obeys $0 \preceq (M_T^{\mathrm{ref}})^{-1/2} G_t(a) (M_T^{\mathrm{ref}})^{-1/2} \preceq \ell I$.

Theorem 4.3 (Conditional target-normalized concentration)**.** *For $\delta_0 \in [0,1)$ and $\delta \in (0,1)$, under the preceding drift and leverage conditions, for $x = \log(2d/\delta)$, with probability at least $1 - \delta_0 - \delta$,*

$$\|(M_T^{\mathrm{ref}})^{-1/2}(M_T - M_T^{\mathrm{ref}})(M_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}} \leq \eta := b + \sqrt{2\ell(1+b)x} + 2\ell x/3. \tag{4.6}$$

If $\eta < 1$, then every c obeys

$$\frac{c^\top (M_T^{\mathrm{ref}})^{-1} c}{1 + \eta} \leq c^\top M_T^{-1} c \leq \frac{c^\top (M_T^{\mathrm{ref}})^{-1} c}{1 - \eta}. \tag{4.7}$$

With unit noise, the campaign reference $M_T^{\mathrm{ref}} = \mathrm{diag}(1200, 150, 150)$ assigns variance 0.0075 to both targets. Theorem 4.3 turns that number into the explicit multiplicative interval $[0.0075/(1+\eta), 0.0075/(1-\eta)]$, an explicit directional guarantee based on realized geometry.

4.3 FROM THE PROPENSITY LOG TO VALID INFERENCE

The matrix certificate establishes Gaussian acquisition geometry. Valid confidence intervals under adaptive randomization require a design-based argument. Can the same log support exact finite-sample statements and a studentized limit even when evaluation decays? The chain is progressive: Theorem 4.4 gives exact design moments, Proposition F.1 gives finite-horizon coverage, Corollary F.1 replaces the deterministic envelope with observable studentization, and Corollary F.2 exposes the

resulting welfare–precision rate. This subsection uses the design-based ledger, whose fixed-potential-outcome variance differs from the model-based 0.0075.

Condition throughout this subsection on the fixed potential-outcome array; all probabilities and expectations are over the assignment mechanism. For one arm-space contrast $c^{\mathrm{arm}} \in \mathcal{C}_{\mathrm{arm}}$, define the predictable residual score $\psi_t(a) = c_a^{\mathrm{arm}}\{Y_t(a) - m_t(a)\}$, the finite-population target $\tau_T = T^{-1}\sum_{t,a} c_a^{\mathrm{arm}} Y_t(a)$, and

$$\widehat{\tau}_T = T^{-1}\sum_t \left\{ \sum_a c_a^{\mathrm{arm}} m_t(a) + \frac{\psi_t(I_t)}{\pi_t(I_t)} \right\}. \tag{4.8}$$

A ratio with $\psi_t(a) = \pi_t(a) = 0$ is defined as zero.

Theorem 4.4 (Exact unbiasedness and finite-sample variance)**.** *If m_t, π_t are $\mathcal{F}_{t-1}$-measurable, $\pi_t(a) > 0$ whenever $\psi_t(a) \neq 0$, and $\mathbb{E}\sum_{t,a} \psi_t(a)^2/\pi_t(a) < \infty$, then $\mathbb{E}\widehat{\tau}_T = \tau_T$ and*

$$\mathrm{Var}(\widehat{\tau}_T) = T^{-2}\mathbb{E}\sum_t \left\{ \sum_a \frac{\psi_t(a)^2}{\pi_t(a)} - \Big(\sum_a \psi_t(a)\Big)^2 \right\}. \tag{4.9}$$

If $|\psi_t(a)| \le U_t$, define $p_t^{\min} = \min_{a:\psi_t(a)\neq 0} \pi_t(a)$ when the index set is nonempty and set $p_t^{\min} = 1$, $U_t = 0$ otherwise. Then $\mathrm{Var}(\widehat{\tau}_T) \le KT^{-2}\mathbb{E}\sum_t U_t^2/p_t^{\min}$.

For the deterministic campaign outcomes $Y(A) = 0.3, Y(B) = Y(C) = 0.2$, $m_t = 0$, and target $B - A$, the logged $(0.8, 0.1, 0.1)$ gives design variance 0.000335 and standard error 0.0183. These numbers average over assignment and correspond to a different probability law from the unit-noise Gaussian acquisition variance 0.0075.

Beyond exact moments, Proposition F.1 gives propensity-sensitive finite-horizon coverage, Corollary F.1 proves observable studentization, and Corollary F.2 shows that $\rho_t \asymp t^{-\alpha}$ yields regret $T^{1-\alpha}$ and s.e. $T^{-(1-\alpha)/2}$, hence $T^{2/3}$ and $T^{-1/3}$ at $\alpha = 1/3$. Statements and proofs are in Section F; the model-based Wald route is separate in Theorem H.1.

4.4 Reference-design attainability and adaptive necessity

Could another adaptive policy avoid the weak-direction price while respecting the same campaign budget? The answer is local and reference-based. Theorem 4.5 gives a two-point risk floor for every budget-feasible policy under its information and local-alternative conditions; rounded fixed designs and ordinary least squares (OLS) attain the reference oracle up to rounding. Corollary 4.1 then solves the unregularized reference design exactly and returns the four campaign numbers, closing that loop with Theorem 3.1. Theorem 4.1 separately certifies the regularized learned deployment.

Fix x_a, $\sigma^2 > 0$, and $Y_t \mid (\mathcal{F}_{t-1}, I_t = a) \sim \mathrm{N}(x_a^\top\theta, \sigma^2)$ for $\theta \in \Theta \subseteq \mathbb{R}^d$, with θ-independent action kernels. At θ_0, let $\Delta_a = \max_b x_b^\top\theta_0 - x_a^\top\theta_0$, and let $\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})$ be the welfare-constrained directional oracle in Equation (G.1).

Theorem 4.5 (Reference-instance local geometric bound)**.** *For such an adaptive policy with $\mathbb{E}_{\theta_0} R_T \le \mathfrak{B}$ and estimators $(\widehat{\tau}_j)$, put $n_a = \mathbb{E}_{\theta_0} N_a(T)$. Assume $M(n) \succ 0$, choose $j^\star \in \arg\max_j (c_j^{\mathrm{par}})^\top M(n)^{-1} c_j^{\mathrm{par}}$, set $v = (c_{j^\star}^{\mathrm{par}})^\top M(n)^{-1} c_{j^\star}^{\mathrm{par}} > 0$ and $\theta_1 = \theta_0 + M(n)^{-1}c_{j^\star}^{\mathrm{par}}/\sqrt{v}$, and assume this local alternative belongs to Θ. Then*

$$\max_{i\in\{0,1\}} \max_j \mathbb{E}_{\theta_i}(\widehat{\tau}_j - (c_j^{\mathrm{par}})^\top\theta_i)^2 \ge \Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})/16. \tag{4.10}$$

If continuous minimizers $n_T^\star$ exist along a $T \to \infty$ sequence, fixed-design OLS at any feasible integer allocation $\widetilde{n}_T$ has worst-contrast variance $\max_j (c_j^{\mathrm{par}})^\top M(\widetilde{n}_T)^{-1} c_j^{\mathrm{par}}$. Any feasible sequence of roundings such that $\|M(n_T^\star)^{-1/2}\{M(\widetilde{n}_T) - M(n_T^\star)\}M(n_T^\star)^{-1/2}\|_{\mathrm{op}} = o(1)$ attains $(1 + o(1))\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})$.

Corollary 4.1 (Closed-form three-arm frontier)**.** *In Example 2.1, with unit noise and common challenger gap $g > 0$, if $0 < \mathfrak{B}/g < \{\sqrt{2}/(1+\sqrt{2})\}T$, the continuous reference optimum is*

$$n_A = T - \mathfrak{B}/g, \qquad n_B = n_C = \mathfrak{B}/(2g), \qquad \Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}}) = \frac{1}{T - \mathfrak{B}/g} + \frac{2g}{\mathfrak{B}}. \tag{4.11}$$

For $(T, g, \mathfrak{B}) = (1500, 0.1, 30)$, this is $(1200, 150, 150)$ and 0.0075.

The balanced integer reference attains 0.0075; under Theorem 4.5's conditions every budget-feasible policy has risk at least $0.0075/16$ at one local point. The weak-direction price is therefore intrinsic.

5 Experiments: Evidence Across the Chain

Questions and protocol. We test four implications under a shared schedule or matched regret: clock transfer, design-based validity, directional geometry, and pilot-to-held-out transfer. Precision is the worst prespecified contrast; full protocols are in Section I.

Synthetic theorem checks. In the Gaussian MAB, both two-clock wrappers track the additive reference, whereas one-clock feedback drives the ratio to 0.85; targeting moves (regret, s.e.) from $(171.0, 0.0607)$ to $(157.2, 0.0531)$ (Figure 2a). On a frozen $T = 1500$ population, width moves 0.0246→0.0223 at 94.0% coverage (panel b). At matched regret, exact A/G-optimal designs attain a $6.3\times$ larger $\lambda_{\min}$, while target-max CS–IF attains a 28.5% lower target s.e. (panel c), the reversal in Proposition 2.1.

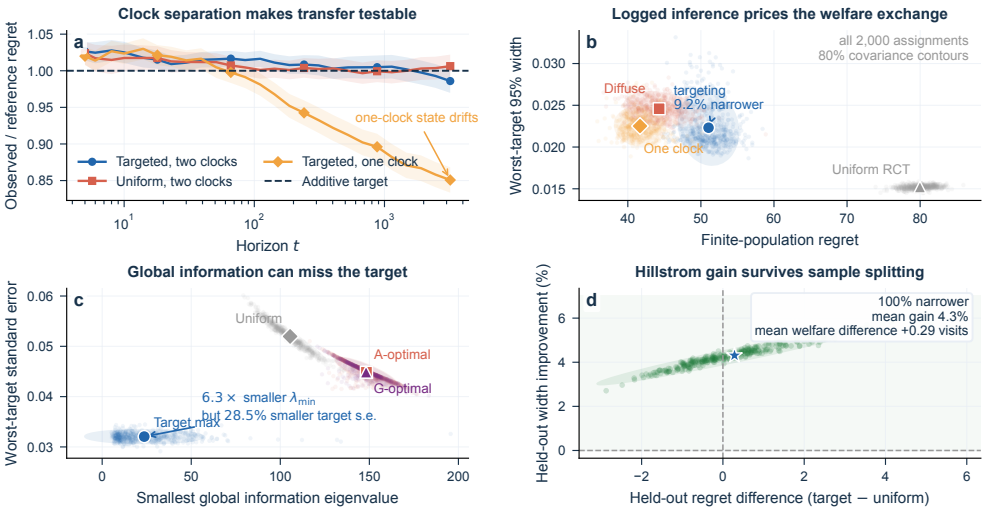


Figure 2: Evidence along the welfare-to-inference chain: (a) clock transfer; (b) finite-population regret–width; (c) global versus target geometry; and (d) Hillstrom split transfer. Pale points are repetitions; bands are 95% Monte Carlo intervals and ellipses are descriptive contours. Full validity and certificate diagnostics are in Figures 5, 6 and 13.

Baselines, sensitivity, and randomized data. At equal evaluation mass, targeted CS–IF lowers s.e. by 20.0% versus MAD–UCB (Liang & Bojinov, 2023), at a 14.7% regret cost; the finite grid gives lower target s.e. at every budget from 120 to 300. Pilot design error scales as $m^{-0.54}$, and all pilots satisfy the radius event and identify the best action (Section I.8). On Hillstrom's $64{,}000$-customer randomized campaign (Hillstrom, 2008), all 300 splits favor targeting: prospective plug-in width for a $12{,}000$-customer experiment falls by 4.3% at $+0.29$ expected lost visits (panel d). Additional baselines, calibration checks, rate studies, ablations, and sensitivity analyses appear in Section I.

6 Discussion

CS–IF treats evaluation as a target-specific information portfolio. It prices prespecified directions, separates the reward and evaluation clocks, and carries a reference design through finite-sample transfer and logged inference. At fixed evaluation mass or welfare budget, targeting can sharpen weak-contrast estimation even when a diffuse design has larger armwise counts or a larger minimum eigenvalue. Classical frontiers price how much to explore; ours prices where information flows.

AI USE STATEMENT

Generative AI tools were used for concept generation and stress testing; framework, method, and theorem ideation; literature discovery; mathematical proof development and Lean formalization; experimental design; code implementation and debugging; simulation and result interpretation; figure design; and substantial manuscript drafting and language editing. Model output was not treated as factual or bibliographic evidence. The released artifact records citation checks, Lean compilation, numerical regeneration, and the proof status of each formal claim. The authors retain responsibility for the final text, claims, code, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The empirical study reuses an existing randomized marketing experiment and reports aggregate outcomes. The method is intended to make the welfare and inferential costs of adaptive allocation explicit. Deployments involving high-stakes outcomes require domain-specific oversight, prespecified target contrasts, logged propensities, and minimum-probability safeguards.

REPRODUCIBILITY STATEMENT

The supplement contains detailed proofs, theorem assumptions, data-generating processes, baseline definitions, seeds, optimization tolerances, and the boundary of the Lean formalization. The accompanying artifact contains the raw randomized-data file, repetition-level CSV outputs, vector figures, and a single command that regenerates every numerical result and verifies finite outputs.

REFERENCES

Debangshu Banerjee, Avishek Ghosh, Sayak Ray Chowdhury, and Aditya Gopalan. Exploration in linear bandits with rich action sets and its implications for inference. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 8233–8262, 2023. URL https://proceedings.mlr.press/v206/banerjee23b.html.

Aurélien Bibaut, Maria Dimakopoulou, Nathan Kallus, Antoine Chambaz, and Mark van der Laan. Post-contextual-bandit inference. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28548–28559, 2021.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Jessica Dai, Paula Gradu, and Christopher Harshaw. CLIP-OGD: An experimental design for adaptive neyman allocation in sequential experiments. In *Advances in Neural Information Processing Systems*, volume 36, pp. 32235–32269, 2023. doi: 10.52202/075280-1400. URL https://doi.org/10.52202/075280-1400.

Vitor Hadad, David A. Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15):e2014602118, 2021.

Peter Hall and Christopher C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.

Dae Woong Ham, Iavor Bojinov, Michael Lindon, and Martin Tingley. Design-based inference for multi-arm bandits, 2023. arXiv:2302.14136.

Qiyang Han, Koulik Khamaru, and Cun-Hui Zhang. UCB algorithms for multi-armed bandits: Precise regret and adaptive inference, 2024. arXiv:2412.06126.

Kevin Hillstrom. The MineThatData e-mail analytics and data mining challenge. MineThatData challenge page, 2008. URL https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html. Posted March 20, 2008; accessed for the empirical analysis in this artifact.

Jack Kiefer and Jacob Wolfowitz. Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30(2):271–294, 1959.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Xinran Li and Anqi Zhao. Design-based theory for causal inference from adaptive experiments, 2026. arXiv:2602.21998.

Biyonka Liang and Iavor Bojinov. An experimental design for anytime-valid causal inference on multi-armed bandits, 2023. arXiv:2311.05794.

Subhojyoti Mukherjee, Qiaomin Xie, Josiah P. Hanna, and Robert Nowak. SPEED: Experimental design for policy evaluation in linear heteroscedastic bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2962–2970, 2024. URL https://proceedings.mlr.press/v238/mukherjee24a.html.

Georgy Noarov, Riccardo Fogliato, Martin Andres Bertran, and Aaron Roth. Stronger neyman regret guarantees for adaptive experimental design. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 46735–46761, 2025. URL https://proceedings.mlr.press/v267/noarov25a.html.

Friedrich Pukelsheim. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.

David Simchi-Levi and Chonghuan Wang. Multi-armed bandit experimental design: Online decision-making and adaptive inference. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 3086–3097, 2023. URL https://proceedings.mlr.press/v206/simchi-levi23a.html.

Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Hongrui Xie, Junyu Cao, and Kan Xu. Adaptive combinatorial experimental design: Pareto optimality for decision-making and inference. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 3466–3474, 2026. URL https://proceedings.mlr.press/v300/hongrui26a.html.

Shunxing Yan and Han Zhong. Optimism stabilizes thompson sampling for adaptive inference. In *Proceedings of the 39th Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6885–6886, 2026. URL https://proceedings.mlr.press/v336/yan26a.html.

APPENDIX CONTENTS

Item	Contents	Page
A	**Notation, assumptions, and proof map**	**13**
B	**Closest work and contribution boundary**	**14**
C	**Auxiliary matrix and martingale tools**	**14**
D	**Proof of the motivating failures**	**16**
E	**Acquisition, design, and clock proofs**	**16**
E.1	Target-normalized concentration	16
E.2	A finite-horizon drift certificate	17
E.3	A pilot-to-deployment certificate	17
E.4	KKT design certificate	18
E.5	Reference clock transfer	19
F	**Proofs for design-based inference**	**20**
G	**Welfare–information lower bounds**	**22**
H	**Structured linear inference with two-stage filtration**	**24**
I	**Experimental and computational details**	**25**
I.1	Theory-to-evidence map	25
I.2	Computing environment and file contract	26
I.3	Gaussian MAB	26
I.4	Fixed finite population	28
I.5	Linear structured reward	30
I.6	Hillstrom randomized experiment	31
I.7	Sensitivity experiments	32
I.8	Closest baselines and independent-pilot plug-in stress test	33
I.9	Key output inventory	35
J	**Lean 4 formalization boundary**	**37**

A NOTATION, ASSUMPTIONS, AND PROOF MAP

For symmetric M, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ are its extreme eigenvalues, $\|M\|_{\mathrm{op}}$ is its operator norm, and $M^{\dagger}$ is the Moore–Penrose inverse. We write $M \preceq N$ for Loewner order, $\|x\|_M = (x^\top M x)^{1/2}$ for $M \succ 0$, Δ_K for the simplex, and e_a for a canonical vector. Potential outcomes and covariates are fixed in the design-based results; probability then comes only from assignment. The acquisition ledger uses parameter directions $\mathcal{C}_{\mathrm{par}} \subset \mathbb{R}^d$, whereas the design-based ledger uses arm coefficients $\mathcal{C}_{\mathrm{arm}} \subset \mathbb{R}^K$. They coincide numerically in the three-arm mean model but remain distinct in a general structured model.

Table 1: Recurring conditions, available diagnostics or sensitivity bounds, and explicit fallbacks.

Condition	Role	Diagnostic	Fallback
Logged predictability	identifies the assignment martingale	replay $\pi_t(I_t)$; test same-period leakage	enlarge the filtration or model logging error
Target positivity	finite IPW scores in requested arms	minimum and harmonic propensity by target	restrict the estimand or use partial identification
Target nondegeneracy	relative matrix comparison	smallest target-span eigenvalue	generalized inverse on the estimable span
Normalized leverage	no unit dominates a target	largest one-step target-normalized contribution	retain finite-sample leverage term; truncate
Drift certificates	links implementation to reference	acquisition drift and KKT residual; evaluate drift only against a preregistered or independent reference	report conditional guarantee or redesign estimator
Lindeberg and score moments	Gaussian studentization	maximum standardized score; envelope bounds for H_T, J_T (oracle values only in simulation)	bounded-score methods or confidence sequences

Table 2: Dependency map for the headline results.

Result	Main ingredient	Location
Proposition 2.1	Bernoulli variance; three-arm diagonal comparison	Section D
Theorem 4.3	Condition C.1 and Lemmas C.1 to C.4	Sections C and E
Theorem E.1	Lemma C.5	Sections C and E
Theorem 4.1	Condition E.1, Gaussian pilot concentration, and Theorem 4.3	Section E
Theorem 3.1	convex epigraph KKT system	Section E
Theorem 4.2	two-clock decomposition and Lemma C.6	Sections C and E
Theorem 4.4	conditional score identities and martingale orthogonality	Section F
Proposition F.1	Condition F.1 and Lemma C.6	Sections C and F
Corollary F.1	Condition F.2 and Lemma F.1	Section F
Corollary F.2	Corollary F.1 and power-sum asymptotics	Section F
Theorem 4.5 and Corollary 4.1	adaptive KL and Le Cam; symmetric convex allocation	Section G
Proposition G.1	Cauchy–Schwarz and Jensen	Section G
Theorem H.1	two-stage filtration, martingale CLT, and Lemmas C.4 and C.7	Sections C and H

All logarithms are natural and numerical constants are not optimized. Equation (4.4) assumes a uniform base-policy certificate. One-horizon precise-pull results such as Han et al. (2024) motivate this input; the displayed simultaneous event additionally requires uniformization over its integer window. The clock-transfer implication is proved here. The formalization boundary is reported separately in Section J.

B CLOSEST WORK AND CONTRIBUTION BOUNDARY

Two closely related regret–inference frontiers are Simchi-Levi & Wang (2023), for ordinary MABs, and Xie et al. (2026), for combinatorial actions. They both control regret and worst-case error over all arm or super-arm gaps. CS–IF fixes the scientifically requested contrasts before sampling and optimizes their finite-horizon quadratic risk. In a homoskedastic Gaussian MAB, taking all pairwise contrasts recovers the same directional criterion; the target-restricted formulation also permits structured contributions $G_t(a)$. Global volume, trace, or minimum eigenvalue can improve while a requested direction remains weak, and design-based precision still depends on the full reciprocal-propensity path.

Adaptive-inference methods analyze a supplied history-dependent log (Hadad et al., 2021; Bibaut et al., 2021; Ham et al., 2023; Li & Zhao, 2026). Liang & Bojinov (2023) mix any bandit with a randomized evaluation channel and tune a scalar schedule for anytime-valid treatment-effect inference. CS–IF optimizes the channel distribution ν_t for prespecified targets and advances the base learner only on decision rounds. Its clock theorem transfers a base-policy guarantee to a directional welfare and information reference; the logged mixture supports exact AIPW moments, a finite-sample interval, and studentization. Adaptive Neyman designs instead learn a variance-optimal allocation for a treatment effect (Dai et al., 2023; Noarov et al., 2025); CS–IF couples multiple target directions to a separate reward clock and a welfare price.

For structured linear models, Mukherjee et al. (2024) learn a heteroscedastic design for one target-policy value; with full evaluation, one target direction, and zero welfare and entropy prices, our acquisition oracle reduces algebraically to that objective. Classical optimal-design criteria protect global geometry (Kiefer & Wolfowitz, 1959; Pukelsheim, 2006); for locally rich action sets, regret-induced Gram growth does as well (Banerjee et al., 2023). Precise UCB pull-count results provide one-horizon inputs for a uniform base-clock certificate, while stabilized Thompson-sampling results provide its asymptotic analogue (Han et al., 2024; Yan & Zhong, 2026); CS–IF prices the requested directions and controls pilot-to-reference and planned-to-realized drift. Table 3 records these axes.

Table 3: Closest theories differ in their target set, information state, and coupling between decisions and inference.

Work	Information state	Inferential target	Decision–information coupling
Han et al. (2024)	arm pulls	arm means	precise base-policy pull curve
Liang & Bojinov (2023)	scalar propensity mixture	treatment effect	fixed channel, anytime-valid schedule
Neyman designs (Dai et al., 2023; Noarov et al., 2025)	outcome-variance loss	treatment effect	approach the hindsight allocation
Mukherjee et al. (2024)	weighted Gram matrix	one target-policy value	learned heteroscedastic oracle
Pareto designs (Simchi-Levi & Wang, 2023; Xie et al., 2026)	gap estimation error	all arm or super-arm gaps	regret–error frontier
Design based (Ham et al., 2023; Li & Zhao, 2026)	full propensity path	finite-population effects	inference under a supplied design
This work	target Gram and full propensity path	prespecified contrasts	directional channel, two-clock transfer, local price

C AUXILIARY MATRIX AND MARTINGALE TOOLS

Condition C.1 (Target-normalized acquisition)**.** *Let $\mathcal{I}_0 \subseteq \mathcal{F}_0$ be an initial sigma-field, enlarge the working filtration by $\mathcal{I}_0$, and let $M_T^{\mathrm{ref}} \succ 0$ be $\mathcal{I}_0$-measurable. The deterministic-reference case takes $\mathcal{I}_0$ to be trivial. For deterministic $b, \ell \geq 0$, suppose*

$$\|(M_T^{\mathrm{ref}})^{-1/2}(\overline{M}_T - M_T^{\mathrm{ref}})(M_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}} \leq b \tag{C.1}$$

with probability at least $1-\delta_0$, and almost surely $0 \preceq (M_T^{\mathrm{ref}})^{-1/2} G_t(a) (M_T^{\mathrm{ref}})^{-1/2} \preceq \ell I$ for all t, a.

Lemma C.1 (Target-normalized variance bound)**.** *Let H be a random self-adjoint matrix satisfying $0 \preceq H \preceq \ell I$ almost surely, and let $X = H - \mathbb{E}(H \mid \mathcal{G})$. Then*

$$\mathbb{E}(X^2 \mid \mathcal{G}) = \mathbb{E}(H^2 \mid \mathcal{G}) - \mathbb{E}(H \mid \mathcal{G})^2 \preceq \ell \mathbb{E}(H \mid \mathcal{G}).$$

Moreover, $\|X\|_{\mathrm{op}} \le \ell$.

Proof. The spectral calculus gives $H^2 \preceq \ell H$. Expanding the conditional square and using $\mathbb{E}[H - \mathbb{E}(H \mid \mathcal{G}) \mid \mathcal{G}] = 0$ yields the equality. Since $\mathbb{E}(H \mid \mathcal{G})^2 \succeq 0$, the first inequality follows. Both H and $\mathbb{E}(H \mid \mathcal{G})$ lie in the Loewner interval $[0, \ell I]$. Hence $-\ell I \preceq H - \mathbb{E}(H \mid \mathcal{G}) \preceq \ell I$, which proves the norm bound. □

Lemma C.2 (Self-adjoint matrix Freedman)**.** *Let (X_t) be a self-adjoint matrix martingale-difference sequence of dimension d with $\|X_t\|_{\mathrm{op}} \le R$ almost surely. Let $W_T = \sum_{t=1}^{T} \mathbb{E}(X_t^2 \mid \mathcal{F}_{t-1})$. For all $u, v > 0$,*

$$\mathbb{P}\left\{ \lambda_{\max}\left(\sum_{t=1}^{T} X_t \right) \ge u,\ \lambda_{\max}(W_T) \le v \right\} \le d \exp\left\{ -\frac{u^2}{2(v + Ru/3)} \right\}.$$

The same bound applied to $-X_t$ controls the minimum eigenvalue; consequently, the corresponding operator-norm probability is bounded by twice the right-hand side.

Proof. This is the matrix Freedman inequality of Tropp (2011, Theorem 1.2). We restate it to fix constants and the filtration convention used below. □

Lemma C.3 (Convenient inversion of the Freedman tail)**.** *For $v, R, x > 0$, set $u = \sqrt{2vx} + 2Rx/3$. Then $u^2 \ge 2x(v + Ru/3)$.*

Proof. Write $a = \sqrt{2vx}$ and $b = 2Rx/3$, so $u = a + b$. Then

$$u^2 - 2x(v + Ru/3) = a^2 + 2ab + b^2 - a^2 - \frac{2Rx}{3}(a + b) = ab \ge 0,$$

because $b = 2Rx/3$. □

Lemma C.4 (Loewner inversion)**.** *If $A, B \in \mathbb{S}^d_{++}$ and $(1 - \eta)B \preceq A \preceq (1 + \eta)B$ for $0 \le \eta < 1$, then*

$$\frac{1}{1 + \eta} B^{-1} \preceq A^{-1} \preceq \frac{1}{1 - \eta} B^{-1}. \tag{C.2}$$

Proof. Conjugate by $B^{-1/2}$ to obtain $(1 - \eta)I \preceq B^{-1/2} A B^{-1/2} \preceq (1 + \eta)I$. Inversion reverses the Loewner order on positive-definite matrices. Conjugating the resulting inequality by $B^{-1/2}$ proves Equation (C.2). □

Lemma C.5 (Discrete Gronwall with cumulative perturbation)**.** *Let a nonnegative sequence $(v_t)_{t=0}^{T}$ satisfy $v_0 \le c$ and*

$$v_t \le c + \sum_{s=1}^{t} \Lambda_s^{\mathrm{dyn}} v_{s-1} \quad (t \le T),$$

where $c \ge 0$ and $\Lambda_s^{\mathrm{dyn}} \ge 0$. Then

$$\max_{t \le T} v_t \le c \prod_{s=1}^{T} (1 + \Lambda_s^{\mathrm{dyn}}) \le c \exp\left(\sum_{s=1}^{T} \Lambda_s^{\mathrm{dyn}} \right).$$

Proof. Induction gives $v_t \le c \prod_{s=1}^{t} (1 + \Lambda_s^{\mathrm{dyn}})$. The second inequality follows from $1 + x \le e^x$. □

Lemma C.6 (Scalar martingale Bernstein bound with deterministic variance envelope)**.** *Let (X_t) be a scalar martingale-difference sequence with $|X_t| \le 1$ and predictable variance $V_T = \sum_t \mathbb{E}(X_t^2 \mid \mathcal{F}_{t-1})$. For every $x, v > 0$,*

$$\mathbb{P}\left\{ \left| \sum_{t=1}^{T} X_t \right| > \sqrt{2vx} + \frac{2x}{3},\ V_T \le v \right\} \le 2e^{-x}.$$

Proof. Apply the scalar Freedman inequality to X_t and $-X_t$, then use Lemma C.3 with $R = 1$ and a union bound. □

Lemma C.7 (Perturbation of a normalized inverse). *Let $H = I + E \succ 0$ with $\|E\|_{\mathrm{op}} \le \eta < 1$. Then*

$$\left\|H^{-1} - I\right\|_{\mathrm{op}} \le \frac{\eta}{1-\eta}, \qquad \left\|H^{-1}\right\|_{\mathrm{op}} \le \frac{1}{1-\eta}.$$

Proof. The spectrum of H lies in $[1-\eta, 1+\eta]$. The claims follow by maximizing $|\lambda^{-1} - 1|$ and λ^{-1} on that interval. □

D Proof of the motivating failures

Proof of Proposition 2.1. For part (i), let Z_t indicate assignment to challenger B, with $Z_t \sim \mathrm{Ber}(\beta_t)$ independently. The HT mean satisfies

$$\mathbb{E}\widehat{\mu} = T^{-1}\sum_t 0.2\mathbb{E}(Z_t)/\beta_t = 0.2, \qquad \mathrm{Var}(\widehat{\mu}) = 0.04T^{-2}\sum_t \frac{1-\beta_t}{\beta_t}.$$

The stationary path $\beta_t = 0.1$ has variance $0.36/T$. Under the second campaign path, half of the propensities equal ε and half equal $0.2 - \varepsilon$, so the variance is

$$\frac{0.04}{2T}\left\{\frac{1-\varepsilon}{\varepsilon} + \frac{0.8+\varepsilon}{0.2-\varepsilon}\right\}.$$

Both paths have $\sum_t \beta_t = 0.1T$, hence the same expected B count, while the variance ratio diverges as $\varepsilon \downarrow 0$.

For part (ii), unit-noise Gaussian acquisition gives $M(n) = \mathrm{diag}(n_A, n_B, n_C)$. Both allocations sum to 1500, have trace 1500, and incur welfare $0.1(n_B + n_C) = 30$. The balanced design has worst-target variance $1/1200 + 1/150 = 0.0075$. Under $(1200, 299, 1)$, the $C - A$ variance is $1/1200 + 1 = 1.00083\ldots$, so the ratio is $1.00083\ldots/0.0075 = 133.44\ldots > 133$. □

E Acquisition, design, and clock proofs

E.1 Target-normalized concentration

Proof of Theorem 4.3. Work with the filtration enlarged by the initial field $\mathcal{I}_0$ from Condition C.1. Because M_T^{ref} is $\mathcal{I}_0$-measurable, normalization by it preserves predictability. Write $M^{\mathrm{ref}} = M_T^{\mathrm{ref}}$ and

$$H_t(a) = (M^{\mathrm{ref}})^{-1/2} G_t(a) (M^{\mathrm{ref}})^{-1/2}, \qquad \overline{H}_t = \sum_a \pi_t(a) H_t(a), \qquad X_t = H_t(I_t) - \overline{H}_t.$$

Predictability gives $\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = 0$. By Lemma C.1,

$$\|X_t\|_{\mathrm{op}} \le \ell, \qquad \mathbb{E}(X_t^2 \mid \mathcal{F}_{t-1}) \preceq \ell \overline{H}_t. \tag{E.1}$$

On the event in Equation (C.1), $(M^{\mathrm{ref}})^{-1/2}\overline{M}_T(M^{\mathrm{ref}})^{-1/2} \preceq (1+b)I$. Since $M_0 \succeq 0$, the predictable matrix variance in Equation (E.1) is at most $\ell(1+b)I$. Matrix Freedman (Lemmas C.2 and C.3), applied to X_t and $-X_t$, yields

$$\mathbb{P}\left(\left\|\sum_t X_t\right\|_{\mathrm{op}} > \sqrt{2\ell(1+b)x} + \frac{2\ell x}{3},\ \textit{Equation (C.1)} \text{ holds}\right) \le 2de^{-x} = \delta.$$

This intersection bound requires no independence between the drift event and the matrix martingale. Adding the probability at most δ_0 of the drift-event complement gives total failure probability at most $\delta_0 + \delta$. Finally,

$$(M^{\mathrm{ref}})^{-1/2}(M_T - M^{\mathrm{ref}})(M^{\mathrm{ref}})^{-1/2} = (M^{\mathrm{ref}})^{-1/2}(\overline{M}_T - M^{\mathrm{ref}})(M^{\mathrm{ref}})^{-1/2} + \sum_t X_t.$$

The triangle inequality proves Equation (4.6). When $\eta < 1$, its eigenvalues lie in $[1-\eta, 1+\eta]$; Lemma C.4 then gives Equation (4.7). □

E.2 A FINITE-HORIZON DRIFT CERTIFICATE

Theorem E.1 (Finite-horizon perturbation bound). *In a finite-dimensional normed space, suppose*

$$S_t = S_{t-1} + F_t(S_{t-1}) + \beta_t + \xi_t, \qquad S_t^{\mathrm{ref}} = S_{t-1}^{\mathrm{ref}} + F_t(S_{t-1}^{\mathrm{ref}}) + d_t^{\mathrm{ref}},$$

where F_t is Λ_t^{dyn}-Lipschitz. With $\mathcal{M}_t = \sum_{s\le t}\xi_s$, pathwise

$$\max_{t\le T}\|S_t - S_t^{\mathrm{ref}}\| \le \exp\left(\sum_{t\le T}\Lambda_t^{\mathrm{dyn}}\right)\left\{\|S_0 - S_0^{\mathrm{ref}}\| + \sum_{t\le T}\|\beta_t - d_t^{\mathrm{ref}}\| + \max_{t\le T}\|\mathcal{M}_t\|\right\}. \quad \text{(E.2)}$$

Proof. For $\mathcal{E}_t = S_t - S_t^{\mathrm{ref}}$, summation and Lipschitz continuity give

$$\|\mathcal{E}_t\| \le \|\mathcal{E}_0\| + \max_{r\le t}\|\mathcal{M}_r\| + \sum_{s\le t}\|\beta_s - d_s^{\mathrm{ref}}\| + \sum_{s\le t}\Lambda_s^{\mathrm{dyn}}\|\mathcal{E}_{s-1}\|.$$

Set

$$c_T = \|\mathcal{E}_0\| + \sum_{s\le T}\|\beta_s - d_s^{\mathrm{ref}}\| + \max_{r\le T}\|\mathcal{M}_r\|, \qquad v_t = \max_{r\le t}\|\mathcal{E}_r\|.$$

Taking the maximum over $r \le t$ in the preceding display yields $v_t \le c_T + \sum_{s\le t}\Lambda_s^{\mathrm{dyn}} v_{s-1}$. Apply Lemma C.5. The factor in Equation (E.2) is explicit; a useful long-horizon consequence requires that it and the bracket of $\mathcal{M}_t$ be controlled for the policy under study. □

E.3 A PILOT-TO-DEPLOYMENT CERTIFICATE

Condition E.1 (Independent-pilot frozen deployment). *Let $\max_a\|x_a\| \le L_x$ and observe a size-m pilot $Y_i = x_{I_i}^\top\theta + \varepsilon_i$, where the actions are fixed independently of $\varepsilon_i \overset{\mathrm{iid}}{\sim} \mathrm{N}(0,\sigma^2)$ and $V_m = \sum_{i\le m} x_{I_i}x_{I_i}^\top \succ 0$. Write $\mathcal{H}_{\mathrm{pil}} = \sigma\{(I_i, Y_i) : i \le m\}$, let $a^\star$ uniquely maximize $x_a^\top\theta$, let $\gamma = \min_{a\neq a^\star}(x_{a^\star} - x_a)^\top\theta > 0$, and form pilot OLS $\widehat{\theta}_m$, its maximizer $\widehat{a}$, and the corresponding true and plug-in gap vectors $\Delta, \widehat{\Delta}$.*

Let ω have full support, $\kappa_{\mathrm{w}} \ge 0$, $\tau > 0$, and let ϕ be a common deterministic convex criterion for which the following argmins exist:

$$r_m(\delta_0) = \frac{\sigma\{\sqrt{d} + \sqrt{2\log(1/\delta_0)}\}}{\sqrt{\lambda_{\min}(V_m)}}, \qquad C_m = \frac{4\kappa_{\mathrm{w}}L_x r_m(\delta_0)}{\tau}. \quad \text{(E.3)}$$

$$\nu^{\mathrm{ref}} \in \operatorname*{arg\,min}_{\nu\in\Delta_K}\{\phi(\nu) + \kappa_{\mathrm{w}}\langle\nu,\Delta\rangle + \tau\,\mathrm{KL}(\nu\|\omega)\}, \quad \widehat{\nu} \in \operatorname*{arg\,min}_{\nu\in\Delta_K}\{\phi(\nu) + \kappa_{\mathrm{w}}\langle\nu,\widehat{\Delta}\rangle + \tau\,\mathrm{KL}(\nu\|\omega)\}.$$

Thus ν^{ref} is the positive-τ regularized reference for this certificate; the closed-form $\tau = 0$ frontier is a separate benchmark. No pilot-growth condition is needed for the finite-m inequalities. Along a pilot sequence with $\lambda_{\min}(V_m) \ge \kappa_{\mathrm{pil}} m$ for fixed $\kappa_{\mathrm{pil}} > 0$, however, $r_m(\delta_0), C_m = O(m^{-1/2})$ at fixed δ_0. On fresh deployment rounds $t = m+1,\dots,T$, use $p_t = e_{\widehat{a}}$, $\nu_t = \widehat{\nu}$, and deterministic (ρ_t); put $R_{\mathrm{eval}} = \sum_{t>m}\rho_t$,

$$\pi_t^{\mathrm{ref}} = (1-\rho_t)e_{a^\star} + \rho_t\nu^{\mathrm{ref}}, \qquad \pi_t^{\mathrm{plug}} = (1-\rho_t)e_{\widehat{a}} + \rho_t\widehat{\nu},$$

$$\varepsilon_T^{\mathrm{eval}} = \max_a\left|\sum_{t>m}\rho_t\{\widehat{\nu}(a) - \nu^{\mathrm{ref}}(a)\}\right|.$$

For deterministic $G(a) \succeq 0$, define

$$M_T^{\mathrm{ref}} = M_0 + \sum_{i\le m}G(I_i) + \sum_{t>m}\sum_a \pi_t^{\mathrm{ref}}(a)G(a) \succ 0,$$

$$M_T^{\mathrm{plug}} \equiv \overline{M}_T = \mathbb{E}(M_T \mid \mathcal{H}_{\mathrm{pil}}) = M_0 + \sum_{i\le m}G(I_i) + \sum_{t>m}\sum_a \pi_t^{\mathrm{plug}}(a)G(a),$$

$$M_T = M_0 + \sum_{i \le m} G(I_i) + \sum_{t>m} G(I_t),$$
$$\ell_T^{\mathrm{ref}} = \max_a \|(M_T^{\mathrm{ref}})^{-1/2} G(a) (M_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}},$$
$$\varepsilon_T^{\mathrm{info}} = \|(M_T^{\mathrm{ref}})^{-1/2}(M_T^{\mathrm{plug}} - M_T^{\mathrm{ref}})(M_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}}.$$

The equality $M_T^{\mathrm{plug}} = \overline{M}_T$ is specific to this frozen-pilot deployment; it introduces no additional plug-in-to-predictable drift. Finally let $R_{\mathrm{dep}} = \sum_{t>m} \Delta_{I_t}$ and $R_{\mathrm{dep}}^{\mathrm{ref}} = \sum_{t>m} \langle \pi_t^{\mathrm{ref}}, \Delta \rangle$.

Proof of Theorem 4.1. Conditional on the fixed pilot actions, $V_m^{1/2}(\widehat{\theta}_m - \theta)/\sigma \sim \mathrm{N}(0, I_d)$. Gaussian norm concentration therefore gives the radius in Equation (E.3): $\|\widehat{\theta}_m - \theta\| \le r_m(\delta_0)$ with probability at least $1 - \delta_0$. On this event,

$$|(x_a - x_b)^\top (\widehat{\theta}_m - \theta)| \le 2L_x r_m(\delta_0) \quad \text{for every } a, b.$$

The gap condition therefore preserves the unique best arm and yields $\|\widehat{\Delta} - \Delta\|_\infty \le 2L_x r_m(\delta_0)$.

For a gap vector d, define

$$F_d(\nu) = \phi(\nu) + \kappa_{\mathrm{w}} \langle \nu, d \rangle + \tau\, \mathrm{KL}(\nu \| \omega).$$

Pinsker's inequality for the Bregman divergence of negative entropy shows that $\nu \mapsto \mathrm{KL}(\nu\|\omega)$ is one-strongly convex in ℓ_1 on the simplex. Thus F_d is τ-strongly convex. Since ν^{ref} minimizes F_Δ,

$$F_\Delta(\widehat{\nu}) - F_\Delta(\nu^{\mathrm{ref}}) \ge \frac{\tau}{2} \|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1^2,$$

whereas optimality of $\widehat{\nu}$ gives $F_{\widehat{\Delta}}(\widehat{\nu}) \le F_{\widehat{\Delta}}(\nu^{\mathrm{ref}})$. Subtracting the two objective comparisons gives

$$\frac{\tau}{2} \|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1^2 \le \kappa_{\mathrm{w}} \langle \widehat{\nu} - \nu^{\mathrm{ref}}, \Delta - \widehat{\Delta} \rangle \le \kappa_{\mathrm{w}} \|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1 \|\widehat{\Delta} - \Delta\|_\infty,$$

which proves $\|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1 \le C_m$. Because the design is frozen,

$$\max_a \left| \sum_{t=m+1}^{T} \rho_t \{\widehat{\nu}(a) - \nu^{\mathrm{ref}}(a)\} \right| \le R_{\mathrm{eval}} \|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1,$$

giving Equation (4.1). The two predictable acquisition targets differ by $\sum_{t>m} \rho_t \sum_a \{\widehat{\nu}(a) - \nu^{\mathrm{ref}}(a)\} G(a)$. If $H(a) = (M_T^{\mathrm{ref}})^{-1/2} G(a) (M_T^{\mathrm{ref}})^{-1/2}$, then

$$\varepsilon_T^{\mathrm{info}} \le \sum_{t>m} \rho_t \sum_a |\widehat{\nu}(a) - \nu^{\mathrm{ref}}(a)|\, \|H(a)\|_{\mathrm{op}} \le R_{\mathrm{eval}} \ell_T^{\mathrm{ref}} C_m.$$

On the same pilot event, $\widehat{a} = a^\star$, so the base-policy terms in the two conditional deployment regrets agree. Consequently,

$$|\mathbb{E}(R_{\mathrm{dep}} \mid \mathcal{H}_{\mathrm{pil}}) - R_{\mathrm{dep}}^{\mathrm{ref}}| = R_{\mathrm{eval}} |\langle \widehat{\nu} - \nu^{\mathrm{ref}}, \Delta \rangle| \le \|\Delta\|_\infty R_{\mathrm{eval}} C_m,$$

which completes Equation (4.2).

For each realized pilot in the preceding good event, condition on $\mathcal{H}_{\mathrm{pil}}$ and apply Theorem 4.3 with M_T^{ref} as the reference, $b = R_{\mathrm{eval}} \ell_T^{\mathrm{ref}} C_m$, and $\ell = \ell_T^{\mathrm{ref}}$; the drift bound is then deterministic, so the theorem is used with $\delta_0 = 0$. This gives Equation (4.3) except on a deployment event of conditional probability at most δ. The tower property bounds the joint failure probability by $\mathbb{P}(\text{pilot event}^{\mathsf{c}}) + \delta \le \delta_0 + \delta$. □

E.4 KKT DESIGN CERTIFICATE

Proof of Theorem 3.1. Write $g_a = \lambda_t h_t \rho_t \Delta_t^{\mathrm{ref}}(a)$ and introduce an epigraph variable z. On the common effective domain, the per-update problem is

$$\min_{\nu, z} \; z + \sum_a g_a \nu_a \quad \text{subject to} \quad \phi_{tj}(\nu) \le z \; (j \le q), \quad \nu_a \ge 0, \quad \sum_a \nu_a = 1.$$

Relative Slater feasibility and closed convexity make the KKT conditions necessary and sufficient. With epigraph multipliers $\alpha_j \geq 0$, nonnegativity multipliers $\mu_a \geq 0$, and equality multiplier χ, the Lagrangian is

$$\mathcal{L}(\nu, z, \alpha, \mu, \chi) = z + \sum_a g_a \nu_a + \sum_j \alpha_j \{\phi_{tj}(\nu) - z\} - \sum_a \mu_a \nu_a + \chi \left(\sum_a \nu_a - 1 \right).$$

Stationarity in z gives $\sum_j \alpha_j = 1$. Epigraph complementary slackness gives $\alpha_j \{\phi_{tj}(\nu) - z\} = 0$, so α_j is supported on the active contrasts. For acquisition information,

$$\frac{\partial}{\partial \nu_a} (c_j^{\mathrm{par}})^\top M(\nu)^{-1} c_j^{\mathrm{par}} = -h_t \rho_t (c_j^{\mathrm{par}})^\top M(\nu)^{-1} G_t^{\mathrm{ref}}(a) M(\nu)^{-1} c_j^{\mathrm{par}};$$

for design variance,

$$\frac{\partial}{\partial \nu_a} \sum_b \frac{w_{tjb}^{\mathrm{ref}}}{\pi_t^{\mathrm{ref}}(\nu, b)} = -\rho_t \frac{w_{tja}^{\mathrm{ref}}}{\pi_t^{\mathrm{ref}}(\nu, a)^2}.$$

For a zero load at zero propensity, both the summand and its marginal are zero under the lower-semicontinuous convention in Equation (3.2). Stationarity in ν_a is

$$\sum_j \alpha_j \partial_{\nu(a)} \phi_{tj}(\nu) + g_a - \mu_a + \chi = 0.$$

Therefore $\mathsf{MB}_t(a; \nu) - g_a = \chi - \mu_a$. Nonnegativity complementary slackness, $\mu_a \nu_a = 0$, gives equality to the common value $\zeta_t = \chi$ when $\nu_a > 0$ and an inequality when $\nu_a = 0$. Conversely, the displayed equalization defines nonnegative μ_a satisfying stationarity; together with primal feasibility and both complementary-slackness relations, convex KKT sufficiency proves optimality. □

E.5 Reference clock transfer

Proof of Theorem 4.2. The decision-clock martingale is

$$D_T - d_T^{\mathrm{ref}} = \sum_t \{(1 - E_t) - (1 - \rho_t)\}.$$

Because $E_t \mid \mathcal{F}_{t-1} \sim \mathrm{Ber}(\rho_t)$, its increments are bounded by one and its predictable variance is the deterministic $v_D = \sum_t \rho_t(1 - \rho_t)$. If $v_D = 0$, this martingale is identically zero; otherwise Lemma C.6 gives $|D_T - d_T^{\mathrm{ref}}| \leq u_D$ except on an event of probability $2e^{-x}$.

Let $N_{a,T}^{\mathrm{eval}} = \sum_t E_t \mathbf{1}\{I_t = a\}$ and $Z_{t,a} = E_t \mathbf{1}\{I_t = a\} - \rho_t \nu_t(a)$. Then predictability of ν_t and the two-stage draw imply $\mathbb{E}(Z_{t,a} \mid \mathcal{F}_{t-1}) = 0$, $|Z_{t,a}| \leq 1$, and

$$\sum_t \mathbb{E}(Z_{t,a}^2 \mid \mathcal{F}_{t-1}) \leq \sum_t \rho_t \nu_t(a) \leq r_T^{\mathrm{eval}}.$$

If $r_T^{\mathrm{eval}} = 0$, every evaluation martingale is identically zero. Otherwise, Lemma C.6 and a union bound show that, except on a probability $2Ke^{-x}$ event, simultaneously for all arms,

$$\left| N_{a,T}^{\mathrm{eval}} - \sum_t \rho_t \nu_t(a) \right| \leq \sqrt{2 r_T^{\mathrm{eval}} x} + \frac{2x}{3}.$$

On the evaluation-allocation event in Theorem 4.2, the left centering differs from $\bar{N}_{a,T}^{\mathrm{eval}}$ by at most $\varepsilon_T^{\mathrm{eval}}$. On Equation (4.4) and the decision-clock event, D_T is an integer in $[0, T]$ satisfying $|D_T - d_T^{\mathrm{ref}}| \leq u_D$, and hence

$$\left| N_a^{\mathrm{base}}(D_T) - n_a^\star(d_T^{\mathrm{ref}}) \right| \leq e_{a,T} + \Lambda_{a,T} u_D.$$

The pathwise base-subsequence coupling gives $N_a(T) = N_a^{\mathrm{base}}(D_T) + N_{a,T}^{\mathrm{eval}}$. Summing the last two bounds proves Equation (4.5). The choice of x makes $2(K+1)e^{-x} = \delta$. Adding the failure probabilities of the two certificate events proves the stated probability. Finally, $R_T = \sum_a \Delta_a N_a(T)$ gives the regret claim by the triangle inequality. □

F PROOFS FOR DESIGN-BASED INFERENCE

Throughout, the potential outcomes and covariates are fixed and expectations are over the adaptive assignment mechanism. Whenever a score and its propensity are both zero, every ratio involving that score is defined as zero.

Condition F.1 (Deterministic score envelope)**.** *In addition to Theorem 4.4, suppose deterministic* $U_t \geq 0$ *and* $\underline{p}_t > 0$ *satisfy* $|\psi_t(a)| \leq U_t$ *and* $\pi_t(a) \geq \underline{p}_t$ *on every nonzero score. Define*

$$v_T^{\mathrm{env}} = K \sum_t U_t^2 / \underline{p}_t, \qquad b_T^{\mathrm{env}} = \max_t U_t (K + \underline{p}_t^{-1}),$$

$$w_T(\delta) = T^{-1}\{\sqrt{2 v_T^{\mathrm{env}} \log(2/\delta)} + (2/3) b_T^{\mathrm{env}} \log(2/\delta)\}.$$

Proposition F.1 (Finite-sample envelope interval)**.** *For* $\delta \in (0,1)$*, suppose the conditions of Theorem 4.4 hold and deterministic* $U_t \geq 0$*,* $\underline{p}_t > 0$ *satisfy* $|\psi_t(a)| \leq U_t$ *and* $\pi_t(a) \geq \underline{p}_t$ *on every nonzero score. Set* $v_T^{\mathrm{env}} = K \sum_t U_t^2/\underline{p}_t$*,* $b_T^{\mathrm{env}} = \max_t U_t(K + \underline{p}_t^{-1})$*, and* $w_T(\delta) = T^{-1}\{\sqrt{2 v_T^{\mathrm{env}} \log(2/\delta)} + (2/3) b_T^{\mathrm{env}} \log(2/\delta)\}$*. Then* $\mathbb{P}\{|\widehat{\tau}_T - \tau_T| \leq w_T(\delta)\} \geq 1 - \delta$*.*

This interval charges every rare assignment through its propensity. It therefore detects the ε-propensity failure in Proposition 2.1(i), even though expected counts do not.

Proof of Theorem 4.4. Conditioning on $\mathcal{F}_{t-1}$ and using predictability,

$$\mathbb{E}\left[\sum_a c_a^{\mathrm{arm}} m_t(a) + \frac{\psi_t(I_t)}{\pi_t(I_t)} \middle| \mathcal{F}_{t-1}\right] = \sum_a c_a^{\mathrm{arm}} m_t(a) + \sum_a \pi_t(a) \frac{c_a^{\mathrm{arm}}\{Y_t(a) - m_t(a)\}}{\pi_t(a)}$$
$$= \sum_a c_a^{\mathrm{arm}} Y_t(a).$$

Summation proves exact unbiasedness. Moreover, $T(\widehat{\tau}_T - \tau_T) = \sum_t \xi_t$, where $(\xi_t, \mathcal{F}_t)$ is a martingale difference sequence. The assumed second-score integrability makes every increment square-integrable. Direct expansion gives

$$\mathbb{E}(\xi_t^2 \mid \mathcal{F}_{t-1}) = \sum_a \frac{\psi_t(a)^2}{\pi_t(a)} - \left(\sum_a \psi_t(a)\right)^2.$$

Orthogonality of martingale increments and an outer expectation yield Equation (4.9). Dropping the nonnegative squared term and using $\psi_t(a)^2/\pi_t(a) \leq U_t^2/p_t^{\min}$ gives

$$\mathrm{Var}(\widehat{\tau}_T) \leq \frac{K}{T^2} \mathbb{E} \sum_t \frac{U_t^2}{p_t^{\min}}.$$

The expectation is necessary when the predictable envelope is random. □

Proof of Proposition F.1. As above, $T(\widehat{\tau}_T - \tau_T) = \sum_t \xi_t$ is a martingale. Its conditional variance satisfies

$$\mathbb{E}(\xi_t^2 \mid \mathcal{F}_{t-1}) \leq \sum_a \frac{\psi_t(a)^2}{\pi_t(a)} \leq \frac{K U_t^2}{\underline{p}_t},$$

and its increment obeys

$$|\xi_t| \leq \frac{|\psi_t(I_t)|}{\pi_t(I_t)} + \sum_a |\psi_t(a)| \leq U_t(\underline{p}_t^{-1} + K) \leq b_T^{\mathrm{env}}.$$

If $b_T^{\mathrm{env}} = 0$, the result is immediate. Otherwise apply Lemma C.6 to $\xi_t / b_T^{\mathrm{env}}$ with deterministic variance envelope $v_T^{\mathrm{env}}/(b_T^{\mathrm{env}})^2$ and $x = \log(2/\delta)$. Multiplication by b_T^{env}/T gives the stated interval. □

For studentization, let $\xi_t = \psi_t(I_t)/\pi_t(I_t) - \sum_a \psi_t(a)$, $s_T^2 = \sum_t \mathbb{E}(\xi_t^2 \mid \mathcal{F}_{t-1})$, $Q_T^{\text{score}} = \sum_{t,a} \psi_t(a)^2/\pi_t(a)$, and $\widehat{Q}_T^{\text{score}} = \sum_t \psi_t(I_t)^2/\pi_t(I_t)^2$. The remaining primitive conditions control residual imbalance and the fourth-order score; see Condition F.2. Every ratio whose numerator and denominator are both zero uses the zero convention.

Condition F.2 (Primitive studentization)**.** *For the triangular array in Theorem 4.4, define*

$$\xi_t = \psi_t(I_t)/\pi_t(I_t) - \sum_a \psi_t(a), \qquad s_T^2 = \sum_t \mathbb{E}(\xi_t^2 \mid \mathcal{F}_{t-1}),$$
$$Q_T^{\text{score}} = \sum_{t,a} \psi_t(a)^2/\pi_t(a), \qquad \widehat{Q}_T^{\text{score}} = \sum_t \psi_t(I_t)^2/\pi_t(I_t)^2,$$
$$H_T = \sum_t (\sum_a \psi_t(a))^2, \qquad J_T = \sum_{t,a} \psi_t(a)^4/\pi_t(a)^3.$$

There is a deterministic $\sigma_T^2 \to \infty$ such that (i) $s_T^2/\sigma_T^2 \to_{\mathbb{P}} 1$; (ii), for every $\epsilon > 0$,

$$\sigma_T^{-2} \sum_t \mathbb{E}[\xi_t^2 \mathbf{1}\{|\xi_t| > \epsilon\sigma_T\} \mid \mathcal{F}_{t-1}] \to_{\mathbb{P}} 0;$$

(iii) $H_T/\sigma_T^2 \to_{\mathbb{P}} 0$; and (iv) $J_T/\sigma_T^4 \to_{\mathbb{P}} 0$.

Corollary F.1 (Primitive studentized limit)**.** *Under Theorem 4.4, assume the variance stabilization, conditional Lindeberg, vanishing residual-imbalance, and fourth-score conditions in Condition F.2. Define the following pivot as zero on $\{\widehat{Q}_T^{\text{score}} = 0\}$. Then $T(\widehat{\tau}_T - \tau_T)/\sqrt{\widehat{Q}_T^{\text{score}}} \rightsquigarrow \mathrm{N}(0,1)$.*

Lemma F.1 (Conservative-score decomposition)**.** *With the notation in Condition F.2,*

$$Q_T^{\text{score}} - s_T^2 = H_T \geq 0, \qquad \mathbb{E}(\widehat{Q}_T^{\text{score}} - Q_T^{\text{score}}) = 0,$$

and $W_T = \widehat{Q}_T^{\text{score}} - Q_T^{\text{score}}$ is a locally square-integrable martingale with predictable quadratic variation at most J_T.

Proof. The conditional-variance calculation above gives the first identity. For

$$W_t = \frac{\psi_t(I_t)^2}{\pi_t(I_t)^2} - \sum_a \frac{\psi_t(a)^2}{\pi_t(a)},$$

we have $\mathbb{E}(W_t \mid \mathcal{F}_{t-1}) = 0$ and

$$\mathbb{E}(W_t^2 \mid \mathcal{F}_{t-1}) \leq \mathbb{E}\left\{ \frac{\psi_t(I_t)^4}{\pi_t(I_t)^4} \middle| \mathcal{F}_{t-1} \right\} = \sum_a \frac{\psi_t(a)^4}{\pi_t(a)^3}.$$

The conditional second moment is finite almost surely because, given $\mathcal{F}_{t-1}$, the assignment has finite support and every loaded propensity is positive. Thus the partial sums are locally square-integrable. Summing the last display proves $\langle W \rangle_T \leq J_T$ and completes the claim. □

Proof of Corollary F.1. Conditions (i)–(ii) are the variance-stabilization and conditional Lindeberg conditions for the martingale CLT (Hall & Heyde, 1980, Chapter 3); hence $\sum_t \xi_t/\sigma_T \rightsquigarrow \mathrm{N}(0,1)$. By Lemma F.1 and condition (iii), $Q_T^{\text{score}}/\sigma_T^2 = s_T^2/\sigma_T^2 + H_T/\sigma_T^2 \to_{\mathbb{P}} 1$.

Localize W at the first time its predictable quadratic variation exceeds an integer L. Lenglart's inequality applies to each stopped square-integrable martingale; letting $L \to \infty$ and using $\langle W \rangle_T \leq J_T$ gives, for any $\epsilon, \eta > 0$,

$$\mathbb{P}\{|\widehat{Q}_T^{\text{score}} - Q_T^{\text{score}}| > \epsilon\sigma_T^2\} \leq \frac{\eta}{\epsilon^2} + \mathbb{P}\{J_T > \eta\sigma_T^4\}.$$

Condition (iv), followed by $\eta \downarrow 0$, shows $(\widehat{Q}_T^{\text{score}} - Q_T^{\text{score}})/\sigma_T^2 \to_{\mathbb{P}} 0$. Thus $\widehat{Q}_T^{\text{score}}/\sigma_T^2 \to_{\mathbb{P}} 1$; in particular, its denominator is positive with probability tending to one. Slutsky's theorem then proves Corollary F.1. □

Corollary F.2 (Decaying randomized evaluation in Example 2.1)**.** *Fix the potential outcomes at* $Y_t(A) = 0.3$ *and* $Y_t(B) = Y_t(C) = 0.2$ *for every* t*. Let* $m_t = 0$*, let the base select* A*, let evaluation split equally between* B, C*, and let* $\rho_t = \rho_0 t^{-\alpha}$ *for* $0 < \rho_0, \alpha < 1$*. For target* $B - A$*, the exact moments are in Equation (F.1), and* $\mathrm{Var}(\widehat{\tau}_T) \sim 0.08T^{-(1-\alpha)}/\{\rho_0(1+\alpha)\}$*, while* $\mathbb{E}R_T \sim 0.1\rho_0 T^{1-\alpha}/(1-\alpha)$*. The limit in Corollary F.1 also holds. Hence* $\alpha = 1/3$ *yields regret* $T^{2/3}$ *and standard error* $T^{-1/3}$*.*

Here α is a managerial rate choice: faster evaluation decay reduces near-term welfare loss and slows effect learning by the matched exponent. The model-based Wald route is kept separate in Theorem H.1.

Proof of Corollary F.2. For contrast $B - A$, the three residual scores are $(-0.3, 0.2, 0)$ and their sum is -0.1. The implemented propensities are $\pi_t(A) = 1 - \rho_t$ and $\pi_t(B) = \pi_t(C) = \rho_t/2$. Substitution in Equation (4.9), and the fact that an evaluation round always selects a suboptimal arm, yield

$$\mathrm{Var}(\widehat{\tau}_T) = T^{-2} \sum_{t \le T} \{0.09/(1-\rho_t) + 0.08/\rho_t - 0.01\}, \qquad \mathbb{E}R_T = 0.1 \sum_{t \le T} \rho_t. \tag{F.1}$$

Since $\sum_{t \le T} t^\beta \sim T^{1+\beta}/(1+\beta)$ for $\beta > -1$,

$$s_T^2 \sim \frac{0.08}{\rho_0(1+\alpha)} T^{1+\alpha}, \qquad \mathrm{Var}(\widehat{\tau}_T) = T^{-2} s_T^2,$$

which gives both asymptotic rates stated in the corollary.

For studentization take the deterministic scale $\sigma_T^2 = s_T^2$. Direct calculation gives

$$H_T = 0.01T, \qquad J_T = \sum_{t \le T} \left\{ \frac{0.3^4}{(1-\rho_t)^3} + \frac{0.2^4}{(\rho_t/2)^3} \right\} = O(T^{1+3\alpha}).$$

Hence $H_T/s_T^2 = O(T^{-\alpha})$ and $J_T/(s_T^2)^2 = O(T^{\alpha-1})$. Moreover $\max_t |\xi_t|/s_T = O(T^{(\alpha-1)/2})$, so the conditional Lindeberg condition holds. Corollary F.1 proves the studentized limit. □

G Welfare–Information Lower Bounds

The welfare-constrained directional oracle used in Theorem 4.5 is

$$\begin{aligned} &\mathcal{D}_T(\mathfrak{B}) = \{n \ge 0 : \mathbf{1}^\top n = T,\ \Delta^\top n \le \mathfrak{B},\ M(n) \succ 0\}, \\ &M(n) = \sigma^{-2} \sum_a n_a x_a x_a^\top, \qquad \Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}}) = \inf_{n \in \mathcal{D}_T(\mathfrak{B})} \max_j (c_j^{\mathrm{par}})^\top M(n)^{-1} c_j^{\mathrm{par}}. \end{aligned} \tag{G.1}$$

Proof of Theorem 4.5. Let $H_t = (I_1, Y_1, \ldots, I_t, Y_t)$, and let P_i be the law of H_T under θ_i, and write $n_a = \mathbb{E}_{\theta_0} N_a(T)$. A possibly randomized adaptive policy is represented by kernels $q_t(\cdot \mid H_{t-1})$. These kernels are parameter-free: conditional on the same realized history, the policy uses the same action law under P_0 and P_1. Given H_{t-1} and $I_t = a$, the reward laws are $N(x_a^\top \theta_0, \sigma^2)$ and $N(x_a^\top \theta_1, \sigma^2)$. Consequently, the chain rule for relative entropy, followed by conditioning on I_t, yields, for $h = \theta_1 - \theta_0$,

$$\begin{aligned} \mathrm{KL}(P_0, P_1) &= \sum_{t=1}^T \mathbb{E}_{\theta_0} \mathrm{KL}(P_0((I_t, Y_t) \mid H_{t-1}), P_1((I_t, Y_t) \mid H_{t-1})) \\ &= \sum_{t=1}^T \mathbb{E}_{\theta_0} \sum_a q_t(a \mid H_{t-1}) \mathrm{KL}\big(N(x_a^\top \theta_0, \sigma^2), N(x_a^\top \theta_1, \sigma^2)\big) \\ &= \frac{1}{2\sigma^2} \sum_a n_a (x_a^\top h)^2 = \frac{1}{2} h^\top M(n) h. \end{aligned} \tag{G.2}$$

The action kernels contribute zero conditional KL; all information in Equation (G.2) therefore comes from the rewards actually observed under the policy.

Choose $j^* \in \arg\max_j (c_j^{\mathrm{par}})^\top M(n)^{-1} c_j^{\mathrm{par}}$, let $v = (c_{j^*}^{\mathrm{par}})^\top M(n)^{-1} c_{j^*}^{\mathrm{par}}$, and take

$$h = \frac{M(n)^{-1} c_{j^*}^{\mathrm{par}}}{\sqrt{v}}, \qquad \theta_1 = \theta_0 + h. \tag{G.3}$$

The theorem assumes $\theta_0, \theta_1 \in \Theta$. Since $M(n) \succ 0$ and $v > 0$, the choice in Equation (G.3) is well defined and $h^\top M(n) h = 1$. Hence Equation (G.2) equals $1/2$, whereas the two values $\mu_i = (c_{j^*}^{\mathrm{par}})^\top \theta_i$ of the target satisfy $\mu_1 - \mu_0 = (c_{j^*}^{\mathrm{par}})^\top h = \sqrt{v}$.

Fix any estimator $\widehat{\tau}_{j^*}$, and form the midpoint test $\phi = \mathbf{1}\{\widehat{\tau}_{j^*} \geq (\mu_0 + \mu_1)/2\}$. The two error events obey

$$\{\phi = 1\} \subseteq \{|\widehat{\tau}_{j^*} - \mu_0| \geq \sqrt{v}/2\} \quad (P_0), \qquad \{\phi = 0\} \subseteq \{|\widehat{\tau}_{j^*} - \mu_1| \geq \sqrt{v}/2\} \quad (P_1).$$

Therefore

$$\mathbb{E}_{\theta_0}(\widehat{\tau}_{j^*} - \mu_0)^2 \geq \frac{v}{4} P_0(\phi = 1), \qquad \mathbb{E}_{\theta_1}(\widehat{\tau}_{j^*} - \mu_1)^2 \geq \frac{v}{4} P_1(\phi = 0).$$

Every test satisfies $P_0(\phi = 1) + P_1(\phi = 0) \geq 1 - \mathrm{TV}(P_0, P_1)$. Taking the larger of the two risks gives

$$\max_{i \in \{0,1\}} \mathbb{E}_{\theta_i}(\widehat{\tau}_{j^*} - (c_{j^*}^{\mathrm{par}})^\top \theta_i)^2 \geq \frac{v}{8}\{1 - \mathrm{TV}(P_0, P_1)\}.$$

Pinsker's inequality and $\mathrm{KL}(P_0, P_1) = 1/2$ imply $\mathrm{TV}(P_0, P_1) \leq 1/2$, giving $v/16$.

It remains to connect this policy-dependent v to the oracle. Since one arm is selected at every round, $n \geq 0$ and $\mathbf{1}^\top n = T$. Moreover, linearity of expectation and the definition of the gaps at θ_0 give $\Delta^\top n = \mathbb{E}_{\theta_0} R_T \leq \mathfrak{B}$. Together with the assumed positive definiteness of $M(n)$, these facts show $n \in \mathcal{D}_T(\mathfrak{B})$. Hence

$$v = \max_j (c_j^{\mathrm{par}})^\top M(n)^{-1} c_j^{\mathrm{par}} \geq \Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}}),$$

which proves Equation (4.10).

We finally prove the attainability claim. Consider the T-indexed sequence in the theorem, and abbreviate $M_T^\star = M(n_T^\star)$, $\widetilde{M}_T = M(\widetilde{n}_T)$, and

$$E_T = (M_T^\star)^{-1/2}(\widetilde{M}_T - M_T^\star)(M_T^\star)^{-1/2}, \qquad \epsilon_T = \|E_T\|_{\mathrm{op}} = o(1).$$

For all sufficiently large T, $\epsilon_T < 1$, and

$$(1 - \epsilon_T) I \preceq I + E_T \preceq (1 + \epsilon_T) I.$$

Thus $\widetilde{M}_T \succ 0$, inversion reverses the Loewner order, and

$$\frac{1}{1 + \epsilon_T}(M_T^\star)^{-1} \preceq \widetilde{M}_T^{-1} \preceq \frac{1}{1 - \epsilon_T}(M_T^\star)^{-1}. \tag{G.4}$$

Under the fixed integer allocation $\widetilde{n}_T$, Gaussian OLS is unbiased and has covariance $\widetilde{M}_T^{-1}$. Applying Equation (G.4) to each c_j^{par}, then taking the maximum over j, gives

$$\frac{\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})}{1 + \epsilon_T} \leq \max_j (c_j^{\mathrm{par}})^\top \widetilde{M}_T^{-1} c_j^{\mathrm{par}} \leq \frac{\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})}{1 - \epsilon_T}.$$

Here $\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})$ is evaluated at the corresponding horizon in the T-indexed sequence. Since $\epsilon_T \to 0$, the displayed quantity is $(1 + o(1))\Phi_{\mathfrak{B}}(\mathcal{C}_{\mathrm{par}})$, as claimed. □

The lower bound quantifies adaptive necessity at the reference instance; the matching fixed-design construction identifies the corresponding information price after integer rounding.

Proof of Corollary 4.1. For the unit-variance Gaussian MAB, $M(n) = \mathrm{diag}(n_A, n_B, n_C)$ and the two contrast variances are $n_A^{-1} + n_B^{-1}$ and $n_A^{-1} + n_C^{-1}$. Put $s = n_B + n_C$, so $n_A = T - s$ and the welfare constraint is $s \leq \mathfrak{B}/g$. For fixed s,

$$\max\left\{\frac{1}{T - s} + \frac{1}{n_B}, \frac{1}{T - s} + \frac{1}{n_C}\right\} \geq \frac{1}{T - s} + \frac{2}{s},$$

with equality exactly at $n_B = n_C = s/2$. The derivative of $f(s) = (T - s)^{-1} + 2s^{-1}$ is $f'(s) = (T - s)^{-2} - 2s^{-2}$, so f decreases until $s = \sqrt{2}T/(1 + \sqrt{2})$. Under the stated tight-budget condition the optimum therefore uses $s = \mathfrak{B}/g$. Substitution gives Equation (4.11) and the numerical instance. □

The same obstruction has a purely design-based scalar form.

Proposition G.1 (Finite-population regret–variance product)**.** *In a two-arm experiment, suppose arm 1 is better than arm 2 by a constant $g > 0$ and $Y_t(2) = y \neq 0$. If $0 < p_t \leq 1$ almost surely is the predictable probability of arm 2, $S_T = \sum_t p_t$, and $0 < \mathbb{E}S_T \leq T/2$, then the centered HT estimator of arm 2 has predictable quadratic variation $V_T = y^2 T^{-2} \sum_t (1 - p_t)/p_t$ and*

$$\mathbb{E}R_T \, \mathbb{E}V_T \geq gy^2/2.$$

Proof. Cauchy–Schwarz gives $\sum_t 1/p_t \geq T^2/S_T$, and Jensen gives $\mathbb{E}(1/S_T) \geq 1/\mathbb{E}S_T$. Therefore

$$\mathbb{E}V_T \geq \frac{y^2}{T^2} \left\{ \frac{T^2}{\mathbb{E}S_T} - T \right\} \geq \frac{y^2}{2\mathbb{E}S_T}.$$

Since $\mathbb{E}R_T = g\mathbb{E}S_T$, multiplication proves the result. □

H Structured linear inference with two-stage filtration

Let $x_t = x_{t,I_t}$ be $\mathcal{G}_t$-measurable and $Y_t = x_t^\top \theta + \varepsilon_t$, where $\mathbb{E}(\varepsilon_t \mid \mathcal{G}_t) = 0$ and $\mathbb{E}(\varepsilon_t^2 \mid \mathcal{G}_t) = \sigma^2$. Assume the spanning initialization makes Γ_T positive definite almost surely for every reported horizon, and set $\Gamma_T = \sum_t x_t x_t^\top$ and $\widehat{\theta}_T = \Gamma_T^{-1} \sum_t x_t Y_t$. In the unregularized Fisher normalization of Section 2, $M_0 = 0$ and $M_T = \sigma^{-2}\Gamma_T$, so $c^\top M_T^{-1} c = \sigma^2 c^\top \Gamma_T^{-1} c$.

Theorem H.1 (Information stability implies adaptive Wald inference)**.** *Let $\Gamma_T^{\mathrm{ref}} \succ 0$ be deterministic and d fixed. Suppose*

(i) $\|(\Gamma_T^{\mathrm{ref}})^{-1/2}(\Gamma_T - \Gamma_T^{\mathrm{ref}})(\Gamma_T^{\mathrm{ref}})^{-1/2}\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$*;*

(ii) $\max_{t \leq T} \|(\Gamma_T^{\mathrm{ref}})^{-1/2} x_t\| = o_{\mathbb{P}}(1)$*;*

(iii) for some $\delta_{\mathrm{mom}} > 0$ and $C_{\mathrm{mom}} < \infty$, $\sup_{t \leq T} \mathbb{E}(|\varepsilon_t/\sigma|^{2+\delta_{\mathrm{mom}}} \mid \mathcal{G}_t) \leq C_{\mathrm{mom}}$ almost surely.

Then for every deterministic nonzero contrast sequence c_T,

$$\frac{c_T^\top(\widehat{\theta}_T - \theta)}{\sigma\sqrt{c_T^\top \Gamma_T^{-1} c_T}} \rightsquigarrow \mathrm{N}(0,1).$$

On any event on which the operator norm in (i) is at most $\eta < 1$, the realized self-normalizing proxy $\sigma^2 c_T^\top \Gamma_T^{-1} c_T$ lies between $(1+\eta)^{-1}$ and $(1-\eta)^{-1}$ times $\sigma^2 c_T^\top (\Gamma_T^{\mathrm{ref}})^{-1} c_T$.

Proof. Let

$$S_T = \sum_t x_t \varepsilon_t, \quad \mathsf{H}_T = (\Gamma_T^{\mathrm{ref}})^{-1/2} \Gamma_T (\Gamma_T^{\mathrm{ref}})^{-1/2}, \quad z_T = (\Gamma_T^{\mathrm{ref}})^{-1/2} S_T/\sigma,$$

and $u_T = (\Gamma_T^{\mathrm{ref}})^{-1/2} c_T$. Define the interlaced filtration $\mathcal{H}_{2t-2} = \mathcal{F}_{t-1}$, $\mathcal{H}_{2t-1} = \mathcal{G}_t$, and $\mathcal{H}_{2t} = \mathcal{F}_t$, with a zero increment at assignment revelation and increment $x_t\varepsilon_t$ at outcome revelation. Because x_t is known in $\mathcal{G}_t$ before ε_t is revealed, the nonzero summands in

$$Z_T^{\mathrm{ref}} = \frac{u_T^\top z_T}{\|u_T\|} = \sum_t \frac{u_T^\top (\Gamma_T^{\mathrm{ref}})^{-1/2} x_t}{\|u_T\|} \frac{\varepsilon_t}{\sigma}$$

are martingale differences for $(\mathcal{H}_s)$. Their predictable quadratic variation is $u_T^\top \mathsf{H}_T u_T/\|u_T\|^2 = 1 + o_{\mathbb{P}}(1)$. Put $v_T = u_T/\|u_T\|$. For $a_{t,T} = v_T^\top (\Gamma_T^{\mathrm{ref}})^{-1/2} x_t$ and every $\varepsilon > 0$, conditional Markov and (iii) bound the Lindeberg sum by

$$C_{\mathrm{mom}} \varepsilon^{-\delta_{\mathrm{mom}}} \max_{t \leq T} |a_{t,T}|^{\delta_{\mathrm{mom}}} \sum_{t \leq T} a_{t,T}^2 = o_{\mathbb{P}}(1),$$

because (ii) controls the maximum and (i) makes the final sum $v_T^\top \mathsf{H}_T v_T = 1 + o_{\mathbb{P}}(1)$. Hence the martingale CLT yields $Z_T^{\mathrm{ref}} \rightsquigarrow \mathrm{N}(0,1)$.

Assumption (i) gives $\|\mathsf{H}_T - I\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$ and, by Lemma C.7, $\|\mathsf{H}_T^{-1} - I\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$. We next justify the vector tightness used below. For each canonical vector e_k, the martingale $e_k^\top z_T$ has predictable quadratic variation $e_k^\top \mathsf{H}_T e_k = 1 + o_{\mathbb{P}}(1)$. A localized Lenglart inequality gives, for every $A > 0$,

$$\mathbb{P}(|e_k^\top z_T| > A) \le \mathbb{P}(e_k^\top \mathsf{H}_T e_k > 2) + \frac{2}{A^2} = o(1) + \frac{2}{A^2}.$$

Hence each coordinate is tight, and fixed d implies $\|z_T\| = O_{\mathbb{P}}(1)$.

With $v_T = u_T/\|u_T\|$, the actual pivot equals

$$\frac{v_T^\top \mathsf{H}_T^{-1} z_T}{\sqrt{v_T^\top \mathsf{H}_T^{-1} v_T}}.$$

On $\{\|\mathsf{H}_T - I\|_{\mathrm{op}} \le 1/2\}$, the denominator is bounded away from zero and elementary norm inequalities give

$$\left| \frac{v_T^\top \mathsf{H}_T^{-1} z_T}{\sqrt{v_T^\top \mathsf{H}_T^{-1} v_T}} - v_T^\top z_T \right| \le C \|\mathsf{H}_T^{-1} - I\|_{\mathrm{op}} \|z_T\| = o_{\mathbb{P}}(1)$$

for a universal constant C. The event has probability tending to one, so the pivot is $Z_T^{\mathrm{ref}} + o_{\mathbb{P}}(1)$. This proves the limit by Slutsky. The proxy comparison follows from Lemma C.4. □

If σ is unknown, the same conclusion holds after replacing it by a predictable or cross-fitted estimator $\widehat{\sigma}_T$ that is positive with probability tending to one and satisfies $\widehat{\sigma}_T/\sigma \to_{\mathbb{P}} 1$; information stability alone does not imply that scale consistency.

I Experimental and computational details

Table 4 links each theorem to its empirical diagnostic. The main text presents the synthetic and real-data results; this appendix gives specifications, repeated-sampling diagnostics, sensitivity grids, and additional figures.

I.1 Theory-to-evidence map

Table 4: Evidence is organized by theoretical claim. "Diagnostic criterion" is the prespecified pattern used to assess the implication; complete estimands and assignment laws appear in the cited subsection.

Claim	Diagnostic criterion	Evidence and metric
Proposition 2.1 and Theorem 3.1	Pull counts or global $\lambda_{\min}$ correctly rank every target design	Target-max versus exact A/G/uniform designs; worst-target s.e. and global spectrum (Section I.5)
Theorem 4.1	Pilot radius or the induced design bound fails to contract, or the stated finite-m radius is below one but its sandwich fails	Pilot-size slope, design distance, theorem-level radius, and separate ex-post realized-matrix diagnostics (Section I.8)
Theorem 4.2	The wrapped process cannot be decomposed into the decision and evaluation clocks	Two-clock arm totals and the one-clock counterfactual in Figure 2; unwrapped base-policy tracking in Section I.3
Theorem 4.3	A realized acquisition matrix exceeds the radius built from its declared drift and leverage	Target-normalized operator error, ex-post radius, and directional quadratic forms (Section I.8)
Theorem 4.4	Monte Carlo bias statistically inconsistent with zero, or empirical variance outside the exact-bracket tolerance	Contrastwise bias, empirical/exact variance ratio, width, and marginal coverage (Section I.4)
Proposition F.1 and Corollary F.1	Rare propensities break finite-horizon coverage or studentized pivots depart systematically from normality	Propensity-floor grid, interval coverage, horizon paths, and Q–Q diagnostics (Section I.4)
Corollary F.2	Log–log regret and s.e. slopes disagree with $1-\alpha$ and $-(1-\alpha)/2$	Multi-horizon Monte Carlo rate study (Section I.4)
Theorem 4.5 and Corollary 4.1	A feasible rounded reference design improves on the oracle value, or the finite-T rounding error exceeds its analytic bound	Exact three-arm frontier and integer-rounding verification (Section G)
Theorem H.1	Requested adaptive OLS pivots depart systematically from the Gaussian benchmark	Directionwise Q–Q and coverage with the unregularized terminal Gram matrix (Section I.5)

The main composite Figure 2 is a descriptive map across four settings: panel (a) visualizes clock separation, panel (b) is a finite-population regret–width comparison, panel (c) uses the smooth evaluation-only linear surrogate under a common schedule, and panel (d) is the sample-splitting analogue of pilot-to-deployment transfer. The logged-validity checks are reported in Figures 5 and 6, and the finite-m pilot and matrix certificates are evaluated in Section I.8.

Table 5: Evaluation-matched comparisons: worst-target s.e. (MAB/linear) or 95% width (fixed population), with targeted-design marginal coverage.

Setting	Repetitions	Diffuse regret	Targeted regret	Diffuse precision	Targeted precision / validity
Gaussian MAB	500	171.0	157.2	0.0607	0.0531 / 94.4%
Fixed population	500	44.3	51.0	0.0246	0.0223 / 94.0%
Linear rewards	500	428.2	383.8	0.0518	0.0320 / 94.8%

I.2 Computing environment and file contract

The code is written for Python 3.11 or later and uses NumPy, pandas, SciPy, and Matplotlib. Compatible package lower bounds are listed in `requirements.txt`, while `ENVIRONMENT.txt` records the current validation environment. Each subsection below names its raw and summary files; the plotting routines export the corresponding figures in PDF and PNG. The script `code/make_results_macros.py` imports selected headline entries directly into `results/results_macros.tex`; `code/verify_outputs.py` and `code/verify_closest_and_plugin.py` check the remaining reported comparisons.

The complete pipeline is

```
python -m pip install -r requirements.txt
python code/run_all.py
python code/verify_outputs.py
latexmk -pdf -interaction=nonstopmode -halt-on-error main.tex
```

Every experiment records an explicit deterministic seed. The MAB, design-based, linear, finite-grid, pilot, envelope, and rate studies use NumPy `SeedSequence` streams at the configuration or repetition level; the Hillstrom bootstrap and sensitivity study use the explicit generators reported in their subsections.

I.3 Gaussian MAB

The file `code/simulate_mab.py` uses

$$K = 4, \qquad T = 3200, \qquad \mu = (0.30, 0.255, 0.14, 0.02), \qquad \sigma = 1,$$

with 500 repetitions and master seed 2407. Every arm is pulled once on the decision clock. Thereafter the base index is

$$\widehat{\mu}_a(t-1) + \frac{\sigma\sqrt{2\log T}}{\sqrt{N_a(t-1)}}.$$

For a wrapped policy, the base counts and empirical means in this index are updated only when $E_t = 0$, while evaluation rewards enter the terminal contrast estimator. The explicitly named "naive one-clock targeted" ablation updates the base state after every reward, so evaluation outcomes also advance its internal state.

The unwrapped target count at each checkpoint is computed by root finding for $\lambda_t > \max_a \mu_a$ in

$$\sum_{a=1}^{K} \frac{\sigma^2(2\log T)}{(\lambda_t - \mu_a)^2} = t.$$

For Figure 2, however, the deterministic wrapped target is the literal two-clock quantity

$$N_{a,t}^{\mathrm{ref}} = n_a^*(d_t^{\mathrm{ref}}) + \bar{N}_{a,t}^{\mathrm{eval}}, \qquad d_t^{\mathrm{ref}} = t - \sum_{s=K+1}^{t} \rho_s, \qquad \bar{N}_{a,t}^{\mathrm{eval}} = \nu_a \sum_{s=K+1}^{t} \rho_s.$$

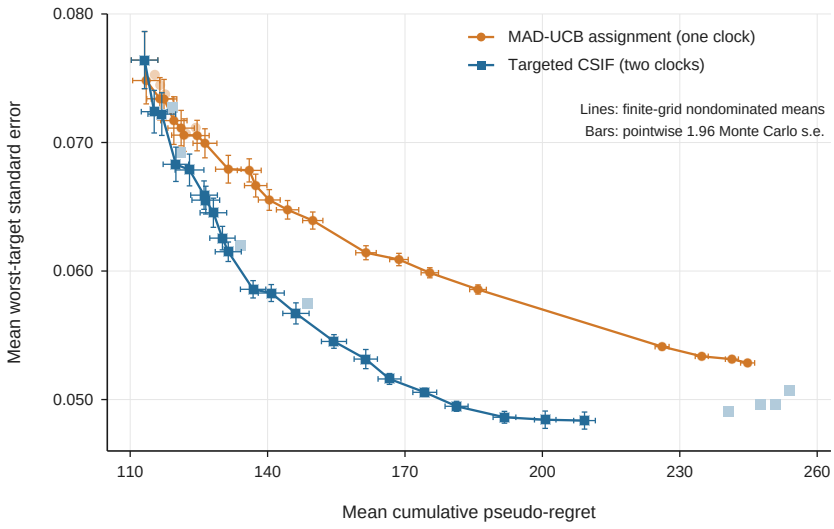


Figure 3: Targeted information shifts the prespecified finite-grid frontier toward lower worst-target error. Lines join method-specific nondominated sample means; bars are pointwise 1.96 Monte Carlo s.e. Every simulated schedule, including dominated ones, is retained in `results/final_frontier_all_grid.csv`.

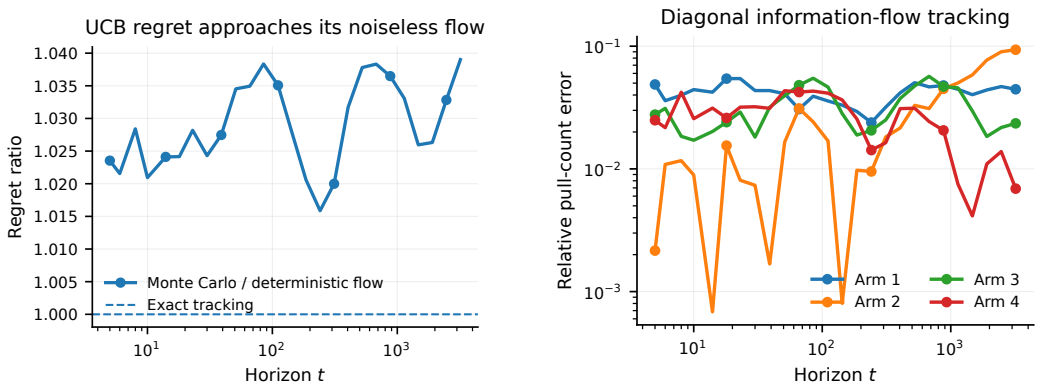
Thus the first K warm-start rounds have $\rho_s = 0$ in the clock calculation. The file `results/mab_csif_target.csv` saves both clocks, every armwise component of $N_{a,t}^{\mathrm{ref}}$, the observed Monte Carlo means, and their relative errors. The inferential contrasts are $(-1, 1, 0, 0)$ and $(-1, 0, 1, 0)$, matching the paper-wide $B - A$ and $C - A$ convention. The optimized evaluation distribution is, up to numerical tolerance,

$$\nu = (0, 0.5, 0.5, 0),$$

which makes the design ablation transparent. It is optimized once at the reference mixture intensity 0.10 and then used with the decaying implementation schedule $\rho_t = \min\{0.55, 1.60t^{-0.32}\}$. The diffuse comparator uses exactly the same schedule and a uniform evaluation distribution. Contrast standard errors use the known σ and all realized arm counts. The primary precision metric takes the larger of the two target standard errors within each repetition; their mean is retained as a supplementary column.

The finite-grid comparison independently prespecifies the 30 positive schedules $\rho_t = \min\{0.55, ct^{-\alpha}\}$ over $c \in \{0.2, 0.4, 0.8, 1.6, 3.2, 6.4\}$ and $\alpha \in \{0.16, 0.24, 0.32, 0.40, 0.48\}$, together with the zero-evaluation reference. These schedules start at $t = 1$, whereas the headline implementation starts after the four warm-start pulls, so the headline schedule is not a literal grid cell. Capped aliases reduce the 31 specified configurations to 29 unique schedules per method and 29,000 repetition-level outcomes. At each prespecified regret budget $120, 140, \ldots, 300$, the selected targeted schedule has lower mean worst-target s.e. than the selected MAD assignment; Figure 3 reports all method-specific nondominated means and pointwise Monte Carlo intervals. This finite-grid assignment-law comparison is empirical; the analytic lower bound and integer-rounding guarantee are verified separately in Section G.

The companion curves in Figure 4 are unwrapped-UCB base-tracking diagnostics: repetition paths from `results/mab_paths.csv` are compared with the common-index reference in `results/mab_target.csv`. The two-clock totals are stored separately in `results/mab_csif_target.csv` and summarized in Figure 2a; the one-clock ablation also advances the base state with evaluation outcomes.



(a) Regret relative to the unwrapped UCB reference. (b) Armwise count error relative to unwrapped UCB.

Figure 4: Unwrapped-UCB and base-policy tracking diagnostics. Curves combine every checkpoint and all 500 repetitions in `results/mab_paths.csv` with the deterministic reference in `results/mab_target.csv`; the two-clock quantities used for Theorem 4.2 are reported separately in `results/mab_csif_target.csv` and Figure 2a.

I.4 Fixed finite population

The file `code/simulate_design_based.py` creates one population of $T = 1500$ units using seed 1931. Let $t = 1, \dots, T$, $X_{1t} \sim \mathrm{Unif}[-1, 1]$, $X_{2t} \sim \mathrm{N}(0, 1)$, and $P_t = \sin(2\pi t/173)$. Define

$$\begin{aligned} S_t &= 0.36 + 0.09X_{1t} - 0.05X_{2t} + 0.04P_t, \\ R_t &= 0.05\sin(2\pi t/31) + 0.025\,\mathrm{sign}\{\cos(2\pi t/47)\}, \\ Y_t(1) &= S_t + R_t, \\ Y_t(2) &= S_t + 0.095 + 0.045X_{1t} - 0.012P_t + 0.75R_t, \\ Y_t(3) &= S_t + 0.050 - 0.035X_{1t} + 0.020P_t + 1.25R_t. \end{aligned}$$

All potential outcomes are clipped to $[0.02, 0.98]$ and then held fixed. Let I_t^{base} denote the arm proposed by the base policy, which uses empirical means and bonus $0.35\sqrt{\log(T+1)/N_a}$. Its counts, means, and internal state are updated only on $E_t = 0$ rounds for both CSIF methods. Separately, predictable arm-specific ridge regressions use features $(1, X_{1t}, X_{2t}, P_t)$ and ridge 2; these outcome models use all observed assignments, including evaluation rounds. Predictions are computed before assignment, clipped to $[0, 1]$, and updated only after the selected outcome is observed. The naive one-clock ablation feeds evaluation outcomes to both states.

The plug-in evaluation design estimates each arm's residual variance from past squared prediction errors. For the two contrasts arm 2 minus arm 1 and arm 3 minus arm 1, it solves at every round the same worst-contrast problem reported in the main text,

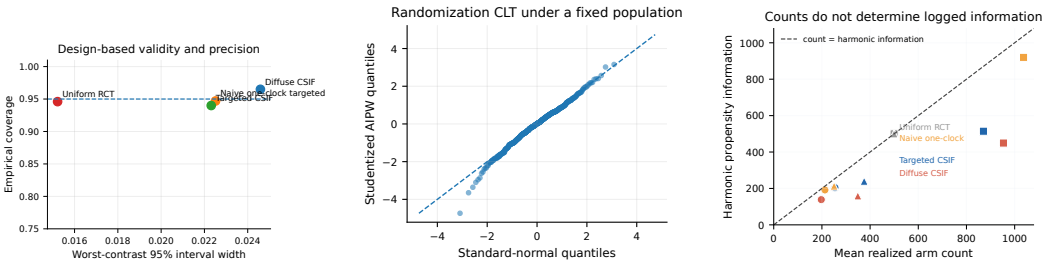
$$\min_{\nu \in \Delta_3} \max \left\{ \frac{\widehat{s}^2_{1,t}}{p_1} + \frac{\widehat{s}^2_{2,t}}{p_2}, \frac{\widehat{s}^2_{1,t}}{p_1} + \frac{\widehat{s}^2_{3,t}}{p_3} \right\}, \qquad p = (1 - \rho_t)e_{I_t^{\mathrm{base}}} + \rho_t \nu.$$

The implementation enumerates the finite collection of stationary points and active-boundary transitions, thereby solving the exact three-arm worst-contrast objective. The A-optimal sum of variances is a distinct criterion. The decaying schedule is $\rho_t = \min\{0.85, 2.4t^{-0.34}\}$. The diffuse comparator uses the identical schedule with uniform ν, and the uniform-RCT reference sets $\rho_t = 1$. We use 500 assignment repetitions and seed 7321.

The observable conservative variance proxy for contrast j is

$$(\widehat{\mathbf{Q}}_T)_{jj}/T^2 = T^{-2} \sum_t \frac{c^2_{j,I_t}\{Y_t(I_t) - m_t(I_t)\}^2}{\pi_t(I_t)^2},$$

while the reference width uses the exact predictable bracket computed from the fixed, fully known simulation array. Each reported width is the maximum of the two contrast widths, matching the



(a) Marginal coverage and width. (b) Targeted AIPW studentization. (c) Counts versus harmonic information.

Figure 5: Logged propensities, rather than pull counts alone, govern design-based inference. Panels (a)–(b) test Theorem 4.4 and Corollary F.1; panel (c) directly probes the count-versus-information failure in Proposition 2.1. The Q–Q plot uses every targeted repetition and the count plot compares each method–arm pair with the equality line.

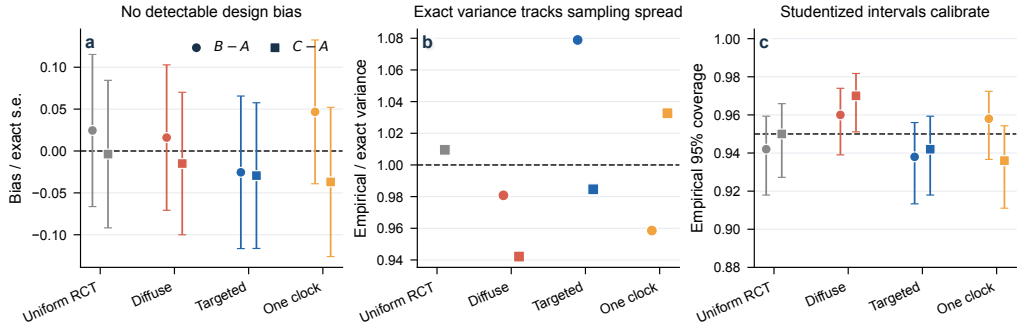


Figure 6: Exact design moments match repeated-assignment behavior. Bias bars are 95% Monte Carlo intervals after scaling by the exact standard error; variance ratios compare the repeated-assignment variance with the mean exact bracket; coverage bars are 95% Monte Carlo intervals. Both contrasts and all 500 repetitions per method are retained.

minimax target. The harmonic information reported for arm a is

$$H_a = \frac{T^2}{\sum_{t=1}^{T} 1/\pi_t(a)}.$$

The joint diagnostic in Figure 5 shows the resulting chain: targeted AIPW pivots follow Gaussian quantiles and retain near-nominal coverage, while comparable pull counts can encode different harmonic information.

The contrastwise calibration reconstructs each estimation error from the saved studentized statistic and standard error, and reconstructs the exact assignment variance from the predictable bracket. Across the eight method–contrast cells, the absolute bias is below 1.068 Monte Carlo s.e., the ratio of empirical to exact variance lies in $[0.942, 1.079]$, and marginal coverage lies in $[0.936, 0.970]$ (Figure 6). These are direct numerical checks of both identities in Theorem 4.4; their benchmark is the predictable design bracket.

We separately stress the deterministic envelope in Proposition F.1 over horizons, evaluation shares $\rho \in \{0.08, 0.20, 0.40\}$, and targeted versus diffuse evaluation. All 144,000 intervals cover; the smallest cellwise 95% Wilson lower bound is 0.9990. The price of this finite-horizon guarantee is visible in Figure 7b: the envelope half-width is 2.87–4.90 times the Gaussian reference half-width as the loaded-arm propensity becomes rare.

The rate study uses 3,000 independent repetitions at each of six horizons and four decay exponents. For $\alpha = 1/3$, the fitted regret slope is 0.6683 against $2/3$, and the empirical s.e. slope is -0.3415

Table 6: Complete configuration for the finite-envelope and decay-rate studies. Both use score vector $(-0.30, 0.20, 0)$, truth -0.10, and master seed 20260926.

Study	Horizons and repetitions	Assignment law	Confidence, fitting, and seed stream
Envelope	$T = 75, 150, 300, 600, 1200, 2400$; 4,000 repetitions per cell	$p = (1-\rho)e_A + \rho\nu$, $\rho \in \{0.08, 0.20, 0.40\}$; $\nu = (0, 1/2, 1/2)$ or $(1/3, 1/3, 1/3)$	$\delta = 0.05$; child streams from $(20260926, 1)$
Decay rate	$T = 400, 800, 1600, 3200, 6400, 12800$; 3,000 repetitions per cell	$\rho_t = 0.80t^{-\alpha}$, $\alpha \in \{0.20, 1/3, 0.50, 0.65\}$; evaluation splits equally between B, C	child streams from $(20260926, 2)$; log–log slopes fit on the tail four horizons 1600–12800

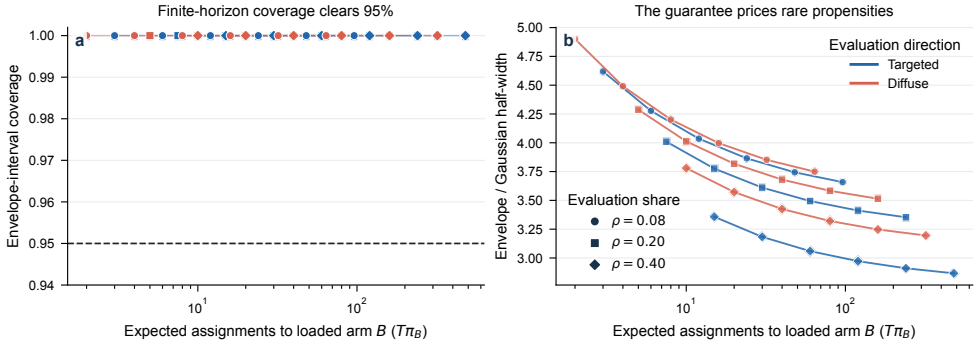


Figure 7: The finite-horizon interval preserves coverage while explicitly pricing rare assignments. Panel (a) reports every design cell against the 95% target; panel (b) compares the theorem's half-width with a Gaussian reference. Marker shape is evaluation share and color is evaluation direction.

against $-1/3$. Across all exponents, the maximum absolute slope discrepancies are below 0.0164 for regret and below 0.0260 for s.e. Thus Corollary F.2 jointly predicts both sides of the tradeoff, with every fitted-versus-theoretical comparison collected in Figure 8.

I.5 Linear structured reward

The file `code/simulate_linear.py` uses 18 actions in $\mathbb{R}^4$. Eight are the signed coordinate vectors; ten are normalized Gaussian vectors generated with seed 9103. The parameter and noise are

$$\theta = (0.62, 0.28, -0.22, 0.13), \qquad \sigma = 0.55,$$

with horizon 2400, 500 repetitions, and master seed 5319. LinUCB uses ridge 1 and exploration coefficient 1.35. For $c_1 = e_2$ and $c_2 = e_3$, the information-flow evaluation minimizes $\log \sum_{j=1}^{2} \exp[c_j^\top \{\sum_a \nu_a x_a x_a^\top + 10^{-5} I\}^{-1} c_j]$. It places approximately one quarter of its mass on each of $\pm e_2$ and $\pm e_3$ and negligible mass elsewhere. Both decaying evaluation streams use $\rho_t = \min\{0.60, 1.8t^{-0.30}\}$. This directional-geometry ablation fixes the evaluation schedule and isolates the smooth evaluation-only surrogate from the full welfare-priced online oracle in Equation (3.3). For two targets, log-sum-exp exceeds the exact maximum by at most $\log 2$; every reported precision metric uses the exact realized maximum. For both CSIF methods, the ridge-regularized base LinUCB Gram matrix, score vector, inverse, and parameter estimate advance only when $E_t = 0$. Terminal inference instead uses the unregularized all-history matrix $\Gamma_T = \sum_{t \le T} X_t X_t^\top$, checks that $\operatorname{rank}(\Gamma_T) = 4$, and reports $\widehat{\theta}_T = \Gamma_T^{-1} \sum_{t \le T} X_t Y_t$ with standard error $\sigma\sqrt{c^\top \Gamma_T^{-1} c}$. The primary precision metric is the maximum over $c \in \{e_2, e_3\}$ within each repetition; the mean is saved as a supplementary column. Sherman–Morrison updates are used only for the base LinUCB inverse. The naive one-clock targeted ablation uses every reward in the base state as well.

For each direction c_j, the saved pivot is $Z_{j,T} = c_j^\top (\widehat{\theta}_T - \theta)/\{\sigma\sqrt{c_j^\top \Gamma_T^{-1} c_j}\}$. In Figure 9, the global eigenvalue ranking reverses the requested target-precision ranking, while both terminal directional pivots from every repetition closely follow the Gaussian benchmark. The Q–Q panel is a finite-horizon consequence diagnostic for the Wald limit. The simulation records the rank and condition

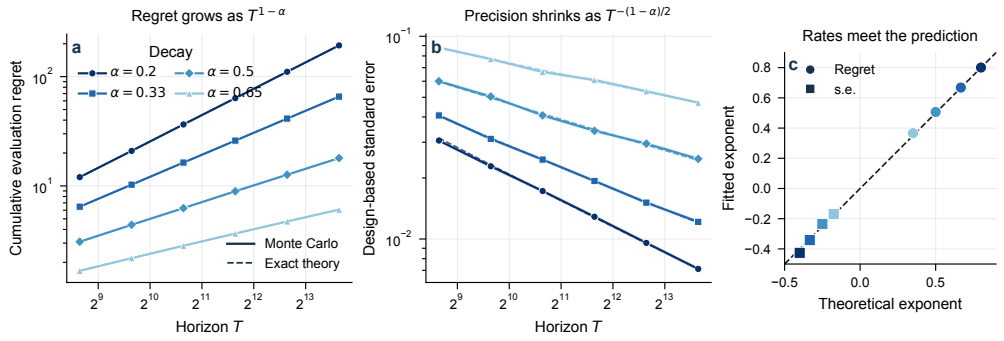
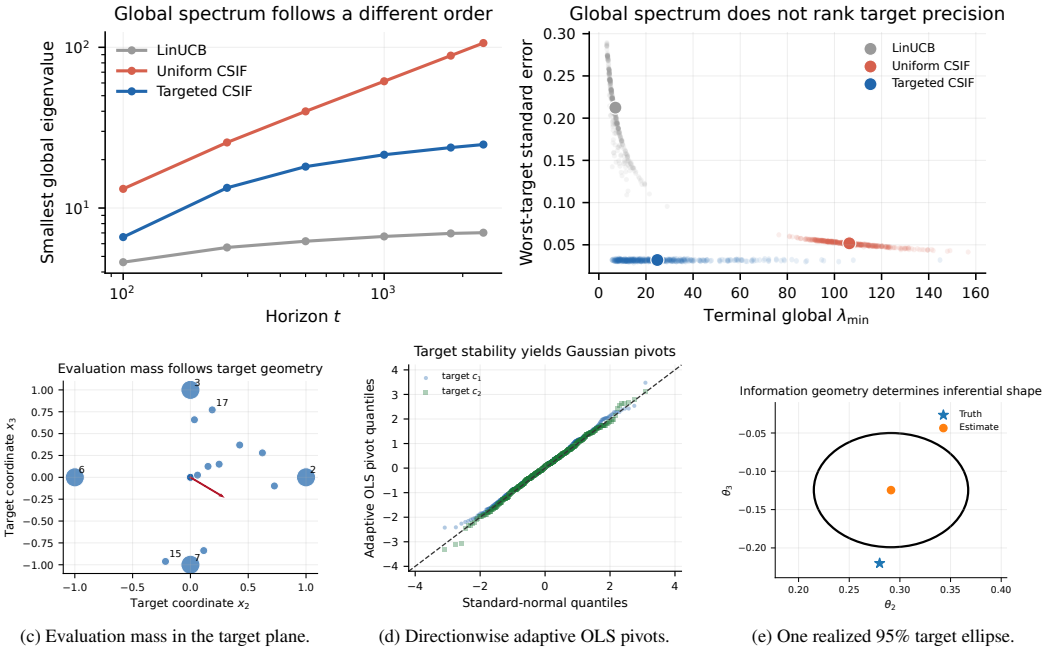


Figure 8: Multi-horizon Monte Carlo rates recover the exponents in Corollary F.2. Solid curves are Monte Carlo means or standard errors, dashed curves are exact moments, and panel (c) plots fitted against theoretical exponents for every decay choice.



(c) Evaluation mass in the target plane. (d) Directionwise adaptive OLS pivots. (e) One realized 95% target ellipse.

Figure 9: Target geometry defeats scalar spectral ranking. Panels (a)–(b) show that a larger global $\lambda_{\min}$ can accompany a larger requested standard error, the reversal in Proposition 2.1; panel (c) shows the target-directed allocation, and panel (d) diagnoses the Gaussian consequence of Theorem H.1. Panel (e) illustrates one 95% ellipse, for which an occasional miss is part of nominal calibration; aggregate targeted marginal coverage is 94.8% across all repetitions.

number of every terminal Gram matrix; the reference-normalized stability conditions in Theorem H.1 remain the analytic assumptions connecting these diagnostics to the limit theorem.

I.6 Hillstrom randomized experiment

The raw file `data/hillstrom.csv` contains 64,000 rows and has SHA-256 checksum

`27bab8c5d3669f26ec08ebb50a0a78317542f29501156f2e2af6781fab4cd7e2`.

It is the public MineThatData e-mail analytics challenge data described by Hillstrom (2008). The three randomized groups have sizes 21,306, 21,307, and 21,387. We use the binary `visit` outcome.

For a prospective experiment of size 12,000, the uniform benchmark at intensity ρ is

$$p^{\mathrm{unif}}(\rho) = (1-\rho)e_{a^*} + \rho(1/3, 1/3, 1/3).$$

Let $b(\rho) = \left\langle p^{\mathrm{unif}}(\rho), \widehat{\Delta} \right\rangle$ be its empirical per-unit regret budget. Since the empirically best arm is the men's e-mail, the cost-matched target design solves

$$\min_{p \in \Delta_3} \quad \max_{j \in \{\mathrm{mens-control}, \mathrm{womens-control}\}} \sum_{a=1}^{3} \frac{c_{ja}^2 \widehat{\sigma}_a^2}{p_a}$$
$$\text{subject to} \quad \left\langle p, \widehat{\Delta} \right\rangle \le b(\rho), \qquad p_a \ge 10^{-5}.$$

With three arms and one zero-gap arm, the regret constraint reduces the problem to a one-dimensional convex minimization, implemented by `scipy.optimize.minimize_scalar`.

Because the file has no event-time field, its randomized-arm structure supports an exchangeable repeated-split analysis. At the focal intensity $\rho = 0.11534$, we run 300 arm-stratified splits using seed 20260916. Training halves estimate the best arm, gaps, and arm variances and fit the cost-matched target allocation; disjoint held-out halves measure its plug-in worst-contrast variance and diagnostic welfare cost. This construction propagates design-estimation error into the held-out comparison. The armwise parametric bootstrap (seed 5527, 120 draws per selected intensity) refits the information-flow allocation in every draw at $\rho \in \{0.03, 0.08, 0.18, 0.40\}$, producing four point clouds of moment and design-selection uncertainty.

All 300 held-out splits select the same best arm and yield a positive worst-width gain. The mean gain is 4.31%, with split quantiles [3.41%, 5.46%]; the signed held-out welfare difference is +0.285 expected lost visits per prospective 12,000 customers. These paired results form panels (c)–(d) of Figure 10.

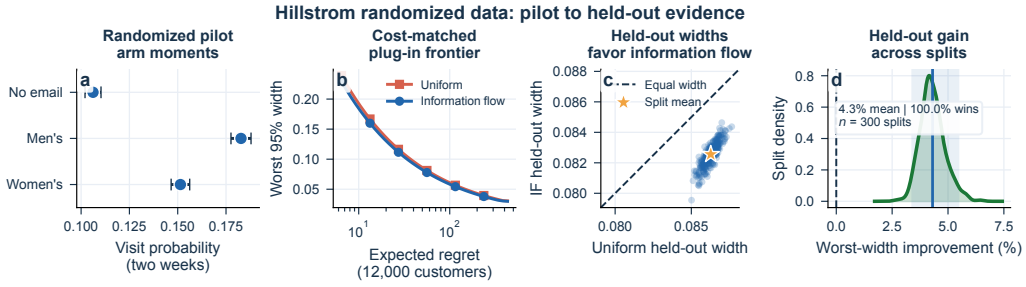


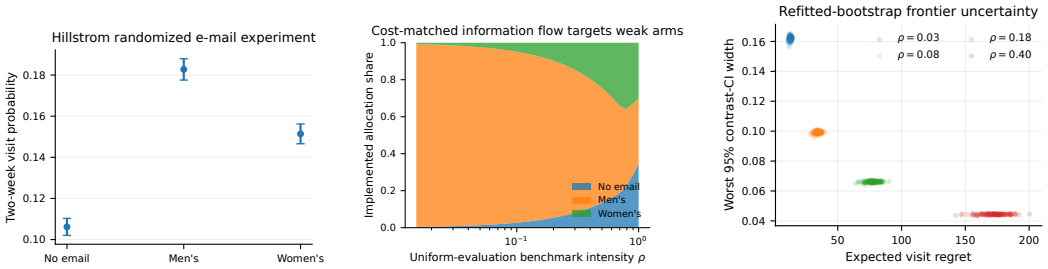
Figure 10: Hillstrom prospective and held-out evidence. Panels (a)–(b) use the full randomized sample to estimate arm moments and the prospective regret–width frontier; panels (c)–(d) use the 300 stratified train/held-out splits for paired widths and their gain distribution. Dashed lines mark equality; the blue band and line in (d) mark the central 95% and mean gain.

The additional panels in Figure 11 show how full-sample arm moments generate the allocation path and how refitting propagates moment uncertainty at four selected intensities.

I.7 Sensitivity experiments

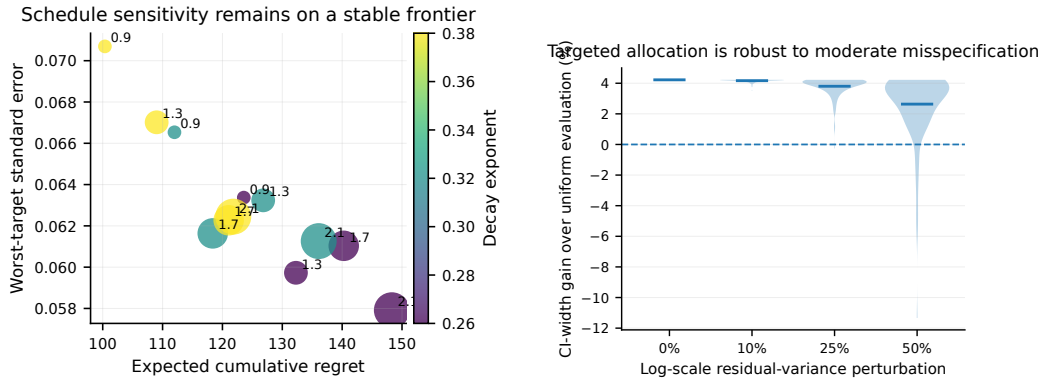
The schedule phase portrait uses horizon 2200, 35 repetitions per setting, seed 8123, and $\rho_t = \min\{0.55, ct^{-\alpha}\}$ for $c \in \{0.9, 1.3, 1.7, 2.1\}$ and $\alpha \in \{0.26, 0.32, 0.38\}$. The second study uses the full-sample Hillstrom arm moments and prospective intensity $\rho = 0.12$. With seed 9012, it forms $\widetilde{\sigma}_a^2 = \widehat{\sigma}_a^2 \exp(\epsilon_a - s^2/2)$, where $\epsilon_a \sim \mathrm{N}(0, s^2)$, for $s \in \{0, 0.10, 0.25, 0.50\}$ and 160 repetitions per level. Each fitted design is evaluated under the original full-sample empirical variances.

The schedule grid fills a stable welfare–precision tradeoff region across all prespecified tunings. Under residual-variance perturbations of 0%, 10%, 25%, 50%, mean prospective plug-in width gains are respectively 4.22%, 4.17%, 3.80%, 2.64%, and the corresponding win rates are



(a) Randomized arm moments. (b) Cost-matched allocation path. (c) Re-fitted bootstrap point clouds.

Figure 11: Hillstrom full-sample prospective design: (a) randomized arm moments, (b) the cost-matched allocation path, and (c) 120 refitted parametric-bootstrap points at each of $\rho \in \{0.03, 0.08, 0.18, 0.40\}$. Panel (c) displays point clouds rather than a continuous frontier or uncertainty ribbon.



(a) Schedule scale and decay exponent. (b) Residual-variance misspecification.

Figure 12: Prospective sensitivity across implementation and plug-in model-error grids. Bubble area in (a) is evaluation scale and color is the decay exponent; (b) uses full-sample Hillstrom moments and shows all variance-perturbation repetitions as violins with the mean overlaid. The horizontal zero line marks no gain over uniform evaluation.

$100\%, 100\%, 99.4\%, 92.5\%$. At the largest perturbation the lower 2.5% quantile is -3.80%, which locates the empirical robustness boundary. Figure 12 shows the complete schedule grid and every variance-perturbation repetition behind these summaries.

I.8 Closest baselines and independent-pilot plug-in stress test

The script `code/benchmark_closest_and_plugin.py` evaluates two additional baseline families alongside the headline methods. In the MAB, the propensity-level MAD–UCB mapping of Liang & Bojinov (2023) is $\pi_t = (1 - \delta_t)p_t^{\mathrm{UCB}} + \delta_t\omega$, with a uniform ω, one-clock UCB updates, and $\delta_t = 0.922713t^{-0.24}$. Its cumulative mixture mass is 559.304738, exactly equal to that of CSIF; the roundwise schedules differ. Scoring every method by the same terminal sample-mean metric isolates the precision effect of the assignment law; MAD's HT confidence sequence addresses a distinct inferential target. In this independent 500-repetition rerun, MAD–UCB, targeted CSIF, and uniform CSIF have (regret, worst-target s.e.) equal to $(141.20, 0.06620)$, $(161.99, 0.05295)$, and $(172.25, 0.06113)$.

For the linear experiment, A-optimality minimizes $\mathrm{tr}\{M(\nu)^{-1}\}$ and G-optimality is solved with an exact epigraph for $\max_a x_a^\top M(\nu)^{-1}x_a$. Both use the same two-clock wrapper, schedule, horizon, and OLS estimator as the smooth target surrogate. The target surrogate, A-optimal, G-optimal, and uniform evaluations have worst-target standard errors 0.03205, 0.04482, 0.04488, and 0.05197,

with regrets 382.89, 382.91, 382.66, and 429.18. SPEED (Mukherjee et al., 2024) has a different scalar target-policy-value estimand; when $\rho = 1$, there is one direction $c = \sum_a \pi(a)x_a$, and $G(a) = x_a x_a^\top/\sigma^2(a)$, our acquisition oracle with $\lambda = \tau = 0$ reduces algebraically to its PE-optimal objective. This exact containment establishes the estimand-aligned connection to SPEED, while the direct comparisons above share our pair of directional contrasts.

The pilot specialization uses the same 18 actions, independent balanced pilots $m \in \{72, 144, 288, 576\}$, parameter $\theta = 3(0.62, 0.28, -0.22, 0.13)$, noise s.d. 0.20, welfare coefficient $\kappa_{\mathrm{w}} = 0.80$, entropy coefficient $\tau = 1.50$, oracle mixture $\rho = 0.30$, look-ahead $h = 80$, deployment horizon $T = 2400$, and 300 repetitions per pilot size. The master seed is 8061; the pilot and deployment failure levels are $\delta_0 = 0.10$ and $\delta = 0.05$. Here ν^{ref} is the $\tau = 1.50$ regularized reference, distinct from the unregularized three-arm frontier. With $R_{\mathrm{eval}} = \sum_{t \le T-m} \min\{0.60, 1.8t^{-0.30}\}$, it records

$$M_T^{\mathrm{ref}}(m) = I + V_m + (T - m - R_{\mathrm{eval}})x_* x_*^\top + R_{\mathrm{eval}} \sum_a \nu^{\mathrm{ref}}(a) x_a x_a^\top,$$
$$M_T^{\mathrm{plug}}(m) = I + V_m + (T - m - R_{\mathrm{eval}})x_{\widehat{*}} x_{\widehat{*}}^\top + R_{\mathrm{eval}} \sum_a \widehat{\nu}(a) x_a x_a^\top.$$

Thus the reported drift is the full-deployment quantity $\|\big(M_T^{\mathrm{ref}}(m)\big)^{-1/2}\big[M_T^{\mathrm{plug}}(m) - M_T^{\mathrm{ref}}(m)\big]\big(M_T^{\mathrm{ref}}(m)\big)^{-1/2}\|_{\mathrm{op}}$. The unique-best margin is 0.65754. On the 0.9 OLS radius event, entropy strong convexity gives $\|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1 \le 4\kappa_{\mathrm{w}} L_x r_m/\tau$, which feeds directly into $\varepsilon_T^{\mathrm{eval}} \le R_{\mathrm{eval}}\|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1$ and $\varepsilon_T^{\mathrm{info}} \le R_{\mathrm{eval}}\ell_T^{\mathrm{ref}}\|\widehat{\nu} - \nu^{\mathrm{ref}}\|_1$. All repetitions satisfy the radius event and identify the best action. Median design error decreases from 0.01947 to 0.00632 with log–log slope -0.54; median target-normalized drift falls from 0.00963 to 0.00165.

The finite-m radius in Theorem 4.1 substitutes the pilot bound $b_m = R_{\mathrm{eval}}\ell_T^{\mathrm{ref}} C_m$ into the matrix concentration formula; it is saved as `eta_pilot_certificate`. For $m = (72, 144, 288, 576)$, this gives respectively $\eta_T^{\mathrm{cert}} \approx (2.738, 1.876, 1.223, 0.739)$. Hence the theorem's $\eta_T^{\mathrm{cert}} < 1$ condition, and therefore its inverse-quadratic sandwich, is certified at $m = 576$ in this finite experiment.

The same 1,200 pilot–deployment repetitions also draw complete deployments from the frozen plug-in mixtures. The saved ex-post diagnostic η_T^{post} (`eta_expost`) replaces b_m by the observed oracle-normalized plug-in drift. As m increases from 72 to 576, the median normalized realized operator error falls from 0.09035 to 0.04975, while the median η_T^{post} falls from 0.30341 to 0.24055; the largest observed operator-error-to-η_T^{post} ratio is 0.76812. Figure 13 assembles the matched baselines, pilot contraction, and this realized-geometry diagnostic, while the theorem-level radius above identifies the pilot size with a formal directional certificate.

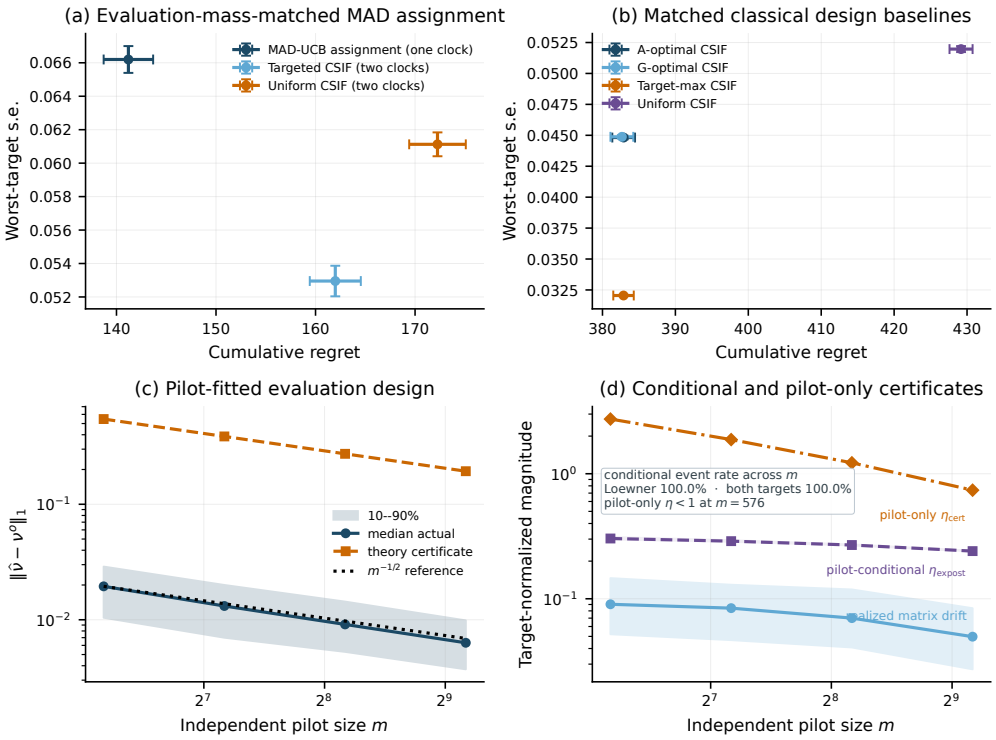


Figure 13: Evaluation-mass-matched MAD assignment, schedule-matched classical linear designs, and independent-pilot plug-in transfer. Panel (c) compares realized design error with the theorem's C_m bound. Panel (d) separates realized matrix drift, the conditional ex-post radius η_T^{post}, and the pilot-only radius η_T^{cert}; only the $m = 576$ pilot-only radius is below one. Error bars are 95% Monte Carlo intervals and ribbons are repetition quantiles.

I.9 KEY OUTPUT INVENTORY

File	Content
`results/mab_final.csv`	Repetition-level MAB regret, coverage, precision, and arm counts.
`results/mab_target.csv`	Noiseless common-index target counts at every checkpoint.
`results/mab_csif_target.csv`	Two-clock deterministic targets, observed checkpoint means, and armwise relative tracking errors.
`results/final_frontier_raw.csv`	All 29,000 repetition-level outcomes on the prespecified MAB schedule grid.
`results/final_frontier_summary.csv`	Grid-cell means and Monte Carlo standard errors for regret and worst-target precision.
`results/final_frontier_all_grid.csv`	All 31 prespecified parameter aliases per method, mapped to their unique schedule IDs, including dominated settings.
`results/final_frontier_nondominated.csv`	Method-specific nondominated finite-grid frontier points.
`results/final_frontier_budget.csv`	Cost-matched selections used for direct welfare-budget comparisons.
`results/final_frontier_grid.csv`	All 31 prespecified scale–decay aliases, their schedule IDs, and cumulative masses.
`results/final_frontier_metadata.json`	Grid construction, repetition count, seed, and frontier-selection metadata.
`figures/final_frontier.pdf` / `figures/final_frontier.png`	Vector and raster renderings of the complete finite-grid frontier.
`results/design_final.csv`	Fixed-population regret, AIPW coverage, widths, average propensities, and harmonic information.

File	Content
results/design_studentized.csv	Terminal studentized AIPW statistics.
results/design_calibration.csv	Contrastwise bias, exact-versus-empirical variance, interval width, and coverage calibration.
results/envelope_evidence_raw.csv	Finite-horizon envelope repetitions over horizon, evaluation share, and evaluation direction.
results/envelope_evidence_summary.csv	Cellwise coverage, Wilson bounds, and envelope-to-Gaussian width ratios.
results/decay_rate_raw.csv	Repetition-level regret and estimation error over horizons and decay exponents.
results/decay_rate_summary.csv	Exact and empirical regret/standard-error paths for the rate study.
results/decay_rate_slopes.csv	Fitted log–log exponents and their theoretical counterparts.
figures/design_calibration.pdf / figures/design_calibration.png	Exact design-moment calibration figure.
figures/envelope_evidence.pdf / figures/envelope_evidence.png	Finite-horizon coverage and width-price figure.
figures/decay_rate_evidence.pdf / figures/decay_rate_evidence.png	Multi-horizon regret–precision rate figure.
results/linear_final.csv	Linear regret, coverage, targetwise standard errors and pivots, eigenvalues, condition numbers, and counts.
figures/linear_adaptive_qq.pdf / figures/linear_adaptive_qq.png	Directionwise Q–Q diagnostic built from both target pivots in every repetition.
results/hillstrom_frontier.csv	Cost-matched uniform and information-flow allocations over the full budget path.
results/hillstrom_split_evaluation.csv	Repetition-level stratified train/held-out design evaluation.
results/hillstrom_split_summary.csv	Out-of-sample width gain, quantiles, win rate, and diagnostic regret summary.
results/hillstrom_bootstrap.csv	Refitted parametric-bootstrap point clouds at four prospective intensities.
results/sensitivity_schedule.csv	MAB evaluation-schedule grid.
results/sensitivity_misspecification.csv	Full-sample prospective Hillstrom variance-perturbation repetitions.
results/mab_closest_baseline.csv	Evaluation-mass-matched MAD assignment and two-clock comparisons.
results/mab_closest_baseline_summary.csv	Summary and Monte Carlo errors for the closest MAB assignment baseline.
results/linear_geometry_baselines.csv	Target-max, A-optimal, G-optimal, and uniform linear evaluations.
results/linear_geometry_baseline_summary.csv	Summary and Monte Carlo errors for the linear design baselines.
results/linear_geometry_designs.csv	Evaluation distributions and their defining design criteria.
results/pilot_plugin.csv	Pilot error, reference drift, realized matrices, eta_expost, eta_pilot_certificate, and their distinct event indicators for all 1,200 repetitions.
results/pilot_plugin_summary.csv	Pilot-size summaries of realized operator error, both radii, and their corresponding matrix/target event rates.
results/pilot_plugin_metadata.csv	Pilot and deployment constants, margin, schedule, empirical log–log slope, confidence levels, and aggregate ex-post checks.
figures/closest_baselines_plugin.pdf / figures/closest_baselines_plugin.png	Closest-baseline comparisons, pilot contraction, and the realized-matrix diagnostic path.
results/run_metadata.csv	Seeds, commands, validation runtime, and explicit measured or metadata-only timing status.

File	Content
`results/results_macros.tex`	Automatically generated manuscript macros.

J LEAN 4 FORMALIZATION BOUNDARY

The artifact contains a pinned Lean 4/mathlib project in `LeanProofs/`. A full `lake build` kernel-checks eleven named declarations: four in `InformationFlows/Basic.lean`, three in `InformationFlows/FiniteSums.lean`, and four paper-facing declarations in `InformationFlows/VerifiedCore.lean`. The first seven certify the mixture-regret, two-clock information, target-deviation, regret–variance, finite-sum regret, regret-error, and probability-mass identities. The four paper-facing declarations are:

- `ht_finite_sum_unbiased`: the finite HT score identity, requiring nonzero propensity only on nonzero scores;
- `aipw_finite_sum_unbiased`: the corresponding AIPW identity;
- `ht_residual_variance_decomposition`: the algebraic decomposition underlying $Q_T^{\mathrm{score}} - s_T^2 = H_T$; and
- `two_clock_arm_count_decomposition`: realized counts equal decision-clock plus evaluation-clock counts.

The pinned commands

```
cd LeanProofs
lake env lean --error=warning InformationFlows/VerifiedCore.lean
lake build
```

complete successfully. For these four paper-facing declarations only, the explicit `#print axioms` report lists Lean's standard `propext`, `Classical.choice`, and `Quot.sound`; there is no `sorryAx` or user axiom. A source check also rejects `sorry`, `admit`, and declared `axiom` placeholders. The exact commands and declaration inventory are maintained in `LeanProofs/README.md` and `PROOF_STATUS.md`.

The Lean layer kernel-checks the algebraic cores listed above. The mathematical appendices establish the filtration and predictability arguments, problem-specific concentration reductions, optimization regularity, and asymptotic limits, using Gaussian concentration, matrix and scalar Freedman inequalities, martingale CLTs and Lenglart's inequality, convex KKT theory, and adaptive KL, Pinsker, and Le Cam bounds. These reductions include Theorems 3.1 and 4.1 to 4.3, the design-based chain Theorem 4.4, Proposition F.1, and Corollary F.1, and the local lower-bound, reference-frontier, and linear-model conclusions in Theorems 4.5 and H.1 and Corollary 4.1. The artifact kernel-checks the finite algebra; the appendices prove the probabilistic, optimization, and asymptotic results.