# BLOCK-OSRB: EXACT OUTPUT-STATISTIC RAO–BLACKWELLIZATION FOR AGGREGATE-INVARIANT VECTOR BLACK-BOX OPTIMIZATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Optimizing differentiable models against vector-valued black-box scores is expensive when each query evaluates thousands of binary decisions. Many utility and fairness scores depend only on public aggregates, despite an unknown mapping from aggregates to values. We ask whether this structure can reduce gradient noise without increasing oracle queries. Block Output-Statistic Rao–Blackwellization (Block-OSRB) conditions a two-query antithetic estimator on paired block aggregates, integrating out row identities invisible to the evaluator. The same two oracle values suffice for every row, without counterfactual calls. For independent Bernoulli decisions, the estimator is unbiased for the unsmoothed objective and has no greater covariance than its DisARM parent for each output column. Nested blocks provide a variance–compute path; common refinement gives joint covariance control, and weighted-total conditioning dominates weight-category count conditioning. On Adult, ACSIncome, Waterbirds, and UCI Credit, count Block-OSRB has lower mean one-time holdout hypervolume regret in all eight task–budget cells than the evaluated DisARM and true UGC arms: 2.0–19.7% and 2.2–23.8% reductions, respectively. The first three DisARM holdouts use the ordinary parent; UCI Credit uses relevance masking. A separately frozen Student Dropout trainable-MLP test shows an estimator gain, while a deterministic group-ranking policy performs better. On weighted UCI Credit development data, conditioning on totals removes 59.4–60.7% of the covariance trace remaining after category-count conditioning. The approach targets declared aggregate interfaces when oracle queries dominate local computation.

## 1 INTRODUCTION

A differentiable model may produce thousands of binary decisions, while an evaluator returns only a few non-differentiable scores. Utility and group-fairness metrics often have a useful public structure: exchanging decisions within a label–group cell leaves their confusion counts, and hence their scores, unchanged. The function that maps those counts to scores can remain a numerical black box.

Likelihood-ratio (LR) estimators optimize these scores without differentiating the evaluator. DisARM uses an antithetic Bernoulli pair to reduce variance (Dong et al., 2020); other couplings use multiple correlated samples (Dimitriev & Zhou, 2021; Dong et al., 2021). Yet the gradient estimate still changes when two rows exchange decisions while every score remains fixed. That row-identity variation can be integrated out.

Block-OSRB conditions DisARM on the paired aggregates observed in each public block. Two prediction pairs with identical confusion counts have the same metric feedback even if different rows produced those counts. A bivariate generating polynomial averages their remaining row coefficients, reusing the two observed oracle values for counts and bounded integer-weighted sums. This is the query-saving step: the conditional coefficients, rather than counterfactual metric values, are computed locally.

We split large public cells into blocks of at most $b$ rows. Larger aligned blocks integrate more identity noise, yielding a nested covariance order at $O(Mnb^2)$ local arithmetic. Averaging estimates from



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.

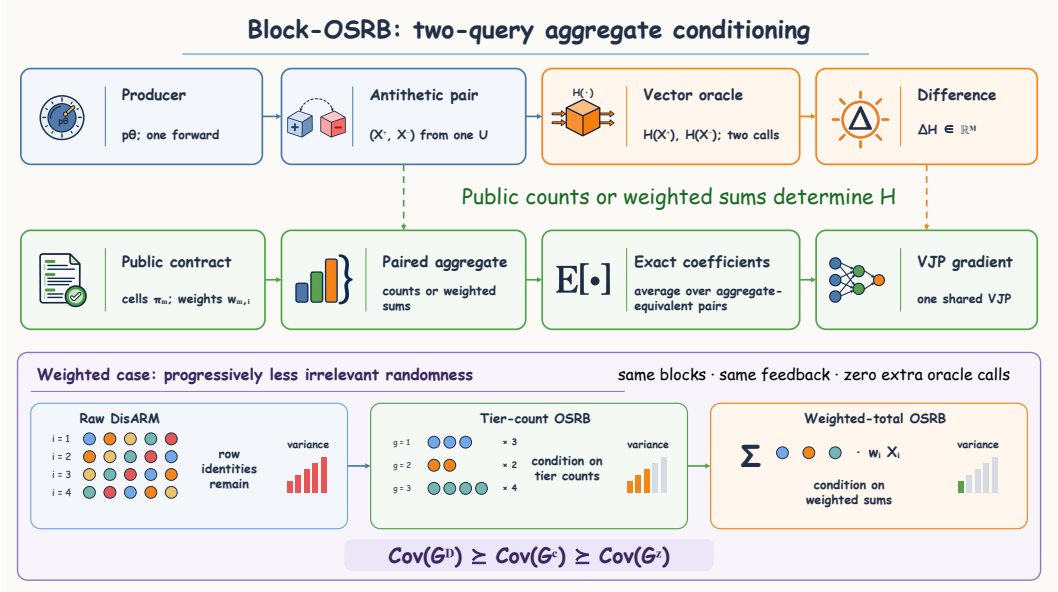



Figure 1: Access contract and exact conditioning. The antithetic pair supplies two oracle values; public aggregates condition the row coefficients before one VJP. Bottom: bounded weighted totals remove additional variance beyond tier counts using the same pair and blocks. Dashed arrows carry public information or reused feedback, not extra oracle calls.

several random block partitions retains the two oracle calls and makes the estimator equivariant in distribution to row order.

The paper establishes:

- exact objective-wise and common-refinement estimators, with $L^2$ projection and columnwise or joint PSD covariance guarantees for their specified DisARM parents;
- exact conditioning on bounded weighted totals, which dominates conditioning on weight-category counts, together with block and multi-partition implementations; and
- finite-state checks, local-cost measurements, and low-query optimization on four fixed-representation tasks and a separately frozen trainable-MLP task, including one-time count-based holdouts.

Each conditional gradient is the best count-measurable approximation to a fixed DisARM parent in $L^2$. We also characterize bias under a misspecified aggregate contract and the SGD noise consequence of the covariance order. The experiments separate estimator gains from task-level policy choice and report local cost alongside oracle queries.

## 2 Problem Setup

Let independent binary decisions satisfy $X_i \sim \text{Bernoulli}(p_i)$ with $p_i \in (0,1)$ for $i = 1, \dots, n$, where $p = p_\theta$ is produced by a differentiable model. A vector oracle $H : \{0,1\}^n \to \mathbb{R}^M$ returns $M$ objectives in one call. We minimize a scalarization $v^\top \mathbb{E}[H(X)]$ for a fixed preference $v$, or estimate all columns before selecting a preference.

For each objective $m$, assume a known partition $\pi_m$ and an unknown function $h_m$ such that

$$H_m(X) = h_m(C_m(X)), \qquad C_{m,g}(X) = \sum_{i:\pi_m(i)=g} X_i. \tag{1}$$

Rows irrelevant to an objective may be omitted from its partition. The partition is known, while $h_m$ is evaluated through the black box. Utility and fairness can therefore use different public partitions.

Table 1: Representative public count contracts. The evaluator may apply an unknown nonlinear function to these counts. Multiple rows can be returned together by one vector oracle.

| Oracle output family | Public row cells whose positive counts suffice |
|---|---|
| Confusion-matrix utility (F-score, balanced accuracy, MCC) | ground-truth label $y$ |
| Demographic-parity or selection-rate gap | protected group $a$ |
| Equal-opportunity or equalized-odds component | $(y, a)$; irrelevant labels may be omitted |
| Intersectional or policy-specific score | any declared fixed strata $(y, a_1, a_2, \ldots)$ |

The hidden function of these counts may be nonsmooth. Table 1 lists common metrics; fixed stratum sizes let one positive-decision count per cell determine both positive and negative counts.

For the experimental oracle $H = [1 - \mathrm{macroF1}, |\mathrm{DP}|, |\mathrm{EO}|]$, this contract is concrete. Macro-F1 uses cells indexed by $y$: the positive prediction count in each cell, together with its fixed size, recovers every confusion count. DP uses cells indexed by $a$. EO uses only $(y = 1, a)$ cells, so all $y = 0$ rows are publicly irrelevant to that column. The common refinement is $(y, a)$. Consequently the paired cell counts determine $H(X^+) - H(X^-)$ exactly while leaving unresolved which rows produced those counts; Block-OSRB integrates precisely that remaining identity randomness.

Write $d_i = \nabla_\theta p_i$. The desired gradient column is

$$g_m^\star = \sum_i d_i \frac{\partial}{\partial p_i} \mathbb{E}[H_m(X)]. \tag{2}$$

Query cost counts complete-vector evaluations of $H$; we report local algebra and the model vector–Jacobian product separately.

## 3 Output-statistic Rao–Blackwellization

### 3.1 The parent antithetic estimator

Draw $U_i \overset{\mathrm{iid}}{\sim} \mathrm{Uniform}(0, 1)$ and form the standard antithetic Bernoulli pair $X_i^+ = \mathbb{I}[U_i < p_i]$ and $X_i^- = \mathbb{I}[1 - U_i < p_i]$. Define $T_i = (X_i^+, X_i^-)$ and

$$q_i(T_i) = \frac{1}{2}(-1)^{X_i^-} \mathbb{I}[X_i^+ \neq X_i^-] \frac{\max(p_i, 1 - p_i)}{p_i(1 - p_i)}. \tag{3}$$

The probability-coordinate DisARM column is

$$G_m^{\mathrm{D}} = \Delta H_m \sum_i d_i q_i(T_i), \qquad \Delta H_m = H_m(X^+) - H_m(X^-), \tag{4}$$

and uses exactly two vector-oracle calls (Dong et al., 2020).

### 3.2 Conditioning on paired output counts

For a public cell $g$, let $A_g = C_{m,g}(X^+)$ and $B_g = C_{m,g}(X^-)$. Equation (1) makes $\Delta H_m$ measurable given all paired cell counts. We therefore replace every $q_i$ with

$$\phi_{i,m} = \mathbb{E}[q_i(T_i) \mid \{(A_g, B_g)\}_g] \tag{5}$$

and define

$$G_m^{\mathrm{OS}} = \Delta H_m \sum_i d_i \phi_{i,m}. \tag{6}$$

The expectation contains no oracle evaluation. For a cell, the four probabilities of $T_i$ are $P_i(0, 1) = P_i(1, 0) = \min(p_i, 1 - p_i)$, $P_i(0, 0) = (1 - 2p_i)_+$, and $P_i(1, 1) = (2p_i - 1)_+$. Define

$$P_g(s, t) = \prod_{i \in g} \sum_{a,b \in \{0,1\}} P_i(a, b) s^a t^b. \tag{7}$$

**Algorithm 1** One Block-OSRB update (objective-wise)

**Require:** Interior probabilities $p_\theta$, VJP closure, public partitions and fixed block refinements, preference $v$

1: Draw shared $U_i \sim \mathrm{Uniform}(0,1)$ and form the DisARM pair $(X^+, X^-)$
2: Query the complete vector oracle exactly twice: $H^+ = H(X^+)$ and $H^- = H(X^-)$
3: **for** each objective $m$ and its public block $g$ **do**
4: Record the paired count $(A_g, B_g)$ (or weighted total $(Z_g^+, Z_g^-)$)
5: Build prefix–suffix coefficient tables for the block polynomial
6: Evaluate Equation (8) for every row $i \in g$
7: **end for**
8: Form the output-gradient columns using $H^+ - H^-$, then scalarize by $v$
9: Apply one model VJP and return the parameter-gradient estimate

If $P_{g,-i}$ deletes factor $i$, coefficient extraction gives the exact value

$$\phi_i = \frac{\sum_{a,b} q_i(a,b) P_i(a,b) [s^{A_g - a} t^{B_g - b}] P_{g,-i}}{[s^{A_g} t^{B_g}] P_g}. \tag{8}$$

The realized denominator is positive; the implementation checks this condition.

Algorithm 1 shows the two oracle calls and the blockwise coefficient calculation. Appendix E gives implementation details.

**Theorem 1** (Exact columnwise dominance)**.** *Under Equation* (1) *and finite second moments,* $G_m^{\mathrm{OS}} = \mathbb{E}[G_m^{\mathrm{D}} \mid \{(A_g, B_g)\}_g]$. *Consequently,* $\mathbb{E} G_m^{\mathrm{OS}} = g_m^\star$ *and*

$$\mathrm{Cov}(G_m^{\mathrm{D}}) - \mathrm{Cov}(G_m^{\mathrm{OS}}) = \mathbb{E}\big[\mathrm{Cov}(G_m^{\mathrm{D}} \mid \{(A_g, B_g)\}_g)\big] \succeq 0. \tag{9}$$

*For every fixed linear map* $A$*, the same PSD order holds after applying* $A$*.*

**Corollary 1** ($L^2$ parent-projection characterization)**.** *Let* $S_m = \{(A_g, B_g)\}_g$*. For every square-integrable,* $S_m$*-measurable random vector* $Z$ *with the same dimension as* $G_m^{\mathrm{D}}$*,*

$$\mathbb{E}\|G_m^{\mathrm{D}} - Z\|^2 = \mathbb{E}\|G_m^{\mathrm{D}} - G_m^{\mathrm{OS}}\|^2 + \mathbb{E}\|G_m^{\mathrm{OS}} - Z\|^2. \tag{10}$$

*Thus Block-OSRB is the unique* $L^2$*-best count-measurable approximation of this fixed DisARM parent, up to almost-sure equality.*

**Proof sketch.** The feedback difference in Equation (4) is measurable with respect to the paired counts, so conditioning acts only on each $q_i$. Equation (8) evaluates that conditional expectation. The tower property proves the mean identity, total covariance proves Equation (9), and congruence by $A$ preserves PSD order. Moreover, for any fixed direction $a$, the scalar variance gap is $\mathbb{E}[\mathrm{Var}(a^\top G_m^{\mathrm{D}} \mid \{(A_g, B_g)\}_g)]$. It is positive exactly when identity noise remains after observing the counts, and zero exactly when the projected parent is already count-measurable. A complete finite-dimensional proof is in Appendix D.

Corollary 1 is the conditional-expectation Pythagorean identity: $G_m^{\mathrm{D}} - G_m^{\mathrm{OS}}$ is orthogonal in $L^2$ to every $S_m$-measurable vector. Its comparison class is the set of $S_m$-measurable approximations to the specified DisARM parent.

### 3.3 Beyond counts: bounded additive statistics

Counts are the unit-weight case of a broader exact interface. Suppose a public block $g$ has positive integer weights $w_i$ and the oracle column depends on

$$Z_g(X) = \sum_{i \in g} w_i X_i \tag{11}$$

through an otherwise hidden function. A common factor in the weights is removed without changing the statistic's sigma-algebra. For paired totals $(Z_g^+, Z_g^-)$, replace the cell polynomial by

$$P_g^w(s,t) = \prod_{i \in g} \sum_{a,b \in \{0,1\}} P_i(a,b) s^{w_i a} t^{w_i b}. \tag{12}$$

Deleting factor $i$ and shifting the requested exponents by $(w_i a, w_i b)$ gives the exact analogue of Equation (8). It needs no new oracle value.

For the same fixed physical blocks, let $C$ contain the paired positive counts separately for every distinct weight, and let $Z$ contain only the paired weighted totals. Since $Z$ is a deterministic function of $C$, define $G^C = \mathbb{E}[G^{\mathrm{D}} \mid C]$ and $G^Z = \mathbb{E}[G^{\mathrm{D}} \mid Z]$.

**Theorem 2** (Additive-statistic dominance)**.** *Under Equation* (11) *and finite second moments, both estimators are unbiased, reuse the parent's two vector-oracle calls, and*

$$G^Z = \mathbb{E}[G^C \mid Z], \qquad \mathrm{Cov}(G^{\mathrm{D}}) \succeq \mathrm{Cov}(G^C) \succeq \mathrm{Cov}(G^Z). \tag{13}$$

*The second gap equals $\mathbb{E}[\mathrm{Cov}(G^C \mid Z)]$ and is strict in a direction exactly when the category-count estimator is not determined by the weighted total in that direction.*

If a block contains at most $b$ rows and its normalized weights are bounded by $A$, the direct weighted-support dynamic program costs $O(nb^2A^2)$ over all rows for fixed output dimension. The dependence on $A$ is pseudo-polynomial; large-range weights can make the finer category-count fallback computationally preferable despite its larger covariance.

**Proposition 1** (Contract misspecification)**.** *For any declared statistic $S$, let $R_m = \sum_i d_i q_i(T_i)$ and suppose the same coefficient calculation forms $\widehat{G}_m = \Delta H_m \mathbb{E}[R_m \mid S]$, without assuming that $\Delta H_m$ is $S$-measurable. Then*

$$\mathbb{E}[\widehat{G}_m] - g^\star_m = -\mathbb{E}[\mathrm{Cov}(\Delta H_m, R_m \mid S)], \tag{14}$$

$$\|\mathbb{E}[\widehat{G}_m] - g^\star_m\| \le \sqrt{\mathbb{E}\,\mathrm{Var}(\Delta H_m \mid S)\;\mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(R_m \mid S)}. \tag{15}$$

*Thus an approximate contract has a bias determined jointly by within-count oracle variation and residual parent noise.*

A practical contract check derives cells from the published metric or evaluator specification, verifies exact invariance when evaluator code and state space permit, and otherwise swaps decisions within declared cells while holding the statistic fixed. Any discrepancy beyond the evaluator's numerical tolerance selects the original DisARM parent. Repeated occurrences of the same statistic also permit empirical estimates of the two conditional-variance factors in Equation (15). Section 7 states the certification boundary of finite probes.

Theorem 1 gives columnwise PSD dominance for objective-specific partitions. The following common-refinement construction gives a joint PSD order for the vectorized gradient.

Let $I_m$ be the rows relevant to objective $m$ and let $\widetilde{G}^{\mathrm{D}}_m = \Delta H_m \sum_{i \in I_m} d_i q_i(T_i)$ be the masked parent. Form the coarsest common row partition whose cells refine every objective partition, and let $S_\cap$ collect its paired counts. Define $G^\cap = \mathbb{E}[\widetilde{G}^{\mathrm{D}} \mid S_\cap]$ jointly over all columns.

**Theorem 3** (Joint common-refinement dominance)**.** *Under the count contract and finite second moments, $G^\cap$ is unbiased, uses two vector-oracle calls, and*

$$\mathrm{Cov}(\mathrm{vec}\,\widetilde{G}^{\mathrm{D}}) - \mathrm{Cov}(\mathrm{vec}\,G^\cap) \succeq 0. \tag{16}$$

*For every fixed linear map $L$, including a fixed objective scalarization followed by a fixed model VJP, the same PSD order holds after applying $L$.*

The same coefficient extraction in Equation (8) is run once per common cell and shared across output columns. Splitting each common cell into fixed blocks of size at most $b$ preserves the theorem and costs $O(nb^2 + Mn)$ local arithmetic. A finer common partition can retain more identity noise than separate objective-wise conditioning, so this option supplies joint covariance control with a measurable marginal-variance trade-off. The experiments use the objective-wise estimator for the original three-task holdouts, a Waterbirds development diagnostic for common refinement, and a prospectively frozen UCI Credit development-and-holdout evaluation for both variants.

**Corollary 2** (SGD noise term)**.** *Let $f(\theta) = v^\top \mathbb{E}[H(X)]$ be $L$-smooth and lower bounded by $f_{\inf}$. Apply SGD for $T$ steps with an unbiased single-objective Block estimator, or with the common-refinement estimator under a fixed $v$, and constant $0 < \eta \le 1/L$. If its conditional noise trace is uniformly bounded by $\sigma^2_{\mathrm{B}}$, then*

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\|\nabla f(\theta_t)\|^2 \le \frac{2(f(\theta_0) - f_{\inf})}{\eta T} + L\eta\sigma^2_{\mathrm{B}}. \tag{17}$$

*Theorems 1 and 3 allow $\sigma_{\mathrm{B}}^2 \le \sigma_{\mathrm{D}}^2$ for the corresponding parent, so the estimator-dependent term is no larger at the same step size and horizon.*

The covariance order thus lowers the estimator-dependent noise term for the corresponding objective or fixed scalarization.

## 4 Block conditioning at population scale

Exact full-cell dynamic programming has unfavorable memory and time at large $n$. Let $\rho_b$ split each public cell into ordered blocks containing at most $b$ rows. We apply Equation (8) separately inside every block while retaining the same complete-vector feedback difference. This is valid because the block paired counts refine the public cell counts and hence still determine $H_m(X^+)$ and $H_m(X^-)$.

With aligned boundaries, refinements for $b = 1, 2, 4, \dots$ form nested sigma-algebras. A singleton block observes the complete pair state of each row and therefore recovers DisARM after masking oracle-irrelevant rows. A full public cell recovers Equation (6).

**Theorem 4** (Nested block path)**.** *For aligned nested block refinements $\mathcal{S}_1 \supseteq \mathcal{S}_2 \supseteq \cdots \supseteq \mathcal{S}_{\mathrm{full}}$,*

$$\mathrm{Cov}(G_m^{\mathrm{D}}) \succeq \mathrm{Cov}(G_{m,1}^{\mathrm{B}}) \succeq \mathrm{Cov}(G_{m,2}^{\mathrm{B}}) \succeq \cdots \succeq \mathrm{Cov}(G_{m,\mathrm{full}}^{\mathrm{B}}). \tag{18}$$

*Every member is exactly unbiased and uses the same two calls to $H$. If a block of size $b$ uses the direct prefix–suffix bivariate recurrence, total local time is $O(Mnb^2)$ and peak auxiliary state depends on $b$, not $n$.*

Larger aligned blocks lower covariance along this path while increasing local arithmetic and memory.

**Choosing the block size and checking amortization.** We choose the largest aligned $b$ allowed by local time, memory, and precision; $b = 64$ is the measured operating point here. Let $T_{\mathrm{B}}$ and $T_{\mathrm{D}}$ be the measured local update times and let $L$ be the latency of one $H$ call. Since both estimators use two calls, requiring Block's extra local time to be at most a fraction $r$ of DisARM's end-to-end time gives

$$L \ge \frac{(T_{\mathrm{B}} - T_{\mathrm{D}})/r - T_{\mathrm{D}}}{2}. \tag{19}$$

The target device determines the feasible $b$ and whether the added local computation is small relative to the shared $2L$ oracle time.

### 4.1 Multi-partition Block-OSRB

An ordered block refinement depends on an arbitrary row order. For objective $m$, let $G_m^{(1)}, \dots, G_m^{(K)}$ be exact block estimators computed from the same antithetic pair and feedback, using refinements fixed independently of that pair. For fixed nonnegative weights summing to one, define $\bar{G}_m = \sum_k w_k G_m^{(k)}$.

**Theorem 5** (Convex multi-partition estimator)**.** *$\bar{G}_m$ is unbiased, uses two oracle calls, and*

$$\mathrm{Cov}(\bar{G}_m) \preceq \sum_k w_k \,\mathrm{Cov}(G_m^{(k)}) \preceq \mathrm{Cov}(\widetilde{G}_m^{\mathrm{D}}). \tag{20}$$

*If refinements are generated by i.i.d. continuous row keys independent of model state and feedback, the estimator is equivariant in distribution to a row permutation.*

The first inequality follows because the difference between its right and left sides equals $\frac{1}{2}\sum_{k,\ell} w_k w_\ell \mathbb{E}[(G_m^{(k)} - G_m^{(\ell)})(G_m^{(k)} - G_m^{(\ell)})^\top] \succeq 0$. The second applies Theorem 4 and convexity. These claims are columnwise for the objective-specific implementation used in our experiments. If every member instead conditions the masked vector parent on a common refinement, Theorem 3 gives the same two inequalities for the vectorized gradient. Computing $K$ objective-wise members costs $O(KMnb^2)$ local arithmetic but no additional $H$ queries.

**Estimator choice.** Use objective-wise conditioning for separate output gradients and common refinement when a fixed scalarization needs joint covariance control. Uniform MP4 averages four random refinements to remove dependence on a chosen row order in distribution. Additive Block handles bounded integer-weighted totals. Appendix A compares their guarantees and local costs.

## 5 RELATION TO PRIOR WORK

DisARM removes augmentation noise from an antithetic Bernoulli estimator and supplies our fixed two-query parent (Dong et al., 2020); ARMS and UGC instead change the coupling or coordinate estimator (Dimitriev & Zhou, 2021; Kunes et al., 2023). Rao–Blackwell and local-expectation methods integrate discrete randomness, often needing function values at retained or counterfactual outcomes (Liu et al., 2019; Titsias & Lázaro-Gredilla, 2015; De Smet et al., 2023; Kool et al., 2020). Here the public aggregates determine both observed feedback values. Exact coefficient extraction therefore conditions every row of the fixed DisARM pair while preserving its two-query cost.

ULR and OVLR cache model outputs to handle general gradient-agnostic losses (Jiang et al., 2024; Zou et al., 2026). Block-OSRB uses the same output-level perspective for a Bernoulli count interface and derives a covariance order for a fixed antithetic parent. ESG studies general scalar product-Bernoulli oracles through a separate query construction (Sornwanee, 2026). Appendix B gives the detailed comparisons.

## 6 EXPERIMENTS

Table 2: One-time holdout hypervolume regret (mean $\pm$ seed standard deviation; lower is better). Bold marks the lowest mean among the three displayed methods. Block-OSRB denotes objective-wise count Block-OSRB-64; gain is its relative regret reduction versus DisARM/true UGC (%), computed from unrounded means. DisARM is relevance-masked on UCI Credit and unmasked on the other tasks. All runs use fixed task-specific holdouts and hard query ceilings.

| Task | Budget | Block-OSRB | DisARM | True UGC | Gain vs D/U (%) |
|---|---|---|---|---|---|
| Adult | 24H | $\mathbf{0.388} \pm 0.012$ | $0.400 \pm 0.030$ | $0.401 \pm 0.028$ | 3.07/3.32 |
| | 96H | $\mathbf{0.360} \pm 0.032$ | $0.367 \pm 0.021$ | $0.368 \pm 0.023$ | 1.95/2.21 |
| ACSIncome | 24H | $\mathbf{0.313} \pm 0.026$ | $0.348 \pm 0.066$ | $0.354 \pm 0.067$ | 10.24/11.52 |
| | 96H | $\mathbf{0.278} \pm 0.008$ | $0.299 \pm 0.006$ | $0.298 \pm 0.007$ | 6.76/6.56 |
| Waterbirds | 24H | $\mathbf{0.564} \pm 0.007$ | $0.608 \pm 0.013$ | $0.617 \pm 0.010$ | 7.21/8.53 |
| | 96H | $\mathbf{0.439} \pm 0.011$ | $0.546 \pm 0.018$ | $0.576 \pm 0.015$ | 19.66/23.82 |
| UCI Credit | 24H | $\mathbf{0.384} \pm 0.006$ | $0.402 \pm 0.011$ | $0.404 \pm 0.012$ | 4.51/5.12 |
| | 96H | $\mathbf{0.361} \pm 0.005$ | $0.385 \pm 0.012$ | $0.386 \pm 0.006$ | 6.33/6.56 |

### 6.1 DESIGN AND ACCESS ACCOUNTING

We evaluate Adult (Becker & Kohavi, 1996), ACSIncome (Ding et al., 2021), Waterbirds (Sagawa et al., 2020), and prospectively selected UCI Credit (Yeh, 2009): two census tasks, a natural-image robustness task, and a financial-risk task. The oracle returns $[1 - \mathrm{macroF1}, |\mathrm{DP}|, |\mathrm{EO}|]$ from public confusion-count partitions. Three fixed preferences yield three feature-based policies evaluated on unseen rows; we report their dominated hypervolume with reference $(1, 1, 1)$ as $1 - \mathrm{HV}$ (lower is better). Each task uses a train-only frozen predictor and optimizes a 32-dimensional linear logit residual; Waterbirds uses frozen ResNet-18 features. Holding the predictor, optimizer, partitions, and development oracle fixed isolates the effect of the gradient estimator. A separate process receives each complete prediction vector and returns the three objectives.

Block-OSRB and DisARM use two $H$ calls per update, giving 12/48 updates at 24H/96H. We also compare true UGC, ordinary antithetic LR, SPSA, Quotient-FD, and budget-matched AS/FW adaptations; their trajectories stay within the same hard query ceilings. True UGC uses 22–24H/94–96H because its update costs either 2H or 4H; Quotient-FD uses 17–24H/88–96H, and the FW-style adaptation uses 18H/90H. The count-based final evaluations use Adult official test, a sealed 10,000-row ACSIncome split, Waterbirds official test, and the prospectively fixed UCI Credit holdout. Appendix H gives the protocols, seeds, and query ledgers.

### 6.2 LOW-QUERY OPTIMIZATION

Table 2 shows that objective-wise count Block-OSRB achieves the lowest mean one-time holdout regret of the three displayed methods in all eight task–budget cells. The largest reduction is on

Waterbirds at 96H: 19.66% versus DisARM and 23.82% versus true UGC. Across the three original tasks, fixed-task stratified seed bootstrap intervals for pooled absolute comparator-minus-Block improvement are $[0.03078, 0.04539]$ against ordinary DisARM and $[0.03797, 0.05283]$ against true UGC. On the prospectively selected UCI Credit holdout, the paired intervals are $[0.01428, 0.02864]$ against masked DisARM and $[0.01686, 0.02972]$ against true UGC. These intervals quantify seed variation conditional on the fixed tasks and holdout splits; the Adult 96H mean direction against ordinary DisARM corresponds to four Block wins in eight paired seeds.

Table 3: Block-64 versus relevance-masked DisARM: relative reduction in mean hypervolume regret (%, higher is better). Each row uses eight paired seeds. Development comparisons and one-time held-out evaluations are identified separately; wins count lower Block regret at 96H.

| Task | Evaluation | 24H gain | 96H gain | 96H wins |
|---|---|---|---|---|
| Adult | Development | 0.20 | 3.40 | 7/8 |
| ACSIncome | Development | 5.53 | 3.93 | 8/8 |
| Waterbirds | Development | 4.57 | 13.04 | 8/8 |
| UCI Credit | One-time holdout | 4.51 | 6.33 | 8/8 |
| Student Dropout | One-time test | 0.73 | 2.93 | 8/8 |

Table 3 isolates the benefit over the stronger parent that removes publicly irrelevant rows. The Adult 24H development gain is small, with a paired interval spanning zero; the 96H mean gain is positive in all five matched settings. Only UCI Credit and Student Dropout provide one-time held-out matched-parent comparisons. In the separately frozen Waterbirds development audit, common-refinement Block-64 improves on masked DisARM by 3.82%/12.24% at 24H/96H; on the UCI Credit holdout the gains are 3.25%/6.20%, with 3.87%/6.44% versus true UGC. On weighted UCI Credit development data, additive Block-64 lowers 24H/96H regret by 5.77%/8.97% versus masked DisARM and 0.29%/2.46% versus the stronger tier-first count implementation. Appendix C gives the full paired intervals, block-size curves, and row-order analyses.

To test a trainable nonlinear policy, we separately froze a Student Dropout comparison using an 801-parameter MLP, masked DisARM, true UGC, and a deterministic baseline that transfers development-selected within-group ranking fractions. On its previously unopened test split, 96H mean regrets are .4381, .4513, .4526, and .3455, respectively. Block-64 reduces regret by 2.93%/3.20% versus the two stochastic estimator baselines, with positive paired eight-seed intervals. Ranking wins in every seed: this task supports an estimator gain within randomized MLPs, while its simple structured policy is better overall. Appendix Table 9 reports both budgets.

### 6.3 Mechanism and operating cost

Exhaustive count enumerations achieve maximum mean error $1.08 \times 10^{-15}$; the tested covariance gaps have minimum eigenvalue $-2.73 \times 10^{-17}$, at numerical precision. A separate unequal-partition enumeration verifies the joint theorem to $1.67 \times 10^{-16}$ mean error. At the initial Waterbirds development state, common refinement reduces vectorized gradient covariance trace by 76.70% versus masked DisARM; objective-wise conditioning reduces the three fixed scalarizations by 77.82–78.93%. The weighted $A = 4$ UCI Credit enumeration verifies both additive PSD gaps to $1.55 \times 10^{-15}$ mean error. At the weighted development state, additive conditioning removes 59.4–60.7% of the parameter-gradient covariance trace remaining after matched exact tier-category conditioning.

At $n = 3000, b = 64$ on an RTX 4090, a complete local Block update takes 58.28 ms and 262.5 MiB incremental peak allocation, versus 0.47 ms for DisARM. Equation (19) puts the single-$H$ latency at 288.82 ms for the extra Block computation to be 10% of DisARM end-to-end time on that setup. The separate Student run has median 96H wall times of 55.37 s for Block and 1.02 s for masked DisARM, yielding a 5.64 s/$H$ threshold under the same hypothetical criterion. These are hardware-specific algebraic thresholds, not measured external-oracle latencies. A deployment can measure its own $H$ latency, apply Equation (19), and compare available structured policies before choosing a gradient estimator. Weighted additive trajectories take 7.3–9.3 times the local runtime of exact count fallbacks in the evaluated CUDA implementation. Appendix C gives the scaling and precision audits.

## 7 LIMITATIONS

The method targets independent Bernoulli decisions and metrics with a declared count or bounded weighted-total interface. Its variance order requires the aggregate contract: Swap tests can detect violations, while certification needs an evaluator specification or exhaustive verification; Proposition 1 gives the bias under misspecification. Weighted conditioning has pseudo-polynomial cost in the weight range, and common refinement trades marginal reduction for joint covariance control. The theorem orders gradient covariance; finite-budget regret is measured empirically.

The primary study uses four fixed tasks, one holdout split per task, 24H/96H budgets, and 32-dimensional residuals over frozen representations. The separate Student test trains every parameter of a small MLP; its deterministic group-ranking arm outperforms the gradient estimators. Weighted optimization is evaluated on development data; at Adult 96H, Quotient-FD has the lowest development mean among the compared methods. The results document low-query estimator effects under this interface and do not establish task-population generalization or superiority to every structured policy class.

## 8 CONCLUSION

In vector black-box optimization, a declared aggregate interface identifies decision vectors that an evaluator cannot distinguish. Conditioning on those aggregates turns this equivalence into a lower-variance gradient estimate without additional oracle queries. Block-OSRB does so exactly for independent Bernoulli decisions: it remains unbiased for the unsmoothed objective and follows a nested PSD covariance path as block size grows. Common refinement controls the joint output gradient, while weighted totals remove variance beyond exact category counts. Across four fixed-representation tasks and a separate trainable-MLP test, this structure improves low-query regret against the compared stochastic estimators; on Student Dropout, a structured ranking policy performs better still. The weighted development study shows a further gain at higher local cost. Aggregate-aware conditioning is most useful when oracle calls dominate that cost.

## AI ASSISTANCE

Generative AI assisted with research ideation and proof development, experiment design and analysis, code, literature discovery, and drafting and polishing manuscript text and figures. The authors checked the final claims, results, citations, and artifacts and take responsibility for the paper.

### REPRODUCIBILITY STATEMENT

The appendix contains complete proofs, implementation details, protocols, query accounting, and additional results. The supplementary ZIP contains reference estimator code and a compact, split-labeled summary of the reported numerical results. The raw datasets are available from their cited sources.

## REFERENCES

Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996. URL `https://doi.org/10.24432/C5XW20`.

Lennert De Smet, Emanuele Sansone, and Pedro Zuidberg Dos Martires. Differentiable sampling of categorical distributions using the CatLog-derivative trick. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Aleksandar Dimitriev and Mingyuan Zhou. ARMS: Antithetic-REINFORCE-multi-sample gradient for binary variables. In *International Conference on Machine Learning*, pp. 2717–2727, 2021.

Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring adult: New datasets for fair machine learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Zhe Dong, Andriy Mnih, and George Tucker. DisARM: An antithetic gradient estimator for binary latent variables. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Zhe Dong, Andriy Mnih, and George Tucker. Coupled gradient estimators for discrete latent variables. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Gaurush Hiranandani, Jatin Mathur, Harikrishna Narasimhan, Mahdi Milani Fard, and Sanmi Koyejo. Optimizing black-box metrics with iterative example weighting. In *International Conference on Machine Learning*, pp. 4239–4249, 2021.

Jinyang Jiang, Zeliang Zhang, Chenliang Xu, Zhaofei Yu, and Yijie Peng. One forward is enough for neural network training via likelihood ratio method. In *International Conference on Learning Representations*, 2024.

Qijia Jiang, Olaoluwa Adigun, Harikrishna Narasimhan, Mahdi Milani Fard, and Maya Gupta. Optimizing black-box metrics with adaptive surrogates. In *International Conference on Machine Learning*, pp. 4784–4793, 2020.

Wouter Kool, Herke van Hoof, and Max Welling. Estimating gradients for discrete random variables by sampling without replacement. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rklEj2EFvB.

Russell Z. Kunes, Mingzhang Yin, Max Land, Doron Haviv, Dana Pe'er, and Simon Tavaré. Gradient estimation for binary latent variables via gradient variance clipping. In *AAAI Conference on Artificial Intelligence*, volume 37, pp. 8405–8412, 2023.

Runjing Liu, Jeffrey Regier, Nilesh Tripuraneni, Michael I. Jordan, and Jon McAuliffe. Rao-blackwellized stochastic gradients for discrete distributions. In *International Conference on Machine Learning*, pp. 4023–4031, 2019.

Shaocong Ma and Heng Huang. On the optimal construction of unbiased gradient estimators for zeroth-order optimization. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/7f7eade8b69c853e3137cab80df3ccf6-Abstract-Conference.html.

Valentim Realinho, Mónica Vieira Martins, Jorge Machado, and Luís Baptista. Predict students' dropout and academic success. UCI Machine Learning Repository, 2021. URL https://doi.org/10.24432/C5MC89.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*, 2020.

Jiaxin Shi, Yuhao Zhou, Jessica Hwang, Michalis K. Titsias, and Lester W. Mackey. Gradient estimation with discrete stein operators. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25829–25841, 2022.

Thanawat Sornwanee. Unbiased single-queried gradient for combinatorial objective. *arXiv preprint arXiv:2602.05119*, 2026. URL https://arxiv.org/abs/2602.05119.

James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

Michalis Titsias and Miguel Lázaro-Gredilla. Local expectation gradients for black box variational inference. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Klas Wijk, Ricardo Vinuesa, and Hossein Azizpour. SFESS: Score function estimators for $k$-subset sampling. In *International Conference on Learning Representations*, 2025.

I-Cheng Yeh. Default of credit card clients. UCI Machine Learning Repository, 2009. URL https://doi.org/10.24432/C55S3H.

Table 4: Estimator choice under the public-interface contract. Every Block variant retains two vector-oracle calls; the distinctions concern guarantees and local computation.

| Choice | Use when | Guarantee and operating scope |
|---|---|---|
| DisARM parent | no verified aggregate contract is available | general unbiased estimation without structural conditioning |
| Objective-wise Block | objectives have different public partitions and marginal efficiency matters | unbiasedness and columnwise PSD dominance |
| Common-refinement Block | a joint vector or fixed-scalarization guarantee matters and intersection cells remain useful | joint PSD dominance over masked DisARM; marginal efficiency follows the shared-cell resolution |
| Uniform MP4 | row order has no natural meaning and four local DPs fit the compute budget | permutation equivariance in distribution and PSD upper bounds; incremental gain follows member diversity |
| Additive Block | the oracle exposes bounded normalized integer-weighted totals | exact additive PSD ladder with pseudo-polynomial support cost in $A$ |

Minhao Zou, Tao Ren, Jinyang Jiang, Rui Tao, Zehao Li, Jiale Fu, Hui Shao, Xianhua Liu, and Yijie Peng. OVLR: Efficient, scalable, and robust training via output-level variance-reduced likelihood ratio. In *International Conference on Machine Learning*, 2026.

## A Estimator operating guide

Table 4 separates proved guarantees from operating choices. Objective-wise conditioning is the default when low marginal variance per output is the priority. Common refinement is appropriate when a joint PSD statement for the vectorized gradient or a fixed cross-objective linear map is required. The tested MP4 uses $K = 4$ and uniform weights fixed before sampling; theory permits any fixed convex weights. Member correlation determines the incremental gain from increasing $K$.

## B Detailed relation to prior work

DisARM conditions ARM's antithetic estimator and supplies the two-query parent used here (Dong et al., 2020). ARMS and other coupled estimators vary the binary sampling scheme (Dimitriev & Zhou, 2021; Dong et al., 2021). UGC switches between independent bit-flip and DisARM regimes at the published $1/(2n)$ threshold (Kunes et al., 2023); we evaluate that switch on independent seeds. Our "ordinary antithetic LR" arm averages two score estimators and is reported separately from true UGC.

Rao–Blackwell and local-expectation methods integrate discrete randomness analytically (Liu et al., 2019; Titsias & Lázaro-Gredilla, 2015). IndeCateR sums categorical states coordinatewise (De Smet et al., 2023), while the unordered-set estimator sums retained distinct outcomes after conditioning on a sample set (Kool et al., 2020). Discrete Stein control variates learn a correction to leave-one-out REINFORCE (Shi et al., 2022); SFESS computes scores and control variates for a fixed-cardinality sampler (Wijk et al., 2025). Block-OSRB conditions a fixed DisARM pair after its public aggregate counts are observed. Those counts determine both feedback values, leaving analytic row coefficients to average with the same two black-box calls. This output-statistic query collapse combines the established Rao–Blackwell principle with generating-function coefficient extraction.

ULR and OVLR apply LR perturbations to cached model outputs and use a VJP for general gradient-agnostic losses (Jiang et al., 2024; Zou et al., 2026). Block-OSRB operates at the binary output as well, with the aggregate interface supporting an exact conditional expectation for the unsmoothed Bernoulli objective. ESG constructs unbiased single-query gradients for general scalar product-Bernoulli objectives (Sornwanee, 2026); Appendix G records a scaling audit of its three published raw variants. Black-box metric optimization also uses adaptive surrogates and example weighting (Jiang et al., 2020; Hiranandani et al., 2021). SPSA targets scalar parameter-space oracles (Spall, 1992), and randomized telescoping gives unbiased one- to four-evaluation gradients for smooth continuous

objectives (Ma & Huang, 2025). These approaches use different oracle inputs and gradient targets; our covariance comparisons keep the Bernoulli objective and its DisARM pair fixed.

# C Detailed experimental evidence

## C.1 Design

We study binary classifiers on Adult, ACSIncome, Waterbirds, and UCI Credit. ACSIncome comes from Folktables, which documents why repeated reliance on the older Adult benchmark is problematic (Ding et al., 2021); UCI Credit adds a prospectively selected financial-risk task. Waterbirds changes the modality and group semantics: a frozen ImageNet ResNet-18 representation predicts land versus water birds under spurious land/water backgrounds (Sagawa et al., 2020). Each black-box call returns three count-derived objectives spanning utility, balanced performance, and equality-of-opportunity style fairness (Hardt et al., 2016). We optimize three fixed preference vectors. Their three endpoints form a minimization front; the primary metric is $1 - \mathrm{HV}$ with reference point $(1, 1, 1)$, so lower is better.

The experimental estimand is the effect of replacing a two-query LR gradient estimator while holding the predictive representation, optimizer, query budget, and oracle fixed. Every task therefore freezes a train-only base predictor and optimizes a 32-dimensional linear residual on its logits. Adult, ACSIncome, and UCI Credit use an intercept, standardized base logit, centered protected-group coordinate, and 29 fixed projected features. Waterbirds uses an intercept, standardized base logit, and 30 fixed projected ResNet coordinates; the residual uses projected image features while protected background labels define evaluation groups. The oracle vector is $[1 - \mathrm{macroF1}, |\mathrm{DP}|, |\mathrm{EO}|]$, all three components being functions of public confusion counts.

The paper prints the formulas for scientific audit. Optimization sends a complete prediction vector to a separate oracle process and receives the three objective values; the estimator separately receives the declared row partitions. This process boundary implements the stated numerical-oracle access model.

Every Block-OSRB and DisARM update receives the same two full-vector oracle evaluations. Query budgets are expressed in $H$ calls; thus 24H and 96H mean 12 and 48 such updates for these methods. These low budgets target expensive remote or policy evaluators. The columns are hard query ceilings: Quotient-FD uses 17–24H/88–96H, FW-style 18H/90H, and true UGC 22–24H/94–96H, with every trajectory inside its cap. Adult and ACSIncome broad comparisons use five and ten seeds; true-UGC comparisons use eight new seeds. Waterbirds uses eight seeds for all eight methods after a frozen acquisition, representation, and covariance preflight. All tasks use common initializations within a dataset, a fixed development oracle, and independent common-query evaluation. We compare Block-OSRB-64 with DisARM, true UGC, ordinary antithetic LR, parameter SPSA, Quotient-FD, and budget-matched AS-style and FW-style local-probing adaptations. True UGC's independent bit-flip branch makes an update cost 2H or 4H. Methods, parameters, seeds, and success gates were frozen before the one-time final evaluations on Adult official test, 10,000 sealed ACSIncome rows, and Waterbirds official test. Appendix H records the chronology and actual utilization.

## C.2 Exactness and scaling

Exhaustive enumeration of six $n = 8$ settings gives maximum absolute mean error $1.08 \times 10^{-15}$. The smallest eigenvalue in all tested member-versus-parent PSD differences is $-2.73 \times 10^{-17}$, at numerical precision. A separate 243-state enumeration with two unequal objective partitions verifies Theorem 3: maximum conditional-expectation error is $8.88 \times 10^{-16}$, mean error is $1.67 \times 10^{-16}$, the minimum joint covariance-difference eigenvalue is $-3.25 \times 10^{-16}$, and covariance trace falls by 15.8%. At the initial Waterbirds development state, common-refinement Block-OSRB-64 lowers the vectorized $3 \times 32$ parameter-gradient covariance trace by 76.70% relative to masked DisARM. Reductions after the three fixed scalarizations are 76.18–76.65%, with paired-bootstrap lower bounds of 72.09–74.18%; objective-wise conditioning lowers them by 77.82–78.93%, quantifying the marginal-variance price of the shared statistic. All estimators reuse the same 512 development calls and two $H$ calls per pair. A frozen RTX 4090 audit over $n \in \{256, 1000, 3000, 10000\}$ and $b \in \{16, 32, 64\}$ completes without OOM. At $n = 3000, b = 64$, a full local Block update takes 58.28 ms versus 0.47 ms for DisARM and 262.5 MiB incremental peak allocation; at $n = 10000, b = 64$ it takes 75.49 ms and 889.2 MiB. By Equation (19), a single-H latency of 115.39, 288.82, or 577.88 ms

Table 5: Tier-weighted UCI Credit development hypervolume regret (mean ± seed standard deviation; lower is better). The matched tier-count arm uses exactly the same physical blocks as additive conditioning. Additive's initial covariance-trace reduction versus that arm is 59.4–60.7%.

| Method | 24H | 96H |
|---|---|---|
| Masked DisARM | .4342±.0096 | .4170±.0123 |
| Tier-first count Block-64 | .4103±.0068 | .3892±.0070 |
| Matched tier-count Block-64 | .4180±.0122 | .3966±.0097 |
| Additive Block-64 | **.4091±.0072** | **.3796±.0034** |

at $n = 3000, b = 64$ limits Block's additional local cost to 25%, 10%, or 5% of DisARM end-to-end time. These measurements identify expensive count-oracle settings as the target runtime regime.

The additive extension is evaluated on UCI Credit development data. An exhaustive 1,024-state unequal-weight test gives maximum mean error $1.55 \times 10^{-15}$ and verifies both PSD gaps in Equation (13). Four public credit-exposure tiers define ordinal weights $1, 2, 3, 4$ ($A = 4$) and an exposure-policy metric. With identical physical blocks, 1,024 antithetic pairs, saved feedback, and zero new oracle calls, additive conditioning reduces parameter-gradient covariance trace by 60.72%, 60.02%, and 59.45% for the utility, balanced, and fairness preferences relative to exact tier-category counts. The pooled paired-bootstrap interval for the count-minus-additive squared-deviation difference is $[0.08170, 0.09825]$. This initial-state diagnostic validates the mechanism on a natural bounded-weight contract.

A separately frozen optimization test reuses the weighted development oracle with eight new seeds, three fixed preferences, and the 24H/96H budgets. The precision policy checks strict interiority after casting and uses float64 coefficients for a block when float32 reaches a boundary. Independent CUDA regressions give maximum float32-to-float64 coefficient errors of $1.50 \times 10^{-7}$ and $1.37 \times 10^{-7}$ on GTX 1650 and RTX 4090, respectively, for batch 8, $b = 64$, and $A = 4$; Appendix F reports the scaling audit and stopped precursor run. Table 5 reports the optimization result. Additive Block-64 lowers regret relative to masked DisARM by 5.77% and 8.97%, winning all eight seeds at both budgets; the pooled absolute-improvement interval is $[0.02596, 0.03735]$. Relative to the matched-block tier-count fallback, it lowers regret by 2.11%/4.28% with interval $[0.00789, 0.01827]$. Relative to the stronger tier-first count implementation, the reductions are 0.29%/2.46%, with pooled interval $[0.00077, 0.01053]$ and mixed seedwise direction at 24H. Additive's 24 trajectories take 1,860.7 seconds on the RTX 4090, versus 200.9 and 254.1 seconds for the two count fallbacks. All 96 trajectories have exact ledgers and separate oracle processes.

### C.3 Optimization results

Table 6 reports mean regret and seed standard deviation. Block-OSRB improves DisARM on both datasets and budgets: 6.08%/8.19% on Adult and 6.01%/5.74% on ACSIncome at 24H/96H. On ACSIncome it also improves the per-seed best strong-baseline regret by 5.42% at 24H and 2.17% at 96H; the paired 24H improvement interval is $[0.01196, 0.02115]$. At Adult 96H, Quotient-FD records the lowest mean among the compared methods.

On Waterbirds, Block-OSRB improves DisARM regret by 6.59% at 24H and 17.78% at 96H, winning all eight seeds at both budgets. Even after selecting the best of all seven non-Block methods separately for every seed and budget, its mean regret is lower by 5.34% and 12.67%. The 16 paired DisARM differences average $0.07342$ with a seed-block bootstrap 95% interval $[0.06878, 0.07791]$.

A separately frozen development audit completes the relevance-masked DisARM comparison on Adult and ACSIncome using the original R123 settings. At 96H, objective-wise Block-64 lowers mean regret by 3.40% and 3.93%, respectively; the equal-task-weight paired absolute improvement has a seed-block interval of $[0.00717, 0.01666]$. Adult 24H is nearly tied (0.20% lower Block mean). Appendix Table 8 gives all four cells, including ordinary DisARM and seed variation.

Removing publicly irrelevant rows from DisARM strengthens the parent once the count partition is known. A separately frozen Waterbirds development audit compares both Block variants with this masked parent. Objective-wise Block-64 reduces mean regret by 4.57% and 13.04% at 24H and 96H; the theorem-covered common-refinement variant reduces it by 3.82% and 12.24%, winning all

Table 6: Hypervolume regret under hard query ceilings (mean ± seed standard deviation; lower is better). Bold marks the lowest mean in each column. "Ordinary LR" denotes the two-score antithetic arm; true UGC is reported separately. Appendix 13 reports actual calls for variable-cost methods.

| | Adult | | ACSIncome | |
|---|---|---|---|---|
| Method | 24H | 96H | 24H | 96H |
| Block-OSRB-64 | **.3883±.0349** | .3607±.0384 | **.2925±.0082** | **.2685±.0119** |
| DisARM | .4135±.0335 | .3929±.0204 | .3112±.0152 | .2849±.0105 |
| Ordinary LR | .4835±.0637 | .4475±.0910 | .4282±.1076 | .4156±.1019 |
| Parameter SPSA | .4031±.0110 | .3799±.0134 | .3173±.0122 | .2852±.0081 |
| Quotient-FD | .3958±.0121 | **.3509±.0085** | .3158±.0043 | .2794±.0108 |
| AS-style (8H) | .3980±.0057 | .3815±.0156 | .3253±.0157 | .3057±.0097 |
| FW-style (9H) | .4017±.0065 | .3571±.0087 | .3148±.0049 | .2908±.0090 |

Table 7: Eight-seed development regret under hard query ceilings (top; mean ± standard deviation, lower is better) and one-time final-holdout regret reduction (bottom; %, higher is better).

| | Adult | | ACSIncome | | Waterbirds | |
|---|---|---|---|---|---|---|
| Method | 24H | 96H | 24H | 96H | 24H | 96H |
| Block-OSRB-64 | **.3922±.0153** | **.3610±.0341** | **.3042±.0220** | **.2691±.0057** | **.6053±.0037** | **.4815±.0083** |
| True UGC | .4127±.0315 | .3712±.0223 | .3503±.0679 | .2905±.0100 | .6574±.0089 | .6151±.0140 |
| DisARM | .4102±.0337 | .3705±.0203 | .3462±.0668 | .2918±.0105 | .6480±.0113 | .5856±.0091 |
| Holdout Δ vs. DisARM (%) | 3.07 | 1.95 | 10.24 | 6.76 | 7.21 | 19.66 |
| Holdout Δ vs. true UGC (%) | 3.32 | 2.21 | 11.52 | 6.56 | 8.53 | 23.82 |

eight seeds at both budgets. Common refinement is 0.79% and 0.92% above objective-wise Block-64. Its 16 paired absolute improvements over masked DisARM average $0.04601$, with frozen bootstrap interval $[0.03874, 0.05348]$. Appendix Table 14 reports the development values.

UCI Credit was prospectively fixed as the fourth task before Block-OSRB trajectories were observed on its new stratified split. On the 3,000-row development oracle, objective Block-64 reduces masked-DisARM regret by 4.46%/6.85% at 24H/96H, and common Block-64 reduces it by 3.82%/6.60%. Common refinement remains within 0.67% of the objective-wise variant at both budgets. Its 16 paired improvements average $0.02134$, with frozen interval $[0.01498, 0.02824]$. All 72 trajectories have exact ledgers and separate oracle processes. True UGC and the holdout gate were then frozen before the one-time evaluation on 10,000 sealed rows. On holdout, objective Block-64 reduces regret by 4.51%/6.33% versus masked DisARM and 5.12%/6.56% versus true UGC; common Block-64 reduces it by 3.25%/6.20% and 3.87%/6.44%, respectively. Common's paired intervals are $[0.01188, 0.02624]$ versus masked DisARM and $[0.01252, 0.02737]$ versus true UGC. Appendix Table 15 gives the absolute values.

### C.4 Published UGC comparison

We evaluate published UGC on eight new seeds per dataset. Its coordinate switch uses an independent uniformly selected bit-flip-1 branch when $\min(p_i, 1 - p_i) < 1/(2n)$ and an independent DisARM pair for the remaining coordinates. A complete update costs 2H or 4H; reserving 4H before an update yields checkpoints at 22H/94H when the final branch is indivisible. Exhaustive $n = 8$ enumeration gives maximum mean error $2.22 \times 10^{-16}$ and exercises both query branches.

Block-OSRB improves the true-UGC mean in all six cells. The Adult/ACSIncome reductions are 4.98%, 2.76%, 13.16%, and 7.36%; their 32 paired differences average $0.02456$ with 95% interval $[0.01250, 0.03824]$. Waterbirds reductions are 7.92% and 21.71%, with 8/8 seed wins at both budgets; its 16 paired differences average $0.09282$ with interval $[0.08703, 0.09924]$. Waterbirds is boundary-rich: 379/759 UGC updates use the 4H branch and boundary counts range from 689 to 727 of 1199 rows, so the visual result exercises both published UGC regimes extensively. Adult records 6/8 and 4/8 seed wins at 24H and 96H.

Table 8: Adult and ACSIncome development hypervolume regret for objective-wise Block-64 and both DisARM parents (mean ± seed standard deviation; lower is better). R316 adds only the relevance-masked DisARM trajectories; the Block and ordinary DisARM rows are unchanged R123 checkpoints.

| | Adult | | ACSIncome | |
|---|---|---|---|---|
| Method | 24H | 96H | 24H | 96H |
| Block-OSRB-64 | **.3922±.0153** | **.3610±.0341** | **.3042±.0220** | **.2691±.0057** |
| Masked DisARM | .3930±.0064 | .3737±.0262 | .3220±.0257 | .2801±.0060 |
| Ordinary DisARM | .4102±.0337 | .3705±.0203 | .3462±.0668 | .2918±.0105 |

### C.5 Relevance-masked parent on Adult and ACSIncome

The additional R316 development comparison keeps the R123 data, three preferences, seeds, antithetic streams, optimizer, and 24H/96H ceilings fixed. It changes only DisARM's output coefficients on rows publicly irrelevant to an objective, using the same two complete-vector feedback calls per update. The R123 development scores reproduce exactly before the new arm is scored. Table 8 reports the complete matched comparison.

At 96H, the paired masked-minus-Block differences average $0.01270$ on Adult and $0.01100$ on ACSIncome, with task-specific seed-block 95% intervals $[0.00501, 0.02096]$ and $[0.00617, 0.01628]$. Their equal-task-weight mean is $0.01185$, interval $[0.00717, 0.01666]$. At 24H, Adult's difference is $0.00079$, interval $[-0.00665, 0.00760]$; ACSIncome's is $0.01781$, interval $[0.01219, 0.02338]$. These are development comparisons on previously studied tasks. All 48 new trajectories used exactly 96 training oracle calls, for 4,608 total; the 96 checkpoints used 49,152 development-evaluation calls. No final holdout was rescored for R316.

The saved three-arm means clarify the two gaps: on Adult at 24H, ordinary-to-masked DisARM accounts for $0.01721$ of the $0.01800$ ordinary-to-Block regret difference, while masked-to-Block is $0.00079$. At 96H, masked DisARM is $0.00321$ worse than ordinary DisARM, and Block improves on masked DisARM by $0.01270$. These are arithmetic contrasts between separately optimized trajectories, not a causal decomposition.

All eight dataset–budget holdout directions per comparator—16 across the two matched parents—are positive. On the first three tasks, the 48 paired dataset–seed–budget differences average $0.03777$ against DisARM, with dataset-seed block-bootstrap interval $[0.02438, 0.05119]$; against true UGC they average $0.04526$, with interval $[0.02903, 0.06165]$. Those original intervals describe seed variation conditional on these three fixed tasks. The separately preregistered UCI Credit intervals are $[0.01188, 0.02624]$ and $[0.01252, 0.02737]$. These intervals summarize variation over the fixed observed task–seed blocks.

### C.6 Trainable nonlinear policy and a structured comparator

Student Dropout (Realinho et al., 2021) supplies a separately frozen transfer test. A previously prepared split provides 2,500 pretraining rows, 1,400 development rows, and 524 test rows; the test split was unopened when the R356 protocol and training code were fixed. The 23 enrollment-time numeric inputs exclude Gender and later curricular outcomes. All 801 parameters of the pretrained $23 \to 32$ ReLU MLP are updated by the black-box gradient estimators; Gender defines public count partitions. The same eight seeds and three preferences are used for objective-wise Block-64, relevance-masked DisARM, and true UGC at 24H/96H. The fourth arm ranks test rows by the frozen pretrained score within each Gender group and transfers two group-specific selection fractions selected from 24 or 96 distinct development queries. It is a different deterministic policy class, included to test task-level competitiveness. Every stochastic checkpoint uses 256 common antithetic test pairs; the structural checkpoint needs one deterministic test call.

Block improves on both stochastic estimators in all eight seeds at both budgets. At 96H, its absolute masked-DisARM and true-UGC improvements are $.01322$ and $.01448$, with predeclared paired bootstrap intervals $[.01106, .01577]$ and $[.01250, .01667]$; the predeclared estimator-success gate therefore passes. At 24H the corresponding relative improvements are 0.73% and 0.89%. Group

Table 9: Student Dropout one-time test hypervolume regret, mean ± seed standard deviation over eight seeds (lower is better). The group-ranking arm transfers development-selected fractions and is not a gradient estimator.

| Method | 24H | 96H |
|---|---|---|
| Block-OSRB-64 | $.4523 \pm .0027$ | $.4381 \pm .0029$ |
| Masked DisARM | $.4557 \pm .0017$ | $.4513 \pm .0039$ |
| True UGC | $.4564 \pm .0014$ | $.4526 \pm .0033$ |
| Group ranking | $\mathbf{.3310 \pm .0107}$ | $\mathbf{.3455 \pm .0195}$ |

ranking beats Block in every seed and has lower mean regret by .12131 at 24H and .09264 at 96H. Its 96H test mean is slightly worse than its own 24H mean, showing that additional development selection did not help on this split. The training ledger contains 9,168 oracle calls across the 96 runs; Block, masked DisARM, and group ranking use 24/96H per checkpoint, while true UGC uses 22/94H. Another 73,776 test responses score the frozen checkpoints. Median full 96H trajectory times on the local inexpensive oracle are 55.37 s for Block, 1.02 s for masked DisARM, 1.12 s for UGC, and 0.53 s for group ranking.

A saved-response sensitivity check resamples the same 256 antithetic pair indices jointly across all stochastic checkpoints. At 24H, the observed absolute Block improvements over masked DisARM and true UGC are .00334 and .00407; 2,000 common-panel resamples give intervals $[.00292, .00383]$ and $[.00358, .00456]$. At 96H the corresponding intervals are $[.01217, .01424]$ and $[.01338, .01541]$. These post-hoc intervals quantify Monte Carlo evaluation variation conditional on the fixed test rows and trained policies, not variation over new students. The same saved-run audit pairs each Block trajectory with its masked-DisARM counterpart. Applying Equation (19) to their full 96H elapsed times gives a median hypothetical additional single-$H$ latency of 5.64 s for a 10% overhead target on the Student setup, compared with 288.82 ms from the separate RTX 4090 update benchmark. Neither value is an observed external-service latency.

The one-time test scorer's oracle subprocess lost its Windows pipe after 67,572 logged responses, including 500 responses in the interrupted checkpoint. A documented continuation verified the hashes of those 500 predictions, resumed the frozen query sequence, and independently audited all 73,776 recorded responses and 64 fronts. The failed send produced no response; one unobserved request cannot be excluded from the physical query count. No model, split, panel, or gate changed after test access. Earlier development-only trainable-policy studies give complementary context: the R270 Adult 3,521-parameter MLP showed 96H additive Block regret .38052 versus masked DisARM's .39932, while the R176 Waterbirds 4.7-million-parameter trainable residual experiment failed its gate, with 96H count Block regret .62639 versus DisARM's .53108. These three settings do not support a monotone claim that estimator gains grow with model size.

### C.7 Mechanism and block size

Figure 2 connects the covariance theorem to optimizer behavior. At the shared initial state, $b = 64$ reduces natural-parameter gradient covariance trace relative to $b = 1$ by 86.5–88.5% across Adult preferences and 70.1–83.2% across ACSIncome preferences. The reduction grows with block size as predicted by Theorem 4. Relative to $b = 1$, $b = 64$ lowers regret by 4.80%/11.86% on Adult and 9.70%/4.77% on ACSIncome at 24H/96H; the preregistered Adult 24H mechanism threshold was 5%.

Across $b \in \{16, 32, 64\}$, $b = 64$ has the lowest mean regret in seven of eight dataset–budget cells; at ACSIncome 8H it is 1.62% above $b = 32$. We therefore use $b = 64$ as the practical point for the tested scale range.

### C.8 Order sensitivity and multi-partition averaging

Physical, fixed-random, and base-logit row orders agree within 5% of the best order in seven of eight cells. At ACSIncome 8H, physical order is 6.76% above logit order; all 24H–96H differences are within 3%. This order effect motivates the permutation-symmetric extension.

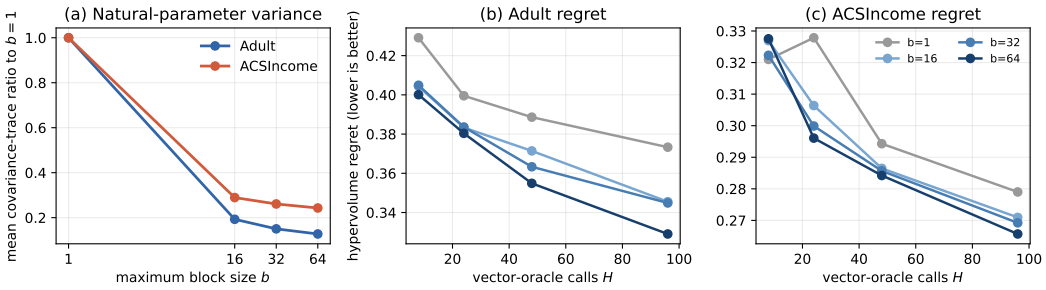



Figure 2: Block-size mechanism audit. Left: covariance-trace ratio relative to singleton blocks for three preferences. Middle/right: regret over oracle budget. Larger aligned blocks sharply reduce one-step variance, while finite-budget optimizer gains are smaller and task-dependent.

A prospectively frozen four-random-partition estimator (MP4) remains within 3% of physical Block-OSRB-64 in every cell and is better in six of eight. Relative to singleton blocks, it reduces covariance trace by 88.21% on Adult and 74.33% on ACSIncome on average. Relative to one random partition, the extra reduction is 10.01% on Adult and 3.81% on ACSIncome; pooled 24H and 96H regret are 0.40% and 0.29% higher. MP4 supplies permutation equivariance, and the measured incremental gain tracks member correlation.

## D Proof Details

For a fixed objective column, write $S$ for all paired block counts. By the count contract, $\Delta H_m$ is $S$-measurable. Linearity of conditional expectation gives

$$\mathbb{E}[G_m^{\mathrm{D}} \mid S] = \mathbb{E}\left[\Delta H_m \sum_i d_i q_i(T_i) \mid S\right] \tag{21}$$

$$= \Delta H_m \sum_i d_i \mathbb{E}[q_i(T_i) \mid S] = G_m^{\mathrm{B}}. \tag{22}$$

The coefficient of $s^{A_g} t^{B_g}$ in $P_g$ is exactly the probability of the observed paired counts. Conditioning on $T_i = (a, b)$ deletes factor $i$, yielding Equation (8). This proves exact computation of the conditional expectation.

The tower property gives $\mathbb{E}G_m^{\mathrm{B}} = \mathbb{E}G_m^{\mathrm{D}} = g_m^\star$. Total covariance gives

$$\mathrm{Cov}(G_m^{\mathrm{D}}) = \mathrm{Cov}(\mathbb{E}[G_m^{\mathrm{D}} \mid S]) + \mathbb{E}[\mathrm{Cov}(G_m^{\mathrm{D}} \mid S)], \tag{23}$$

which proves Theorem 1. For nested sigma-algebras $\mathcal{S}_j \supseteq \mathcal{S}_{j+1}$, apply the same identity to $\mathbb{E}[G_m^{\mathrm{D}} \mid \mathcal{S}_j]$ conditioned on $\mathcal{S}_{j+1}$. Congruence proves the fixed-linear-map statement.

For Corollary 1, let $R = G_m^{\mathrm{D}} - G_m^{\mathrm{OS}}$. For every square-integrable $S$-measurable $Z$, $\mathbb{E}[R^\top (G_m^{\mathrm{OS}} - Z)] = \mathbb{E}[\mathbb{E}[R \mid S]^\top (G_m^{\mathrm{OS}} - Z)] = 0$. Expanding $\|R + (G_m^{\mathrm{OS}} - Z)\|^2$ proves Equation (10); equality with the minimum occurs only when $Z = G_m^{\mathrm{OS}}$ almost surely.

For Proposition 1, the DisARM parent remains unbiased without a count contract. Conditional expansion gives

$$\mathbb{E}[\widehat{G}_m - G_m^{\mathrm{D}} \mid S] = \mathbb{E}[\Delta H_m \mid S]\mathbb{E}[R_m \mid S] - \mathbb{E}[\Delta H_m R_m \mid S] \tag{24}$$

$$= -\mathrm{Cov}(\Delta H_m, R_m \mid S), \tag{25}$$

which proves Equation (14) after taking expectations. Conditional vector Cauchy–Schwarz bounds the covariance norm by $\sqrt{\mathrm{Var}(\Delta H_m \mid S)\,\mathrm{tr}\,\mathrm{Cov}(R_m \mid S)}$; Cauchy–Schwarz over $S$ then proves Equation (15).

For Theorem 3, each common cell lies inside one cell of every objective partition (including the irrelevant cell). Its paired counts therefore reconstruct every objective's public counts, making the

full feedback difference vector measurable with respect to $S_\cap$. Applying the preceding conditional-expectation identity simultaneously to all entries gives $G^\cap = \mathbb{E}[\widetilde{G}^{\mathrm{D}} \mid S_\cap]$. Total covariance applied to the vectorization proves Equation (16); for fixed $L$, the covariance difference becomes $L(\mathrm{Cov}(\mathrm{vec}\,\widetilde{G}^{\mathrm{D}}) - \mathrm{Cov}(\mathrm{vec}\,G^\cap))L^\top \succeq 0$. Masking preserves the parent mean because $H_m$ is independent of rows outside $I_m$ and their score terms have zero mean.

For Corollary 2, condition on $\theta_t$ and write $G_t = \nabla f(\theta_t) + \xi_t$, where $\mathbb{E}[\xi_t \mid \theta_t] = 0$. Smoothness gives

$$\mathbb{E}[f(\theta_{t+1}) \mid \theta_t] \le f(\theta_t) - \eta\|\nabla f(\theta_t)\|^2 + \frac{L\eta^2}{2}\mathbb{E}[\|G_t\|^2 \mid \theta_t] \tag{26}$$

$$\le f(\theta_t) - \frac{\eta}{2}\|\nabla f(\theta_t)\|^2 + \frac{L\eta^2}{2}\sigma_{\mathrm{B}}^2, \tag{27}$$

where $\eta \le 1/L$ and $\mathbb{E}\|\xi_t\|^2 = \mathrm{tr}\,\mathrm{Cov}(G_t \mid \theta_t)$. Taking expectations, summing, and using the lower bound proves Equation (17). The pointwise PSD order survives the deterministic scalarization and model VJP at each $\theta$, so taking a supremum over $\theta$ yields $\sigma_{\mathrm{B}}^2 \le \sigma_{\mathrm{D}}^2$ whenever the corresponding uniform parent bound is finite.

For Theorem 5, fix objective $m$; unbiasedness follows by linearity. For centered $Y_k = G_m^{(k)} - \mathbb{E}G_m^{(k)}$,

$$\sum_k w_k\mathbb{E}[Y_kY_k^\top] - \mathbb{E}[(\sum_k w_kY_k)(\sum_\ell w_\ell Y_\ell)^\top] = \frac{1}{2}\sum_{k,\ell} w_kw_\ell\mathbb{E}[(Y_k - Y_\ell)(Y_k - Y_\ell)^\top] \succeq 0. \tag{28}$$

Each objective-$m$ member is a conditional expectation of the same masked parent $\widetilde{G}_m^{\mathrm{D}}$, so its covariance is PSD-bounded by that parent; their convex combination is as well. For common-refinement members, the same argument applies to the vectorized masked parent by Theorem 3. Finally, i.i.d. continuous random keys induce a uniformly random relative order. Permuting rows only permutes the keys and decisions jointly, proving distributional equivariance.

For the complexity statement, consider one objective and one block of size $r \le b$. Extending the bivariate prefix or suffix table after $j$ Bernoulli pairs touches $O(j^2)$ coefficients, so constructing all prefix and suffix tables costs $\sum_{j=1}^{r} O(j^2) = O(r^3)$. For each of the $r$ rows, the four-state leave-one coefficient needs $O(r^2)$ prefix–suffix products, again $O(r^3)$ per block. If block sizes are $r_k \le b$ and $\sum_k r_k = n$, then $\sum_k r_k^3 \le b^2\sum_k r_k = nb^2$. Repeating over $M$ objective partitions gives $O(Mnb^2)$. The direct implementation stores $O(b^3)$ prefix/suffix coefficients for one active block and releases them before the next; its auxiliary block state is independent of $n$.

For Theorem 2, Equation (12) enumerates the joint law of the two weighted totals: selecting state $(a, b)$ for row $i$ contributes $(w_ia, w_ib)$ to the exponents. Deleting that factor therefore gives each exact conditional coefficient. Let $\mathcal{C}$ be the sigma-algebra generated by paired counts separated by weight category inside the fixed blocks, and let $\mathcal{Z}$ be generated by their paired weighted totals. Then $\mathcal{Z} \subseteq \mathcal{C}$, while the additive contract makes the feedback difference measurable under both. Hence $G^C = \mathbb{E}[G^{\mathrm{D}} \mid \mathcal{C}]$ and $G^Z = \mathbb{E}[G^{\mathrm{D}} \mid \mathcal{Z}] = \mathbb{E}[G^C \mid \mathcal{Z}]$ by the tower property. Two applications of total covariance prove Equation (13) and its stated equality condition. With maximum normalized weight $A$, the paired support has side length $O(bA)$; replacing the count table's $O(b^2)$ support by $O(b^2A^2)$ throughout the same block accounting yields $O(nb^2A^2)$ local time for fixed output dimension.

For the common-refinement variant, the dynamic program is run once for the shared cells rather than once per output. The same bound gives $O(nb^2)$ coefficient work, and writing/scalarizing $M$ output columns costs $O(Mn)$, yielding $O(nb^2 + Mn)$.

## E Implementation recipe

Given interior probabilities, their model Jacobian/VJP closure, public partitions, a fixed block refinement, and a fixed preference: (1) draw i.i.d. uniforms and construct the DisARM antithetic pair; (2) call the complete vector oracle once at each outcome; (3) inside every objective–block, build bivariate prefix and suffix coefficient tables; (4) evaluate Equation (8) for every row at the realized paired counts; (5) multiply each coefficient by the corresponding feedback difference and

preference weight; and (6) sum the columns and apply one VJP. For an additive contract, step 3 uses the weighted exponents in Equation (12), and step 4 reads the realized paired weighted totals after dividing all weights by their common factor. The batched implementation deduplicates repeated total pairs and differentiates the tilted log-normalizer for each unique pair. A multi-partition update repeats steps 3–5 locally for pre-fixed refinements and weights, reusing both queried feedback vectors. A common-refinement update instead constructs one shared refinement, computes one coefficient per row, masks objective-irrelevant rows, and reuses that coefficient across columns. No branch performs an uncharged counterfactual query.

## F Local cost and numerical precision

Table 10 reports the frozen count implementation grid. It measures the complete local update on one RTX 4090 in float64; it excludes the two shared oracle calls. Batching makes time depend strongly on the largest block and only weakly on $n$ over this grid, while memory still grows with the number of blocks. The $n = 3000, b = 64$ incremental peak allocation is 262.5 MiB; at $n = 10000, b = 64$ it is 889.2 MiB. Applying Equation (19) at the latter point gives single-$H$ latency thresholds 149.76, 374.76, and 749.77 ms for 25%, 10%, and 5% extra local cost.

Table 10: Median complete local-update time in milliseconds. Parentheses give the extra time over DisARM at the same $n$. Hardware-specific timings guide deployment measurement rather than define an algorithmic guarantee.

| $n$ | DisARM | $b = 16$ | $b = 32$ | $b = 64$ |
|---|---|---|---|---|
| 256 | 0.47 | 15.59 (15.12) | 36.99 (36.52) | 66.57 (66.10) |
| 1,000 | 0.56 | 22.36 (21.81) | 38.29 (37.73) | 71.55 (70.99) |
| 3,000 | 0.47 | 14.68 (14.21) | 29.53 (29.06) | 58.28 (57.81) |
| 10,000 | 0.49 | 15.54 (15.06) | 35.11 (34.62) | 75.49 (75.00) |

For additive conditioning, a CPU reference sweep over $b \in \{8, 16, 32, 64\}$ and $A \in \{1, 2, 4\}$ produced finite coefficients in all 12 cases. At $b = 64$, median times were 0.0745, 0.1430, and 0.4664 seconds as $A$ increased from 1 to 2 to 4, with realized paired-state upper bounds 4,225, 8,464, and 28,561. Removing a common $10{,}000$-fold weight factor changed no coefficient. Independent batched CUDA checks at $b = 64, A = 4$ matched the NumPy reference to $6.66 \times 10^{-16}$ and gave the precision results in Table 11. The optimization failure reported in the main text arose before this recurrence, when a float32 probability cast reached the boundary. Production code therefore checks cast probabilities for strict interiority and recomputes affected blocks in float64 rather than clipping or adding an epsilon.

Table 11: Independent additive coefficient regressions for batch 8, $b = 64$, $A = 4$. Times are implementation and hardware specific.

| Environment | float64 CUDA–CPU max error | float32–float64 max error | median time |
|---|---|---|---|
| GTX 1650, Torch 2.5.1 | $5.55 \times 10^{-16}$ | $1.50 \times 10^{-7}$ | 0.1973 s |
| RTX 4090, Torch 2.13.0 | $2.84 \times 10^{-16}$ | $1.37 \times 10^{-7}$ | 0.0158 s |

## G Raw ESG scaling audit

ESG's published Arch, Spike, and LongJump constructions form a stochastic value $V = Q(K) \prod_{j=1}^{n} F_j$, then differentiate it pathwise (Sornwanee, 2026). We independently implemented their stated $f$, noise law, and encoding CDF. For a three-bit arbitrary oracle, factorized numerical integration agrees with the enumerated multilinear gradient to maximum absolute error $1.33 \times 10^{-15}$; the signed one-dimensional moment identities agree to $4.44 \times 10^{-16}$. Thus the following scaling check is not based on a failed unbiasedness reproduction.

**Proposition 2** (Product-moment diagnostic)**.** *Suppose the ESG coordinate noises are independent,* $\mathbb{E}F_j = 1$*, and* $D_i = \partial F_i/\partial x_i$*. For the constant oracle* $Q \equiv c$*,*

$$\frac{\mathrm{Var}(V)}{c^2} = \prod_{j=1}^{n} \mathbb{E}[F_j^2] - 1, \tag{29}$$

$$\mathbb{E}[G_i^2] = c^2 \mathbb{E}[D_i^2] \prod_{j \neq i} \mathbb{E}[F_j^2]. \tag{30}$$

*Consequently, repeated nondegenerate factors give exponential second-moment growth.*

The result follows directly from independence, $\mathbb{E}V = c$, and $\mathbb{E}D_i = 0$, the last identity being the derivative of $\mathbb{E}F_i = 1$. It diagnoses the raw published estimator; a new baseline, importance sampler, or other stabilization can change these moments.

Table 12 reports exact products of one-dimensional quadrature moments. For Arch and Spike, $\mathbb{E}F^2$ is respectively $\pi^2/8$ and $4/3$, independent of the interior probability. LongJump has $\mathbb{E}F^2 = 1 + 4(x - 1/2)^2$: it is degenerate and stable when every $x_j = 1/2$, but loses that exception away from the exact center.

Table 12: Base-10 log of $\mathbb{E}[V^2]$ for $Q \equiv 1$ under raw ESG. Alternating profiles differ by coordinate. The stochastic value has mean one, so these values also determine relative variance up to subtracting one.

| Variant | Probability profile | $n = 1199$ | $n = 3000$ |
|---|---|---|---|
| Arch | all .5 | 109.36 | 273.63 |
| Spike | all .5 | 149.80 | 374.82 |
| LongJump | alternating $.4/.6$ | 20.42 | 51.10 |

We evaluated the published raw ESG variants before natural-task runs. The prespecified gate required at least one variant to keep both the value and constant-oracle coordinate-gradient second moments below $10^6$ at $n \in \{1199, 3000\}$ under centered and alternating $.4/.6$ profiles. All three are unbiased and exceed the gate; LongJump's corresponding gradient logs, for example, are 21.01 and 51.69. We report this formula-level scaling result without using task data or holdouts. A stabilized ESG construction would be a new estimator and calls for a separate comparison.

# H Experimental audit details

## H.1 Frozen runs and query accounting

Adult and ACSIncome broad comparisons use five and ten seeds; census UGC uses eight new seeds, and Waterbirds uses eight seeds for three preferences and eight methods (192 valid trajectories after a fixed smoke). UCI Credit uses eight prospectively frozen seeds for masked DisARM, true UGC, objective Block-64, and common Block-64. Every development checkpoint uses a separate 512-call common-random-number panel excluded from training. All runs require finite trajectories, exact ledgers, no budget overrun, independent oracle processes, and zero holdout calls during training. Block-OSRB and DisARM cost 2H/update; true UGC uses 2H when its independent selected row is outside the boundary mask and 4H otherwise. R126 reconstructs every train-only transform and evaluates the first three tasks; R151 applies the same hash, reconstruction, and 256-pair one-time protocol to UCI Credit. The cost audit uses no natural H calls and reports explicit $T_{\mathrm{local}} + 2L$ extrapolations.

The three fixed scalarization weights are utility $(.8, .1, .1)$, balanced $(.5, .25, .25)$, and fairness $(.2, .4, .4)$. Hypervolume uses reference point $(1, 1, 1)$. Every residual starts at zero and uses the same Adam configuration: learning rate .03, $(\beta_1, \beta_2) = (.9, .999)$, gradient-norm cap 10, and parameter-radius cap 5. The shared forward map applies $30\tanh(z/30)$ before the Bernoulli sigmoid to prevent machine-exact probabilities; analytic estimators include its chain derivative.

Adult fits a logistic base model on 29,561 pretrain rows and optimizes on 3,000 development rows. ACSIncome uses 182,665 pretrain, 3,000 development, and 10,000 sealed holdout rows from

Table 13: Actual optimization calls under the nominal hard ceilings. Ranges are over all saved seeds and preferences; a slash separates the 24H and 96H columns.

| Method | Adult | ACSIncome | Waterbirds | UCI Credit |
|---|---|---|---|---|
| Block-OSRB / DisARM | 24 / 96 | 24 / 96 | 24 / 96 | 24 / 96 |
| Ordinary LR / SPSA / AS-style | 24 / 96 | 24 / 96 | 24 / 96 | – |
| FW-style | 18 / 90 | 18 / 90 | 18 / 90 | – |
| Quotient-FD | 17–24 / 89–96 | 18 / 88–96 | 21 / 91–92 | – |
| True UGC | 22 / 94 | 22 / 94–96 | 22–24 / 94–96 | 22 / 94 |

Table 14: Waterbirds development alignment audit (hypervolume regret, mean $\pm$ seed standard deviation; lower is better). Every method uses exactly two vector calls per update; final-holdout values appear in Table 16.

| Method | 24H | 96H |
|---|---|---|
| Masked DisARM | $.6343 \pm .0125$ | $.5537 \pm .0156$ |
| Objective Block-64 | $.6053 \pm .0037$ | $.4815 \pm .0083$ |
| Common Block-64 | $.6101 \pm .0062$ | $.4859 \pm .0101$ |

California 2018 one-year ACS. UCI Credit uses 17,000 pretraining, 3,000 development, and 10,000 sealed holdout rows, stratified jointly by label and protected group. Waterbirds fits its logistic head on 4,795 train embeddings and optimizes on 1,199 official-validation embeddings; R126 uses all 5,794 official-test images. All standardization, projections, and base fitting use training data only.

The historical artifact key `ugc_antithetic` calls ordinary antithetic LR, with no UGC boundary switch. Table 7 uses a separate true-UGC implementation and seeds under the corrected contract.

True UGC starts an update when at least 4H remains, allowing either its 2H or 4H branch under the same hard ceiling. Its 24H and 96H checkpoints can therefore contain 22H and 94H of optimization, respectively; Block-OSRB and DisARM consume the full nominal budgets. Table 13 gives actual calls.

### H.2 Comparator implementations

Ordinary antithetic LR averages two Bernoulli score estimators on the same antithetic pair and costs 2H. Parameter SPSA uses one 32-dimensional Rademacher direction, radius $.05$, common random numbers, and 2H. Quotient-FD samples one prediction, queries it once, flips one representative row for each unique public objective–cell–state intervention, and shares the resulting finite difference across rows whose interventions are count-equivalent; every distinct vector query is charged, and an update is skipped if it cannot fit in the remaining budget.

The AS-style arm uses four antithetic parameter directions of radius $.1$ (eight probes, 8H), fits the local scalarized metric in the four public label–group logistic-loss profiles with ridge $10^{-4}$, and maps that fit through the analytic profile Jacobian. The FW-style arm uses a center and positive/negative shifts of $.15$ in each of four public-cell logits (nine probes, 9H), fits the metric in cell mean-probability space, and applies its rate-profile Jacobian. These are low-query adaptations under the common estimand and budget; the published high-query training recipes address a different compute regime. All methods use independent replayable random streams, while paired LR estimators share uniforms within a seed and step.

### H.3 One-time final-holdout audit

Table 16 gives the absolute values underlying the relative reductions in Table 7. The Adult official test contains 16,281 rows, the sealed ACSIncome and UCI Credit splits contain 10,000 rows each, and Waterbirds official test contains 5,794 images. Adult training-data and development-index hashes, all ACSIncome and UCI Credit split hashes, and the Waterbirds tarball, feature-array, and backbone-state hashes were checked before scoring. Reconstructed development logits and design coordinates had to match saved artifacts before scoring Waterbirds or UCI Credit.

Table 15: Prospective UCI Credit development replication and one-time holdout (hypervolume regret, mean ± seed standard deviation; lower is better). The fourth task was fixed before any Block-OSRB outcome.

| | Development | | Holdout | |
|---|---|---|---|---|
| Method | 24H | 96H | 24H | 96H |
| Masked DisARM | $.4199 \pm .0115$ | $.4040 \pm .0088$ | $.4019 \pm .0105$ | $.3850 \pm .0120$ |
| True UGC | $.4220 \pm .0104$ | $.4044 \pm .0046$ | $.4045 \pm .0119$ | $.3859 \pm .0058$ |
| Objective Block-64 | $\mathbf{.4012 \pm .0064}$ | $\mathbf{.3763 \pm .0034}$ | $\mathbf{.3838 \pm .0058}$ | $\mathbf{.3606 \pm .0051}$ |
| Common Block-64 | $.4039 \pm .0087$ | $.3773 \pm .0064$ | $.3888 \pm .0122$ | $.3611 \pm .0084$ |

Table 16: One-time final-holdout hypervolume regret (mean ± seed standard deviation; lower is better). All parameters and success gates were frozen before holdout access.

| | Adult | | ACSIncome | | Waterbirds | |
|---|---|---|---|---|---|---|
| Method | 24H | 96H | 24H | 96H | 24H | 96H |
| Block-OSRB-64 | **.3881±.0123** | **.3602±.0323** | **.3128±.0261** | **.2783±.0079** | **.5641±.0069** | **.4390±.0115** |
| DisARM | .4003±.0300 | .3674±.0208 | .3485±.0656 | .2985±.0061 | .6080±.0132 | .5464±.0182 |
| True UGC | .4014±.0282 | .3684±.0232 | .3535±.0666 | .2979±.0068 | .6167±.0103 | .5762±.0153 |

All eight dataset–budget mean directions per comparator are positive; none triggers a preregistered 3% regression boundary. The original three-task bootstrap resamples 24 dataset–seed blocks while retaining both budgets and gives 95% intervals $[0.02438, 0.05119]$ against DisARM and $[0.02903, 0.06165]$ against true UGC. The separately frozen UCI Credit common-Block intervals are $[0.01188, 0.02624]$ and $[0.01252, 0.02737]$. Adult is heterogeneous across seeds: Block-OSRB is no worse in 5/8 seeds at 24H and 4/8 at 96H against each comparator. The reported effect is the pooled mean over the fixed task–seed blocks.

A sensitivity analysis on the original three tasks resamples seeds within each fixed dataset and gives intervals $[0.03078, 0.04539]$ against DisARM and $[0.03797, 0.05283]$ against true UGC; hierarchical dataset–seed resampling gives the wider $[0.00930, 0.07356]$ and $[0.01092, 0.09366]$. Both remain positive. These intervals quantify uncertainty across the observed dataset–seed blocks; Section 7 states the population-level boundary.

No R126-scored checkpoint was retrained, reselected, or rescored after holdout access. The later R316 masked-DisARM trajectories are a separate development-only comparator with zero final-holdout calls. The original saved artifacts contain all 144 checkpoint evaluations and 48 fronts per dataset; the separate UCI Credit artifact contains 192 checkpoint evaluations and 64 fronts. Both include finite-output and 512H gates, manifests, hashes, and per-seed regret differences.

### H.4 Interpreting fixed-preference hypervolume

The three preferences are utility $(.8, .1, .1)$, balanced $(.5, .25, .25)$, and fairness $(.2, .4, .4)$. Their union hypervolume measures the scored set of three policies, not the quality of every component or every preference. For example, on the Adult official-test 24H fairness policy, Block has mean objective vector $(.2834, .1259, .0776)$ versus ordinary DisARM's $(.2899, .1407, .0475)$: macro-F1 loss and demographic-parity gap improve, while equalized-odds gap increases. The fairness-weighted losses are .13808 and .13329, respectively. In the separate R316 Adult development comparison, Block's fairness-weighted loss is .13869 versus masked DisARM's .12906 at 24H and .12496 versus .12183 at 96H. These component and preference results coexist with the front-level hypervolume comparisons.

As a post-hoc reference-point check, we recomputed HV from the saved scored vectors using common references $(r, r, r)$ for $r \in \{.5, .75, 1, 1.25, 2\}$. All scored coordinates are below .5. At every reference, Block's mean HV exceeds the corresponding DisARM parent and true UGC in each of the eight task–budget cells. The preregistered metric remains $1 - \text{HV}$ at $(1, 1, 1)$.

Table 17: Shape of the saved Block-OSRB holdout fronts. Each budget has eight seeds and three fixed-preference points. "Singleton" counts seeds with one nondominated point; extra HV is the mean fraction of union HV beyond the best single point. This is a post-hoc reanalysis of saved objective vectors, with no new oracle calls.

| | 24H | | 96H | |
|---|---|---|---|---|
| Task | Singleton | Extra HV | Singleton | Extra HV |
| Adult | 0/8 | 3.27% | 0/8 | 5.91% |
| ACSIncome | 2/8 | 0.69% | 0/8 | 3.16% |
| Waterbirds | 4/8 | 0.17% | 6/8 | 0.09% |
| UCI Credit | 3/8 | 0.20% | 2/8 | 0.27% |