WHEN HONEST EVALUATION FAILS TO COMPOSE: LEARNER-PROTECTED TARGETING CURVES

Anonymous authors
Paper under double-blind review

ABSTRACT

Evaluating each learned treatment rule on fresh data does not ensure valid inference when the evaluations are combined. Earlier evaluation noise can influence later model selection and change the variance of later errors, invalidating a bootstrap that freezes the learned path. A two-candidate characterization identifies when this failure persists and when frozen inference remains valid. We construct asymptotically valid confidence bounds for the value of the realized rules, simultaneous over treatment budgets and version weights. Versionwise bands combine rectangularly even with overlapping evaluation samples. For disjoint evaluations with Gaussian limits, studentization gives radial bounds with a sharp common critical value at each budget. A matched four-route experiment separates the benefits of data reuse from those of calibration geometry: using all later cohorts can improve selected utility, while radial bounds improve precision for balanced mixtures. A matched selection–decision simulation shows a concentrated benefit of protection. At one predeclared cost, it leaves 99.6% of decisions unchanged and improves average utility by preventing harmful deployment in the remaining 0.4%; at a lower cost, it forgoes useful treatment. On Hillstrom, reuse increases the certified net value, while exact-budget targeting advantages remain uncertified. A finite-batch correction supports deployment with integer treatment quotas.

1 INTRODUCTION

A team must decide whom to offer a treatment when it can treat only part of a population. It trains a model to rank people, then estimates the benefit of treating, say, the top 10% or 20%. To avoid optimism from training, it evaluates the model on fresh randomized-trial data. This separation of training and evaluation is often called *honest evaluation*.

As new data arrive, the team updates its model. It may use the first evaluation to choose between two candidates and evaluate the selected one on the next group. Every evaluation still uses fresh data, but the results are now connected: noise in the first evaluation helps choose the next model, and that model determines how variable the next evaluation will be. A bootstrap that holds all fitted models fixed when combining their estimates can miss this dependence. The problem can persist even at a fixed treatment budget and fixed averaging weights.

The final decision adds a second selection problem. The team may compare budgets and model versions, or divide future recipients among several versions, after seeing all the evaluations. An interval for one prespecified choice does not protect this search. We therefore construct confidence bounds that cover every eligible choice together, then allow the team to select from them. The target is the value of the rules actually learned.

We study two ways to construct these bounds. The first builds a band for each version and averages the endpoints; it needs no independence between version errors. The second divides each error by its conditional standard deviation before combining disjoint evaluations. This removes the random scale that caused the failure above and leads to an exact calculation of the worst error over version weights. The mathematical details follow the two-candidate example in Section 2.

Protection widens bounds and can prevent a useful rule from being certified. Our main experiment measures this cost on the same fitted models and asks how much precision is recovered by using all later cohorts for evaluation. Evaluation on a final untouched cohort provides another comparison.

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.

A matched deployment experiment uses an independent selection history to isolate the effect of feedback. It measures both the frequency of changed decisions and their utility gains or losses. At predeclared costs near the deployment threshold, protection improves mean utility by stopping a small fraction of harmful proposals while leaving most decisions unchanged. Section 5 extends the bounds to integer treatment quotas.

Related work. Treatment-prioritization measures and policy evaluation describe benefits across ranks and budgets (Yadlowsky et al., 2025; Sverdrup et al., 2025; Imai & Li, 2023); other methods summarize treatment effects for model-defined groups (Chernozhukov et al., 2025; Imai & Li, 2025). Sequential validation and adaptive-combination tests protect evidence collected in stages (Wager, 2025; Bauer & Köhne, 1994). Scaling errors by their conditional standard deviations also supports adaptive policy evaluation (Bibaut et al., 2021; Luedtke & van der Laan, 2016). We build on inference for learned treatment rules (van der Laan & Luedtke, 2015) to cover their combined value when earlier evaluations affect later fits and the final weights are chosen from the results. The fitted scores may come from forests, meta-learners, or other methods (Athey et al., 2019; Nie & Wager, 2021; Künzel et al., 2019).

2 How one evaluation changes the next

Consider two candidates with similar true gains at the same treatment budget. The team compares them on the first cohort and sends the better-looking one to a second, unopened cohort. A random fluctuation that raises the first reported gain can also change which candidate wins. If the candidates have different evaluation variances, this choice changes the scale of the second error. Fresh second-cohort data keep that error conditionally centered, but do not remove its dependence on the first evaluation.

The following expression makes the dependence explicit. Z is the first gain error, $V + h$ is the difference used to select the candidate, and W is independent noise from the second cohort. The population difference h is measured in standard-error units; σ_+ and σ_- are the two candidate error scales. The "frozen" analysis holds the selected models fixed when resampling and treats their pooled, studentized error as standard normal.

Proposition 1 (Two-candidate feedback)**.** *Let (Z, V) be standard bivariate normal with correlation ρ, and let $W \sim N(0,1)$ be independent. Set $\sigma(v) = \sigma_+ > 0$ for $v > 0$ and $\sigma(v) = \sigma_- > 0$ otherwise. For finite h, define*

$$U_h = \frac{Z + \sigma(V + h)W}{\sqrt{1 + \sigma^2(V + h)}}. \tag{1}$$

Then $U_h \sim N(0,1)$ if and only if $\rho = 0$ or $\sigma_+ = \sigma_-$. Moreover,

$$\sup_x |P(U_h \le x) - \Phi(x)| \le \Phi(-|h|). \tag{2}$$

Thus feedback matters when the selection noise is correlated with the earlier value error and the candidates have different scales. When the true gain difference is large relative to its standard error, the same candidate is selected with high probability. The proof in Appendix B.1 also realizes this limit in one i.i.d. randomized trial by selecting between the scores X and $-X$. When the true gains are equal and the scale ratio is four, limiting frozen coverage is 94.54% instead of 95%; 10,000 replications at each cohort size give 94.51% and 94.49% for sizes 400 and 1600. Equal-scale controls recover approximately 95% (Appendix B.2). This is a persistent coverage deficit of 0.46 percentage points.

This example suggests two constructions. Combining valid component bands avoids modeling their joint error law. Alternatively, dividing each error by its conditional scale recovers the independent pair (Z, W). The next section uses these constructions to protect subsequent choices of budget and version weights.

3 CONFIDENCE BOUNDS FOR CHOOSING A RULE AND BUDGET

3.1 GAINS AND VERSION MIXTURES

We observe i.i.d. data (X, A, Y) with binary treatment and conditional treatment effect $\tau(x) = E\{Y(1) - Y(0) \mid X = x\}$. Consistency, no interference, conditional exchangeability, and overlap identify this effect. Order the sample into K folds. Before opening fold I_k, fit a score s_k using the history $\mathcal{H}_{k-1}$, for $k = 2, \ldots, K$; I_1 is burn-in.

Let $\pi_{s_k,q}$ treat the top population fraction q under s_k, breaking ties with independent uniform keys. On a prespecified finite grid $\mathcal{Q} \subset (0, 1)$, define

$$G_k(q) = E\{\pi_{s_k,q}(X, U)\tau(X) \mid \mathcal{H}_{k-1}\}, \qquad G^\lambda(q) = \sum_{k=2}^{K} \lambda_k G_k(q), \tag{3}$$

where λ belongs to the simplex $\Lambda = \{\lambda_k \geq 0 : \sum_{k=2}^{K} \lambda_k = 1\}$. A weight of one selects a single version; otherwise λ_k is the fraction of future recipients assigned to version k. These targets depend on the fitted scores and hence on sample size; we omit the index n.

Write $\widehat{G}_k(q)$ for a gain estimate from data that were unopened when version k was learned, and average these estimates with the same weights to obtain $\widehat{G}^\lambda(q)$. Appendix A constructs the augmented inverse-probability estimator and its foldwise inference, including re-estimation of the rank threshold. The first combination principle needs only a valid band for each version.

3.2 COMBINE CONFIDENCE BANDS FOR INDIVIDUAL VERSIONS

Consider two learned versions at one budget. If their gains lie in $[0.02, 0.08]$ and $[0.04, 0.10]$, their 50/50 mixture lies in $[0.03, 0.09]$. The same averaging argument holds for every mixture on that event, so choosing the weights afterwards cannot invalidate it. If the input bands cover all eligible budgets, the argument also holds across budgets.

For the general construction, let $[L_k, U_k]$ cover version k on $\mathcal{Q}$ with conditional noncoverage at most $\alpha_k + o_p(1)$ given $\mathcal{H}_{k-1}$. To protect all versions together, allocate $\sum_k \alpha_k \leq \alpha$ on every path, fixing each allocation before its evaluation outcomes are opened. Average the band endpoints:

$$B_\lambda^{\mathrm{rect}}(q) = \left[\sum_{k=2}^{K} \lambda_k L_k(q), \sum_{k=2}^{K} \lambda_k U_k(q)\right]. \tag{4}$$

Theorem 2 (Rectangular protection)**.** *For fixed K, the conditional band guarantees above imply*

$$\liminf_n P\{G^\lambda(q) \in B_\lambda^{\mathrm{rect}}(q)\ \forall (q, \lambda) \in \mathcal{Q} \times \Lambda\} \geq 1 - \alpha. \tag{5}$$

The evaluation samples may overlap, and later learners may use all earlier evaluation outcomes. The same result holds on a prespecified simplex face when every eligible component receives protection.

The only probabilistic step is a union bound across versions. Averaging on the resulting coverage event is deterministic, so dependence between version errors needs no joint model.

Reusing all later cohorts. Evaluate each fitted version on every cohort that was unopened when it was learned. Conditional on its training history, these observations form an i.i.d. evaluation sample. Older versions use more observations; the resulting samples overlap across versions. Since the theorem needs no independence between their bands, this reuse preserves the guarantee. It requires more predictions but no additional labels or refitting. Appendix C.2 gives the conditions at the enlarged evaluation-sample scale.

3.3 STANDARDIZE EACH ERROR BEFORE COMBINING EVALUATIONS

Rectangular protection allows all component errors to move against us together. Disjoint next-fold evaluations offer more structure. Recall the two-candidate example: the second error's scale depends on the first evaluation, but dividing by its own scale recovers independent standard-normal noise. We apply that step separately to every version before combining their values.

Let $\widehat{\mathrm{se}}_k(q)$ estimate the conditional standard error of $\widehat{G}_k(q)$. Under A1–A4, iterated conditioning makes the standardized errors asymptotically independent standard normals. We measure each mixture's error on the scale $\{\sum_k \lambda_k^2 \widehat{\mathrm{se}}_k^2(q)\}^{1/2}$ and protect all weights at once, including choices that depend on the observed scales. Because weights are nonnegative, the worst upward error is driven by positive standardized components; their Euclidean norm gives the critical value below.

Fix the set S of versions before evaluation and write $d = |S|$. Let $c_d^+(u)$ be the u quantile of $\sum_{j=1}^d (Z_j^+)^2$, where the Z_j are independent standard normals and $z^+ = \max(z, 0)$. With $\sum_{q\in\mathcal{Q}} \beta_q \le \alpha$, define the lower bound

$$L_\lambda^{\mathrm{rad}}(q) = \widehat{G}^\lambda(q) - \left[c_d^+(1-\beta_q) \sum_{k\in S} \lambda_k^2 \widehat{\mathrm{se}}_k^2(q)\right]^{1/2}. \tag{6}$$

Theorem 3 (Radial protection)**.** *Under A1–A4, with consistent conditional standard errors,*

$$\liminf_n P\{G^\lambda(q) \ge L_\lambda^{\mathrm{rad}}(q)\ \forall (q,\lambda) \in \mathcal{Q} \times \Lambda(S)\} \ge 1-\alpha, \tag{7}$$

where $\Lambda(S)$ is the simplex supported on S. At each nondegenerate budget with $0 < \beta_q < 1 - 2^{-d}$, the limiting coverage over the entire face is exactly $1-\beta_q$. The critical multiplier is the smallest possible within the family using a common deterministic multiplier of $\{\sum_k \lambda_k^2 \widehat{\mathrm{se}}_k^2(q)\}^{1/2}$.

The following identity gives the exact worst error across all weights. For errors $e_k = \widehat{G}_k(q) - G_k(q)$ and $Z_k = e_k/\widehat{\mathrm{se}}_k(q)$,

$$\sup_{\lambda\in\Lambda(S)} \frac{(\sum_{k\in S} \lambda_k e_k)_+}{\{\sum_{k\in S} \lambda_k^2 \widehat{\mathrm{se}}_k^2(q)\}^{1/2}} = \|Z_+\|_2. \tag{8}$$

When $Z_+ \neq 0$, weights proportional to $Z_k^+/\widehat{\mathrm{se}}_k(q)$ attain equality. The conditional Gaussian limit then gives the stated critical value. The finite-grid error allocation can still be conservative; sharpness is for this radial family at one budget.

The rectangular bound adds the individual error allowances. The radial bound uses their root-sum-of-squares scale and can be tighter when several versions receive substantial weight. The rectangle can be tighter when one version receives most of the weight. Independence after standardization requires disjoint sequential evaluations; the rectangular construction handles overlapping samples. Appendix C.3 gives two-sided bounds and the result for separated finite-level scores.

3.4 COSTS AND COMPARISON WITH RANDOM TREATMENT

A simultaneous bound holds at any data-selected pair $(\widehat{q}, \widehat{\lambda})$. A known treatment cost $C(q)$ subtracts from both the target and the bound. The maximized value of $L_G(q,\lambda) - C(q)$ remains a lower bound on the selected rule's population utility; treat-none has value zero when $C(0) = 0$.

Qini asks a different question: does ranking improve on random allocation at the same budget? Writing $\theta = E\tau(X)$,

$$\Delta^\lambda(q) = G^\lambda(q) - q\theta, \qquad \mathcal{U}^\lambda(q) = G^\lambda(q) - C(q). \tag{9}$$

The mean $\theta = G_k(1)$ must enter the joint calibration for Qini; it also permits treat-all to be included as a decision. In each resampling draw, subtract q times the estimated mean from the estimated gain. Jointly calibrate the resulting gain and Qini coordinates within the same error budget.

The regularity requirement concerns the score's budget boundary. Smooth boundaries and budgets separated from finite-level cumulative masses permit radial protection. At an atomic boundary with unequal adjacent effects, the limit is directional; a valid subsampling fold band still permits

rectangular composition under the conditions in Appendix D. Calibrating a raw score into separated levels can recover the Gaussian case, but it changes the ranking and can lose value. Stable frozen pooling is also possible when the predictable influence covariance approaches a matrix determined before evaluation, with weights also fixed at that time (Appendix B.3).

4 Benefits and Costs of Protection

4.1 Evaluating the Same Fitted Models

Table 1 applies four evaluation procedures to the same four fitted ridge models. The first two evaluate each model on its next cohort ("diagonal" in the table) and compare rectangular and radial bounds. The third uses rectangular bounds with all later cohorts ("all-suffix"), isolating the benefit of data reuse. The fourth evaluates all models on the last cohort, after every model has been fixed, and uses their joint bootstrap covariance (Appendix E.2).

All four procedures use the same total data, original nuisance fits, 18-budget grid including treat-all, full version simplex, and cost $C(q) = 0.08q$. Each has nominal 95% simultaneous coverage and maximizes its lower utility, returning treat-none if the maximum is nonpositive. Safe U is the selected *population* utility with this fallback. There are 500 paired trials at each sample size, 999 resampling draws, and 30,000 independent covariates for evaluating each realized target.

Table 1: Evaluation of the same fitted models. Coverage, selected-bound validity, and deployment are percentages; brackets are 95% Wilson Monte Carlo intervals. Safe utility includes the treat-none fallback.

n	Evaluation	Coverage	Selected cov.	Deploy	Safe U
800	Diagonal rectangle	95.2 [93.0,96.8]	97.0	28.0	0.0138
800	Diagonal radial	97.8 [96.1,98.8]	98.8	32.0	0.0157
800	All-suffix rectangle	96.2 [94.1,97.6]	97.8	37.8	0.0186
800	Common last fold	96.6 [94.6,97.9]	97.4	23.2	0.0106
1600	Diagonal rectangle	95.6 [93.4,97.1]	96.6	65.4	0.0398
1600	Diagonal radial	98.0 [96.4,98.9]	99.2	79.8	0.0489
1600	All-suffix rectangle	97.2 [95.4,98.3]	98.8	81.6	0.0492
1600	Common last fold	97.2 [95.4,98.3]	98.2	49.4	0.0290

With $m = n/5$, using only the next cohort gives (m, m, m, m) rows per version; using all later cohorts gives $(4m, 3m, 2m, m)$. Reuse increases version–observation predictions from $4m$ to $10m$ on the same $4m$ distinct rows. The suffix route's utility advantage over diagonal radial bounds is 0.00292 (95% paired Monte Carlo interval $[0.00098, 0.00485]$) at $n = 800$ and 0.00023 ($[-0.00167, 0.00212]$) at $n = 1600$.

For an equal mixture, radial bounds have smaller radii than rectangular bounds using all later cohorts: 0.1148 versus 0.1328 at $n = 800$, and 0.0803 versus 0.0918 at $n = 1600$, averaged across budgets. The rectangular lower bound is linear in the weights, so its maximum can be achieved by selecting one version. Radial bounds can favor mixtures. Appendix E provides the protocols and summarizes additional benchmarks retained with the code. In those benchmarks, a $50/50$ honest split gives narrower bands in every tested learner setting for its own separately trained policy. This comparison evaluates the full learning pipeline; Table 1 holds the fitted policies fixed.

4.2 When Does Protection Change Deployment?

We compare deployment decisions for the same proposed policy, deploying when its lower utility bound is positive. Protection stops additional proposals, avoiding losses for harmful policies and forgoing gains for useful ones. We measure how often decisions change and the resulting utility difference. Appendix C.4 derives the exact utility difference and its concentration near the decision threshold.

We refit two ridge treatment-effect models in each of 20,000 simulated randomized trials per noise condition, then evaluate four successive versions on fresh cohorts. The models use different feature groups with similar gains. The first is the default: use it when the running estimated gain exceeds a

fixed review hurdle C_0; otherwise use the alternative. In the unequal-noise setting, the alternative's conditional error scale averages 2.48 times that of the default. Observed candidate gains are within one paired standard error in 68.3% of trials, and the first review retains the default in 48.2%.

All three bounds use the same realized version sequence, half-population budget, equal weights, and deployment gate: treat only when the lower gain bound exceeds the cost. An independent replay generates the selection history on four additional cohorts, then evaluates its versions on the original cohorts. It preserves the distribution of branch paths while removing feedback from those evaluations. The equal-noise control removes the variance contrast. Appendix E.3 gives the simulation protocol and exact values of the fitted policies.

With unequal noise, frozen coverage is 94.03% (95% Monte Carlo interval $[93.69, 94.35]$), compared with 95.19% under independent replay. The paired deficit is 1.17 percentage points ($[0.82, 1.51]$). Equal-noise feedback/replay coverage is $95.16/95.22\%$. The radial bound protects the full weight simplex and covers the fixed mixture in 99.44% of unequal-noise feedback trials.

Table 2: When radial protection helps or hurts, on the same fitted proposal in 20,000 unequal-noise feedback trials. The three decision categories partition the trials; ΔU is protected minus frozen population utility, averaged over all trials. $C_0 = 1/\sqrt{2\pi}$ is the review hurdle; $C_+ = C_0 + 0.02$ is the primary harm cost and $C_- = C_0 - 0.04$ measures useful deployment. All three costs were fixed before the original run. The decomposition is a secondary analysis of those trials.

Cost	Same decision (%)	Avoided harm (%)	Blocked benefit (%)	$\Delta U \times 10^3$ [95% MC interval]
C_-	76.560	0.000	23.440	-8.530 [-8.744, -8.315]
C_0	95.845	4.155	0.000	+0.126 [+0.115, +0.138]
C_+	99.595	0.405	0.000	+0.097 [+0.076, +0.118]

At C_+, decisions agree in 99.595% of trials. The remaining 0.405% (81 trials) are harmful deployments stopped by protection, with mean avoided loss 23.94 per 1,000 recipients when a correction occurs. Averaged over *all* trials, protection improves utility by 0.097 per 1,000 (95% paired interval $[0.076, 0.118]$). At C_-, protection blocks useful deployments and lowers mean utility by 8.530 per 1,000.

At C_0, feedback raises frozen harmful deployment from 3.82% under replay to 4.48%, and radial protection lowers it to 0.325%. The replay contrasts in Appendix E.3 measure the feedback effect with each critical value held fixed.

4.3 Precision and ranking loss on Hillstrom

On Hillstrom's randomized email data (Hillstrom, 2008), we evaluate four sequential boosted T-learners using seven independently calibrated score levels and an illustrative cost of 0.01 visit-equivalents per email. For an equal mixture, radial bounds have smaller average radii than rectangular bounds using all later cohorts (0.01106 versus 0.01213). After selection, the all-suffix rectangle certifies a higher net value than the radial bound (0.05244 versus 0.04963). All four procedures select treat-all; their exact-budget Qini lower bounds remain nonpositive.

Coarsening has a measurable cost. An independent audit compares raw and seven-level policies fixed by a separate calibration sample. At nominal budget 0.05, their observed masses are $0.0502/0.0512$ and Qini estimates are $0.00774/0.00265$. The paired Qini loss is 0.00509, with simultaneous 95% interval $[0.00076, 0.00942]$. Appendix E.4 specifies the calibrated policies and the independent evaluation protocol.

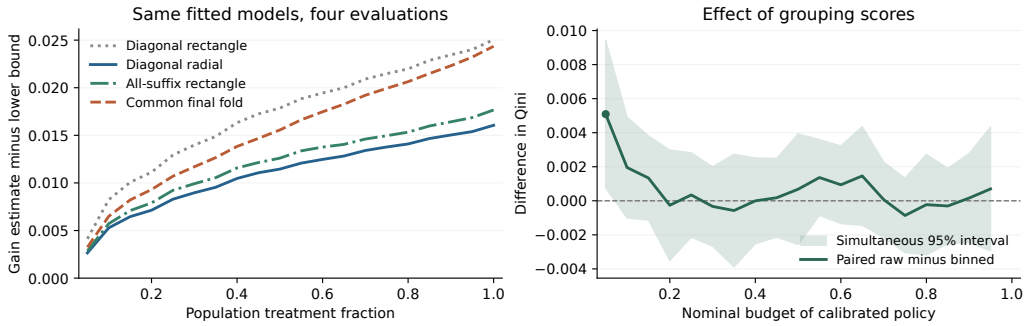


Figure 1: Hillstrom. Left: evaluations of the same rules at exact population budgets. Right: paired Qini loss from coarsening independently fixed calibrated policies, with a simultaneous 95% interval. The right panel evaluates each policy at its own treatment mass.

5 From inference to deployment

For deployment, the chosen treatment fraction must become an integer number of people. Split a future batch among the selected versions and treat the top-ranked people up to each version's quota. Rounding and ranking within a finite batch can change expected gain. Proposition 12 bounds this difference using a known bound on the treatment effect. After subtracting its correction R_{quota} and the cost at the actual treated fraction $\bar{q}_B$, the lower bound on expected batch net gain is

$$L_{\text{serve}} = L_G(\widehat{q}, \widehat{\lambda}) - R_{\text{quota}} - C(\bar{q}_B). \tag{10}$$

Deploy only if $L_{\text{serve}} > 0$; otherwise treat no one. Fix the rules and counts before serving the batch. On the simultaneous coverage event, every deployed rule has positive expected batch net gain. Algorithm 1 combines evaluation, selection, and deployment; Appendix F gives the quota bound and implementation details.

Algorithm 1 Selecting a rule and applying integer treatment quotas

Require: Ordered folds, budget grid, eligible versions, fixed cost C,
error budget α, batch size B, and effect envelope M

Fix the eligible decisions, choice of confidence bound, and outcome-free tie keys
Allocate error and fix calibration tuning before each evaluation is opened
for $k = 2, \ldots, K$ **do**
    Fit $(s_k, \widehat{\eta}_k)$ from $\mathcal{H}_{k-1}$; retain the fitted pair
    Open I_k; update $\mathcal{H}_k$ with its data and recorded evaluation
end for
Complete each version's evaluation on its declared honest sample
Form L_G by rectangular or eligible diagonal-radial protection
$(\widehat{q}, \widehat{\lambda}) \leftarrow \arg\max_{q,\lambda}\{L_G(q, \lambda) - C(q)\}$
Round to component counts B_k and treatment quotas t_k
Compute R_{quota} from equation (81)
$\bar{q}_B \leftarrow B^{-1} \sum_k t_k$
$L_{\text{serve}} \leftarrow L_G(\widehat{q}, \widehat{\lambda}) - R_{\text{quota}} - C(\bar{q}_B)$
if $L_{\text{serve}} \leq 0$ **then**
    return treat-none, zero quotas, and certificate 0
else
    Lock the scores and counts; treat the top t_k ranks within each component
    return the serving rule and certificate L_{serve}
end if

On YouWiN!'s business-plan competition data, all four procedures retain positive expected gain after this correction under the stated assignment and missing-at-random working model. With a 30% capacity cap, $B = 1000$, $M = 1$, and zero treatment cost, the all-suffix rule assigns 300 places and certifies an increase of at least 6.04 percentage points in operating probability per applicant. The other three serving lower bounds range from 2.85 to 3.88 points. Selected Qini lower bounds are negative. Appendix F.4 gives the application design and sensitivity checks; Appendix F.5 reports calibration on its observed covariates.

6 Conclusion

Sequential feedback connects the errors of otherwise honest evaluations. Combining per-version confidence bands, or standardizing errors from disjoint evaluations, permits valid inference after choosing budgets and version weights. Reusing later cohorts can improve selected utility, while radial bounds improve precision for balanced mixtures. In the matched feedback experiment, protection improves mean utility at predeclared costs by stopping harmful proposals while leaving most decisions unchanged. Its value is concentrated near the deployment threshold, where the cost determines whether withholding treatment avoids a loss or sacrifices a useful gain. The YouWiN! application illustrates positive certified gain after finite-batch adjustment.

AI use statement

In this work, we used generative AI tools to clean and reformat datasets, generate synthetic datasets, design and provide feedback on research methodology and experiments, implement methods, interpret results, and assist with translation. Additionally, we used generative AI tools to edit the manuscript for readability and polish mathematical phrasing. We have not used generative AI tools to develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, or propose or refine hypotheses; support for qualitative and thematic data analysis is not applicable to this work. We have reviewed all AI-assisted work, including manually verifying all mathematical derivations and testing implementation code. We take full responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

Reproducibility statement

Appendices A–D give the statistical conditions and proofs. Appendix E contains the evaluation comparisons; Appendix F collects the finite-batch theory, implementation details, and capacity audits. The anonymous supplement contains code, environment specifications, tests, and recorded outputs. Its README maps each exhibit to its inputs and gives verification and full rerun commands.

References

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019. doi: 10.1214/18-AOS1709.

Peter Bauer and Karl Köhne. Evaluation of experiments with adaptive interim analyses. *Biometrics*, 50(4):1029–1041, 1994. doi: 10.2307/2533441. Corrigendum: Biometrics, 52(1):380, 1996, doi:10.2307/2533187.

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. Curran Associates, Inc., 2021.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Iván Fernández-Val. Fisher–Schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments,

with an application to immunization in India. *Econometrica*, 93(4):1121–1164, 2025. doi: 10.3982/ECTA19303.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412, 2019. doi: 10.1093/restud/rdy049.

Peter Hall and Christopher C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. ISBN 9780123193506.

Kevin Hillstrom. The MineThatData e-mail analytics and data mining challenge. MineThatData blog and public challenge data, 2008.

Kosuke Imai and Michael Lingzhi Li. Experimental evaluation of individualized treatment rules. *Journal of the American Statistical Association*, 118(541):242–256, 2023. doi: 10.1080/01621459.2021.1923511.

Kosuke Imai and Michael Lingzhi Li. Statistical inference for heterogeneous treatment effects discovered by generic machine learning in randomized experiments. *Journal of Business & Economic Statistics*, 43(1):256–268, 2025. doi: 10.1080/07350015.2024.2358909.

Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019. doi: 10.1073/pnas.1804597116.

Alexander R. Luedtke and Mark J. van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2):713–742, 2016. doi: 10.1214/15-AOS1384. Correction by A. R. Luedtke, A. Bibaut, and M. J. van der Laan: The Annals of Statistics, 49(4):2429–2430, 2021, doi:10.1214/20-AOS2031.

David McKenzie. Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. *American Economic Review*, 107(8):2278–2307, 2017a. doi: 10.1257/aer.20151404.

David McKenzie. Replication data for: Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition. American Economic Association; distributed by ICPSR, 2017b. Version 1. doi:10.3886/E113101V1.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021. doi: 10.1093/biomet/asaa076.

Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer, New York, 1999. doi: 10.1007/978-1-4612-1554-7.

Jens Præstgaard and Jon A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, 21(4):2053–2086, 1993. doi: 10.1214/aop/1176989011.

Erik Sverdrup, Han Wu, Susan Athey, and Stefan Wager. Qini curves for multi-armed treatment rules. *Journal of Computational and Graphical Statistics*, 34(3):948–960, 2025. doi: 10.1080/10618600.2024.2418820.

Mark J. van der Laan and Alexander R. Luedtke. Targeted learning of the mean outcome under an optimal dynamic treatment rule. *Journal of Causal Inference*, 3(1):61–95, 2015. doi: 10.1515/jci-2013-0022.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996. doi: 10.1007/978-1-4757-2545-2.

Stefan Wager. A comment on: "Fisher–Schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India" by Victor Chernozhukov, Mert Demirer, Esther Duflo, and Iván Fernández-Val. *Econometrica*, 93(4):1171–1176, 2025. doi: 10.3982/ECTA23293.

Steve Yadlowsky, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. Evaluating treatment prioritization rules via rank-weighted average treatment effects. *Journal of the American Statistical Association*, 120(549):38–51, 2025. doi: 10.1080/01621459.2024.2393466.

APPENDIX CONTENTS

A Statistical setup and conditional inference **11**

A.1 Signals, ranks, and targets 11

A.2 Conditional assumptions 11

A.3 Regular fold inference 12

B Feedback failure and stable pooling **14**

B.1 Near-tied selection and its RCT realization 14

B.2 Validation-based candidate selection in one i.i.d. RCT 15

B.3 When frozen pooling is valid 15

C Proofs of protection and selection **16**

C.1 Proof of Theorem 2 16

C.2 Evaluating each version on all later cohorts 17

C.3 Radial geometry and proof of Theorem 3 17

C.4 Post-selection and known transformations 18

C.5 Optimization and computational cost 19

D Atomic boundaries and subsampling **19**

E Evaluation comparisons and ranking costs **24**

E.1 Data generation and evaluation of the same fitted models 24

E.2 Common-terminal comparator 24

E.3 Fitted learners, feedback, and deployment 25

E.4 Hillstrom: evaluation and score resolution 26

F From inference to deployment **26**

F.1 Finite-batch exact-quota bridge 26

F.2 Implementation details 27

F.3 Finite-batch decision audit 28

F.4 YouWiN! capacity application 28

F.5 Calibration on YouWiN! covariates 29

A STATISTICAL SETUP AND CONDITIONAL INFERENCE

A.1 SIGNALS, RANKS, AND TARGETS

For $O = (X, A, Y)$, write $\mu_a(x) = E(Y \mid A = a, X = x)$, $e(x) = P(A = 1 \mid X = x)$, and $\tau(x) = \mu_1(x) - \mu_0(x)$. Under the identifying conditions in the main text, the augmented signal is

$$\Gamma(O; \eta) = \mu_1(X) - \mu_0(X) + \frac{A\{Y - \mu_1(X)\}}{e(X)} - \frac{(1 - A)\{Y - \mu_0(X)\}}{1 - e(X)}. \tag{11}$$

The fitted signal uses only the training history. In an RCT with known e, every history-measurable pair of outcome regressions satisfies

$$E\{\Gamma(O; \widetilde{\mu}, e) \mid X, \mathcal{H}\} = \tau(X). \tag{12}$$

Its actual conditional variance, including nuisance misspecification, is the variance relevant to inference.

For a fixed score with CDF F_s and an independent $U \sim \mathrm{Unif}(0, 1)$, set

$$R_s(X, U) = F_s\{s(X)^-\} + U\big[F_s\{s(X)\} - F_s\{s(X)^-\}\big], \tag{13}$$

and

$$\pi_{s,q}(X, U) = \mathbf{1}\{R_s(X, U) > 1 - q\}. \tag{14}$$

Within an atom $S = s$, the randomized rank fills $[F_s(s^-), F_s(s)]$ uniformly. Together with the probability integral transform on the continuous part, this makes the rank uniform and the treated population mass exactly q. The key selects a binary action within each tie.

Use the finite interior grid $\mathcal{Q} = \{q_1, \ldots, q_J\}$, $\mathcal{Q}^+ = \mathcal{Q} \cup \{1\}$, and $\mathcal{D} = \mathcal{Q} \cup \{0, 1\}$ when both endpoints are eligible. At $q = 1$ the gain is the mean $\theta = E\tau(X)$; at $q = 0$ it is known to be zero. Joint inference is required if cost is estimated rather than fixed.

For a diagonal evaluation, let $a_{ik}(q) \in [0, 1]$ select exactly a q fraction of the empirical rank mass in I_k, allowing a fractional boundary record. With the history-fitted signal $\widehat{\Gamma}_i$, estimate

$$\widehat{G}_k(q) = \frac{1}{|I_k|} \sum_{i \in I_k} a_{ik}(q) \widehat{\Gamma}_i, \qquad \widehat{G}^{\lambda}(q) = \sum_{k=2}^{K} \lambda_k \widehat{G}_k(q). \tag{15}$$

The allocation and its rank threshold are recomputed in every weighted draw; Appendix A.3 gives the expansion and calibration.

Write E_{k-1} for conditioning on $\mathcal{H}_{k-1}$ and $\widehat{\Gamma}_k(O) = \Gamma(O; \widehat{\eta}_k)$. The fitted-signal target and its bias are

$$\widetilde{G}_k(q) = E_{k-1}\{\pi_{s_k,q}(X, U) \widehat{\Gamma}_k(O)\}, \qquad \rho_k(q) = \widetilde{G}_k(q) - G_k(q). \tag{16}$$

For $R_k = R_{s_k}(X, U)$, let $\widehat{g}_k(r) = E_{k-1}\{\widehat{\Gamma}_k(O) \mid R_k = r\}$. The regular influence function is

$$\phi^{\dagger}_{k,q} = \mathbf{1}\{R_k > 1 - q\}\{\widehat{\Gamma}_k - \widehat{g}_k(1 - q)\} - \widetilde{G}_k(q) + q\widehat{g}_k(1 - q). \tag{17}$$

The next conditions justify this expansion uniformly along the learned array.

A.2 CONDITIONAL ASSUMPTIONS

Index the triangular array by total sample size n, suppressing it when clear. The number of ordered folds K is fixed. Let $m_k = |I_k|$, $n_{\mathrm{ev}} = \sum_{k=2}^{K} m_k$, and $r_k = m_k / n_{\mathrm{ev}}$. Evaluation shares r_k determine sampling precision; deployment weights lie in Λ. A fresh draw (O, U) is independent of every training history. Conditional probability and expectation given $\mathcal{H}_{k-1}$ are written P_{k-1} and E_{k-1}. The target distribution of O is common across folds; $\mathcal{H}_K$ denotes the completed historical evaluation.

A1–A3 form the *Base* conditions used for causal centering. The *Regular* conditions are Base plus A4; the *Frozen* conditions are Regular plus A5, stated in Appendix B.3. The *Atomic* conditions replace A4 by the finite-level, studentizer/quantile, and stable-geometry requirements in Proposition 7.

A1 (Base: sampling and predictability). The observations are i.i.d.; the ordered partition is generated independently of their values; $\min_{k\geq 2} m_k/n \geq c > 0$ eventually. The score s_k, the nuisance fit $\widehat{\eta}_k$, and every tuning decision used on I_k are $\mathcal{H}_{k-1}$-measurable. The tie keys U_i are i.i.d. uniform and independent of the data and partition. Any outcome-free rank-calibration covariates and recorded keys are drawn independently before I_2, included in $\mathcal{H}_1$, and held fixed across all evaluation gates; they are never used to fit or tune a raw score or nuisance function.

A2 (Base: causal identification and moments). Consistency, no interference, conditional exchangeability, and $\epsilon \leq e(X) \leq 1-\epsilon$ hold. For some $\delta > 0$,

$$\sup_{n,k} E_{k-1}\left|\Gamma(O;\widehat{\eta}_k)\right|^{2+\delta} = O_p(1), \tag{18}$$

Estimated propensities are clipped inside a fixed overlap interval.

A3 (Base: conditional signal bias). Let

$$b_k(x) = E_{k-1}\{\Gamma(O;\widehat{\eta}_k) \mid X = x\} - \tau(x). \tag{19}$$

Set $\mathcal{Q}^+ = \mathcal{Q} \cup \{1\}$ and $\pi_{s_k,1} \equiv 1$. We require

$$\max_{k\geq 2} \max_{q\in\mathcal{Q}^+} \sqrt{m_k}\left|E_{k-1}\{\pi_{s_k,q}(X,U)b_k(X)\}\right| = o_p(1). \tag{20}$$

The $q = 1$ coordinate is $\max_k \sqrt{m_k}|E_{k-1}b_k(X)| = o_p(1)$. It controls the AIPW mean and, together with the $q \in \mathcal{Q}$ coordinates, the Qini pseudo-target bias $\rho_k(q) - q\rho_k(1)$. A3 imposes no ranking-boundary condition at $q = 1$. In an RCT with $\widehat{e}_k = e$ known, $b_k(x) = 0$ exactly for any history-measurable outcome regressions satisfying A2. In an observational study, direct calculation gives

$$b_k(x) = \{\widehat{e}_k(x) - e(x)\}\left\{\frac{\widehat{\mu}_{1,k}(x) - \mu_1(x)}{\widehat{e}_k(x)} + \frac{\widehat{\mu}_{0,k}(x) - \mu_0(x)}{1-\widehat{e}_k(x)}\right\}. \tag{21}$$

Thus overlap and $\|\widehat{e}_k - e\|_2 \max_a \|\widehat{\mu}_{a,k} - \mu_a\|_2 = o_p(m_k^{-1/2})$ are sufficient (Chernozhukov et al., 2018). The expansion below uses the actual fitted signal.

A4 (Regular: actual-signal boundaries). Let $Z_k = (s_k(X), U)$ with lexicographic order and let H_k be its conditional CDF. The distribution is non-atomic and $R_k = H_k(Z_k)$ is uniform. For $r_q = 1 - q$, $q \in \mathcal{Q}$, choose versions of

$$\widehat{g}_{k,n}(r) = E_{k-1}\{\Gamma(O;\widehat{\eta}_k) \mid R_k = r\} \tag{22}$$

such that $\max_{k,q}|\widehat{g}_{k,n}(r_q)| = O_p(1)$ and, for every fixed $M < \infty$,

$$\max_{k,q} \sup_{\substack{|t|\leq M,\\ r_q + t/\sqrt{m_k}\in[0,1]}} \left|\widehat{g}_{k,n}\left(r_q + \frac{t}{\sqrt{m_k}}\right) - \widehat{g}_{k,n}(r_q)\right| = o_p(1). \tag{23}$$

Thus the actual-signal regression is locally stochastically equicontinuous at the root-m_k quantile-fluctuation scale. Ordinary continuity separately for each n is not enough: it would allow a transition whose width shrinks like $m_k^{-1/2}$. A4 concerns the realized learned rank and neither compares s_k with an oracle score nor requires $s_k(X)$ to have a Lebesgue density. Conditional variances of all coordinates included in a studentized maximum are bounded away from zero in probability; deterministically zero coordinates, such as Qini at $q = 1$, are omitted from that maximum and appended exactly.

Bootstrap weights are independent across units and folds, independent of the data, nonnegative, and have mean one, variance one, and a finite $(2+\delta)$ moment. Exponential weights used in the code satisfy these conditions; exchangeable alternatives may be used under Præstgaard & Wellner (1993).

A.3 Regular fold inference

Proposition 4 (Regular conditional fold band)**.** *Fix the finite set $\mathcal{Q}$. Under the Regular conditions (A1–A4), conditionally on $\mathcal{H}_{k-1}$,*

$$\sqrt{m_k}\{\widehat{G}_k(q) - G_k(q) : q \in \mathcal{Q}\} = m_k^{-1/2}\sum_{i\in I_k}\{\phi^{\dagger}_{k,q}(O_i,U_i) : q \in \mathcal{Q}\} + o_p(1). \tag{24}$$

The re-ranking exchangeably weighted law consistently estimates this conditional vector law. Its max-t band $B_k = [L_k, U_k]$ therefore has conditional simultaneous coverage $1 - \alpha_k + o_p(1)$. Under Base at $q = 1$, append the gain coordinate $G_k(1) = E\tau(X)$ using influence $\widehat{\Gamma}_k - E_{k-1}\widehat{\Gamma}_k$; joint max-t calibration after $(G, \theta) \mapsto G - q\theta$ gives simultaneous gain–Qini coverage.

Exact empirical allocation. Let nonnegative record weights sum to W. Sort the records decreasingly by (S, U), take all records until cumulative weight first reaches qW, and give the boundary record fraction $(qW - W_{\text{prior}})/w_{\text{boundary}}$. The allocation lies in $[0, 1]$, is fractional for at most one record, and its weighted sum is exactly qW. Use this allocation for the point estimate and every weighted draw.

The boundary-estimation term in equation (17) follows by differentiating the upper-tail functional: moving its lower rank boundary by dr changes its value by $-\widehat{g}_k(1-q)dr$. The following argument controls this derivative along the learned array.

Proof of Proposition 4. Condition on $\mathcal{H}_{k-1}$, abbreviate $m = m_k$ and $V_i = \widehat{\Gamma}_k(O_i)$, and let $\mathbb{F}_m$ be the empirical CDF of the uniform ranks R_i. For $r_q = 1 - q$, the empirical allocation boundary $\widehat{r}_q$ satisfies, jointly over the finite grid,

$$\sqrt{m}(\widehat{r}_q - r_q) = -\sqrt{m}\{\mathbb{F}_m(r_q) - r_q\} + o_p(1). \tag{25}$$

In particular $\max_q \sqrt{m}|\widehat{r}_q - r_q| = O_p(1)$. The only triangular-array term not controlled by the fixed-law derivative is the population boundary remainder

$$\sqrt{m} \int_{r_q}^{\widehat{r}_q} \{\widehat{g}_{k,n}(r) - \widehat{g}_{k,n}(r_q)\}\, dr. \tag{26}$$

On $\{\sqrt{m}|\widehat{r}_q - r_q| \le M\}$ its absolute value is bounded by $\sqrt{m}|\widehat{r}_q - r_q|$ times the modulus in equation (23); first let $n \to \infty$ and then $M \to \infty$. Hence equation (26) is $o_p(1)$. The centered empirical-process increment over the random interval between r_q and $\widehat{r}_q$ is also $o_p(1)$ by the VC local maximal inequality and A2 (van der Vaart & Wellner, 1996); truncation handles the $(2+\delta)$-moment envelope. Finally, the one fractional boundary record contributes at most $m^{-1/2} \max_{i \in I_k} |V_i| = o_p(1)$. Combining these facts with equation (25) gives

$$\sqrt{m_k}\{\widehat{G}_k(q) - \widetilde{G}_k(q) : q \in \mathcal{Q}\} = \frac{1}{\sqrt{m_k}} \sum_{i \in I_k} \{\phi^{\dagger}_{k,q}(O_i, U_i) : q \in \mathcal{Q}\} + o_p(1). \tag{27}$$

A3 on $\mathcal{Q}$ replaces $\widetilde{G}_k$ by G_k. For the appended mean let $\widehat{\theta}_k = m_k^{-1} \sum_{i \in I_k} V_i$, $\widetilde{\theta}_k = E_{k-1}V_i$, and $\theta = E\tau(X)$. The $q = 1$ coordinate of A3 gives $\sqrt{m_k}|\widetilde{\theta}_k - \theta| = \sqrt{m_k}|E_{k-1}b_k(X)| = o_p(1)$. Hence A2 gives the joint conditional Lindeberg CLT for the gain coordinates and the mean. In an RCT both target replacements are identities, not empirical-process approximations.

For the bootstrap, let $\widehat{r}^{\#}_q$ be the boundary after normalizing the exchangeable weights to total mass one. Conditional multiplier quantile tightness puts both $\widehat{r}^{\#}_q$ and $\widehat{r}_q$ in an $O_{P^\#}(m^{-1/2})$ neighborhood of r_q. Repeating the preceding boundary decomposition, now conditionally on the data, equation (23) and the multiplier local maximal inequality give

$$\max_{q \in \mathcal{Q}} \left| \sqrt{m}\{\widehat{G}^{\#}_k(q) - \widehat{G}_k(q)\} - \frac{1}{\sqrt{m}} \sum_{i \in I_k} (W_i - 1)\phi^{\dagger}_{k,q}(O_i, U_i) \right| = o_{P^\#}(1) \tag{28}$$

in probability. Write $\bar{W} = m^{-1} \sum_{i \in I_k} W_i$ and $\widehat{\theta}^{\#}_k = \sum_{i \in I_k} W_i V_i / \sum_{i \in I_k} W_i$. The appended weighted mean satisfies

$$\begin{aligned} \sqrt{m}(\widehat{\theta}^{\#}_k - \widehat{\theta}_k) &= \frac{1}{\bar{W}\sqrt{m}} \sum_{i \in I_k} (W_i - 1)(V_i - \widehat{\theta}_k) \\ &= \frac{1}{\sqrt{m}} \sum_{i \in I_k} (W_i - 1)\{V_i - E_{k-1}V_i\} + o_{P^\#}(1) \end{aligned}$$

in probability. Here $\bar{W} - 1 = O_{P^\#}(m^{-1/2})$ and $\widehat{\theta}_k - E_{k-1}V_i = O_p(m^{-1/2})$, so weight normalization and sample-mean centering preserve the conditional limit. Jointly with equation (28), this gives the

same augmented Gaussian limit as the original gain–mean vector. Together with A3 at $q = 1$, the fixed map $(G, \theta) \mapsto G - q\theta$ therefore gives joint gain–Qini coverage and the treat-all coordinate centered at $E\tau(X)$.

B FEEDBACK FAILURE AND STABLE POOLING

B.1 NEAR-TIED SELECTION AND ITS RCT REALIZATION

Proof of Proposition 1. Write $Z = \rho V + \sqrt{1-\rho^2}E$ for independent standard normals E, V, W. Conditional on $V = v$, the numerator in (1) has mean ρv and variance $1 - \rho^2 + \sigma^2(v + h)$. Hence

$$F_h(x) = \int_{-\infty}^{-h} \Phi\left(\frac{x\sqrt{1+\sigma_-^2} - \rho v}{\sqrt{1-\rho^2+\sigma_-^2}}\right) \varphi(v)\, dv + \int_{-h}^{\infty} \Phi\left(\frac{x\sqrt{1+\sigma_+^2} - \rho v}{\sqrt{1-\rho^2+\sigma_+^2}}\right) \varphi(v)\, dv. \tag{29}$$

If $\rho = 0$, the integrands are $\Phi(x)\varphi(v)$; if the scales are equal, the numerator divided by its standard deviation is a fixed Gaussian linear combination. These give the two normal recovery cases. Conversely, the difference of the two integrands at zero is odd apart from φ, and symmetry yields

$$F_h(0) - \frac{1}{2} = \int_{|h|}^{\infty} \left[\Phi\left(\frac{-\rho v}{\sqrt{1-\rho^2+\sigma_+^2}}\right) - \Phi\left(\frac{-\rho v}{\sqrt{1-\rho^2+\sigma_-^2}}\right)\right] \varphi(v)\, dv. \tag{30}$$

For finite h, $\rho \neq 0$ and unequal scales, this integrand has a strict, constant sign for $v > |h|$, proving nonnormality. Finally couple U_h with the statistic using its more probable branch on every observation. That statistic is standard normal and differs from U_h only on an event of probability $\Phi(-|h|)$, proving (2). □

A5 below gives sufficient conditions for general frozen pooling; independent selection provides another normal recovery case.

An i.i.d. randomized-trial realization. For both evaluation cohorts, independently sample $X \sim \mathrm{Unif}[-1, 1]$, $A \sim \mathrm{Bernoulli}(1/2)$, and $E \sim N(0, 1)$. Set

$$Y(a) = 2a\delta_m X + \tfrac{1}{2}\{\mathbf{1}(X > 0) + r\mathbf{1}(X \leq 0)\}E, \qquad \delta_m = \frac{h\sqrt{(1+r^2)/2}}{\sqrt{m}}. \tag{31}$$

Both original nuisance regressions are identically zero, so the actual AIPW signal is $\Gamma = 2(2A-1)Y$, with conditional mean $2\delta_m X$. The fixed candidate scores $s_+(X) = X$ and $s_-(X) = -X$ have exact-half-budget gains $G_+ = \delta_m/2$ and $G_- = -\delta_m/2$. Evaluate s_+ on the first cohort. After observing that cohort, select s_+ for the next stage precisely when $\widehat{G}_+ - \widehat{G}_- > 0$; otherwise select s_-. Evaluate that selection on the next unopened cohort. All dependence is generated by validation-based rule selection, while the observation law is identical across cohorts.

As $m \to \infty$ with fixed h, r, the centered first gain and validation contrast have joint Gaussian limits with variances $1/2$ and $(1 + r^2)/2$ and covariance $1/2$. The actual-signal boundary regression is $2\delta_m X$, continuous at both median cutoffs with value zero. Thus the empirical rank expansions give

$$\rho = \frac{1}{\sqrt{1+r^2}}, \qquad \sigma_+ = 1, \qquad \sigma_- = r. \tag{32}$$

The selected second-cohort normalized error is asymptotically standard normal conditionally on the entire first cohort. Centering the equal-weight portfolio at its own realized target therefore gives (1). At $h = 0, r = 4$, the exact limiting one-sided coverage is $F_0(\Phi^{-1}(.95)) = 0.94542149$.

B.2 VALIDATION-BASED CANDIDATE SELECTION IN ONE I.I.D. RCT

The protocol is (31) with $r \in \{1, 4\}$, $h \in \{-3, 0, 3\}$ and two evaluation cohorts of size $m \in \{400, 1600\}$. Each cell has 10,000 replications, sharing random numbers across r, h; the 120,000 setting records are therefore dependent. Candidate selection uses the empirical exact-half-budget gain on the first cohort. Variances use the empirical second moments of the actual signal at the evaluated cuts, consistent with the zero boundary regression in the local limit. No bootstrap tuning is used.

Frozen inference covers the fixed equal-weight portfolio. Radial inference protects the entire two-version simplex using its declared dimension two. The suffix comparator evaluates the first score on both cohorts, re-ranks the pooled sample, evaluates the selected version on the second cohort, and spends 0.025 per version. Its overlapping version estimates are not treated as independent. Table 3 reports full-simplex coverage for the two protected methods, a stronger event than the frozen fixed-weight event.

Table 3: I.i.d. RCT selection diagnostic: coverage in percent. Each empirical cell uses 10,000 replications; the "Frozen limit" column is numerical integration of (29).

m	r	h	Frozen limit	Frozen	Radial	Suffix rectangle
400	1	-3	95.00	95.15	94.80	95.18
400	1	0	95.00	94.91	95.03	95.93
400	1	3	95.00	95.08	95.24	96.09
400	4	-3	94.99	95.05	94.98	95.11
400	4	0	94.54	94.51	95.10	95.78
400	4	3	95.00	95.31	95.71	96.16
1,600	1	-3	95.00	95.30	94.96	95.04
1,600	1	0	95.00	95.24	94.98	95.56
1,600	1	3	95.00	95.19	95.31	95.89
1,600	4	-3	94.99	94.76	94.86	94.94
1,600	4	0	94.54	94.49	94.93	95.43
1,600	4	3	95.00	95.22	94.98	95.96

At $r = 4, h = 0$, frozen coverage is 94.51% (95% Wilson interval $[94.05, 94.94]$) and 94.49% ($[94.03, 94.92]$) for $m = 400, 1600$. Both agree with the 94.54% limit. Equal-scale and separated-gap rows serve as controls. The linear effect in (31) avoids quantile-boundary bias and isolates the selection mechanism.

B.3 WHEN FROZEN POOLING IS VALID

A5 (Frozen: covariance stability for pooling). Fix $\lambda_n \in \Lambda$ that is $\mathcal{H}_1$-measurable. For $q \in \mathcal{Q}$ use equation (17), and append the mean influence $\phi^\dagger_{k,1} = \Gamma(O; \widehat{\eta}_k) - E_{k-1}\Gamma(O; \widehat{\eta}_k)$. For $q, q' \in \mathcal{Q}^+$, let

$$V_n^\lambda(q, q') = n_{\text{ev}} \sum_{k=2}^{K} \frac{\lambda_k^2}{m_k} E_{k-1}\{\phi^\dagger_{k,q}\phi^\dagger_{k,q'}\}. \tag{33}$$

There is a sequence of positive-definite random matrices Σ_n, each measurable with respect to the burn-in sigma-field $\mathcal{H}_1$, whose eigenvalues are bounded away from zero and infinity in probability, such that

$$E\left[1 \wedge \max_{q,q' \in \mathcal{Q}^+} |V_n^\lambda(q, q') - \Sigma_n(q, q')| \,\middle|\, \mathcal{H}_1\right] \to_p 0. \tag{34}$$

A5 concerns the augmented gain–mean vector; Qini is its fixed linear image. Neither Σ_n nor the learner need converge. The condition excludes first-order covariance branches or deployment weights driven by innovations from $I_2, \ldots, I_K$; predictability alone does not imply it.

Theorem 5 (Stable frozen-path pooling)**.** *Under the Frozen conditions (Regular plus A5), the pooled vector* $S_n^\lambda = \{\sqrt{n_{\text{ev}}}(\widehat{G}^\lambda(q) - G^\lambda(q)) : q \in \mathcal{Q}^+\}$*, with* $q = 1$ *appended rather than ranked, and its*

frozen-path weighted analogue $T_n^{\#,\lambda}$ obey

$$\sup_{f\in \mathrm{BL}_1}\left|E\{f(S_n^\lambda)\mid \mathcal{H}_1\}-\int f\,dN(0,\Sigma_n)\right|\to_p 0,$$
$$\sup_{f\in \mathrm{BL}_1}\left|E^{\#}f(T_n^{\#,\lambda})-\int f\,dN(0,\Sigma_n)\right|\to_p 0. \tag{35}$$

Thus aggregating re-ranking weighted fold deviations before max-t calibration is valid under Frozen; the fixed map $(G(q),G(1))\mapsto G(q)-qG(1)$ gives the joint Qini statement. If Σ_n itself converges stably, the display also yields the corresponding mixed-normal limit. The conclusion generally fails without Frozen's stability condition, as Proposition 1 shows.

Proof of Theorem 5. Order all evaluation observations by fold and then within fold. For $i\in I_k$ and $q\in\mathcal{Q}^+$, the scaled vector increment is

$$D_{i,n}^{\lambda}(q)=\frac{\sqrt{n_{\mathrm{ev}}}\,\lambda_k}{m_k}\phi_{k,q}^{\dagger}(O_i,U_i). \tag{36}$$

It is a martingale difference because the rule, nuisance fit, and conditional centering are measurable before I_k is opened. A2 and positive fold fractions imply the conditional Lindeberg condition. Moreover, fixed K, positive fold fractions, and A3 on $\mathcal{Q}^+$ give

$$\max_{q\in\mathcal{Q}^+}\sqrt{n_{\mathrm{ev}}}\left|\sum_{k=2}^{K}\lambda_k\rho_k(q)\right|\le\sum_{k=2}^{K}\lambda_k\sqrt{\frac{n_{\mathrm{ev}}}{m_k}}\max_{q\in\mathcal{Q}^+}\sqrt{m_k}|\rho_k(q)|=o_p(1). \tag{37}$$

Thus the augmented actual-signal vector is centered at the true gain and mean targets. Its predictable quadratic covariation is equation (33). Equation (34) makes that covariation conditionally close to the $\mathcal{H}_1$-measurable proxy Σ_n. The conditional martingale CLT therefore gives the first bounded-Lipschitz approximation in equation (35) (Hall & Heyde, 1980).

Apply the regular fold expansion to every weighted draw. Conditional laws within folds obey the multiplier CLT, and their empirical covariance sum converges to equation (33). A5 then gives the second bounded-Lipschitz approximation in equation (35). The continuous mapping theorem gives max-t calibration whenever the included coordinate variances are nonzero. A5 is exactly the step that fails in the feedback counterexample: conditional on burn-in, its quadratic variation retains the random branch $\mathrm{sign}(Z)$.

For a sufficient condition, suppose the finite-dimensional fitted parameters converge uniformly over stages to a burn-in-measurable $\vartheta_{0,n}$ at rate r_n. If their augmented influence covariance map C_n is locally L_n-Lipschitz, $L_nr_n\to 0$, and $C_n(\vartheta_{0,n})$ has uniformly bounded positive eigenvalues, then A5 holds with $\Sigma_n=(\sum_k\lambda_k^2/r_k)C_n(\vartheta_{0,n})$. Indeed, the covariance difference is $O_p(L_nr_n)=o_p(1)$; bounded truncation and conditional Markov give (34). For fixed-dimensional strongly convex ERM, a root-n parameter expansion and local covariance Lipschitz continuity meet the rate requirement.

C Proofs of Protection and Selection

The main statements use gain on $\mathcal{Q}$. Their extensions below append the mean and treat-none coordinates, giving $\mathcal{D}$, and can include foldwise Qini in the same error budget. Conditional bands must cover the complete declared coordinate list before any selection.

C.1 Proof of Theorem 2

Let $E_k=\{G_k(q)\in B_k(q)\text{ for all }q\in\mathcal{D}\}$. Iterated expectation gives $P(E_k^c)\le E(\alpha_k)+o(1)$ from the stated conditional guarantee. Hence

$$P\left(\bigcap_{k=2}^{K}E_k\right)\ge 1-E\left(\sum_{k=2}^{K}\alpha_k\right)-o(1)\ge 1-\alpha-o(1). \tag{38}$$

On this intersection, for every $\lambda\in\Lambda$, multiplying fold k's inequality by λ_k and summing yields equation (4) over the entire simplex, including any measurable $\widehat{\lambda}$ selected after evaluation. No joint

law for fold errors is used. For a prespecified face, only its components need receive error; an unprotected component cannot be reactivated after evaluation.

For joint inference, let E_k cover both $G_k(q)$ and $\Delta_k(q) = G_k(q) - qG_k(1)$ on $\mathcal{D}$. Propositions 4 and 7, using A3 at $q = 1$, supply its conditional probability. The same union and convex-averaging argument applies coordinatewise, proving Theorem 2's joint statement.

C.2 Evaluating each version on all later cohorts

For version k, retain the original pair $(s_k, \widehat{\eta}_k)$ and evaluate on $J_k = \bigcup_{j=k}^{K} I_j$, with $M_k = |J_k|$. By A1, conditional on $\mathcal{H}_{k-1}$ this is an i.i.d. sample independent of the fitted pair. Apply the single-version construction to the pooled suffix, including its own empirical ranking threshold, AIPW mean, and sample-size scaling. Let $E_k^{\triangle}$ be its simultaneous component-coverage event over the declared coordinates. Conditional validity gives $P\{(E_k^{\triangle})^c \mid \mathcal{H}_{k-1}\} \le \alpha_k + o_p(1)$. Boundedness of probabilities and iterated expectation imply $P\{(E_k^{\triangle})^c\} \le E\alpha_k + o(1)$. Fixed K therefore gives

$$P\left(\bigcap_{k=2}^{K} E_k^{\triangle}\right) \ge 1 - \sum_{k=2}^{K} E\alpha_k - o(1) \ge 1 - \alpha - o(1). \tag{39}$$

On this event all convex combinations are covered, exactly as in Theorem 2. Neither disjoint suffixes nor independence between version estimates is needed. Later learners may use observations in J_k; the argument conditions on $\mathcal{H}_{k-1}$, not the final learned archive.

Under A1's positive limiting fold fractions, $M_k/m_k = O(1)$ uniformly in k. Thus the A3 bias requirement transfers from $\sqrt{m_k}$ to $\sqrt{M_k}$, and A4's local equicontinuity holds on the smaller $M_k^{-1/2}$ neighborhood. Moments and conditional variances concern the same fitted signal. The proof of Proposition 4 then applies with sample size M_k. Atomic inference instead requires its single-version conditions at this size and an associated subsample sequence; the implementation uses $b_k = \lfloor M_k^{0.75} \rfloor$. These overlapping errors do not satisfy the disjoint-increment argument for Theorem 3.

C.3 Radial geometry and proof of Theorem 3

For a two-sided band, let $c_d^{\pm}(u)$ be the u quantile of $\max\{\sum_{j=1}^{d}(Z_j^+)^2, \sum_{j=1}^{d}(Z_j^-)^2\}$, with $z^- = \max(-z, 0)$. Replacing c_d^+ by $c_d^{\pm}$ in equation (6) gives the two-sided radius. The same construction applies under the Atomic conditions on a prespecified finite-valued face separated from every cumulative level mass. Gain, the appended mean, and Qini coordinates may share one finite error allocation.

Fix one scalar coordinate $q \in \mathcal{Q}^+$ and, for $k \in S$, set

$$Z_{k,n}(q) = \frac{\widehat{G}_k(q) - G_k(q)}{\widehat{\mathrm{se}}_k(q)}. \tag{40}$$

For a regular continuous coordinate, Proposition 4, A3, and consistency of the fold studentizer imply, for every fixed $t \in \mathbb{R}$,

$$E\{\exp(itZ_{k,n}(q)) \mid \mathcal{H}_{k-1}\} = \exp(-t^2/2) + o_p(1). \tag{41}$$

At a stable atomic coordinate separated from every cumulative level mass, equation (58) is linear; Lemma 10 and its studentizer conclusion give the same display. Exact equalities use the kinked derivative in equation (59) and are therefore excluded from this radial route. The same statement at $q = 1$ follows from the appended mean coordinate. The left side is bounded, so convergence in probability gives the L_1 control needed under iterated expectation. Starting from the last eligible fold and applying equation (41) recursively shows that, for every fixed vector $(t_k : k \in S)$,

$$E\exp\left\{i\sum_{k\in S} t_k Z_{k,n}(q)\right\} \longrightarrow \exp\left(-\frac{1}{2}\sum_{k\in S} t_k^2\right). \tag{42}$$

Thus $Z_{S,n}(q) \Rightarrow N(0, I_d)$ although later scores and conditional variances may depend arbitrarily on earlier evaluation innovations.

For the two-sided statement, define

$$M_{S,n}(q) = \max\left\{\sum_{k\in S}(Z^+_{k,n}(q))^2, \sum_{k\in S}(Z^-_{k,n}(q))^2\right\}.$$

On the event $M_{S,n}(q) \le c_d^{\pm}(1-\beta_q)$, Cauchy–Schwarz gives, simultaneously for every $\lambda \in \Lambda(S)$,

$$\left|\widehat{G}^{\lambda}(q) - G^{\lambda}(q)\right| \le \left\{c_d^{\pm}(1-\beta_q)\sum_{k\in S}\lambda_k^2\widehat{\mathrm{se}}_k^2(q)\right\}^{1/2}. \tag{43}$$

Conditional on exactly m positive coordinates, the positive and negative squared norms are independent χ^2_m and χ^2_{d-m} variables. Hence

$$P(M_d^{\pm} \le x) = 2^{-d}\sum_{m=0}^{d}\binom{d}{m}F_{\chi^2_m}(x)F_{\chi^2_{d-m}}(x), \tag{44}$$

with $\chi^2_0 = 0$; equation (42) and continuity at the critical value establish the claimed probability. For a lower band, only positive estimation errors matter:

$$\widehat{G}^{\lambda}(q) - G^{\lambda}(q) \le \sum_{k\in S}\lambda_k\widehat{\mathrm{se}}_k(q)Z^+_{k,n}(q) \le \left\{\sum_{k\in S}\lambda_k^2\widehat{\mathrm{se}}_k^2(q)\sum_{k\in S}(Z^+_{k,n}(q))^2\right\}^{1/2}. \tag{45}$$

The bound is attained simultaneously in its normalized form. If $Z^+_{S,n}(q) \neq 0$, choose $\lambda_k = \{Z^+_{k,n}(q)/\widehat{\mathrm{se}}_k(q)\}/\sum_{\ell\in S}Z^+_{\ell,n}(q)/\widehat{\mathrm{se}}_\ell(q)$. Then the normalized positive error equals $\|Z^+_{S,n}(q)\|_2$; if all coordinates are nonpositive both sides of equation (8) are zero. Hence the fixed-budget coverage event for any common multiplier $c \ge 0$ is exactly $\{\|Z^+_{S,n}(q)\|_2 \le c\}$. The limit law is continuous and strictly increasing at the stated positive critical value. Coverage therefore converges to $1-\beta_q$, and any fixed smaller nonnegative multiplier has strictly smaller limiting coverage.

For independent standard normals, the number of positive coordinates is $M \sim \mathrm{Binomial}(d, 1/2)$ and, conditional on $M = m$, their squared sum is χ^2_m (with $\chi^2_0 = 0$). Hence its CDF is

$$F_d^+(x) = 2^{-d}\sum_{m=0}^{d}\binom{d}{m}F_{\chi^2_m}(x), \tag{46}$$

whose quantile is c_d^+. A union bound over the fixed coordinate list costs $\sum_q \beta_q \le \alpha$; each coordinate event already holds for the entire face, so no further multiplicity is paid for data-selected λ. Applying the same argument to each scalar foldwise Qini coordinate proves the joint gain–Qini extension when their error spends are included in the same finite sum.

C.4 Post-selection and known transformations

Corollary 6 (Post-weight and post-budget selection)**.** *Let $L_{\mathcal{U}}(q,\lambda) = L_G(q,\lambda) - C(q)$ and let $\Lambda^\star$ be Λ under Theorem 2, or the prespecified regular face $\Lambda(S)$ under Theorem 3. For any measurable $(\widehat{q},\widehat{\lambda}) \in \mathcal{D}\times\Lambda^\star$ chosen from the scores, estimates, bands, or the entire displayed surface,*

$$\liminf_n \mathbb{P}\{\mathcal{U}^{\widehat{\lambda}}(\widehat{q}) \ge L_{\mathcal{U}}(\widehat{q},\widehat{\lambda})\} \ge 1-\alpha. \tag{47}$$

For Theorem 5, the same statement holds for post-selection of q with λ fixed before evaluation innovations. Treat-none has known gain zero; treat-all uses the jointly calibrated AIPW mean.

Proof. Let $E_n = \{\mathcal{U}^{\lambda}(q) \ge L_{\mathcal{U}}(q,\lambda)$ for every $(q,\lambda)\in\mathcal{D}\times\Lambda^\star\}$. On E_n, the inequality holds at $(\widehat{q},\widehat{\lambda})$ regardless of how the pair was selected. Hence $P\{\mathcal{U}^{\widehat{\lambda}}(\widehat{q}) \ge L_{\mathcal{U}}(\widehat{q},\widehat{\lambda})\} \ge P(E_n) \ge 1-\alpha+o(1)$. The radial construction supplies this event on its prespecified face. A frozen-path band supplies an event only at its burn-in-measurable λ, so the identical argument then selects only q.

When protection changes deployment. Consider the same proposed policy with true net utility u_n, estimate $\widehat{u}_n$, and nested lower bounds $\ell_F = \widehat{u}_n - a_n$, $\ell_P = \widehat{u}_n - b_n$, where $0 \le a_n \le b_n$. This includes the matched frozen/radial comparison at fixed weights. Put $D_n = \{\ell_P \le 0 < \ell_F\}$ and let $W_j = u_n \mathbf{1}\{\ell_j > 0\}$ include the treat-none fallback. Then, exactly,

$$W_P - W_F = -u_n \mathbf{1}(D_n), \qquad D_n \subseteq \{|u_n| \le |\widehat{u}_n - u_n| + b_n\}.$$

The first identity partitions the value of protection into avoided losses and forgone gains. For the second, disagreement requires $a_n < \widehat{u}_n \le b_n$, hence the stated inequality by the triangle inequality. If estimation error and both penalties are $O_p(r_n)$, $r_n \downarrow 0$, and $|u_n|/r_n \to \infty$ in probability, then $P(D_n) \to 0$. Thus separated decisions are asymptotically unchanged.

For a harmful policy near the decision threshold, suppose $u_n = -\kappa r_n$, $\kappa > 0$, and

$$((\widehat{u}_n - u_n)/r_n, a_n/r_n, b_n/r_n) \Rightarrow (T, R_F, R_P), \qquad P(T = \kappa + R_F \text{ or } T = \kappa + R_P) = 0.$$

The identity, convergence of the disagreement event, and boundedness of its indicator give

$$r_n^{-1}\mathbb{E}(W_P - W_F) \longrightarrow \kappa P\{\kappa + R_F < T \le \kappa + R_P\}.$$

This limit is strictly positive when that interval has positive probability. For the two-cohort equal-gain Gaussian feedback experiment ($h = 0$), take $T = Z + \sigma(V)W$ and $(R_F, R_P) = (c_F, c_P)\sqrt{1 + \sigma^2(V)}$, with $c_P > c_F$. Conditional on (Z, V), W has full normal support, so the probability is positive for every finite $\kappa > 0$. Changing the known cost realizes this local margin without changing the candidate comparison. Positive local margins reverse the utility sign. The identity and margin argument apply to any nested bounds on the same proposal, including the matched comparison in Table 2.

C.5 Optimization and computational cost

At fixed q, a rectangular lower surface is linear in λ and attains its maximum at a vertex. For the radial surface, let $a_{kq} = \widehat{G}_k(q) - C(q)$, $s_{kq} = \widehat{\mathrm{se}}_k(q) > 0$, and let $c > 0$ be the critical multiplier of the declared face. The optimum is given by

$$\sum_{k \in S} \left\{ \frac{(a_{kq} - \mu_q)_+}{s_{kq}} \right\}^2 = c^2, \qquad \widehat{\lambda}_{kq} = \frac{(a_{kq} - \mu_q)_+ / s_{kq}^2}{\sum_{\ell \in S} (a_{\ell q} - \mu_q)_+ / s_{\ell q}^2}. \tag{48}$$

The root $\mu_q < \max_k a_{kq}$ is unique and equals the optimized lower utility. Each root evaluation costs $O(d)$; comparison across budgets and with zero completes selection. The declared dimension remains fixed even if some optimized weights are zero.

For one regular evaluation sample of size m, let $N_\#$ be the number of resampling draws and $J_+ = |\mathcal{Q}^+|$. The implementation sorts (S, U) once in $O(m \log m)$ time. Each draw changes cumulative weights and boundary allocations, so re-ranking does not require sorting scores again. The vectorized implementation scans the m cumulative weights for each budget, using $O(m \log m + N_\# m J_+)$ work and $O(N_\#(m + J_+))$ storage, excluding model fitting. These costs add over version–sample pairs. The diagonal radial critical value uses the $d + 1$ chi-square terms in equation (46); the active-set computation in Appendix E.2 concerns the common terminal route.

D Atomic boundaries and subsampling

For the finite-valued atomic route we remove auxiliary-randomization noise. If the ordered score levels are $c_1 > \cdots > c_L$, define $\widehat{p}_{k\ell} = m_k^{-1} \sum_{i \in I_k} \mathbf{1}\{S_i = c_\ell\}$ and $\widehat{h}_{k\ell} = m_k^{-1} \sum_{i \in I_k} \widehat{\Gamma}_i \mathbf{1}\{S_i = c_\ell\}$. With $\widehat{j}_k(q) = \min\{j : \sum_{\ell \le j} \widehat{p}_{k\ell} \ge q\}$ and $\widehat{\zeta}_k(q) = \{q - \sum_{\ell < \widehat{j}_k(q)} \widehat{p}_{k\ell}\}/\widehat{p}_{k,\widehat{j}_k(q)}$, the atomic estimator is

$$\widehat{G}_k^{\mathrm{at}}(q) = \sum_{\ell < \widehat{j}_k(q)} \widehat{h}_{k\ell} + \widehat{\zeta}_k(q) \widehat{h}_{k,\widehat{j}_k(q)}. \tag{49}$$

Integrating the allocation key gives each boundary-level record the same fractional allocation. The estimator targets equation (14), meets the empirical budget exactly, and depends only on level masses and signal totals.

Finite-valued scores create a different boundary. Random tie ranks make population treatment mass exactly q but do not smooth a jump in $\widehat{g}_k(r)$. If q exactly exhausts a positive-mass score level and adjacent levels have different conditional effects, the finite-level map in equation (49) is only directionally differentiable; the ordinary bootstrap can fail (Fang & Santos, 2019). We use i.i.d. subsampling (Politis et al., 1999, Chapter 2): within fold k, draw b_k records without replacement, recompute the level map, and use

$$\sqrt{\frac{b_k}{m_k - b_k}}\{\widehat{G}^{\mathrm{at},\#}_{k,b} - \widehat{G}^{\mathrm{at}}_k\} \tag{50}$$

as a full-fold-scale deviation. The finite-population correction is asymptotically $\sqrt{b_k/m_k}$.

Proposition 7 (Nonregular conditional fold band)**.** *Fix $\mathcal{Q}$. Under Base (A1–A3), suppose, conditionally on $\mathcal{H}_{k-1}$, the score takes $L \le \bar{L}$ ordered values whose positive masses are bounded away from zero. Let $P_{k,\ell,n}$ denote their cumulative masses. Assume the boundary geometry is stable at the subsample scale,*

$$\min_{q\in\mathcal{Q}} \min_{\ell: q\neq P_{k,\ell,n}} \sqrt{b_k}\,|q - P_{k,\ell,n}| \;\to_p\; \infty, \tag{51}$$

where the minimum over an empty set is infinity; exact equalities $q = P_{k,\ell,n}$ are allowed. If $b_k \to \infty$, $b_k/m_k \to 0$, and the conditional coordinate scales are bounded away from zero, and the studentized directional Gaussian maximum satisfies the quantile continuity and identification conditions in Lemma 10, subsampling consistently estimates the conditional law of equation (49). Augmenting by the mean coordinate and applying $(G,\theta) \mapsto G - q\theta$ gives a joint gain–Qini band over $\mathcal{D}$ with conditional coverage $1 - \alpha_k + o_p(1)$.

This covers finite-leaf trees without a score density, score sup-norm rate, or margin exponent. Condition (51) excludes local-to-boundary sequences of order $m_k^{-1/2}$: because $b_k/m_k \to 0$, the subsample would otherwise see a different local kink from the full fold. At an exact equality, the directional subsampling band combines through Theorem 2.

Proposition 8 (Local drift at an atomic budget boundary)**.** *Consider an i.i.d. triangular array with a fixed finite-level score at one budget q, with adjacent level masses bounded away from zero and bounded conditional $(2+\delta)$ signal moments. Let T collect the levels above and including level ℓ, with $p_m = P(T) = q + h_m/\sqrt{m}$, $h_m \to h \in \mathbb{R}$, and let $H_m = E[\Gamma \mathbf{1}(T)]$. Suppose the adjacent conditional signal means converge to a and b, and $\sqrt{m}(\widehat{p} - p_m, \widehat{H} - H_m) \Rightarrow (Z_p, Z_H)$, a centered Gaussian limit. Define $f(t) = -at_+ + b(-t)_+$. For the integrated empirical exact-budget gain,*

$$\sqrt{m}\{\widehat{G}(q) - \widetilde{G}_m(q)\} \Rightarrow Z_H + f(h + Z_p) - f(h). \tag{52}$$

Here $\widetilde{G}_m$ is the actual-signal population value; Base bias control replaces it by G_m without changing the limit. For uniform subsampling without replacement, $b_m \to \infty$ and $b_m/m \to 0$, the conditional centered law instead converges in probability to the law of $Z_H + f(Z_p)$.

Proof. In a neighborhood of this boundary, the population value equals $H_m - a_m^{\mathrm{sig}}(p_m - q)_+ + b_m^{\mathrm{sig}}(q - p_m)_+$, where $a_m^{\mathrm{sig}}, b_m^{\mathrm{sig}}$ are the two adjacent conditional signal means. The integrated empirical value has the same identity with empirical masses, signals, and adjacent means. Adjacent-mean consistency and the root-m size of the boundary terms make the mean-estimation products negligible. Subtraction and the joint CLT give (52). At the subsample scale, $\sqrt{b_m}(p_m - q) = h_m\sqrt{b_m/m} \to 0$; centering at the full-fold estimate also differs by $O_p(\sqrt{b_m/m})$. The finite-population CLT then gives $Z_H + f(Z_p)$. Appending the mean influence Z_θ replaces the gain limits by the corresponding Qini limits minus qZ_θ. □

For $a = 0, b = 1$ and deterministic signals zero on T, the two gain limits are $(-h - Z_p)_+ - (-h)_+$ and $(-Z_p)_+$. For $h > 0$, their atoms at zero have probabilities $\Phi(h/\mathrm{sd}(Z_p))$ and $1/2$. Thus the local exclusion in (51) cannot be removed from that subsampling argument. At $h = 0$ the existing directional law is recovered; large positive and negative drift recover the adjacent linear faces.

Independent rank calibration into fixed, separated levels gives a practical way to meet these conditions; Corollary 11 below states the construction and its requirements.

Condition on $\mathcal{H}_{k-1}$, suppress k, and write $V = \Gamma(O; \widehat{\eta}_k)$. The learned score takes ordered values $c_1 > \cdots > c_L$. Define the level masses, actual-signal totals, and cumulative masses

$$p_\ell = P_{k-1}(S = c_\ell), \qquad h_\ell = E_{k-1}\{V\mathbf{1}(S = c_\ell)\}, \qquad P_j = \sum_{\ell \le j} p_\ell, \tag{53}$$

and let $\theta = (p_1, \ldots, p_L, h_1, \ldots, h_L)$. For $q < 1$, set

$$j_q(p) = \min\{j : P_j \ge q\}, \qquad \zeta_q(p) = \frac{q - P_{j_q(p)-1}}{p_{j_q(p)}}, \tag{54}$$

where $P_0 = 0$, and define the finite-dimensional exact-budget map

$$\Psi_{q,n}(p, h) = \sum_{\ell < j_q(p)} h_\ell + \zeta_q(p) h_{j_q(p)}. \tag{55}$$

The levels and their conditional law may change with n without convergence. Equation (55) integrates the independent allocation key. With

$$Y_i = \big(\mathbf{1}\{S_i = c_\ell\}, V_i\mathbf{1}\{S_i = c_\ell\} : 1 \le \ell \le L\big), \qquad \widehat{\theta} = m^{-1}\sum_{i=1}^{m} Y_i, \tag{56}$$

the estimator in equation (49) is exactly $\Psi_{q,n}(\widehat{\theta})$, with state dimension $2L \le 2\bar{L}$. Below, $\Psi_n(\theta) = \{\Psi_{q,n}(\theta) : q \in \mathcal{Q}\}$ denotes the corresponding finite vector. For joint inference append the linear mean map $\Theta(\theta) = \sum_{\ell=1}^{L} h_\ell = E_{k-1}V$ and use

$$\Psi_n^+(\theta) = \big(\{\Psi_{q,n}(\theta) : q \in \mathcal{Q}\}, \Theta(\theta)\big), \qquad \dot{\Theta}_\theta(a, b) = \sum_{\ell=1}^{L} b_\ell. \tag{57}$$

Qini coordinates are the fixed linear image $\Psi_{q,n} - q\Theta$. This augmentation is uniformly Lipschitz and introduces neither a new face nor a boundary condition.

Lemma 9 (Directional derivative of the finite-level map)**.** *Suppose* $\min_\ell p_\ell > 0$ *and write* $\mu_\ell = h_\ell/p_\ell$. *For a direction* $z = (a, b) \in \mathbb{R}^{2L}$ *with* $\sum_\ell a_\ell = 0$, *let* $A_j = \sum_{\ell \le j} a_\ell$. *If* $P_{j-1} < q < P_j$, *then*

$$\dot{\Psi}_{q,n,\theta}(a, b) = \sum_{\ell < j} b_\ell + \zeta_q b_j - \mu_j\{A_{j-1} + \zeta_q a_j\}. \tag{58}$$

If $q = P_j$ *for* $j < L$, *then, with* $x^+ = \max(x, 0)$ *and* $x^- = \max(-x, 0)$,

$$\dot{\Psi}_{q,n,\theta}(a, b) = \sum_{\ell \le j} b_\ell - \mu_j A_j^+ + \mu_{j+1} A_j^-. \tag{59}$$

The derivative is linear in the first case and continuous piecewise linear in the second. On sets where $\min_\ell p_\ell$ *and the relevant moments are bounded, all these maps have a common finite Lipschitz constant.*

Proof. On an interior face, $\Psi_q = \sum_{\ell < j} h_\ell + \{q - P_{j-1}\}h_j/p_j$. Differentiating this rational expression gives equation (58). At $q = P_j$, a perturbation with $A_j > 0$ leaves excess mass in level j and has derivative $\sum_{\ell \le j} b_\ell - \mu_j A_j$. A perturbation with $A_j < 0$ admits $-A_j$ mass from level $j + 1$ and has derivative $\sum_{\ell \le j} b_\ell - \mu_{j+1} A_j$. The formulas agree when $A_j = 0$, which gives equation (59) and continuity. The Lipschitz claim follows directly from these displays and the lower mass bound. □

Lemma 10 (Uniform conditional delta method and atomic subsampling)**.** *Consider the conditional triangular array above with* $L \le \bar{L}$, *positive level masses bounded away from zero in probability, A2, a fixed finite grid in* $(0, 1)$, *and condition (51). In this lemma,* Ψ_n *may denote the gain vector or any retained coordinate vector obtained from* Ψ_n^+ *by the fixed gain–Qini linear map. Let*

$$\Omega_n = \mathrm{Var}_{k-1}(Y_i), \qquad Z_n \mid \mathcal{H}_{k-1} \sim N(0, \Omega_n), \qquad W_n = \dot{\Psi}_{n,\theta}(Z_n). \tag{60}$$

Neither Ω_n nor W_n need converge. If $b \to \infty$ and $b/m \to 0$, then

$$d_{\mathrm{BL}}\Big(\mathcal{L}[\sqrt{m}\{\Psi_n(\widehat{\theta}) - \Psi_n(\theta)\} \mid \mathcal{H}_{k-1}], \mathcal{L}[W_n \mid \mathcal{H}_{k-1}]\Big) \to_p 0, \tag{61}$$

$$d_{\mathrm{BL}}\left(\mathcal{L}^{\#}\left[\sqrt{\frac{mb}{m-b}}\{\Psi_n(\widehat{\theta}_b^{\#}) - \Psi_n(\widehat{\theta})\}\right], \mathcal{L}[W_n \mid \mathcal{H}_{k-1}]\right) \to_p 0, \tag{62}$$

where $\widehat{\theta}_b^{\#}$ is computed on a simple random subsample without replacement and both displays are vector statements over the grid.

Let $\sigma_{r,n}^2 = \mathrm{Var}_{k-1}(W_{r,n})$ for every retained gain or Qini coordinate r, and suppose these scales are bounded away from zero and infinity in probability. For

$$T_{b,n}^{\#} = \sqrt{\frac{mb}{m-b}}\{\Psi_n(\widehat{\theta}_b^{\#}) - \Psi_n(\widehat{\theta})\}, \qquad (\sigma_{r,n}^{\#})^2 = \mathrm{Var}^{\#}(T_{b,n,r}^{\#}), \tag{63}$$

the ideal conditional subsample standard deviations obey

$$\max_r |\sigma_{r,n}^{\#} - \sigma_{r,n}| \to_p 0. \tag{64}$$

For a one-sided band let $M_n = \max_r W_{r,n}/\sigma_{r,n}$; for a two-sided band replace each coordinate by its absolute value. Let F_n be its conditional CDF and $c_n = \inf\{t : F_n(t) \geq 1 - \alpha_k\}$. Assume that, for every $\eta > 0$,

$$\lim_{\varepsilon \downarrow 0} \limsup_n P\{F_n(c_n + \varepsilon) - F_n(c_n - \varepsilon) > \eta\} = 0, \tag{65}$$

and that, for every fixed $\varepsilon > 0$, some $\delta_\varepsilon > 0$ satisfies, with probability tending to one,

$$F_n(c_n - \varepsilon) \leq 1 - \alpha_k - \delta_\varepsilon, \qquad F_n(c_n + \varepsilon) \geq 1 - \alpha_k + \delta_\varepsilon. \tag{66}$$

Then the ideal studentized subsample quantile converges to c_n in probability, and its conditional simultaneous coverage is $1 - \alpha_k + o_p(1)$. The same conclusion holds for the empirical resampling quantile when the number of Monte Carlo draws tends to infinity.

Proof. Pad the state by zeros if necessary so its dimension is bounded by $2\bar{L}$. A2 gives a conditional $(2+\delta)$-moment bound for $Y_i - E_{k-1}Y_i$. The conditional Lyapunov argument, applied to every Cramér–Wold projection, therefore yields

$$d_{\mathrm{BL}}\{\mathcal{L}(\sqrt{m}(\widehat{\theta} - \theta) \mid \mathcal{H}_{k-1}), N(0, \Omega_n)\} \to_p 0. \tag{67}$$

This approximation allows an n-dependent Gaussian law. Conditional on the full fold, the empirical covariance converges to Ω_n and the empirical Lindeberg condition follows from the same moment bound. The finite-population Cramér–Wold CLT gives

$$d_{\mathrm{BL}}\left\{\mathcal{L}^{\#}\left(\sqrt{\frac{mb}{m-b}}(\widehat{\theta}_b^{\#} - \widehat{\theta})\right), N(0, \Omega_n)\right\} \to_p 0. \tag{68}$$

The factor in this display removes the without-replacement variance factor $1 - b/m$. On an event whose probability tends to one, $\min_\ell p_\ell \geq c$ for some $c > 0$. A union bound then shows that the probability that the full fold or a size-b subsample omits any population level is at most $\bar{L}\{(1-c)^m + (1-c/2)^b\} + o_p(1) = o_p(1)$. Thus every denominator in the empirical level map is well defined on the relevant event.

It remains to verify that the same n-dependent directional map applies to both statistics. At an exact equality $q = P_j$, $\sqrt{b}(\widehat{\theta} - \theta) = O_p(\sqrt{b/m}) = o_p(1)$, so centering a subsample at the full fold does not shift the kink at the subsample scale. At a nonzero gap, condition (51) and tightness in equations (67)–(68) put both the full fold and its subsamples on the same smooth face with probability tending to one. Lemma 9, the lower mass bound, and the finite number of faces then give the two uniform remainders

$$\sqrt{m}\{\Psi_n(\widehat{\theta}) - \Psi_n(\theta)\} - \dot{\Psi}_{n,\theta}\{\sqrt{m}(\widehat{\theta} - \theta)\} = o_p(1), \tag{69}$$

$$\sqrt{\frac{mb}{m-b}}\{\Psi_n(\widehat{\theta}_b^{\#}) - \Psi_n(\widehat{\theta})\} - \dot{\Psi}_{n,\theta}\left\{\sqrt{\frac{mb}{m-b}}(\widehat{\theta}_b^{\#} - \widehat{\theta})\right\} = o_{P^{\#}}(1) \tag{70}$$

in probability. Applying the uniformly Lipschitz derivative to the two Gaussian approximations proves equations (61) and (62).

The $(2+\delta)$ moment bound and the Lipschitz formulas in Lemma 9 give uniform integrability of the squared coordinates. Hence the first two conditional subsample moments converge to those of W_n, proving equation (64); this step also allows W_n to have a nonzero mean at a kink. Studentization and the maximum are continuous once the scales are bounded away from zero. Equations (65)–(66) convert the resulting local CDF approximation into quantile convergence and conditional coverage. A conditional law of large numbers adds Monte Carlo estimation. □

Proof of Proposition 7. Equations (53)–(56) define the state vector and map required by Lemma 10. Applying that lemma gives the conditional law, studentizer, and critical value for the actual fitted signal. A3 on $\mathcal{Q}$ replaces each gain pseudo-target by $G_k(q)$; its $q=1$ coordinate gives $\sqrt{m_k}|\sum_\ell h_\ell - E\tau(X)| = o_p(1)$. Hence every Qini pseudo-target has bias $\rho_k(q) - q\rho_k(1) = o_p(m_k^{-1/2})$. Applying the map in equation (57) therefore yields joint gain–Qini coverage; the deterministic zero-Qini coordinate at $q=1$ is omitted from the maximum. Theorem 2 aggregates fold endpoints rather than subsample draws. □

Corollary 11 (Outcome-free calibration into fixed score levels)**.** *Let $\{(X_t^c, U_t^c) : 1 \le t \le r_n\}$ be i.i.d., independent of the labeled evaluation folds, fixed before I_2, and included in $\mathcal{H}_1$, with $r_n \to \infty$. The sample is not used to fit or tune the predictable raw score s_k. Let $\widehat{H}_{k,r}$ be the empirical CDF of the lexicographically ordered pairs $(s_k(X_t^c), U_t^c)$. Fix deterministic cutpoints $0 = a_0 < a_1 < \cdots < a_H = 1$, with fixed H and $\underline{\pi} = \min_{0\le h<H}(a_{h+1} - a_h) > 0$. For a new independent key U^{bin}, set*

$$B_k(X, U^{\mathrm{bin}}) = \sum_{h=1}^{H-1} \mathbf{1}\{\widehat{H}_{k,r}(s_k(X), U^{\mathrm{bin}}) \ge a_h\}, \tag{71}$$

and write $\pi_h = a_{h+1} - a_h$. Uniformly over the fixed number of folds,

$$\max_{0\le h<H} |P_{k-1}(B_k = h) - \pi_h| = O_p(r_n^{-1/2}). \tag{72}$$

Consequently all positive masses are bounded away from zero. If

$$d_{\mathcal{Q},a} = \min_{q\in\mathcal{Q}} \min_{1\le h<H} |q - (1 - a_h)| > 0, \tag{73}$$

then condition (51) holds for every $b_k \to \infty$. An independent allocation key within B_k defines the exact-budget policy, and equation (49) integrates that second key. Under the coordinate-scale bounds in Proposition 7, equations (65)– (66) then hold automatically.

Proof. Condition on the labeled training history that fixes s_k but not the independent calibration draw. The true lexicographic CDF H_k is continuous and $H_k(s_k(X), U^{\mathrm{bin}})$ is uniform. On the DKW event $\|\widehat{H}_{k,r} - H_k\|_\infty \le \varepsilon$, the event $B_k = h$ is bracketed by uniform-rank intervals of lengths $\pi_h - 2\varepsilon$ and $\pi_h + 2\varepsilon$. Thus its probability differs from π_h by at most 2ε. Conditional DKW and a union bound over fixed K prove equation (72). The same argument puts every cumulative top-bin mass within $O_p(r_n^{-1/2})$ of an ideal edge $1 - a_h$. Under equation (73), all nonzero budget–edge gaps are therefore bounded below by $d_{\mathcal{Q},a}/2$ with probability tending to one, which implies equation (51). Separation also puts every reported budget on an interior face with probability tending to one, so W_n is a finite centered Gaussian vector. Its one-sided or absolute maximum has no atom: a point event is contained in the union of zero-probability marginal hyperplanes. The standardized covariance matrices form a compact set, and continuity plus strict increase at the interior quantile make both properties uniform. This proves the two max-quantile conditions. □

Subsample sensitivity. The supplemental atom diagnostic has level masses $(0.3, 0.4, 0.3)$, signal means $(0, 1, 0)$, and budget $q = 0.3$, giving an exact kink. Across prespecified exponents $\gamma \in \{0.60, 0.70, 0.75, 0.80, 0.90\}$ in $b = \lfloor m^\gamma \rfloor$, coverage is 94.8–95.2% for $\gamma \le 0.80$ and 97.5% for $\gamma = 0.90$; the latter has $b/m > 1/2$. The default 0.75 lies in this stable range. Tuning is fixed before evaluation; selection among bands using evaluation outcomes requires an error allocation across them.

E EVALUATION COMPARISONS AND RANKING COSTS

E.1 DATA GENERATION AND EVALUATION OF THE SAME FITTED MODELS

All scenarios draw $X \in [-1,1]^8$ independently and uniformly. The baseline regression and smooth treatment effect are

$$\mu_0(X) = 0.25X_1 + 0.18X_2^2 - 0.12X_3X_4, \tag{74}$$

$$\tau(X) = \mathrm{clip}[0.04 + 0.31\tanh(2.2X_1 + 0.8X_2) - 0.12X_3 + 0.06X_4X_5, -0.45, 0.45]\,. \tag{75}$$

Treatment is Bernoulli $1/2$ and $Y = \mu_0(X) + A\tau(X) + \varepsilon$ with $\varepsilon \sim N(0, 0.55^2)$. The exploratory-feedback scenario sets $\tau = 0$.

Outcome nuisances are arm-specific ridge regressions on $X_1, \ldots, X_8$, $X_1^2, \ldots, X_4^2$, and X_1X_2, X_1X_3, X_2X_3. Each fit standardizes these features within its training arm and uses an intercept; the ridge penalty is 10 for nuisances and 8 for the score regressions. The smooth-ridge score is the difference between arm-specific ridge predictions on this same basis.

Table 1 uses five folds and four predictable smooth-ridge rules, seed 290905, $n = 800, 1600$, 500 paired trials per size, and 999 draws. Each realized target is evaluated on 30,000 fresh covariates. The positive budgets are $0.10, 0.15, \ldots, 0.90, 1$, with cost $C(q) = 0.08q$ and the full version simplex. The configuration was fixed before the extension ran. Every route maximizes its lower utility and falls back to treat-none when that maximum is nonpositive.

Both rectangles spend 0.0125 per version and calibrate the maximum over its 18 gain coordinates. Diagonal radial calibration spends over coordinates; the terminal route uses the shared covariance calibration below. The suffix route re-ranks all subsequent observations using each original fitted signal; its last version shares the same diagonal draws. Each route targets 95% coverage separately. Selecting across routes would require an allocation between them.

Whole-simplex rectangular coverage is checked at every vertex. Radial coverage uses equation (8), and terminal coverage uses the active-set supremum below. Selected validity checks the chosen rule's bound; safe utility includes the fallback. Binary intervals are Wilson intervals, and utility intervals use paired trial differences.

The suffix-minus-diagonal-rectangle safe-utility differences are 0.00477 (95% Monte Carlo interval $[0.00279, 0.00675]$) at $n = 800$ and 0.00936 ($[0.00697, 0.01174]$) at $n = 1600$. This comparison changes data reuse while holding calibration fixed. Radial versus suffix differences are reported in the main text. The suffix route makes 2.5 times as many predictions without new labels or fits. Stored timings include four-worker contention and shared fitting.

E.2 COMMON-TERMINAL COMPARATOR

Every fitted score and nuisance fit is fixed before I_K opens. Conditional on $\mathcal{H}_{K-1}$, use one common vector of exponential observation weights across versions, re-ranking each version in every draw. For declared gain or Qini coordinate j, let $\widehat{g}_j$ be the vector of estimates, $e_j^{\#} = \widehat{g}_j^{\#} - \widehat{g}_j$, and $\widehat{V}_j = \mathrm{Cov}^{\#}(e_j^{\#})$. The shared weights retain cross-version covariance. Calibrate the $(1-\alpha)$ quantile $\widehat{c}$ of

$$T^{\#} = \max_j \sup_{\lambda \in \Lambda(S)} \frac{(\lambda^\top e_j^{\#})_+}{\sqrt{\lambda^\top \widehat{V}_j \lambda}}, \tag{76}$$

and use

$$L_j^{\mathrm{H}}(\lambda) = \lambda^\top \widehat{g}_j - \widehat{c}\sqrt{\lambda^\top \widehat{V}_j \lambda}. \tag{77}$$

Known zero coordinates are appended exactly. Assume the joint Regular or separated-Atomic expansion, consistent covariance estimation, conditional root-cohort-scale variances bounded away from zero over the simplex, and a continuous identified critical quantile. The joint conditional bootstrap and continuous mapping theorem then give simultaneous coverage: the simplex is compact and the denominators are bounded away from zero. Full covariance nonsingularity is unnecessary. The statement uses the ideal resampling quantile; finite draw counts add Monte Carlo error.

The implementation enumerates the nonempty linearly independent active sets A, retains $\widehat{V}_{AA}^{-1}e_A \geq 0$, and takes the maximum of $\sqrt{e_A^\top \widehat{V}_{AA}^{-1} e_A}$ and zero. At $d = 4$ there are at most 15 sets. Enumeration is exponential in d; for larger sets of models the same supremum equals $\max_{v\geq 0,\, \|Rv\|_2 \leq 1} e^\top v$ with $R^\top R = V$, a second-order cone program. The experiments evaluate only $d \leq 4$. Selection maximizes the concave lower utility on the simplex, checking solver convergence and all vertices.

Additional comparisons retained with the code. Fixed-weight learner stress tests, a weight-selection sample-size path, and training/routing frontiers are supplemental checks. A paired $50/50$ honest split has narrower bands in every tested learner setting, for its own separately trained policy. Equal allocation among the fitted versions is also dominated by an honest split in the smooth-ridge frontier. Conversely, selecting radial weights improves utility over fixed uniform weights at $n = 800, 1600, 3200$ by $0.0094, 0.0139, 0.0019$, with paired 95% intervals $[0.0082, 0.0106]$, $[0.0124, 0.0155]$, and $[0.0014, 0.0025]$. These comparisons change policies or routing; Table 1 compares evaluation procedures on the same fitted models.

E.3 Fitted learners, feedback, and deployment

The simulation protocol in the supplement fixes 20,000 trials per noise condition, refitting both learners each time. The retained 256-trial preflight uses separate seeds and is excluded. Noise conditions share random inputs; the three costs reuse each trial.

Let $X \sim N(0, I_6)$, $b = (1, 1/2, 1/4)^\top/\sqrt{21/16}$, $u = b^\top X_{1:3}$, and $v = b^\top X_{4:6}$. Generate

$$A \sim \text{Bernoulli}(1/2), \qquad Y = (A - 1/2)(u + v) + \sigma(u)\varepsilon, \qquad \varepsilon \sim N(0, 1),$$

with A and ε mutually independent and independent of X. Thus $\tau = u + v$ and $\Gamma = 2(2A - 1)Y$ is unbiased. The equal-noise condition has $\sigma = 0.25$; the unequal condition uses 0.25 for $u > 0$ and 3 otherwise. On 2,000 burn-in rows, regress Γ on each three-feature group separately, using an intercept and ridge penalty one on slopes. Keep these two fitted scores fixed. Four subsequent cohorts have 800 rows each. The first evaluates learner one; each subsequent choice uses only the running mean of previously reported gains and the hurdle $C_0 = (2\pi)^{-1/2}$. This places the hurdle near the best gain in either feature class and produces frequent changes of learner.

For either fitted slope vector w, Gaussian integration gives exact gain $G(w) = C_0 b^\top w/\|w\|$ at budget $1/2$. For the first learner let $\rho = b^\top w/\|w\|$ and for the second let $\rho = 0$; the signal variance for its allocation is

$$V(w) = 1 + 4\{\sigma_{\text{low}}^2 p + \sigma_{\text{high}}^2(1/2 - p)\} - G(w)^2, \qquad p = 1/4 + \arcsin(\rho)/(2\pi),$$

where $\sigma_{\text{low}} = 0.25$ and $\sigma_{\text{high}} = 0.25$ or 3. The estimator re-ranks each cohort and averages Γ times its top-half indicator. Since the conditional mean signal at the population score boundary is zero, that product's sample variance divided by 800 consistently estimates the gain variance. Gaussian integration supplies the critical values. In a stored-output calibration check for unequal-noise feedback, the rectangular component coverages are 98.70–98.79% against 98.75%. Whole-simplex radial/rectangular coverage is $94.86/95.05\%$, with Wilson intervals $[94.54, 95.15]/[94.74, 95.34]$ against 95%.

Replay uses four additional cohorts for selection and the original data for evaluation. Frozen inference uses $z_{.95}$ times the pooled standard error; radial inference uses $\sqrt{c_4^+(.95)}$ and rectangular inference uses $z_{.9875}$ for each component. Each procedure has its own 5% error budget. The protected procedures cover the full weight simplex; the deployment comparison fixes equal weights.

Deploy the equal mixture only if its lower gain exceeds the declared cost. The true population utility is the mean of the four realized $G(w)$ values minus that cost. At C_0, feedback increases frozen harmful deployment by 0.66 percentage points relative to replay (95% paired interval $[0.35, 0.97]$). At this cost the replay contrast attributes 0.017 loss units per 1,000 to feedback ($[0.003, 0.030]$); the total frozen–radial improvement is 0.126. At C_+, frozen harm with feedback/replay is $0.43/1.005\%$. Replay measures the feedback effect with each critical value held fixed. All rates, Wilson intervals, paired contrasts, and all three costs are retained with the code. These experiments evaluate population utility under Gaussian treatment effects.

Table 2 is a secondary decomposition of the original paired differences, using known truth to label helpful and harmful changes. At C_+, the improvement frequency is 0.405% (Wilson interval $[0.326, 0.503]\%$), and $0.00405 \times 23.939 = 0.097$ is the average gain per 1,000 recipients. After deployment decisions are fixed, true policy values label each change as helpful or harmful. All costs, controls, protected methods, and an exploratory cost scan are retained in the supplement, together with the original coverage comparison.

E.4 HILLSTROM: EVALUATION AND SCORE RESOLUTION

We retain Men's E-Mail and No E-Mail from the $64{,}000$-row source, encode pretreatment categoricals, and exclude conversion and spending. Before evaluation, a seed-fixed 20% covariate/key sample is reserved: 4,261 rows calibrate score levels, while 2,130 calibrate diagnostic ranks and 2,131 assess disagreement. All 4,261 rows in the latter two groups supply the independent outcome audit below; their labels are unused by training or the original curve evaluation.

Four predictable boosted T-learners use five folds, 100 iterations, learning rate 0.05, at most nine leaves, minimum leaf size 150, and ℓ_2 penalty 8. Predictions are clipped to $[0, 1]$ and propensity is known to be $1/2$. Calibration cutpoints $(0.125, 0.275, 0.425, 0.575, 0.725, 0.875)$ create seven levels, with ideal minimum mass 0.125 and grid separation 0.025. A second independent key allocates within levels. Corollary 11 gives the separated-score justification; subsampling uses exponent 0.75.

The same fitted models are evaluated on diagonal cohorts of 6,818 rows, suffixes of 27,272, 20,454, 13,636, and 6,818 rows, or the common final cohort. All routes jointly protect gain and Qini. The rectangles spend 0.0125 per version; the radial and terminal critical values are 3.8061 and 2.9010, respectively. Terminal calibration uses 999 shared weighted draws. Uniform weights describe fixed-mixture precision; the cost 0.01 is a sensitivity value in visit equivalents. All routes select treat-all, and exact-budget Qini lower bounds are nonpositive. The complete numerical comparison is retained with the code.

The independent diagnostic fixes both raw and binned allocations $a_{kq}(X, U) \in [0, 1]$ before opening its outcomes. Raw allocations threshold calibrated ranks; binned allocations use ideal level masses. Their true masses $p_{kq} = Ea_{kq}$ need not equal nominal q. Thus its target is

$$D_{kq}^{\mathrm{cal}} = E(a_{kq}\tau) - p_{kq}\theta, \tag{78}$$

estimated by $\overline{a_{kq}\Gamma_k} - \bar{a}_{kq}\bar{\Gamma}_k$ with influence

$$a_{kq}\Gamma_k - E(a_{kq}\Gamma_k) - p_{kq}(\Gamma_k - \theta) - \theta(a_{kq} - p_{kq}). \tag{79}$$

The final term accounts for estimating treatment mass; the product remainder is $O_p(m^{-1})$. No threshold is estimated on these outcomes. One shared Gaussian multiplier jointly calibrates the uniform-mixture raw, binned, and paired-difference curves over all 20 budgets (1,999 draws, critical value 3.1228). Only $q = 0.05$ gives positive simultaneous lower bounds for raw Qini and coarsening loss, as reported in Figure 1.

F FROM INFERENCE TO DEPLOYMENT

F.1 FINITE-BATCH EXACT-QUOTA BRIDGE

For a planned future horizon, fix $\mathcal{H}_K$-measurable component counts $B_k \geq 0$, let $\mathcal{K}_B = \{k : B_k > 0\}$, $B = \sum_k B_k$, and $\lambda_k^B = B_k/B$. For $k \in \mathcal{K}_B$, choose integer treatment quotas $t_k \in \{0, \ldots, B_k\}$ with $q_k^B = t_k/B_k$. Draw B_k i.i.d. future units for component k and let $Q_{ki}^{B_k,t_k}$ select the top t_k lexicographic ranks (s_k, U) in that component. Write $\mathbf{t} = (t_2, \ldots, t_K)$ and $\mathbf{B} = (B_2, \ldots, B_K)$, with $t_k = 0$ when $B_k = 0$. Conditional on the learned rules, define the expected per-unit batch gain

$$G_B^{\mathrm{quota}}(\mathbf{t}, \mathbf{B}) = \frac{1}{B} E\left[\sum_{k=2}^{K}\sum_{i=1}^{B_k} Q_{ki}^{B_k,t_k}\tau(X_{ki}) \,\middle|\, \mathcal{H}_K\right]. \tag{80}$$

Proposition 12 (Population-to-quota bridge). *If $|\tau(X)| \le M$ almost surely, then for every target $(q, \lambda) \in [0,1] \times \Lambda$,*

$$\left|G_B^{\mathrm{quota}}(\mathbf{t}, \mathbf{B}) - G^{\lambda}(q)\right| \le M\|\lambda^B - \lambda\|_1 + M \sum_{k \in \mathcal{K}_B} \lambda_k^B \left\{ |q_k^B - q| + \sqrt{\frac{q_k^B(1-q_k^B)}{B_k}} \right\}. \quad (81)$$

If $B_k = \lambda_k B$ and $t_k = qB_k$ are integral on the K_+ active components, the bound becomes

$$\left|G_B^{\mathrm{quota}}(q, \lambda) - G^{\lambda}(q)\right| \le M\sqrt{\frac{q(1-q)}{B}} \sum_{k:\lambda_k > 0} \sqrt{\lambda_k} \le M\sqrt{\frac{K_+ q(1-q)}{B}}. \quad (82)$$

Therefore a simultaneous population lower band minus the displayed deterministic radius is a simultaneous lower band for expected exact-quota batch gain.

Proof. Fix component k and write $q_k = q_k^B$. Under the population-threshold policy at q_k, the treated count $N_k = \sum_{i=1}^{B_k} \mathbf{1}\{R_{ki} > 1 - q_k\}$ is $\mathrm{Binomial}(B_k, q_k)$. The threshold-treated and top-t_k sets are nested, so their symmetric-difference size is exactly $|N_k - t_k|$. Hence

$$\left|G_{k,B_k}^{\mathrm{quota}}(q_k) - G_k(q_k)\right| \le \frac{M}{B_k} E|N_k - t_k| \le M\sqrt{\frac{q_k(1-q_k)}{B_k}}. \quad (83)$$

Uniform ranks and nested thresholds also give $|G_k(q_k) - G_k(q)| \le M|q_k - q|$. Averaging these bounds at λ_k^B and using $|G_k(q)| \le M$ for the routing-weight difference proves equation (81). In the integral special case, $\lambda^B = \lambda$ and $q_k = q$; then $\sum_k \lambda_k/\sqrt{B_k} = B^{-1/2}\sum_k \sqrt{\lambda_k} \le \sqrt{K_+/B}$. The band-transfer claim follows by subtracting the bound on the population-coverage event. □

Choosing the effect envelope. The quota radius uses a known bound on $|\tau|$. If both potential outcomes lie in a known interval $[y_{\min}, y_{\max}]$, take $M = y_{\max} - y_{\min}$; Hillstrom's binary visit outcome permits $M = 1$. For the smooth simulations, equation (75) gives $M = 0.45$ even though the outcome noise is unbounded. An estimated upper envelope $\overline{M}$ is usable if it has a separate guarantee $P\{\|\tau\|_\infty \le \overline{M}\} \ge 1 - \delta$. For $0 < \delta < \alpha$, calibrating the population surface at error $\alpha - \delta$ and intersecting these two events gives quota coverage $1 - \alpha + o(1)$.

Deploy only when the final lower bound is positive. Normalize $C(0) = 0$. For the selected counts, let $U_B = G_B^{\mathrm{quota}}(\mathbf{t}, \mathbf{B}) - C(\bar{q}_B)$ and $L_{\mathrm{serve}} = L_G(\widehat{q}, \widehat{\lambda}) - R_{\mathrm{quota}} - C(\bar{q}_B)$. On the simultaneous population-coverage event, $U_B \ge L_{\mathrm{serve}}$ for every admissible choice of the counts. Hence the rule that deploys only when $L_{\mathrm{serve}} > 0$, returning treat-none and a zero certificate otherwise, satisfies

$$\limsup_n P\{\text{deploy and } U_B \le 0\} \le \alpha.$$

This bounds the unconditional probability of falsely certifying positive expected batch net gain. For example, with four equal components, $q = 1/2$, $B = 1000$, and $M = 1$, the radius is $\sqrt{0.001} = 0.03162$. A population lower utility of 0.01 therefore yields a negative serving certificate and returns treat-none. The expected utility of the proposal is U_B; the radius adjusts its certificate.

F.2 Implementation details

Algorithm 1 first selects a rule using its population lower bound, then applies rounding and the quota correction. Directly optimizing the serving bound over a declared integer-feasible set uses the same simultaneous coverage event.

Regular inference re-ranks exponential-weight draws. Atomic inference integrates allocation keys and uses the without-replacement correction $\sqrt{b/(m-b)}$, under Proposition 7's conditions. Radial protection uses Gaussian boundaries established before evaluation. All-suffix inference retains the original fits and evaluates each on its combined later cohorts after all models have been fitted. Outcome-free calibration rows are fixed before evaluation and excluded from fitting. If a route is selected after comparing multiple lower surfaces, those routes must share the total error budget.

F.3 FINITE-BATCH DECISION AUDIT

A controlled quota audit adds mean effect 0.08 to the two-candidate RCT in Appendix B.1. The candidate scores are X and $-X$, with exact gains $G_{\pm}(q) = 0.08q \pm 2\delta_m q(1-q)$. The four routes select budgets and weights at costs $C(q) = cq$, $c \in \{0.04, 0.08, 0.12\}$. The 2,000 base trials cover $m \in \{400, 1600\}$, $r \in \{1, 4\}$, and $h \in \{-3, 0, 3\}$, using 999 draws; fixed-portfolio frozen errors at $r = 4, h = 0$ are $3.8/5.0\%$ under bootstrap studentization. The frozen/radial harm comparison fixes $q = 1/2$ and equal version weights.

For each stored proposal, apply largest-remainder routing and nearest within-version quotas at $B \in \{100, 1000, 10000\}$, with known envelope $M = 0.08 + 2|\delta_m|$. For score sign $a_k \in \{-1, 1\}$, order statistics give

$$G_B^{\text{quota}}(\mathbf{t}, \mathbf{B}) = \frac{1}{B} \sum_k \left\{ 0.08 t_k + 2\delta_m a_k \frac{t_k(B_k - t_k)}{B_k + 1} \right\}. \tag{84}$$

At $B = 1000, m = 400, r = 4, h = 0, c = 0.12$, frozen/radial harmful deployment is $1.4/0.4\%$. At $c = 0.04$, radial protection forgoes 0.00118 expected batch utility (95% paired interval $[0.00089, 0.00147]$). The supplement retains all 648 route–setting–batch summaries and 648,000 records, the original selection experiment, and both audit protocols.

F.4 YOUWIN! CAPACITY APPLICATION

We use the public business-plan competition data (McKenzie, 2017a;b). The outcome is whether the business operates at the third follow-up; the analysis uses seed 41817 and two score levels. It targets a future applicant population under working assignment probabilities and missing-at-random response. After removing a singleton stratum, there are 1,840 applicants: 728 assigned treatment and 1,112 controls in 23 strata. Their final treatment fractions, 0.1724–0.6, supply the working propensities. The original experiment used fixed-count lotteries followed by disqualifications and replacements; the supplement documents this design history and the working assumptions used in the analysis.

Use 21 pretreatment features, 368 covariate-only calibration rows, and five ordered folds of the remaining rows. Expanding histories fit four boosted T-learners (100 iterations, learning rate 0.05, at most nine leaves, minimum leaf size 30, ℓ_2 penalty 8). History-only imputation and nuisance fitting precede median calibration into two levels. Arm-specific logistic response models with regularization $C = 1$ and clipping to $[0.05, 0.995]$ adjust the AIPW signal for 279 missing outcomes.

Budgets are $\{0.05, 0.10, \ldots, 0.30\}$, with $q = 1$ appended for the mean. Each route receives error 0.0125 jointly for gain and Qini, with 19,999 draws. Diagonal and suffix subsampling use exponent 0.75; the terminal route uses shared exponential weights. All four achieve their maximum gain lower bound at $q = 0.30$. At $B = 1000$, $M = 1$, and zero cost, all four proposals pass the final quota gate (Table 4). Gain is measured in operating-probability points per eligible applicant.

Table 4: YouWiN! under the assignment/MAR working model, in percentage points per applicant. L_Δ is evaluated at the gain-selected decision. The batch has $B = 1000$, $M = 1$, and zero cost; t is the total quota. Random allocation uses the same population budget or integer quota.

	Population, $q = 0.30$				Exact quota		
Route	$\widehat{G}$	L_G	L_Δ	Random L_G	t	L_{serve}	Random L_B
Diagonal rectangle	11.901	5.330	-4.555	7.362	300	3.881	7.362
Diagonal radial	11.707	6.863	-2.124	7.569	300	3.836	7.569
All-suffix rectangle	12.656	7.494	-2.119	7.362	300	6.044	7.362
Common terminal	12.440	5.481	-3.842	7.114	299	2.848	7.090

For random allocation, the jointly protected mean gives lower gain qL_θ, or $(t/B)L_\theta$ after rounding, where $L_\theta = \max_\lambda L_G(1, \lambda)$. All same-decision and optimized Qini lower bounds are negative. The terminal mixture rounds to 299 places, which also defines its random reference.

Holding scores and decisions fixed, stronger response regularization ($C = 0.1$) preserves four positive batch bounds. Completing missing treated outcomes by zero and missing controls by one yields batch gain lower bounds of 0.421 for diagonal rectangular, -0.711 for radial, 0.251 for all-suffix,

and -0.236 for terminal evaluation, in percentage points. Radial and terminal return treat-none at $B = 1000$, and all four pass at $B = 10000$. The supplement retains all 80 explored settings, the family adjustment, and the full missingness checks.

F.5 CALIBRATION ON YOUWIN! COVARIATES

Five settings each use 1,000 independently refitted trials at $n = 1840$, with covariates sampled from the observed empirical distribution. Three settings combine known Bernoulli assignment with zero, constant, or heterogeneous effects; two sensitivities use known MAR response and fixed stratum counts. Two-level scores give exact population gains, and binomial integration gives exact batch gains. The supplement provides the locked protocols, all 40,000 route–cost records, and full coverage and decision tables.

The allocated route coverage is 98.75%. In the heterogeneous complete-outcome setting, common-terminal coverage is 97.3% (95% Monte Carlo interval $[96.1, 98.1]$) with 1,999 draws and a 294-row last cohort. A paired follow-up with 19,999 draws gives 97.1%; the change is -0.2 percentage points ($[-0.59, 0.19]$). With 19,999 draws and 1,000 new trials per size, coverage is 98.0% at $n = 3680$ and 98.2% at $n = 7360$. The increase from $n = 1840$ to $n = 7360$ is 1.1 percentage points ($[-0.23, 2.43]$). Exact high-bin masses are at least $0.1106, 0.1334, 0.1550$ from the budget grid at the three sizes. The audit measures finite-sample calibration at budgets separated from atomic boundaries. Paired contrasts for draw counts and sample sizes are included in the supplement.