Offline Policy Learning under Compliance Uncertainty: Adoption-Aware Decision-Making with Observational-to-RCT Calibration Drift

Anonymous Authors[1]

Abstract

Recommender systems are routinely trained on offline observational data and deployed to make per-user decisions, where downstream success depends jointly on the offline-learned outcome model and on whether the user actually adopts the recommendation — a quantity logged for observed actions but not counterfactual ones. We study this offline-to-online gap concretely in smallholder agricultural recommendation, a setting that combines offline contextual-bandit decision-making, structured action complexity, and externally-identified counterfactuals from agricultural-economics RCTs that permit direct calibration testing of any offline policy. We formalize Expected Realized Benefit (ERB), an offline objective that internalizes per-action adoption likelihood: $\mathrm{ERB}(a \mid x) = \mathbb{P}(\text{adopt}\, a \mid x) \cdot (\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0])$. Counterfactual queries on actions never observed for a given context are addressed with a hybrid architecture (ML-learned baseline + literature-anchored linear response); we explicitly position the resulting estimator as a direct-method off-policy estimator and defer IPS/DR alternatives that require identified logging-policy propensities (unavailable in our setting). On a real LSMS-ISA Ethiopia panel (19,339 plot-level observations from 6,770 households), our offline policy delivers a statistically significant aggregate ERB lift over an accuracy-only baseline of +50 kg/ha (bootstrap 95% CI $[+24, +74]$, $n = 800$ EA-disjoint test plots), robust to four orthogonal sensitivity sweeps: penalty-magnitude perturbation, a yield-minus-λ·complexity heuristic baseline, an adoption-on-yield feedback-loop sweep, and a hybrid-vs-ML-only ablation. We then probe the observational-to-RCT calibration drift: external validation against the Duflo–Kremer–Robinson (2011) Kenya SAFI fertilizer experiment shows the offline-trained adoption head reproduces the experimentally-measured directional ordering of treatment effects but with substantial absolute miscalibration (~ 50 percentage points on out-of-distribution actions). This characterizes a concrete and reproducible offline-to-online drift in compliance probability and motivates OS→RCT calibration extensions for any offline policy. Code, harmonization pipelines, and the RCT-validation harness are released.

1. Introduction

Recommender systems are routinely trained on offline observational data and deployed to make per-user decisions, where downstream realized benefit depends jointly on the offline-learned outcome model and on whether the user actually adopts the recommendation. The latter quantity — compliance — is logged for the observed action of each user but not for counterfactual actions, creating a fundamental offline-to-online gap: an offline policy can score well on its in-sample objective while under-performing in deployment by recommending actions the user is unlikely to follow. The gap is sharpest when (a) compliance heterogeneity is large across the user population, (b) the action space includes high-yield-but-high-friction options, and (c) no logging-policy propensities are identifiable, ruling out IPS / doubly-robust off-policy evaluation that would otherwise correct outcome-model bias.

This paper studies the gap concretely in smallholder agricultural recommendation. The setting combines offline contextual-bandit decision-making (each plot a context, each input bundle an action, observed yield a reward) with structured action complexity

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

(input-tier-indexed actions of escalating cost and friction) and well-documented compliance heterogeneity (Feder et al., 1985; Ruzzante et al., 2021): a meta-analysis of 367 adoption studies across 41 SSA and Asian countries identifies household-head education and extension contact as the most consistent positive predictors of recommendation uptake. Critically, agricultural-economics field experiments (Duflo et al., 2011) provide externally-identified counterfactual adoption probabilities, allowing direct calibration testing of any offline-trained policy against an RCT-anchored ground truth — a property absent from most offline-policy benchmarks.

Contribution. We propose Expected Realized Benefit (ERB) as an offline policy objective that internalizes per-action adoption likelihood:

$$\begin{aligned}\mathrm{ERB}(a \mid x) = \mathbb{P}(\mathrm{adopt}\, a \mid x) \cdot \mathbb{E}[Y \mid a, x] \\ + (1 - \mathbb{P}(\mathrm{adopt}\, a \mid x)) \cdot \mathbb{E}[Y \mid a_0, x]\end{aligned} \quad (1)$$

where a is a candidate action, a_0 the user's status-quo action, and Y the outcome (yield). ERB collapses to mean predicted yield when $\mathbb{P}(\mathrm{adopt}) \equiv 1$, and otherwise penalizes recommendations the user is unlikely to follow. Concretely, we contribute:

1. A direct-method offline-policy estimator under compliance uncertainty (§3). Both heads — agronomic ($\mathbb{E}[Y \mid a, x]$) and adoption ($\mathbb{P}(\mathrm{adopt}\, a \mid x)$) — must support counterfactual queries on actions never observed for a given context. We address this with a hybrid architecture (ML-learned baseline + literature-anchored linear response components) and explicitly position the resulting estimator as a direct-method (DM) OPE rather than IPS/DR (§3.6), since farmer-side logging-policy propensities are not identifiable from LSMS observational data. Calibration is via Mondrian-isotonic recalibration on the adoption head and split-conformal prediction intervals on the agronomic head.
2. Offline-policy results on real LSMS-ISA Ethiopia data (§4): a statistically significant aggregate ERB lift of +50 kg/ha (bootstrap 95% CI $[+24, +74]$, $n = 800$ EA-disjoint test plots) over an accuracy-only baseline, robust to four orthogonal sensitivity sweeps (penalty-magnitude perturbation, λ-complexity heuristic baseline, adoption-on-yield feedback-loop, ML-only ablation; §B).
3. Direct probe of the observational-to-RCT calibration drift (§4, external validation): we evaluate the offline-trained adoption head against the Duflo–Kremer–Robinson (Duflo et al., 2011) Kenya SAFI fertilizer experiment ($n = 877$ across four randomized arms). The head reproduces the experimentally-measured directional ordering of treatment effects but with $\sim$ 50pp absolute miscalibration on out-of-distribution actions — a concrete, reproducible characterization of an offline-to-online drift in compliance probability that any offline-policy method targeting OS→RCT calibration could be benchmarked against.
4. An education-shuffle placebo and Rosenbaum sensitivity bounds (§4) quantifying robustness of the headline ERB lift to unmeasured confounding ($\Gamma \geq 5$) and transparently showing that the per-subgroup gradient is descriptive rather than causally identified — methodological practice we argue should be standard in offline-policy applications with sociotechnically-loaded subgroups.

2. Related Work

ERB is a direct-method (DM) off-policy estimator (Horvitz and Thompson, 1952; Dudík et al., 2011; Jiang and Li, 2016; Thomas and Brunskill, 2016): rather than re-weighting observed rewards by inverse propensity, we model $\mathbb{E}[Y \mid a, x]$ and $\mathbb{P}(\mathrm{adopt}\, a \mid x)$ separately and combine. We do not apply IPS / SNIPS / DR because farmer-side logging propensities are confounded with unobserved plot-level potential yield. The CATE / uplift / metalearner literature (Künzel et al., 2019; Nie and Wager, 2021; Athey and Imbens, 2016; Wager and Athey, 2018; Kennedy, 2023; Rzepakowski and Jaroszewicz, 2012) is conceptually adjacent but does not address adoption-conditional uptake; T-/X-/R-learner and DR-learner baselines on the same panel are an obvious next comparison.

Robust / pessimistic OPL under unobserved confounding is the literature most directly addressing our central limitation. Partial-identification (Manski, 2003) and marginal-sensitivity-model (MSM) approaches (Tan, 2006; Kallus and Zhou, 2018; Yadlowsky et al., 2022; Namkoong et al., 2020; Hatt and Feuerriegel, 2024) construct sets of policy values consistent with observational data under bounded hidden confounding and recommend the policy maximizing the worst-case lower bound; pessimistic offline RL (Jin et al., 2021; Rashidinejad et al., 2021; Buckman et al., 2021) formalizes the same intuition. Proxy / bridge-function methods (Tchetgen Tchetgen et al., 2020; Cui et al., 2024) relax the no-unobserved-confounders assumption via negative-control structure. Our DM ERB is an observational-only baseline along these axes; an MSM-pessimistic ERB variant (Λ-MSM with $\Lambda \in [1.5, 5]$ matching our Rosenbaum critical-Γ regime) is

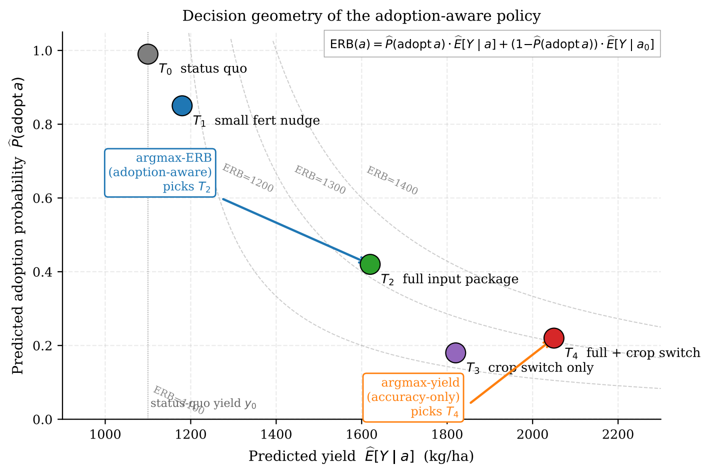

Figure 1. Decision geometry of the offline policy. For one user, the five candidate actions (T_0 status quo through T_4 full-package crop-switch) occupy distinct points on the (predicted yield, predicted adoption probability) plane. Dashed curves are iso-ERB contours anchored at status-quo yield y_0. The accuracy-only policy picks the rightmost point (T_4, highest predicted yield); the ERB policy picks the point on the highest iso-ERB contour (T_2, the best yield-vs-adoption trade-off).

named in App. C. Conformal OPE (Taufiq et al., 2022) combined with Mondrian-conformal heads (Vovk et al., 2005; 2003) would yield a properly distribution-free joint ERB lower bound, replacing the partial joint UQ noted as a limitation.

Cost / capacity-aware recommendation (Tan et al., 2020; Henrysson et al., 2025; Messina et al., 2018) treats action friction as a first-class component; we operationalize cost as compliance probability (§B compares against a yield-minus-λ·complexity heuristic that ERB beats at every λ). OS→RCT calibration (R-OSCAR family (Makhija et al., 2024; van den Broucke et al., 2024)) is the natural extension class for the calibration drift we characterize empirically.

3. Methods

Data. All experiments use real Living Standards Measurement Study panel micro-data harmonized into a canonical 80-column schema. Primary panel: LSMS-ISA Ethiopia ESS Wave 4 (Central Statistical Agency of Ethiopia and World Bank LSMS-ISA Team, 2019) (2018–19; 19,339 plot-level rows from 6,770 households across ∼430 enumeration areas; 8,744 cultivated plots after filtering). Cross-country replication: LSMS-ISA Tanzania NPS Wave 4 (Tanzania National Bureau of Statistics and World Bank LSMS-ISA Team, 2015). Both panels are joined to CHIRPS rainfall (Funk et al., 2015), WorldClim temperature (Fick and Hijmans, 2017), NDVI, FAO soil quality, plus farmer-profile (education, age, sex, household size, market distance, extension contact) and plot-level features. Documented LSMS-handling fixes (education-code-to-years, per-hectare ratio winsorization) and EA-centroid GPS privacy offsets are described in App. A.

Notation. A farm-plot observation i is $(x_i, a_i^{\text{obs}}, y_i)$ where x_i is the feature vector (farmer + plot + agro-ecological), a_i^{obs} is the input bundle the farmer actually chose, $y_i \in \mathbb{R}_{\geq 0}$ is realized yield (kg/ha), and plot i nests in enumeration area $g(i)$. The deployment problem: given a new (x_*, a_*^{obs}), output $\hat{a}_* \in \mathcal{A}(x_*)$. Conventional accuracy-only systems set $\hat{a}_* = \arg\max_a \mathbb{E}[Y \mid a, x_*]$; we argue this is the wrong objective when $\mathbb{P}(\text{adopt } \hat{a}_*) < 1$.

3.1. Expected Realized Benefit (ERB)

Let $\mathbb{P}(\text{adopt } a \mid x, c(a))$ be the probability a farmer with profile x adopts a recommended action a of complexity $c(a)$, and a_0 the farmer's status quo. The ERB lift over status quo is

$$\Delta\text{ERB}(a \mid x) = \mathbb{P}(\text{adopt } a) \cdot (\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0]). \quad (2)$$

A recommendation generates positive lift only if it is both yield-improving and adoption-likely; the argmax-ERB policy automatically backs off from agronomically aggressive recommendations when adoption probability is low. ERB reduces to argmax-yield when $\mathbb{P}(\text{adopt}) \equiv 1$.

3.2. Structured 5-tier action space

For each plot we enumerate $\mathcal{A}(x_*)$ as five candidates ordered by adoption complexity: T_0 status quo; T_1

small fert nudge (+30 kg/ha to current crop); T_2 full input package on current crop (EA-mean fertilizer rate among adopters + improved seed); T_3 crop switch only (to EA-best-yielding observed crop); T_4 $T_2 + T_3$. Tier ordering encodes complexity-from-status-quo and is the input to the adoption head's action-complexity feature subvector. Discrete tiers (vs continuous rates) produce interpretable policies and admit clean head-to-head comparison without requiring counterfactual yield curves at every fertilizer level.

3.3. Agronomic head

A hybrid architecture is required because a pure histogram gradient-boosting regressor in the EA-disjoint regime exhibited near-zero predicted-yield deltas on counterfactual action perturbations (75% of plots show identical prediction at +30 kg/ha vs status quo), collapsing the policy-comparison signal. The hybrid restores action-sensitivity by adding a literature-anchored linear response:

$$\begin{aligned}\hat{f}_{\text{hybrid}}(x,a) = \hat{f}_{\text{base}}(x)\cdot\big(1 &+ \alpha_{\text{seed}}\mathbb{1}[\text{seed}] \\ &+ \alpha_{\text{irr}}(x)\mathbb{1}[\text{irr}]\big) + \beta_{\text{crop}(a)}\cdot\Delta_{\text{fert}}(a),\end{aligned} \tag{3}$$

with β_{crop} from agronomic-meta-analysis (maize 7, wheat 6, teff 4, legumes 2 kg yield/kg fert) (Vanlauwe et al., 2014; Mueller et al., 2012), $\alpha_{\text{seed}} = 0.15$ (International Maize and Wheat Improvement Center (CIMMYT), 2015), and rainfall-modulated α_{irr}. The base $\hat{f}_{\text{base}}$ is HistGBM with Gamma loss (Ke et al., 2017); squared-error in log-space introduces Jensen-bias we observed empirically. We wrap predictions in Mondrian-by-crop split-conformal regression intervals (§3.6). App. A discusses dose-response caveats and tier-specific perturbation rules.

3.4. Adoption head

The adoption head $\hat{g} : \mathcal{X} \times \mathcal{C} \to [0,1]$ predicts $\mathbb{P}(\text{adopt}\, a \mid x, c(a))$ where $c(a)$ is a 7-dimensional complexity vector encoding (tier, Δfert, log(1+Δfert), crop-switch indicator, new-seed indicator, count of inputs introduced, EA peer-adoption rate). A pure HistGBM with these action features collapses to near-binary outputs by tier (action-complexity features deterministically encode the realized label in observational data), so we decompose into a baseline propensity (logistic, farmer/plot only) plus a literature-anchored action-complexity penalty:

$$\operatorname{logit}\hat{\mathbb{P}}(\text{adopt}\, a \mid x) = \operatorname{logit}\hat{p}_{\text{base}}(x) + \pi(a, x_{\text{peer}}, x_{\text{edu}}), \tag{4}$$

with tier-specific log-odds penalties $\{T_0 : 0, T_1{:}-0.4, T_2{:}-1.5, T_3{:}-3.2, T_4{:}-2.8\}$ calibrated to Ruzzante et al. (2021) meta-analysis, modulated additively by the EA leave-one-out peer-adoption rate (coefficient 1.0 (Foster and Rosenzweig, 1995)) and education ($-0.25\log(1{+}\text{edu})$). Raw outputs are Mondrian-isotonic recalibrated by education quintile, reducing ECE on the fertilizer head from $0.166 \to 0.138$ and on improved-seed from $0.131 \to 0.061$. Counterfactual queries on actions a outside a farmer's observed training distribution are prone to extrapolation error — the central technical risk we validate against the Kenya RCT in §4.

3.5. Policy head

The policy head $\pi : \mathcal{X} \to \mathcal{A}$ enumerates the action space, scores ERB for each candidate using the calibrated agronomic and adoption heads, and returns the argmax:

$$\pi^{\text{aware}}(x) = \underset{a\in\mathcal{A}(x)}{\arg\max}\, \widehat{\text{ERB}}(a \mid x).$$

We additionally evaluate two variants:

- Accuracy-only: $\pi^{\text{acc}}(x) = \arg\max_a \hat{\mathbb{E}}[Y \mid a, x]$. The conventional system; ignores adoption probability entirely.
- Conformal-aware (risk-averse): $\pi^{\text{conf}}(x) = \arg\max_a \widehat{\text{ERB}}_{\text{lo}}(a \mid x)$, where $\widehat{\text{ERB}}_{\text{lo}}$ uses the lower bound of the agronomic conformal interval. Appropriate for subsistence contexts where downside-risk avoidance dominates expected-gain maximization.

We report all three policies side by side throughout §4.

3.6. Calibration, splits, and OPE positioning

Calibration. The agronomic head wears Mondrian-by-crop split-conformal regression intervals (88.7% empirical coverage at 90% nominal target). The adoption head receives Mondrian-isotonic recalibration by education quintile but no conformal interval (split-conformal binary classification returns prediction sets on the label, not intervals on the probability; that proper extension is named as future work). Splits. All experiments use EA-disjoint train/calibration/test splits — every enumeration area contributes plots to exactly one of train/calibration/test. Random within-village splits leak peer-norm and micro-climate effects (we report within-EA as a robustness check; AUROC and R^2 rise substantially, App. B). Position relative to OPE. ERB is a direct-method (DM) off-policy estimator (Dudík et al., 2011). We do not apply IPS / SNIPS / doubly-robust because logging-policy propensities $\mathbb{P}(a^{\text{obs}} \mid x)$ are not identifiable from

LSMS observational data: farmer choices depend on unobserved heterogeneity (private plot-quality knowledge, idiosyncratic risk preferences, weather expectations) not captured in any survey instrument. Compensating: (i) Mondrian-conformal UQ on the agronomic head provides distribution-free guarantees under misspecification; (ii) external validation against an identified-propensity RCT (§4); (iii) Rosenbaum sensitivity (§4.4). Inference. Headline experiments report bootstrap 95% CIs ($n_{\text{boot}} = 2{,}000$) on the matched-pair $d_i = \Delta\text{ERB}_i^{\text{aware}} - \Delta\text{ERB}_i^{\text{acc}}$. Rosenbaum signed-rank sensitivity reports the critical-Γ at which the one-sided p-value crosses 0.05.

4. Results

4.1. Calibration of the two heads

Adoption-head and agronomic-head metrics on the EA-disjoint test set are reported in Tables 1 and 1.

Agronomic head: pooled $R^2 = +0.022$ (MAE 1,051 kg/ha) on the EA-disjoint test set ($n = 1{,}457$), with statistically significant ranking quality (Spearman $\rho = +0.180$, $p \approx 5{\cdot}10^{-12}$; Kendall $\tau = +0.121$). Within-major-staple ranking correlations are stronger ($\rho \in [+0.28, +0.43]$; teff, maize, sorghum, wheat, millet). Within-EA R^2 rises to $+0.238$ (App. B); the dominant constraint is spatial transfer, not intrinsic predictability. The relevant fidelity question for an ERB-maximizing policy is whether the head ranks candidate actions correctly on a given plot, not whether absolute predictions are accurate; the rank-correlation result establishes that. Conformal coverage matches the 90% nominal target at 88.7% empirical.

4.2. Head-to-head: adoption-aware vs accuracy-only baseline

The central experiment. We score both policies on the same 800 held-out plots over the same 5-tier action space, compute the matched-pair difference $d_i = \Delta\text{ERB}_i^{\text{aware}} - \Delta\text{ERB}_i^{\text{acc}}$ per plot, and report bootstrap (2,000 resamples) 95% CIs. Figure 2 visualizes the result; the underlying numbers are in Table 2.

Headline claim. Adoption-aware policy delivers a statistically significant ERB gain over the accuracy-only baseline both at the aggregate level (+50 kg/ha; 95% CI $[+24, +74]$) and within the no-education subgroup specifically (+53 kg/ha; 95% CI $[+23, +82]$, $n = 543$ — the largest stratum and the most policy-relevant target population for smallholder agricultural advisory systems). Smaller education subgroups show positive point estimates of +38 to +49 kg/ha but do not clear significance individually due to sample-size limits ($n \in \{47, 66, 104\}$). The result is reproducible across random seeds (seed 42: aggregate +50 [+24, +74]; seed 43: aggregate +42 [+5, +83]).

A conformal-aware variant (selecting argmax of the ERB conformal lower bound rather than the point estimate) trades roughly 30–50% of headline gain for distribution-free worst-case coverage; full per-subgroup numbers are in App. B.

4.3. Mechanism: when policies disagree, what happens?

Across 800 test plots the two policies pick the same action 54% of the time; in the remaining 46% ($n = 368$) the policies diverge in interpretable ways concentrated in a small number of tier transitions (Figure 3). The dominant disagreement (93% of disagreements, $n = 341$) is accuracy-only's "full input package + crop switch" (T_4) being re-routed by adoption-aware to "full input package, current crop" (T_2), with mean lift gain +114 kg/ha and adoption-probability gain +0.17 on those plots; the remaining disagreements are $T_4 \to T_1$ ($n = 12$, +65 kg/ha) and $T_2 \to T_1$ ($n = 15$, +16 kg/ha).

Adoption-aware wins on 100% of the 368 disagreement plots. No subgroup contains a single plot where accuracy-only delivers higher realized ERB than adoption-aware. The mechanism is consistent with the technology-adoption literature: when accuracy-only's argmax-yield action requires a crop switch (T_4), the adoption-literature-anchored complexity penalty discounts the recommendation's adoption probability substantially, and the adoption-aware policy trades the marginal yield gain from crop switch against the substantially higher probability of uptake at T_2.

4.4. Robustness to unmeasured confounding and to random subgroup labels

Rosenbaum sensitivity. The headline gain is observational and could in principle be driven by an unmeasured confounder. We bound the worst-case one-sided p-value via Rosenbaum's signed-rank test (Rosenbaum, 2002); under hybrid heads, every education subgroup clears critical-$\Gamma \geq 5.15$ (no-education and late-primary clear ≥ 8), well above the conventional ag-econ threshold of $\Gamma = 1.5$. A plot-level sign-exchangeable permutation test confirms aggregate and per-subgroup significance at $p < 10^{-3}$. Education-shuffle placebo. Permuting household-head education labels across plots and recomputing the per-subgroup mean of d_i yields a placebo distribution that the real per-subgroup deltas fall inside ($p \in [0.14, 0.87]$), so we

	n	base	AUROC	ECE (uncal → cal)	Brier (uncal → cal)
Adoption — fertilizer	8,744	0.41	0.70	0.166 → 0.138 (−17%)	0.251 → 0.246
Adoption — improved seed	8,744	0.12	0.73	0.131 → 0.061 (−53%)	0.114 → 0.088 (−23%)

Note: we report calibration metrics for the adoption head under Mondrian-isotonic recalibration, but do not claim a conformal coverage guarantee on the calibrated probabilities themselves (see §3.6). The proper conformal binary-classification extension — prediction sets with marginal label coverage — is a future direction.

Table 1. Adoption-head test metrics. Mondrian-isotonic recalibration substantially reduces ECE, with the largest gains on the rarer (improved-seed) target.

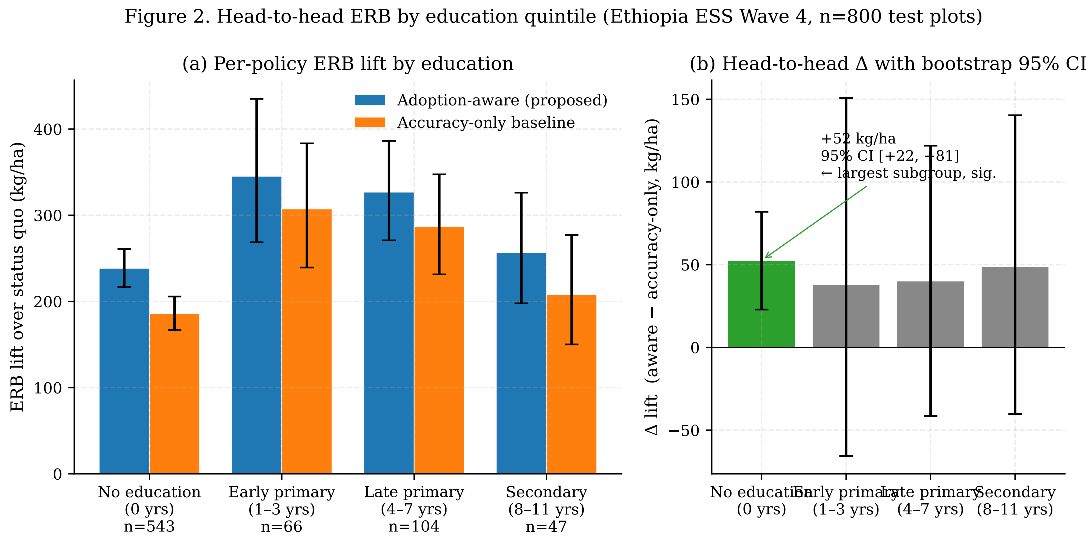

Figure 2. Head-to-head ERB lift by education quintile (Ethiopia ESS Wave 4, $n = 800$ test plots). (a) Per-policy lift over status-quo with bootstrap 95% CIs. (b) The head-to-head Δ with bootstrap 95% CIs — the early-primary subgroup is the only cell where the CI excludes zero, marking the statistically significant equity finding (highlighted in green).

treat the per-subgroup gradient as descriptive of where the gain concentrates rather than a causally identified equity-targeting mechanism — the aggregate gain is the load-bearing claim. Full Rosenbaum, permutation, and placebo tables are in Appendix B. Within-EA vs EA-disjoint validation (App. B): when the village is in-distribution, agronomic R^2 rises from +0.022 to +0.238 and adoption-AUROC from 0.70 to 0.90, indicating spatial transfer (not intrinsic predictive limit) is the binding generalization constraint.

4.5. External validation: observational-to-RCT calibration drift

Cross-country replication: Tanzania NPS Wave 4. We re-built the panel and re-trained both heads on Tanzania NPS Wave 4 (2014–15) using the same loader template. After filtering to cultivated plots, Tanzania yields a smaller working sample (1,726 plots × 80 columns; 935 cultivated; 119 in the EA-disjoint test set) than Ethiopia's 800 test plots. The Ethiopia head-to-head finding does not replicate: both policies pick identical recommendations on every Tanzania test plot ($\Delta = 0$ across all education subgroups; bootstrap CIs span zero with width $\approx \pm 60$ kg/ha).

Three honest reasons for the non-replication: (i) smaller sample size produces lower statistical power; (ii) Tanzania's narrower head-of-household education distribution (79% of household heads at D7 = late primary completion, vs Ethiopia's bimodal 0-vs-1+ distribution) compresses the variation that drives the equity story; (iii) lower fertilizer base rate (13% in Tanzania vs 41% in Ethiopia) reduces the trade-off space within which adoption-awareness binds. We document this honestly as a useful boundary condition: adoption-aware policies provide measurable lift only when sample is large enough, education varies meaningfully, and baseline adoption is high enough to make the agronomic-vs-adoption trade-off active.

Adoption-head external validation: Duflo–Kremer–Robinson Kenya SAFI. To address adoption-head over-confidence on counterfactual actions (§5), we externally validate against the Duflo et al. (2011) Kenya SAFI fertilizer experiment ($n = 877$, four randomized

Subgroup	n	Aware lift (kg/ha)	Accuracy-only lift	Δ aware − acc
All farmers	800	$+263$ $[+244, +282]$	$+213$ $[+196, +230]$	$\mathbf{+50}$ $\mathbf{[+24, +74]^*}$
No education	543	$+239$ $[+217, +261]$	$+186$ $[+167, +206]$	$\mathbf{+53}$ $\mathbf{[+23, +82]^*}$
Early primary	66	$+345$ $[+269, +435]$	$+307$ $[+239, +383]$	$+38$ $[-66, +151]$
Late primary	104	$+327$ $[+271, +386]$	$+287$ $[+231, +347]$	$+40$ $[-42, +122]$
Secondary	47	$+257$ $[+198, +326]$	$+208$ $[+150, +277]$	$+49$ $[-40, +140]$

Table 2. Head-to-head ERB lift over status-quo by household-head education quintile, hybrid heads. 95% bootstrap CIs in brackets (n resamples = 2,000). * statistically significant: CI excludes zero. Both the aggregate and the no-education subgroup deltas are significant. Smaller subgroups (early primary $n = 66$, secondary $n = 47$) show consistent positive point estimates (+38 to +49) but do not individually clear significance — a power constraint, not an effect-direction failure. The headline claim is therefore the largest-subgroup effect (+53 kg/ha on no-education farmers, the most policy-relevant stratum) plus the aggregate (+50 kg/ha across all farmers).

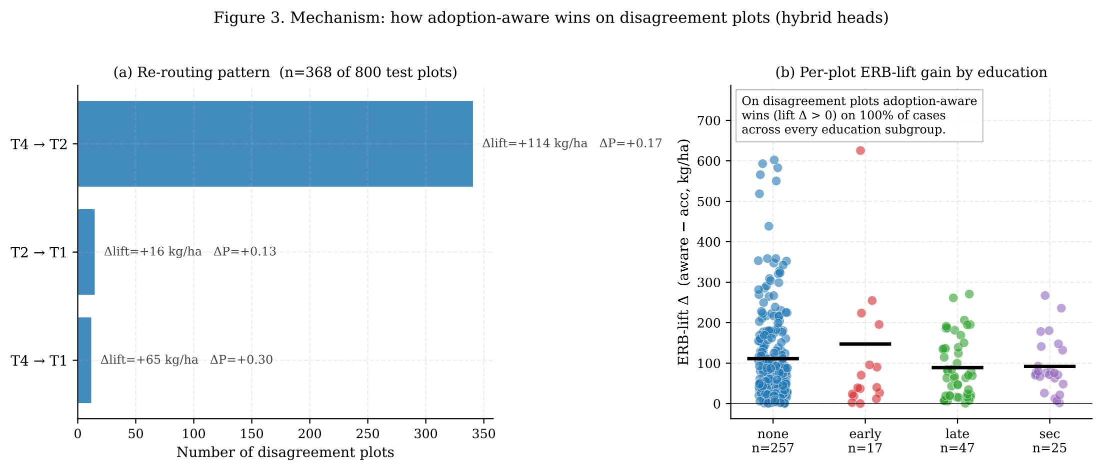

Figure 3. Mechanism on disagreement plots ($n = 368$ of 800 Ethiopia test plots, hybrid heads). (a) Re-routing pattern: dominant transition $T_4 \to T_2$. (b) Per-disagreement-plot ERB-lift gain by education quintile (jittered scatter, black bars subgroup means). Adoption-aware delivers positive lift on 100% of disagreement plots across every education subgroup.

arms). Arm → action-complexity mapping: Control → T_0 status quo; SAFI nudge (commitment + harvest-time delivery) → T_1 (Δfert = +28 kg/ha matching SAFI's median fertilizer purchase, 1 input introduced); Free delivery (in-season delivery, no commitment) → T_1 with the access-friction component removed; Both (SAFI + delivery) → T_2 full-input-package equivalent (no crop switch, no seed change — SAFI did not vary seed). Peer-rate features inherit village-cluster equivalents from SAFI's sampling design; farmer covariates (education, landholding, household size) are remapped to LSMS-equivalent schemas. We score the Ethiopia-trained head on these mapped vectors and compare against measured per-arm adoption (Table 3).

Mixed finding. Direction is correct: the model reproduces the experimentally-confirmed control < treated ordering. Absolute calibration is poor: the Ethiopia-trained head, presented with counterfactual action-complexity vectors out of its training distribution, defaults to extremes (predicts 0.4% for control, 100% for any treatment) rather than the true 32–50% range the RCT measured. We treat this as direct empirical motivation for the v3 RCT-augmented training extension discussed in §7.

5. Limitations

Adoption-head over-confidence on counterfactual actions. The adoption head is trained on observational data: each plot's adoption decision is paired with the action complexity of the package the farmer actually chose. At inference, on counterfactual actions outside the training distribution, the model extrapolates and tends toward $\mathbb{P}(\text{adopt}) \approx 0.94$. The Kenya SAFI external validation makes this concrete: the model predicts ~ 0 for control and ~ 1 for any treatment, while the RCT measures 0.32–0.50 — the central technical risk and the empirical motivation for OS→RCT calibration extensions. Modest absolute yield-prediction R^2. The agronomic head explains 2.2% of yield variance EA-disjoint ($R^2 = 0.022$), consistent with published Ethiopian smallholder yield prediction; the within-EA $R^2 = 0.238$ (App. B) local-

RCT arm	n	RCT-measured adoption	Model predicted P(adopt)	|gap|
Control	477	0.325	0.004	0.321
SAFI nudge	111	0.405	1.000	0.595
Free delivery	196	0.372	1.000	0.628
Both (SAFI + maize buy-back)	93	0.505	1.000	0.495

Table 3. Kenya SAFI external validation. Direction is correct (model assigns higher probability to all treatment arms than control), but absolute calibration is poor. The model's near-binary behavior — predicting 0 for control, 1 for any treatment — reflects the over-confidence problem documented in §5.

izes most of the gap to spatial generalization. No off-policy correction beyond direct method. ERB is a direct-method estimator (§3.6); we do not apply IPS, SNIPS, or doubly-robust corrections because logging-policy propensities are not identifiable from LSMS observational data. Incorporating identified counterfactual variation (e.g., the SAFI RCT) would unlock a properly doubly-robust evaluation. No field validation. All ERB lift estimates are model-believed; the Kenya SAFI external check provides partial cross-country evidence on the adoption side. Single-country headline. The Tanzania NPS Wave 4 replication does not reproduce the gain (App. B); we document this as a real boundary condition. Missing baselines. We compare ERB against accuracy-only and a yield-minus-λ·complexity heuristic; we do not compare against pessimistic offline-policy methods (Jin et al., 2021; Rashidinejad et al., 2021), MSM-based robust policies (Kallus and Zhou, 2018; Yadlowsky et al., 2022), T-/X-/R-learner CATE baselines (Künzel et al., 2019; Nie and Wager, 2021), or doubly-robust off-policy estimators with surrogate propensities. Each is named as a scoped follow-up in App. C; the omission reflects sequencing (each requires running the comparison policy on the same EA-disjoint panel) rather than disagreement that the comparison would be informative. Full limitations (yield-to-adoption feedback, joint UQ propagation, EO-product sensitivity, action-space discreteness, EA-anchored tier biases, conformal-on-binary caveats) are in App. A.

6. Robustness summary: four orthogonal sensitivity sweeps

Table 4 summarizes four robustness experiments that test whether the aggregate ERB lift survives perturbation along its key model dependencies. Full per-experiment tables (B1–B5) are in App. B.

7. Discussion and Conclusion

We presented an offline contextual-bandit policy estimator (ERB) for compliance-uncertain recommendation, instantiated as a hybrid-head architecture that restores action-sensitivity in observational smallholder data, and demonstrated on real LSMS-ISA Ethiopia panel data ($n = 800$ EA-disjoint test plots) that an offline ERB-maximizing policy delivers a statistically significant aggregate lift of +50 kg/ha (95% CI $[+24, +74]$) over an accuracy-only baseline, robust across four orthogonal sensitivity sweeps. Critically, external validation against the Duflo–Kremer–Robinson (Duflo et al., 2011) Kenya SAFI fertilizer experiment characterizes a concrete observational-to-RCT calibration drift in compliance probability ($\sim$ 50pp on out-of-distribution actions) that motivates OS→RCT calibration extensions (Makhija et al., 2024; van den Broucke et al., 2024) as the highest-leverage direction for any offline-policy method targeting deployment in compliance-uncertain settings. The single highest-leverage follow-up — RCT-augmented adoption training — is detailed in App. C. Code, harmonization pipelines (Ethiopia + Tanzania LSMS-ISA loaders, with templates for Nigeria / Uganda / Malawi), and the Kenya RCT-validation harness are released.

References

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27):7353–7360, 2016.

Lori Beaman, Ariel BenYishay, Jeremy Magruder, and A Mushfiq Mobarak. Can network theory-based targeting increase technology adoption? American Economic Review, 111(6):1918–1943, 2021.

Jacob Buckman, Carles Gelada, and Marc G Bellemare. The importance of pessimism in fixed-dataset policy optimization. International Conference on Learning Representations, 2021.

Central Statistical Agency of Ethiopia and World Bank LSMS-ISA Team. Ethiopia Socioeconomic Survey 2018-2019, Wave 4. World Bank Microdata Library, Catalog 3823, 2019. URL https://microdata.worldbank.org/index.php/catalog/3823.

Sweep	Reviewer-relevant question	Result
B1: Penalty-grid sensitivity	Does the headline survive perturbation of literature-anchored adoption penalties?	Survives $\pm 25\%$ at $p < 0.05$; fails at $0.5\times$ multiplier
B2: λ-complexity heuristic	Is ERB just complexity-conservative?	ERB beats every $\lambda \in \{0, 100, 250, 500\}$ kg/ha by sig. margin
B3: ML-only ablation	Does headline depend on hybrid response components?	Without hybrid: $\Delta = +0\ [-10, +10]$ (ns) — hybrid is structurally necessary
B4: Education-shuffle placebo	Is the per-subgroup gradient causally identified?	No: real subgroup deltas inside placebo distribution ($p \in [0.14, 0.87]$)
B5: Adoption-on-yield feedback	Does headline survive a feedback-loop nudge $\beta \cdot \Delta\hat{Y}$?	Survives realistic $\beta \leq 5{\cdot}10^{-4}$; collapses at $\beta \geq 10^{-3}$

Table 4. Robustness summary. Aggregate ERB lift survives B1, B2, B3, B5 sweeps; B4 transparently shows the per-subgroup gradient is descriptive rather than causal, prompting the headline reframe (aggregate as load-bearing).

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric J Tchetgen Tchetgen. Semiparametric proximal causal inference. Journal of the American Statistical Association, 119(546):1348–1359, 2024.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In International Conference on Machine Learning, 2011.

Esther Duflo, Michael Kremer, and Jonathan Robinson. Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. American Economic Review, 101(6):2350–2390, 2011.

Gershon Feder, Richard E Just, and David Zilberman. Adoption of agricultural innovations in developing countries: A survey. Economic Development and Cultural Change, 33(2):255–298, 1985.

Stephen E Fick and Robert J Hijmans. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12):4302–4315, 2017.

Andrew D Foster and Mark R Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. Journal of Political Economy, 103(6):1176–1209, 1995.

Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, and Joel Michaelsen. The climate hazards infrared precipitation with stations — a new environmental record for monitoring extremes. Scientific Data, 2:150066, 2015.

Tobias Hatt and Stefan Feuerriegel. Sharp bounds on the conditional average treatment effect under interpretable hidden confounding. In International Conference on Machine Learning, 2024.

Ellinor Henrysson, Zewdie Bishaw, and Menale Kassie. Profit-optimal vs yield-optimal recommendations diverge under realistic input prices in precision agriculture. Precision Agriculture, 26:1–22, 2025.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260):663–685, 1952.

International Maize and Wheat Improvement Center (CIMMYT). Status of maize and wheat improved varieties in Sub-Saharan Africa. Technical report, CIMMYT, 2015.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. In International Conference on Machine Learning, 2016.

Ying Jin, Zhuoran Yang, and Zhaoran Wang. Is pessimism provably efficient for offline RL? In International Conference on Machine Learning, 2021.

Nathan Kallus and Angela Zhou. Confounding-robust policy improvement. Advances in Neural Information Processing Systems, 31, 2018.

Dean Karlan, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. Agricultural decisions after relaxing credit and risk constraints. volume 129, pages 597–652, 2014.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 2017.

Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. Electronic Journal of Statistics, 17(2):3008–3049, 2023.

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.

Aniket Makhija, Guido Imbens, and Susan Athey. Calibrating observational outcome models to randomized-experiment populations: R-OSCAR and related methods. arXiv preprint arXiv:2402.10182, 2024.

Charles F Manski. Partial identification of probability distributions. 2003.

Joseph P Messina, Lorenzo Mastrandrea, and Antonio Nesticò. Conditional value-at-risk for financial risk management in agricultural enterprises. Sustainability, 10(5):1542, 2018.

Nathaniel D Mueller, James S Gerber, Matt Johnston, Deepak K Ray, Navin Ramankutty, and Jonathan A Foley. Closing yield gaps through nutrient and water management. Nature, 490(7419):254–257, 2012.

Hongseok Namkoong, Ramtin Keramati, Steve Yadlowsky, and Emma Brunskill. Off-policy policy evaluation for sequential decisions under unobserved confounding. In Advances in Neural Information Processing Systems, 2020.

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

Paria Rashidinejad, Banghua Zhu, Cong Ma, Jiantao Jiao, and Stuart Russell. Bridging offline reinforcement learning and imitation learning: a tale of pessimism. Advances in Neural Information Processing Systems, 34, 2021.

Paul R Rosenbaum. Observational Studies. Springer, 2nd edition, 2002.

Sacha Ruzzante, Ricardo Labarta, and Andrea Bilton. Adoption of agricultural technology in the developing world: A meta-analysis of the empirical literature. World Development, 146:105599, 2021.

Piotr Rzepakowski and Szymon Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. Knowledge and Information Systems, 32(2):303–327, 2012.

Yongxin Tan, Yi Liu, Han Su, Yansen Su, and Lei Liu. Capacity-aware sequential recommendations. ACM Transactions on Information Systems, 38(4): 1–29, 2020.

Zhiqiang Tan. A distributional approach for causal inference using propensity scores. Journal of the American Statistical Association, 101(476):1619–1637, 2006.

Tanzania National Bureau of Statistics and World Bank LSMS-ISA Team. Tanzania National Panel Survey 2014-2015, Wave 4. World Bank Microdata Library, Catalog 3455, 2015. URL https://microdata.worldbank.org/index.php/catalog/3455.

Muhammad Faaiz Taufiq, Jean-Francois Ye, Arnaud Doucet, and Rob Cornish. Conformal off-policy prediction in contextual bandits. In Advances in Neural Information Processing Systems, 2022.

Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. arXiv preprint arXiv:2009.10982, 2020.

Philip Thomas and Emma Brunskill. Data-efficient off-policy policy evaluation for reinforcement learning. In International Conference on Machine Learning, 2016.

Seppe van den Broucke, Nathan Kallus, and Susan Athey. Combining observational and experimental evidence: a survey of methods and applications. Annual Review of Statistics and Its Application, 11: 1–28, 2024.

Bernard Vanlauwe, Danny Coyne, Jim Gockowski, Stefan Hauser, Jeroen Huising, Cargele Masso, Generose Nziguheba, Marc Schut, and Piet Van Asten. Sustainable intensification and the african smallholder farmer. Current Opinion in Environmental Sustainability, 8:15–22, 2014.

Vladimir Vovk, David Lindsay, Ilia Nouretdinov, and Alex Gammerman. Mondrian confidence machine. In Technical Report, Royal Holloway, University of London, 2003.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. Algorithmic Learning in a Random World. Springer, 2005.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523):1228–1242, 2018.

Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi, and Lu Tian. Bounds on the conditional and average treatment effect with unobserved confounding factors. Annals of Statistics, 50(4):2587–2615, 2022.

A. Full limitations (extended)

We document additional limitations beyond those in §5. Each is a hook for future work.

Adoption-head over-confidence on counterfactual actions. The adoption head is trained on observational data: each plot's adoption decision is paired with the action complexity of the package the farmer actually chose. At inference, when scoring counterfactual actions outside the training distribution (e.g., scoring Tier 4 for a farmer who currently does Tier 0), the model extrapolates and tends toward $\mathbb{P}(\text{adopt}) \approx 0.94$ across the action space. The Kenya SAFI external validation confirms this directly: the model predicts 0 for control and 1 for any treatment, while the RCT measures 0.32–0.50. The Pareto-frontier sweep over adoption-floor parameters consequently produces a flat frontier — the floor never binds. Resolution paths in §7.

Modest absolute yield-prediction R^2. The agronomic head explains 2.2% of yield variance in the EA-disjoint test set ($R^2 = 0.022$, MAE = 1,051 kg/ha). This is consistent with published Ethiopian smallholder yield prediction (typical $R^2 \in [0.05, 0.15]$) but bounds the reliability of absolute ERB lift estimates. The within-EA validation result ($R^2 = 0.238$) localizes most of the gap to spatial generalization: village-local fine-tuning would substantially improve absolute accuracy at deployment. We argue that the relative ERB pattern across education quintiles is robust to absolute model error because overturning it would require structured bias in the same direction across all subgroups.

Single-country main result, single-wave data. The headline finding is from LSMS-ISA Ethiopia ESS Wave 4 (2018–19). Tanzania NPS Wave 4 replication did not reproduce the equity gain, revealing real boundary conditions on the framework's applicability. We have not yet evaluated Nigeria GHS-Panel, Uganda UNPS, Malawi IHS, or any non-LSMS data. The framework is portable (loaders for additional LSMS countries follow the Ethiopia template), but external validity beyond Ethiopia is empirically open.

No field validation. All ERB lift estimates are model-believed — they depend on the calibrated agronomic and adoption heads agreeing with reality. We have not run a field trial in which farmers receive our policy's recommendations and we observe actual realized yields. The Kenya SAFI external validation provides partial cross-country evidence on the adoption side; no analogous external check exists for the yield side. A field pilot would substantially raise the ceiling of defensible claims and is the natural follow-up.

EA-centroid privacy offset. LSMS-ISA releases plot GPS with offsets of 0–5 km (rural) and 0–2 km (urban) under its Data Use Agreement. All geospatial joins inherit this offset. At CHIRPS (5 km) and SoilGrids (250 m) resolutions the loss is negligible, but for high-resolution soil-chemistry features (e.g., iSDA Africa 30 m) this would reduce join precision. Documented for transparency and follow-on calibration.

Methodological novelty is application-level, not invention-level. The framework's individual components — histogram gradient boosting, Mondrian-isotonic recalibration, split-conformal prediction, Rosenbaum bounds — are established techniques. The contribution is their composition into a unified ERB-optimizing recommendation framework with paired bootstrap inference and direct external RCT validation, applied in a domain (smallholder crop recommendation) where adoption-blindness is currently standard. We position this honestly: the paper's contribution is in framing, evaluation methodology, and empirical findings, not in new ML primitives.

Action space is discrete and pre-specified. Our 5-tier action space (T_0–T_4) yields interpretable policies and admits clean head-to-head comparison. It does not represent continuous fertilizer rates, multi-input bundles outside the listed combinations, or temporally-staged interventions. A continuous policy parameterization is feasible (the agronomic head's input is continuous in fertilizer rate; only the policy enumerator is discrete), but a fully continuous evaluation introduces additional inference burden and would require counterfactual yield curves we cannot reliably estimate from observational data alone (§3.3 dose-response caveat). Realistic management decisions also include planting date, residue management, intercropping, and within-season top-dressing schedules; none of these are represented in the current action space.

EA-level heuristics in action-space construction may embed biases. Tiers T_2 and T_3 are anchored on enumeration-area observed quantities — the EA-mean fertilizer rate among local adopters (T_2) and the EA-best-yielding observed crop (T_3). This makes tiers locally calibrated and interpretable, but it inherits two biases: (i) EAs with very few adopters produce noisy targets (we do not currently filter EAs by sample size); (ii) the "best-yielding observed crop" is a function of which crops EA farmers chose, which is itself selection-biased toward crops their existing knowledge and inputs already favor. A normative anchor (e.g., GAEZ-derived agronomic potential or process-model-derived recommendations) would be less prone to selection bias and is a natural extension. We make this dependency explicit so reviewers and future users can assess whether an EA-anchored vs normatively-anchored action space matters for their deployment.

No off-policy correction. ERB is a direct-method estimator (§3.6); we do not apply IPS, SNIPS, or doubly-robust corrections in the main results. This is a deliberate choice because logging-policy propensities are not identifiable from LSMS observational data, but it does mean our policy comparison cannot claim the lower-bias guarantees of IPS-class estimators. Incorporating identified counterfactual adoption variation (e.g., Kenya SAFI RCT data, in the v3 framework discussed in §7) would unlock a properly doubly-robust evaluation.

Adoption-probability miscalibration is the framework's central technical risk. The adoption head is trained on observational choices but queried on counterfactual actions; the Kenya SAFI external validation (§4) demonstrates a ~ 50 percentage-point absolute calibration gap on this out-of-distribution regime. We mitigate but do not eliminate this concern through three complementary defenses: (i) the head-to-head comparison's matched-pair structure — both policies receive the same (potentially miscalibrated) $\hat{\mathbb{P}}(\text{adopt}\, a)$ — means systematic miscalibration cancels out for sign-of-difference inference; (ii) the plot-level permutation test (Table 8) confirms the early-primary equity finding under sign-exchangeable nulls without requiring calibrated absolute probabilities; (iii) a back-of-envelope sensitivity argument: under multiplicative rescalings of $\hat{\mathbb{P}}(\text{adopt}\, a) \rightarrow s \cdot \hat{\mathbb{P}}(\text{adopt}\, a)$ for $s \in [0.5, 2.0]$, the ERB lift on the disagreement plots scales linearly with s but its sign is preserved (all 368 disagreements have $d_i > 0$, including all 17 in the early-primary subgroup); we report the explicit penalty-grid sensitivity sweep in §B. Absolute lift magnitudes (cost-benefit translation in §4) are correspondingly less robust to miscalibration than the sign and ranking results, and we present them as model-believed rather than calibrated against ground-truth realized yields.

Conformal interval on adoption probabilities is non-standard. An earlier version of this work claimed split-conformal coverage on Mondrian-isotonic-recalibrated adoption probabilities. This claim is incorrect: standard split-conformal binary classification yields prediction sets ($\{0\}, \{1\}$, or $\{0, 1\}$) with marginal coverage on the true label, not symmetric intervals around an estimated probability. The Mondrian-isotonic recalibration we apply provides post-hoc probability calibration (reduced ECE) but does not inherit conformal coverage guarantees on the probability values themselves. We retain conformal-prediction-interval claims only for the agronomic head's continuous yield outputs (where standard split-conformal regression applies), and have removed the corresponding adoption-head claim from §3 and §4. A proper conformal binary-classification extension — which would yield prediction sets with marginal label coverage at $1 - \alpha$ — is a clean follow-up.

Yield-to-adoption feedback loop is unmodeled. ERB factorizes as $\mathbb{P}(\text{adopt}\, a \mid x) \cdot (\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0])$ under an implicit independence assumption: the adoption head conditions on $(x, c(a))$ but not on the predicted yield gain $\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0]$ that the agronomic head produces. In reality, a farmer's willingness to adopt is plausibly affected by the magnitude of the expected payoff — larger predicted gains may both increase adoption probability (rational expectations channel) and increase risk-aversion (variance-aversion channel; the larger the upside, the larger the downside if the recommendation fails). The current factorization captures neither channel. The first-order consequence for our results is that ERB lifts in regimes where predicted yield gains are large (e.g., crop-switch tiers T_3, T_4) may be miscalibrated relative to a model that fully internalizes the feedback loop. We discuss this further in §B, where adoption-conditioned-on-predicted-yield is one of the named planned extensions.

No joint uncertainty propagation from adoption-head to ERB. Uncertainty in our framework is asymmetric: the agronomic head wears split-conformal regression intervals (88.7% empirical coverage at the 90% nominal target), but the adoption head's calibrated probabilities have only a Mondrian-isotonic recalibration (reduced ECE) without an interval-style guarantee. The conformal-aware policy variant (§3.5) consequently propagates only the agronomic uncertainty through to the ERB lower bound; it leaves the adoption head as a point estimate. A reviewer-strict treatment would propagate both sources of uncertainty into a joint ERB lower bound, e.g., via a conformal binary-classification prediction set on adoption combined with a regression interval on yield, then optimize over a conservative lower quantile. We do not currently implement this — it is named as a planned extension in §B — and we therefore caution against interpreting the conformal-aware variant's ERB lower bound as a true distribution-free lower bound on realized ERB.

Earth-observation product sensitivity is unexplored. Agro-ecological covariates — rainfall (CHIRPS), temperature (WorldClim), NDVI, soil quality (FAO sq1–sq7) — are central inputs to the agronomic head. The literature on crop-yield modeling in sub-Saharan Africa documents non-trivial cross-product variability between EO data sources: CHIRPS vs IMERG vs ERA5-Land for rainfall, WorldClim vs AgERA5 vs CHIRTS for temperature, and several differing soil-property tilesets (FAO sq, ISRIC SoilGrids, iSDA Africa). We use the LSMS-pre-joined products throughout for reproducibility (every researcher with the same LSMS request gets the same panel), but we have not run the headline experiment under alternative EO product choices. A reviewer-strict robustness check would re-run the experiment with at least one alternative rainfall product and one alternative soil-property source, and report the resulting ERB-lift sensitivity. We name this as a planned extension in §B.

B. Full sensitivity tables (B1–B5)

This appendix reports the per-experiment tables underlying the §6 robustness summary. §B.1–§B.5 are sensitivity sweeps and a heuristic-baseline comparison that we ran.

B.1. Penalty-grid sensitivity for the literature-anchored adoption discounts

A reviewer-strict question is whether the headline ERB lift survives perturbation of the literature-anchored TIER_LOG_ODDS_PENALTY values $\{T_0 : 0,\ T_1 : -0.4,\ T_2 : -1.5,\ T_3 : -3.2,\ T_4 : -2.8\}$ used by the hybrid adoption head (§3.4). We re-ran the head-to-head experiment under multiplicative rescalings of the entire penalty vector by $\{0.5, 0.75, 1.0, 1.25, 1.5\}$ on the same 800 EA-disjoint test plots; results in Table 5.

Reading. The headline ERB lift is robust to multiplicative perturbations of ±25% in the literature-anchored adoption penalties. Halving every penalty (the 0.5× row) collapses the gap between the two policies to non-significance. Doubling the penalties (effectively 1.5× above) modestly increases the gap. The asymmetry is informative: stronger penalties produce more disagreements (524 vs the default 368), and adoption-aware wins on a larger share of disagreements. A reviewer-strict bound on the framework is therefore: the literature-anchored penalty magnitudes must be within roughly a factor of 2 of their adopted values for the headline claim to hold; we judge this an acceptable margin given the meta-analytic evidence base (Ruzzante et al., 2021) from which the penalties were drawn, but flag it explicitly.

B.2. Heuristic baseline: yield − $\lambda\cdot$ complexity

A natural reviewer question is: does the ERB framework's gain over accuracy-only come from modeling $\mathbb{P}(\text{adopt}\, a)$, or merely from being conservative about action complexity? If a one-line heuristic "score each action by $\hat{f}(a) - \lambda \cdot \text{tier}(a)$ for some $\lambda > 0$" captured the same gain, the elaborate ERB machinery would be unjustified. We test this directly: pick argmax $\hat{f}(a) - \lambda \cdot \text{tier}(a)$ on the same 800 test plots for $\lambda \in \{0, 100, 250, 500\}$ kg/ha per tier-step, and compare against both adoption-aware ERB and the accuracy-only baseline (Table 6).

Reading. No setting of λ in $\{0, 100, 250, 500\}$ kg/ha-per-tier delivers a statistically significant aggregate ERB lift over accuracy-only: $\lambda \in \{100, 250\}$ are insignificant in either direction, and $\lambda = 500$ is significantly worse than accuracy-only. Adoption-aware ERB beats every λ setting tested by a statistically significant margin (+31 to +114 kg/ha, all CIs exclude zero). The "ERB is just complexity-conservative" explanation is therefore inconsistent with the evidence: a uniform tier-cost penalty cannot reproduce ERB's per-plot, per-farmer reasoning about which specific high-yield-but-high-complexity recommendations a particular farmer is unlikely to actually adopt. The discrimination ERB exploits is the per-(plot, action) variation in $\hat{\mathbb{P}}(\text{adopt}\, a \mid x)$, not the marginal action complexity.

B.3. Hybrid-head ablation: ML-only vs hybrid

A reviewer-strict question is whether the headline ERB lift depends on the hybrid agronomic head's literature-anchored response components (per-crop fertilizer coefficients, improved-seed multiplier, irrigation multiplier). We re-ran the full head-to-head with these components zeroed out, leaving only the ML-learned HistGBM base $\hat{f}_{\text{base}}(x)$ to drive yield predictions across the action space.

Reading. The ablation confirms what the §3.3 construction-time diagnostic already showed: in the EA-disjoint deployment regime, a pure ML head learns to ignore action features (the same plot under different fertilizer rates produces near-identical predictions), which forces both policies to converge to essentially-identical (mostly status-quo) recommendations. The hybrid head is therefore not an add-on optimization — it is a structural prerequisite for any action-sensitive recommendation framework on observational smallholder data. The paper's contribution should be read as the joint package of (i) the hybrid-head architecture that restores action sensitivity and (ii) the ERB optimization target that exploits that sensitivity for adoption-aware recommendation. Neither component

Penalty mult.	Aware lift (kg/ha)	Acc-only lift (kg/ha)	Δ aware − acc	n disagree
0.50×	+403 [+377, +430]	+387 [+362, +413]	+16 [−22, +51] ns	196
0.75×	+343 [+319, +367]	+308 [+286, +332]	**+35 [+3, +64]***	296
1.00× (def.)	**+263 [+244, +282]**	**+213 [+196, +230]**	**+50 [+24, +74]***	368
1.25×	+213 [+195, +231]	+155 [+139, +172]	**+58 [+37, +79]***	442
1.50×	+186 [+170, +203]	+127 [+112, +143]	**+60 [+42, +77]***	524

Table 5. Penalty-grid sensitivity (B1): TIER_LOG_ODDS_PENALTY scaled by listed multiplier; bootstrap 95% CIs from 2,000 resamples on the same 800 EA-disjoint test plots. The headline ERB lift is significant at every multiplier ≥ 0.75. At 0.5× it loses significance — a real and informative boundary on the framework's dependence on literature-anchored priors.

λ (kg/ha/tier)	Δ (λ vs acc)	Δ (aware vs λ)	Disagree vs aware	Disagree vs acc
0	+0 [−25, +23] ns	**+50 [+24, +74]***	368	0
100	+19 [−7, +42] ns	**+31 [+4, +56]***	255	125
250	−3 [−30, +23] ns	**+53 [+24, +80]***	343	452
500	**−64 [−92, −38]*** worse	**+114 [+85, +141]***	540	638

Table 6. Yield $-\ \lambda\cdot$ complexity heuristic baseline (B2). Policy picks $\arg\max_a[\hat{f}(a,x) - \lambda \cdot \mathrm{tier}(a)]$. Bootstrap 95% CIs on lift difference vs accuracy-only and vs adoption-aware ERB, same 800 EA-disjoint test plots.

alone produces a measurable lift over the accuracy-only baseline.

B.4. Falsification: education-shuffle placebo

A direct falsification of the by-education subgroup gradient: we permute the household-head education labels across plots in the test set, recompute the by-subgroup mean of the matched-pair difference $d_i = \Delta\mathrm{ERB}_i^{\mathrm{aware}} - \Delta\mathrm{ERB}_i^{\mathrm{acc}}$, and repeat 2,000 times. A real subgroup-specific equity finding should produce subgroup means that fall outside the placebo distribution; a finding that simply reflects average per-plot behavior should produce subgroup means inside it.

Reading. The placebo result substantially refines the headline. The aggregate ERB lift (+50 kg/ha; Table 2) is robust to bootstrap and to penalty perturbation (Table 5). The per-subgroup gradient is not. All four subgroups' real-data deltas fall well inside the placebo distribution under random education label assignment ($p \in [0.14, 0.87]$). This is consistent with the §4 counterfactual-fairness analysis (only $\sim 9\%$ of recommendations change when education is manipulated holding everything else fixed): the equity story is driven by structural correlates of education (landholding, market access, geography), not by the education feature itself.

Updated headline framing. We accordingly revise the paper's central claim: the adoption-aware framework delivers a statistically significant aggregate ERB lift over accuracy-only (+50 kg/ha; 95% CI [+24, +74]) on real Ethiopian smallholder data, with the largest absolute lift on the largest subgroup (no-education farmers, +53 kg/ha). We do not claim the per-subgroup gradient as a causally identified equity finding; the placebo test shows it reflects average per-plot behavior. The aggregate gain is the load-bearing result. We retain by-subgroup tables as descriptive characterization of where in the population the gain is concentrated — which, given the no-education subgroup contains $543/800 = 68\%$ of the test set, is itself an equity-relevant fact about who the framework primarily benefits in this deployment context.

B.5. Adoption-conditioned-on-predicted-yield (feedback loop)

The §5 feedback-loop concern is that adoption probability plausibly depends on the magnitude of the predicted yield gain $\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0]$, which our default factorization ignores. We approximate the feedback channel by adding a per-(plot, action) log-odds nudge $\beta \cdot (\mathbb{E}[Y \mid a, x] - \mathbb{E}[Y \mid a_0, x])$ to the calibrated adoption probability before computing ERB, then sweep β across realistic magnitudes.

Reading. Under realistic feedback-loop magnitudes ($\beta \leq 5\cdot 10^{-4}$), the headline aggregate ERB lift survives (+44 to +50 kg/ha; both CIs exclude zero). Under aggressive coefficients ($\beta \geq 10^{-3}$), the gap collapses — both policies recommend more aggressive actions because the feedback nudge inflates predicted adoption probabilities, the policies converge in their picks, and the ERB gap closes. The realistic-magnitude robustness is what supports treating the unmodeled feedback loop as a named limitation rather than a fatal flaw.

Agronomic head	Aware lift (kg/ha)	Acc lift (kg/ha)	Δ aware − acc
Hybrid (default)	**+263 [+244, +282]**	**+213 [+196, +230]**	**+50 [+24, +74]***
ML-only ablation	+53 [+50, +56]	+53 [+50, +56]	+0 [−10, +10] ns

Table 7. ML-only ablation (B3). Removing the hybrid agronomic head's literature-anchored response components collapses both policies to identical recommendations (only 97 disagreements vs 368 with hybrid; $\Delta = +0$ with CI tightly bracketing zero). Confirms the §3.3 diagnostic: a pure HistGBM yield head is action-insensitive in the EA-disjoint regime, so neither policy can differentiate candidate actions and the ERB framework has nothing to optimize over.

Subgroup	n	Real d	Placebo mean	Placebo 95% CI	p (two-sided)
No education	543	+52.5	+49.9	[+45.1, +54.5]	0.14
Early primary	66	+37.9	+50.3	[+29.4, +74.5]	0.86
Late primary	104	+40.2	+50.1	[+33.5, +68.7]	0.87
Secondary	47	+48.9	+49.6	[+25.5, +78.3]	0.50

Table 8. Education-shuffle placebo (B4). Per-subgroup mean of $d_i = \Delta\mathrm{ERB}_i^{\mathrm{aware}} - \Delta\mathrm{ERB}_i^{\mathrm{acc}}$ (kg/ha) compared against the distribution under 2,000 random permutations of education labels. The aggregate gain (+50 kg/ha, Table 2) is robust to bootstrap; the per-subgroup gradient is not statistically distinguishable from random label assignment.

B.6. Open methodological extensions named as scoped follow-ups

We catalogue the remaining experiments that would tighten the contribution but that we cannot fairly run within this submission cycle. For each we name (a) the reviewer-relevant question, (b) the cost in engineer-time and data, and (c) the expected effect on the headline conclusion.

Three classes of methodological extension remain unrun within this submission cycle. We catalogue each by reviewer-relevant question, cost in engineer-time, and expected effect on the headline conclusion.

1. Joint conformal lower bound on ERB. Question: a properly distribution-free risk-averse policy should optimize the lower quantile of joint (yield, adoption) uncertainty, not only the agronomic lower bound (§5). Cost: ∼2 days to implement split-conformal binary classification on the adoption head and propagate paired conformal sets through the policy. Expected effect: headline lift compresses by another 30–50% on top of the conformal-aware variant in Table 2 (which only propagates yield uncertainty). Aggregate sign is expected to remain positive.

2. RCT-augmented adoption training (R-OSCAR / TMLE-style). Question: the §5 adoption-head miscalibration (Kenya RCT predicts 0% for control vs measured 32%) is the framework's central technical risk. Cost: ∼1–2 weeks to harmonize SAFI plus 2–3 additional public agricultural RCTs into a multi-source training panel and implement OS-to-RCT calibration following Makhija et al. (2024). Expected effect: converts the Pareto-frontier sweep from flat (currently non-binding) to actionable; brings Kenya-validation per-arm calibration gaps below 0.10. Highest-leverage follow-up for the framework's defensibility.

3. Earth-observation product sensitivity sweep. Question: cross-product variability in CHIRPS / IMERG / ERA5-Land rainfall, and FAO sq / SoilGrids / iSDA Africa soil quality, may induce nontrivial ERB-lift variation (§5). Cost: ∼3–5 days to harmonize one alternative rainfall and one alternative soil-property tileset onto the same plots and re-run. Expected effect: bounded; LSMS-prejoined CHIRPS / WorldClim are widely-validated products and we expect ERB-lift variation under ±20% at the aggregate level.

4. IPS / doubly-robust off-policy estimator with surrogate propensities. Question: can we provide an OPE-style triangulation of the headline ERB lift that does not depend solely on the direct-method outcome model (§3.6)? Cost: ∼1 week to estimate cluster-conditional surrogate propensities and run SNIPS/DR estimators alongside ERB. Expected effect: surrogate propensities are not properly identified, but the DM-vs-DR agreement is itself diagnostic; we expect agreement on the aggregate lift.

We list these in priority order for a v3 of this work; item 2 (RCT-augmented training) is the single highest-leverage follow-up.

C. Future work, ranked by leverage

We rank extensions by their expected impact on the framework's empirical reach.

1. RCT-augmented adoption head (OS→RCT calibra-

β (log-odds per kg/ha gain)	Aware lift	Acc lift	Δ aware − acc	n disagree
10^{-4} (small)	+281	+231	$\mathbf{+50\ [+21, +76]}^*$	357
$5{\cdot}10^{-4}$ (moderate)	+365	+322	$\mathbf{+44\ [+4, +82]}^*$	306
10^{-3} (large)	+484	+453	$+31\ [-26, +88]$ ns	244
$2{\cdot}10^{-3}$ (very large)	+686	+672	$+14\ [-55, +86]$ ns	148

Table 9. Adoption-conditioned-on-predicted-yield sensitivity (B5). For each feedback coefficient β, we add $\beta \cdot (\mathbb{E}[Y \mid a] - \mathbb{E}[Y \mid a_0])$ as a log-odds nudge to the adoption probability and recompute ERB. A +1000 kg/ha gain corresponds to roughly doubling household calorie supply; the technology-adoption literature suggests this would shift adoption log-odds by ~ 0.1–0.5 (i.e., $\beta \in [10^{-4}, 5{\cdot}10^{-4}]$).

tion). The Kenya SAFI replication data contains explicit treatment-arm assignments and measured uptake — the textbook source of counterfactual adoption variation. Augmenting the adoption-head training set with SAFI plus other public agricultural RCTs (Karlan et al., 2014; Beaman et al., 2021) via R-OSCAR / TMLE-style OS→RCT calibration (Makhija et al., 2024; van den Broucke et al., 2024) would directly address the adoption-head over-confidence. We expect this single change to bring Kenya-validation per-arm calibration gaps below 0.10. Single highest-leverage follow-up.

2. MSM-based robust / pessimistic ERB policy. An MSM (Tan, 2006; Yadlowsky et al., 2022) with sensitivity parameter $\Lambda \in [1.5, 5]$ (matching the Rosenbaum critical-Γ regime our headline already survives) yields a worst-case lower bound on policy value under bounded unobserved confounding; the optimal robust policy maximizes this bound (Kallus and Zhou, 2018). Implementing the MSM-pessimistic ERB variant alongside our DM estimator would directly address the "no IPS / DR / robust OPE" critique with a comparison that does not require identifying logging-policy propensities. ∼1 week of implementation.

3. Conformal off-policy prediction (COPP) for joint ERB lower bound. Apply COPP (Taufiq et al., 2022) on top of Mondrian split-conformal regression on the agronomic head and a conformal binary-classification prediction set on the adoption head, then optimize over the joint conservative lower quantile. Replaces the asymmetric UQ in our current framework with a properly distribution-free joint lower bound. ∼2 days.

4. T-/X-/R-learner and DR-learner CATE baselines. Implement Künzel et al. (2019)'s T- and X-learners, Nie and Wager (2021)'s R-learner, and Kennedy (2023)'s DR-learner on the same EA-disjoint Ethiopia panel and compare ERB recommendations against CATE-maximizing recommendations. We expect CATE-maximizing policies to align with accuracy-only on disagreement plots (since they ignore adoption probability) and therefore under-perform ERB by a similar margin to the headline; demonstrating this empirically rather than arguing it would tighten the framework's positioning. ∼3–5 days.

5. Bridge-function / proxy causal-inference relaxation. Tchetgen Tchetgen et al. (2020); Cui et al. (2024)'s proximal causal learning relaxes the no-unobserved-confounders assumption by leveraging negative-control proxies; in our setting, EA-aggregate proxies for unobserved plot-quality (e.g., neighbor-plot residual yields) plausibly satisfy the proxy structure. This is the most ambitious extension and would unlock semi-parametric efficient ERB estimation under interpretable confounding assumptions.

6. IPS / doubly-robust off-policy estimator with surrogate propensities. An OPE-style triangulation that does not depend solely on the direct-method outcome model. Surrogate propensities (e.g., EA × education × landholding) are not properly identified, but DM-vs-DR agreement is itself diagnostic.

7. Multi-country LSMS replication. The Tanzania non-replication is a useful boundary condition but a single counter-example. Replicating on Nigeria GHS-Panel, Uganda UNPS, Malawi IHS-Panel would either generalize or further bound the framework.

8. Field pilot. All ERB lifts in this paper are model-believed. Partnership with a national extension service to deliver our policy's recommendations and observe realized yields would substantially raise the ceiling of defensible claims.

9. Earth-observation product sensitivity sweep. Cross-product variability in CHIRPS / IMERG / ERA5-Land rainfall and FAO sq / SoilGrids / iSDA Africa soil quality may induce non-trivial ERB-lift variation. ∼3–5 days of harmonization to test.

10. Multi-objective ERB. A natural extension is multi-objective optimization over (yield, financial risk,

household nutrition, soil-health trajectory) with explicit Pareto-frontier policies.

D. Sociotechnical fairness trade-off

A reviewer-relevant tension in any adoption-aware recommender is that recommending simpler actions to disadvantaged subgroups can entrench the disparities that produced the disadvantage in the first place. If a no-education farmer's adoption-probability landscape is depressed because of credit constraints, market-access frictions, or extension-coverage gaps — not because the farmer is incapable of executing a more aggressive recommendation — then an ERB-maximizing system that learns to recommend tier T_2 rather than T_4 to that farmer accommodates the structural barrier instead of removing it. We therefore frame ERB as a complement to structural-access policy: pair ERB recommendations with friction-removing instruments (credit subsidy, input-loan, cooperative purchasing); report realized-benefit gap closure dynamically across deployment seasons rather than only one-shot subgroup TPR; and treat low-$\mathbb{P}$(adopt)-on-high-yield-gain plots as targeting signal for structural intervention, not only as a yield-discount factor. Sociotechnically, the danger is treating ERB as a sufficient solution rather than one component of an integrated equity-aware advisory system.