WHEN STRATEGIC ADAPTATION IMPROVES DECISIONS: CAUSAL INCENTIVE CERTIFICATES FOR DECISION-FOCUSED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Decision-focused learning optimizes predictions for a downstream decision but normally treats the deployment population as fixed, even when the learned rule changes which actions people take. We prove that two predictors can be identical on all historical observations and have zero historical regret while their post-response deployment losses differ by any prescribed amount. We introduce *causal incentive certificates*, which map downstream-aware action returns to a strongly monotone strategic equilibrium and then to a lower bound on beneficial effort. A primitive causal-decision identity derives the regret slopes from true-state effects and model-induced decision distortion and gives the smallest beneficial-effort threshold that guarantees improvement uniformly over the stated envelope. Separate smooth and hard-decision branches quantify operator misspecification, anisotropic response covariance, and normal-fan switching, while an independent selected-policy test certifies deployment without a candidate-library penalty. Across 5,000 controlled effort instances, the certificate had no violations, the primitive identity closed with maximum error below 9×10^{-11}. In a UCI-based semi-synthetic study, CS-DFL achieved a worst-environment gain lower bound of 0.0496 and a positive alignment bound under the prespecified response model.

1 INTRODUCTION

Predictions are increasingly evaluated by the decisions they produce. Decision-focused learning (DFL) makes this principle operational: it uses gradients through, or otherwise accounts for, a downstream optimization problem and trains the predictive model against decision loss rather than prediction error alone (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022). The usual evaluation nevertheless freezes one consequential object: the population on which the learned rule is deployed. This convention is defensible when predictions are passive inputs. It is fragile when people observe, infer, or repeatedly experience the rule and then change their behavior. In that regime, the predictor is both an estimator and a system of prices on actions.

Credit allocation makes the distinction concrete. A lender predicts repayment-related costs and allocates a limited approval budget. A borrower who anticipates the rule may repay balances, acquire stable income, move debt across accounts, or change only a visible utilization statistic. These actions differ along two axes. They may receive different rewards from the model, and they may have different causal effects on the borrower's future repayment state. A rule can therefore be accurate on historical borrowers while rewarding the wrong action after deployment. Conversely, a rule that is locally sensitive to a manipulable variable may improve future decisions if it rewards an upstream action whose effects propagate to the true cost relevant to allocation. The target is not strategy-proofness. It is *decision-improving adaptation*.

Strategic classification explains best responses to decision rules (Hardt et al., 2016; Kleinberg & Raghavan, 2020); causal strategic-learning formulations distinguish gaming from changes in true outcomes and welfare (Miller et al., 2020; Shavit et al., 2020; Haghtalab et al., 2020; Horowitz & Rosenfeld, 2023; Chen et al., 2023); and performative prediction studies model-induced distributions (Perdomo et al., 2020; Mendler-Dünner et al., 2020). Most closely, Efthymiou et al. (2025) characterize desirable effort under causal feature relations and uncertainty about the classifier and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

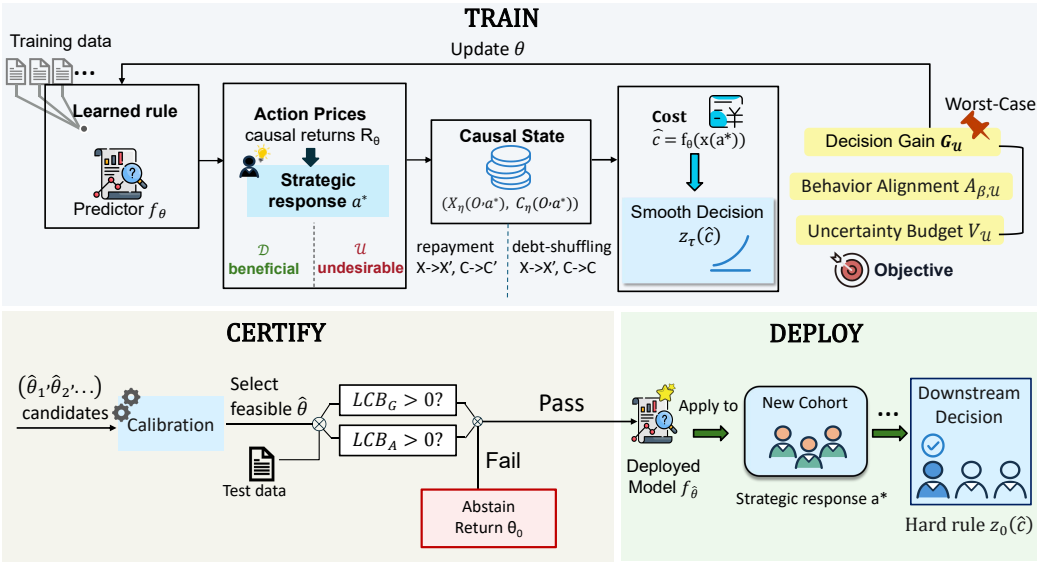

Figure 1: **Overview of Certified Strategic DFL (CS-DFL).** Training accounts for strategic responses and causal state changes, while independent held-out lower confidence bounds (LCBs) determine the deploy-or-abstain verdict.

graph. We instead develop a certificate that maps model-induced behavior to oracle regret for a capacity-constrained optimizer with action-dependent true costs. Section 4 proves why that bridge is necessary: historical predictions and regret can agree exactly while deployment losses differ by an arbitrary amount.

We address this gap with a causal incentive certificate. A downstream-aware marginal utility factors through decision preference, optimizer sensitivity, predictor sensitivity, and the causal action-to-feature map. A variational inequality then converts these returns and possibly coupled effort costs into a stable equilibrium. A separate primitive identity decomposes each regret derivative into a true-state channel and a prediction-induced decision channel. The final slopes are therefore derived from causal and optimization quantities rather than assumed signs. This separation is deliberate: a causal graph says what an intervention changes, whereas the designation "beneficial" records which changes the principal is prepared to value and defend.

The resulting threshold is counterintuitive. More aligned effort need not improve a decision, and stronger causal propagation need not help. Let $\kappa_\mathcal{D}$ be a certified unit reduction in regret from beneficial effort and $\kappa_\mathcal{U}$ an upper bound on the unit harm from other effort. If at least a fraction β of total effort is beneficial, the deterministic regret gain is certified only when

$$\beta > \beta_{\text{crit}} := \frac{\kappa_\mathcal{U}}{\kappa_\mathcal{D} + \kappa_\mathcal{U}}.$$

This is an if-and-only-if statement about the availability of a *uniform guarantee* from the envelope: below the threshold an admissible counterexample exists, although a particular instance may still improve. Under response noise, the gain must also be larger than a direction-aware covariance premium. Thus causal strength is useful only when it widens the right incentive gap and preserves the decision margin. Otherwise it amplifies a mistake. Figure 1 summarizes the certificate and its abstention boundary.

Contributions. Our contribution is a theory of when strategic adaptation changes downstream decision quality, with three results.

- **Historical nonidentification.** We model action-dependent features and true costs, the endogenous law P_θ, and cohort-feasible decisions. Exogenous evaluation cannot rank predictors by deployment regret, even when their historical predictions agree almost surely.
- **A primitive, sharp causal-decision certificate.** We bridge strategic incentives and downstream decision quality by decomposing regret derivatives into true-state and prediction-induced decision channels. This decomposition yields the smallest beneficial effort share supporting a uniform regret improvement guarantee. Strongly monotone variational inequalities characterize equilibrium responses, while operator error, anisotropic covariance, and hard polyhedral switches admit explicit robustness corrections.
- **A finite-sample deploy-or-abstain rule.** Training proposes candidates, calibration selects one, and an untouched test set alone authorizes deployment. We characterize the bounded-data sample complexity and test every observable implication in controlled simulations and a clearly delimited UCI semi-synthetic study.

Section 2 distinguishes adjacent paradigms. Sections 3–5 move from model to certificate and computation; Section 6 tests the certificate; Section 7 states the boundary and research agenda. The appendix gives complete proofs, diagnostics, provenance, and reproducibility details.

2 Related Work

Decision-aware learning and endogenous populations. DFL trains predictions against the loss of their downstream optimizer (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022). Robust DFL protects against coefficient error or exogenous shift (Yamao et al., 2026), and recursive DFL propagates prediction–optimization feedback (Wang et al., 2026); neither makes the rule-induced action and true cost the object of a deployment certificate. Performative prediction instead lets a deployed model change future data (Perdomo et al., 2020; Mendler-Dünner et al., 2020), including under constraints (Yan & Cao, 2023). We give that distribution map a strategic, causal microfoundation and ask whether the induced population improves oracle regret, rather than only whether learning converges.

Strategic effort and causal semantics. Strategic classification studies manipulation (Hardt et al., 2016); effort-allocation work asks which investments a rule induces (Kleinberg & Raghavan, 2020; Haghtalab et al., 2020); and causal strategic learning connects interventions to true outcomes or welfare (Miller et al., 2020; Shavit et al., 2020; Horowitz & Rosenfeld, 2023; Chen et al., 2023). Tsirtsis et al. (2024) optimize utility under strategic feature transitions for an individual policy. Efthymiou et al. (2025) provide the closest comparison, studying desirable effort under causal feature propagation and uncertainty. Our target is oracle regret for a learned predictor feeding a capacity-constrained optimizer with action-dependent true costs, certified uniformly over a prespecified graph-and-response library. Appendix D gives a componentwise comparison.

3 Problem Setup: Endogenous Decision Quality

Running example: two borrowers, one loan. Borrowers A and B are observationally identical in the historical data and compete for one approval. A repayment action changes both the visible feature X and future default cost C; a debt-shuffling action changes X but leaves C unchanged. Two predictors can coincide at the historical state while assigning opposite prices to these off-support actions. After deployment they induce different responses, predicted rankings, and loan recipients. This two-borrower mechanism supplies the counterexample in Theorem 4.1, the two channels in Proposition 4.3, and the ranking switch in Proposition 4.6. When either action effect is unsupported, the held-out rule abstains.

One statistical unit is an independent deployment unit $O \sim P_0$: an individual for an uncoupled decision or a complete cohort when a quota couples agents. A predeclared environment $\eta \in \mathfrak{U}$ specifies a causal graph, effort costs, a response model, and response noise. For a nonnegative action profile $a \in \mathbb{R}_+^{p_M}$, the structural map

$$T_\eta(O, a) = \big(X_\eta(O, a), C_\eta(O, a)\big)$$

returns post-response features and the true cost vector $C_\eta(O, a) \in \mathbb{R}^q$. Dependence of C_η on a separates substantive change from movement of model inputs alone.

The predictor f_θ supplies costs to a downstream optimizer

$$z_\tau(\widehat{c}) \in \arg\min_{z\in\mathcal{S}}\{\widehat{c}^\top z + \tau\Omega(z)\}, \qquad r_\tau(c, \widehat{c}) = c^\top z_\tau(\widehat{c}) - \min_{z\in\mathcal{S}} c^\top z.$$

Here $\mathcal{S} \subset \mathbb{R}^q$ is compact, and is convex whenever strong convexity is invoked to obtain a unique smooth layer; evaluation may instead use the hard rule $\tau = 0$ with fixed tie-breaking. In a quota allocation, for example, $k_{\text{cap}} \in \{1, \dots, N-1\}$ is an integer, $\mathcal{S} = \{z \in [0,1]^N : \mathbf{1}^\top z = k_{\text{cap}}\}$, and the hard selector returns a binary extreme point. One sampled cohort, not one borrower, is the independent unit.

Agents value the implemented joint decision and pay effort costs. Their deterministic response $a^*_{\theta,\eta}(O)$ is a selected Nash equilibrium (an ordinary best response for one agent); a random response $A_{\theta,\eta} \sim Q_{\theta,\eta}(\cdot \mid O)$ has conditional mean $\mu_{\theta,\eta}$ and covariance $\Sigma_{\theta,\eta}$. The resulting law and loss are

$$P_{\theta,\eta} = \mathcal{L}\{X_\eta(O, A_{\theta,\eta}), C_\eta(O, A_{\theta,\eta})\}, \qquad L_\eta(\theta) = \mathbb{E}\, r_\tau\{C_\eta(O, A_{\theta,\eta}), f_\theta(X_\eta(O, A_{\theta,\eta}))\}.$$

Thus the evaluated population depends on the predictor. The complete equilibrium definition, measurability requirements, and smooth-versus-hard boundary are given in Appendix A.

Let $\mathcal{D}$ and $\mathcal{U}$ partition manipulable coordinates into policy-designated beneficial and undesirable actions. This designation is normative input and must be checked against intervention evidence for C_η.With total effort $E(a) = \sum_m a_m$, define the alignment margin $h_\beta(a) = \sum_{j\in\mathcal{D}} a_j - \beta E(a)$; call a β-desirable when $h_\beta(a) \geq 0$. Fix a status quo θ_0 before test access and let $g_{\theta,\eta}$ be its paired deployment loss minus the candidate's. The two quantities that must remain nonnegative are

$$\mathcal{G}_{\mathfrak{U}}(\theta; \theta_0) = \inf_{\eta\in\mathfrak{U}} \mathbb{E} g_{\theta,\eta}, \qquad \mathcal{A}_{\beta,\mathfrak{U}}(\theta) = \inf_{\eta\in\mathfrak{U}} \mathbb{E} h_\beta(A_{\theta,\eta}).$$

The design objective maximizes $\mathcal{G}_{\mathfrak{U}}$ subject to alignment and a curvature-weighted response-uncertainty budget. Positive alignment alone never licenses a decision-improvement claim. Appendix C lists each assumption's role, diagnostic, and fallback.

4 Causal Incentive Certificates

The running example reveals why historical fit cannot certify deployment (Theorem 4.1). We first connect causal returns to equilibrium alignment (Theorem 4.2). Primitive geometry then yields the sharp alignment threshold for uniform improvement (Proposition 4.3;Theorem 4.4). Separate smooth and hard corrections qualify this guarantee (Proposition 4.5- 4.6). Finally, independent held-out testing governs deployment (Theorem 4.7).

4.1 Historical evaluation is insufficient

Theorem 4.1 (Off-support nonidentification)**.** *For every $B > 0$, there exist two affine predictors, a binary action, and a binary decision such that the predictors agree P_0-almost surely and both have zero historical decision regret, while their post-response deployment losses are 0 and B.*

The construction in Appendix E.1 can label its only nonzero action as beneficial. Perfect historical fit and perfectly aligned effort therefore do not identify deployment gain.

4.2 Causal returns and a selection-free equilibrium

For a closed convex action set $\mathcal{A}$, write $\mathrm{VI}(F, \mathcal{A})$ for a point a^* satisfying $\langle F(a^*), a - a^*\rangle \geq 0$ for every $a \in \mathcal{A}$. In a differentiable strategic game, F stacks agents' negative own-payoff gradients; in the potential model, $F = \nabla g - R_\theta$, where R_θ is the downstream-aware causal return(Facchinei & Pang, 2003). Appendix B.3 factors that return through decision preference, optimizer sensitivity, predictor sensitivity, and the intervention Jacobian.

Theorem 4.2 (Strongly monotone equilibrium certificate)**.** *Let $\mathcal{A} \subset \mathbb{R}^{p_M}$ be nonempty, compact, and convex. Suppose $F, \widehat{F} : \mathcal{A} \to \mathbb{R}^{p_M}$ are continuous and μ-strongly monotone, and $\sup_{a\in\mathcal{A}} \|F(a) -$*

$\widehat{F}(a)\|_2 \le \delta$. *Their VI solutions* $a^*, \widehat{a}^*$ *are unique and*

$$\|a^* - \widehat{a}^*\|_2 \le \frac{\delta}{\mu}, \qquad h_\beta(a^*) \ge h_\beta(\widehat{a}^*) - \|w_\beta\|_2 \frac{\delta}{\mu},$$

where $w_\beta = (1-\beta)\mathbf{1}_\mathcal{D} - \beta \mathbf{1}_\mathcal{U}$ *and* $h_\beta(a) = w_\beta^\top a$. *Thus a nominal margin larger than* $\|w_\beta\|_2 \delta/\mu$ *certifies the true equilibrium without an equilibrium-selection rule.*

This result permits heterogeneous, coupled costs and multi-agent interaction when the game pseudo-gradient is strongly monotone. If monotonicity fails, the fallback is a worst-equilibrium certificate over a predeclared response library. Under the homogeneous separable potential model with scalar effort cost ψ, let $\phi(r) = (\psi')^{-1}([r]_+)$ and let $\underline{R}_\mathcal{D}$ and $\overline{R}_\mathcal{U}$ be the worst-case lower desirable and upper undesirable returns. The interpretable special case is

$$\beta_{\text{cert}} = \frac{|\mathcal{D}|\phi(\underline{R}_\mathcal{D})}{|\mathcal{D}|\phi(\underline{R}_\mathcal{D}) + |\mathcal{U}|\phi(\overline{R}_\mathcal{U})},$$

with value 1 when $\overline{R}_\mathcal{U} \le 0$. Theorem B.4 gives the exact boundary conditions. Graph error ε_G replaces the nominal return gap $\widehat{\Gamma} := \min_{j\in\mathcal{D}} \widehat{R}_j - \max_{k\in\mathcal{U}} \widehat{R}_k$ by $\widehat{\Gamma} - 2\varepsilon_G$; the constant 2 is sharp.

4.3 Primitive geometry and the sharp decision threshold

The effort certificate is behavioral. We now derive when an action improves the principal's decision. Define $v_\mathcal{S}(c) = \min_{z\in\mathcal{S}} c^\top z$ and

$$r(a) = C(a)^\top z_\tau\{f_\theta(X(a))\} - v_\mathcal{S}\{C(a)\}, \qquad z_\theta(a) = z_\tau\{f_\theta(X(a))\}.$$

Proposition 4.3 (Primitive causal-decision geometry)**.** *Suppose* C, X, f_θ, z_τ *are differentiable at* a *and the true oracle* $z^*(C(a))$ *is unique. Then*

$$\partial_m r(a) = \underbrace{\{z_\theta(a) - z^*(C(a))\}^\top \partial_m C(a)}_{\textit{true-state channel}} + \underbrace{C(a)^\top D z_\tau \, D f_\theta \, \partial_m X(a)}_{\textit{prediction-decision channel}}. \tag{1}$$

Let $\gamma_\mathcal{D}, \gamma_\mathcal{U}, \delta_\mathcal{D}, \delta_\mathcal{U} \ge 0$. *If every certified nonnegative regret path is absolutely continuous and, almost everywhere along it, the first channel is at most* $-\gamma_\mathcal{D}$ *on* $\mathcal{D}$ *and at most* $\gamma_\mathcal{U}$ *on* $\mathcal{U}$*, while the absolute second channel is bounded by* $\delta_\mathcal{D}$ *and* $\delta_\mathcal{U}$ *respectively, then its path derivative is bounded by*

$$\kappa_\mathcal{D} = \gamma_\mathcal{D} - \delta_\mathcal{D} > 0, \quad \kappa_\mathcal{U} = \gamma_\mathcal{U} + \delta_\mathcal{U}, \qquad \frac{d}{dt} r(ta) \le -\kappa_\mathcal{D} E_\mathcal{D}(a) + \kappa_\mathcal{U} E_\mathcal{U}(a) \quad \textit{a.e.}$$

The bounds must hold along the response path and uniformly over the declared environments. At oracle ties, Appendix B.4 replaces the selected oracle by the worst directional minimizer.

Equation (1) separates the true-state effect from the prediction-induced decision effect. Along beneficial coordinates $j \in \mathcal{D}$, the condition $\gamma_\mathcal{D} > \delta_\mathcal{D}$ ensures that the true-state improvement dominates the prediction-induced distortion.

Theorem 4.4 (Smallest uniform alignment threshold)**.** *Let* $\mathcal{D}$ *and* $\mathcal{U}$ *be nonempty. Suppose that, for every certified nonnegative response* a, $t \mapsto r(ta)$ *is absolutely continuous and, for almost every* t, $\frac{d}{dt} r(ta) \le -\kappa_\mathcal{D} E_\mathcal{D}(a) + \kappa_\mathcal{U} E_\mathcal{U}(a)$, *where* $\kappa_\mathcal{D} > 0$ *and* $\kappa_\mathcal{U} \ge 0$. *Every nonzero* β*-desirable response obeys*

$$r(a) - r(0) \le -m(\beta) E(a), \qquad m(\beta) = \kappa_\mathcal{D}\beta - \kappa_\mathcal{U}(1-\beta).$$

Hence $\beta > \beta_{\text{crit}} := \kappa_\mathcal{U}/(\kappa_\mathcal{D} + \kappa_\mathcal{U})$ *guarantees strict improvement. This is the smallest threshold supporting a uniform guarantee under the envelope: for every* $\beta \le \beta_{\text{crit}}$*, an admissible linear-regret counterexample has nonnegative regret change.*

Substituting Proposition 4.3 gives the primitive threshold

$$\beta_{\text{crit}} = \frac{\gamma_\mathcal{U} + \delta_\mathcal{U}}{\gamma_\mathcal{D} - \delta_\mathcal{D} + \gamma_\mathcal{U} + \delta_\mathcal{U}}.$$

Below it, some instances can still improve; what fails is a universal conclusion from the stated information.

4.4 TWO ROBUSTNESS BRANCHES

Smooth and hard downstream rules require different corrections. They are not combined under incompatible regularity assumptions.

Proposition 4.5 (Anisotropic response uncertainty)**.** *On a smooth decision branch, suppose A is supported on the nonnegative action domain, r is C^2 on every segment joining $\mu = \mathbb{E}A$ to that support, and $\nabla^2 r \preceq H_r$ for a fixed $H_r \succeq 0$. If $\mathrm{Cov}(A) = \Sigma$, then*

$$\mathbb{E}r(A) \le r(\mu) + \tfrac{1}{2}tr(H_r\Sigma).$$

If μ is β-desirable, the same-policy gain survives whenever $m(\beta)E(\mu) > \frac{1}{2}tr(H_r\Sigma)$. Comparison with a distinct status quo additionally retains the no-response cross-policy offset $b_{\theta,\eta}$ from Corollary B.11. The scalar bound is the special case $H_r = L_r I$.

The covariance premium depends on the directions of response uncertainty, not merely its overall magnitude. Uncertainty concentrated along high-curvature directions can consume more of the certified gain than equally variable responses along flatter directions.

Proposition 4.6 (Polyhedral normal-fan exchange)**.** *Let $a \in \mathbb{R}_+^{p_M}$, let $\mathcal{S}$ be a fixed compact polytope, and let the hard optimizer use a vertex-valued fixed tie-breaker. Along $a(t) = ta$, suppose $c(t) = C(ta)$ and $\widehat{c}(t) = f_\theta(X(ta))$ are C^1; the candidate and oracle selections are piecewise constant with one-sided limits and finitely many switches; and the candidate has no endpoint boundary tie. If $\mathcal{T}_\theta \subset (0,1)$ contains its switch times, then*

$$r(1) - r(0) = \int_0^1 \{z_\theta(t) - z^*(t)\}^\top \dot{c}(t)\, dt + \sum_{s \in \mathcal{T}_\theta} c(s)^\top \{z_\theta(s^+) - z_\theta(s^-)\}.$$

Oracle switches have zero value jump. If the continuous part satisfies the envelope, the upper bound is $-m(\beta)E(a) + J_{\mathrm{fan}}$, where J_{fan} is the sum of positive candidate jumps. Hard top-k_{cap}, assignment, matching, and flow over an exact convex-hull representation are special cases.

4.5 HELD-OUT DEPLOYMENT IS AN INFERENCE DECISION

Theorem 4.7 (Selected-policy held-out certificate)**.** *Condition on all training and calibration data. Let the selected policy $\widehat{\theta}$, status quo, and finite environment library be fixed before n independent test units are opened. Write $\widehat{G} = \min_\eta P_n g_{\widehat{\theta},\eta}$ and $\widehat{A} = \min_\eta P_n h_\beta(A_{\widehat{\theta},\eta})$. If the gain and alignment variables have range lengths at most $2B$ and $2B_h$, then with probability at least $1-\alpha$,*

$$\mathcal{G}_{\mathfrak{U}}(\widehat{\theta};\theta_0) \ge \widehat{G} - B\sqrt{\frac{2\log(2/\alpha)}{n}}, \qquad \mathcal{A}_{\beta,\mathfrak{U}}(\widehat{\theta}) \ge \widehat{A} - B_h\sqrt{\frac{2\log(2/\alpha)}{n}}.$$

Neither radius contains the number of candidates or environments. Optimization affects power, not validity: an unsuccessful search may cause abstention, but independent testing does not certify a bad selected candidate as good except on the stated error event.

Raw-cost certification. The certificate applies equally to raw decision cost by replacing the paired regret gain with the paired raw-cost gain, provided that the corresponding boundedness and independence conditions hold.

The absence of an environment factor exploits the one-sided robust minimum after selection; simultaneous comparison of a policy library still pays multiplicity. Appendix B.5 also proves that a bounded-vector test separating $\min_\eta P g_\eta \le 0$ from $\min_\eta P g_\eta \ge \Delta$ with both errors at most α needs and suffices with order $B^2\Delta^{-2}\{\log N_{\mathfrak{U}} + \log(1/\alpha)\}$ samples. If structural coverage or cohort independence is unsupported, the theorem returns no deployment claim.

5 CERTIFIED RESPONSE-IN-THE-LOOP LEARNING

The operational separation is: *training proposes, calibration selects, testing certifies, and deployment obeys the certificate*. We call this procedure *Certified Strategic DFL* (CS-DFL). Training may use a smooth decision layer, exact response oracles, and any optimizer; validity comes only from the untouched hard-rule test in Theorem 4.7.

Algorithm 1 Certified Strategic Decision-Focused Learning

Require: Independent $\mathcal{D}_{\mathrm{tr}}, \mathcal{D}_{\mathrm{cal}}, \mathcal{D}_{\mathrm{te}}$; baseline θ_0; predeclared $\mathfrak{U}$; alignment β, uncertainty budget ρ, test level α

$\Theta_{\mathrm{cand}} \leftarrow \varnothing$
for training iterate $t = 0, \ldots, T-1$ **do**
  for minibatch and $\eta \in \mathfrak{U}$ **do**
    Solve or approximate the strategic response; propagate $(X, C) \leftarrow T_\eta(O, A_{\theta,\eta})$
    Estimate paired gain, alignment margin, and smooth response-uncertainty premium
  end for
  Update the predictor and constraint multipliers; save selected iterates in Θ_{cand}
end for
Use $\mathcal{D}_{\mathrm{cal}}$ to select one feasible $\widehat{\theta} \in \Theta_{\mathrm{cand}}$
On $\mathcal{D}_{\mathrm{te}}$, compute exact hard-rule lower confidence bounds LCB_G and LCB_A
return $\widehat{\theta}$ iff $\mathrm{LCB}_G > 0$ and $\mathrm{LCB}_A \geq 0$; otherwise return θ_0 with an abstention report

For each predeclared environment, training simulates the response, propagates both $X_\eta(O, a)$ and $C_\eta(O, a)$, and evaluates paired status-quo gain. A smooth maximum over environment-specific losses, alignment violations, and covariance premiums supplies the robust Lagrangian detailed in Appendix G. Calibration chooses one saved candidate without examining the test split. The test split evaluates exact decisions and returns the status quo whenever either lower confidence bound fails.

Training and evaluation branches are recorded separately. For top-k_{cap} allocation, a regularized capped-simplex layer supplies training gradients, while exact integer top-k_{cap} with fixed tie-breaking supplies test losses. This separation allows differentiable training while ensuring that deployment certification reflects the exact discrete decision rule. For nonconcave responses, multiple starts form a declared response library. For coupled quadratic effort cost, the exact response is a Q-metric projection, not coordinatewise clipping. Approximate solver and Monte Carlo errors are subtracted from the test bound; Appendix G gives a fixed-objective inexact-hypergradient rate but does not claim convergence of the full stochastic primal–dual game.

Algorithm 1 directly certifies paired gain and expected alignment. Primitive constants such as $(\gamma_{\mathcal{D}}, \delta_{\mathcal{D}})$ remain a complementary mechanism audit: interventions or sensitivity intervals explain why gain persists, while the held-out paired-gain bound alone determines whether deployment is statistically licensed.

6 Empirical Validation of the Certificate

The experiments evaluate each component of the certificate and then execute the full propose-select-certify workflow. All values below were regenerated by the included script with master seed 20260914. Controlled experiments expose ground-truth primitives; the UCI study uses real covariates and labels but declared synthetic actions and responses.

Effort, threshold, primitive geometry, and operator error. We drew 5,000 quadratic-cost response instances with three beneficial and two undesirable actions. Panel 2a compares the endpoint certificate with exact equilibrium effort: the violation rate was 0, median slack was 0.209, and fifth-percentile slack was 0.0118. Across 4,500 independently drawn decision envelopes, every point above β_{crit} reduced regret (panel b). For 1,800 smooth four-person allocation instances, the two terms in (1) matched central finite differences with maximum error below 9×10^{-11} (panel c). Finally, 1,800 random positive-definite VI systems respected the δ/μ displacement bound, with maximum tightness 0.993 (panel d).

Anisotropic noise and hard decisions. Panel 3a compares two response covariances with the same trace, 0.32. Rotating uncertainty from a low- to a high-curvature direction increased the exact quadratic premium from 0.1325 to 1.2035, a factor of 9.1; 300,000 Monte Carlo responses agreed within 2.7×10^{-4}. Panel b uses a 4×4 assignment polytope. Three candidate and two oracle switches give decomposition error 2.8×10^{-16}; oracle switches are continuous in value, while candidate normal-fan crossings create the charged jumps.

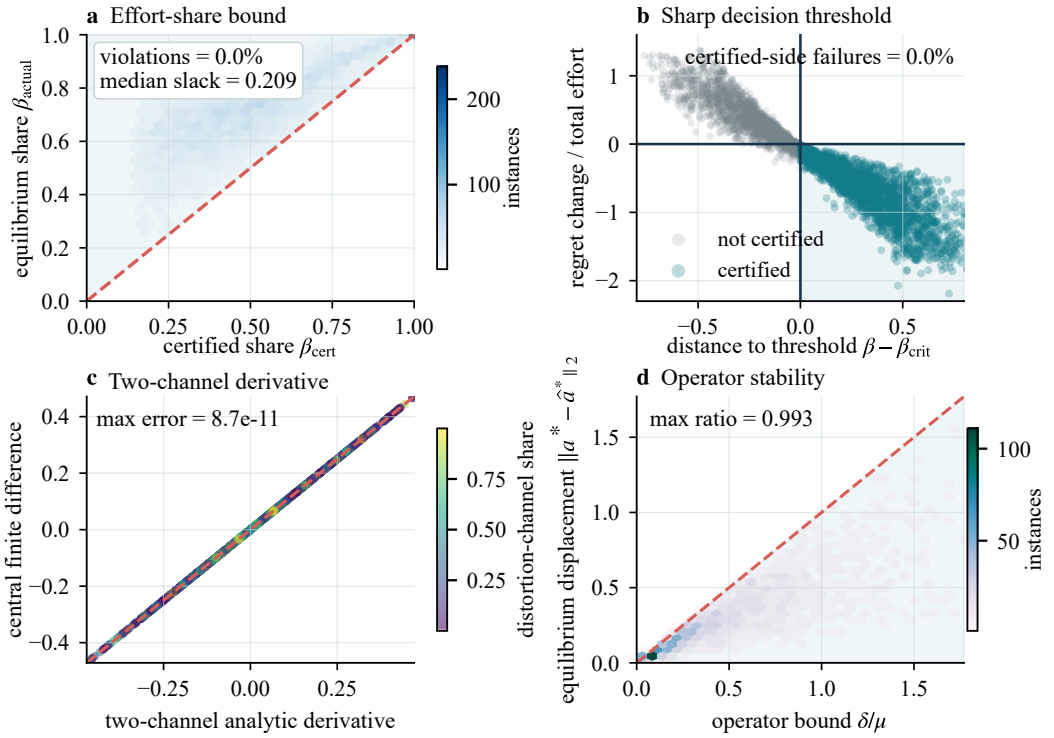

Figure 2: **The four structural inequalities hold instance by instance in the controlled benchmark.** Points represent exact simulated responses; dashed coral lines mark the corresponding theoretical boundaries.

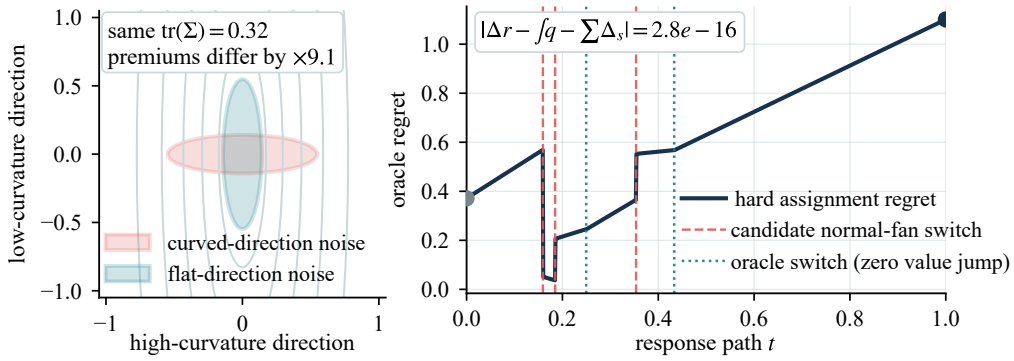

Figure 3: **Smooth uncertainty is directional, whereas hard decisions pay discrete exchange costs.** Left: equal total variance can have a nine-fold different regret premium. Right: the signed candidate jumps plus continuous drift recover the exact hard-assignment regret path; oracle switches have zero value jump.

Finite-sample safety. We fixed a policy before test access, sampled bounded gain vectors for eight environments, and repeated each experiment 5,000 times at $\alpha = 0.05$. Under a boundary null, the rule falsely certified at most 0.10% across the tested sample sizes. Under a safe alternative with robust gain 0.32, abstention fell from 96.0% at $n = 40$ to 0 by $n = 600$, while the conditional gain lower bound increased. Panel 4a also isolates the power saved by testing one fixed policy rather than simultaneously reporting a library.

UCI semi-synthetic train-calibrate-test study. We trained a logistic base risk model on 15,000 records from UCI Default of Credit Card Clients (Yeh, 2009), selected from 84 response-aware candidates on 6,000 calibration records, and held 9,000 records out from fitting and selection as an

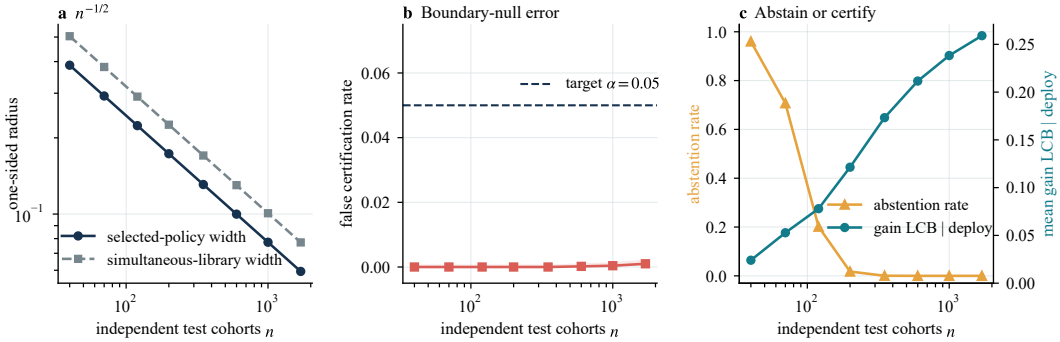

Figure 4: **Independent evidence converts abstention into certified gain.** a, selected-policy versus simultaneous width; b, 95% Wilson intervals over 5,000 null repetitions; c, abstention and conditional gain LCB under the safe alternative.

Table 1: **UCI semi-synthetic certification separates improvement from permission to deploy.** The 95% test requires positive gain and alignment LCBs.

Method	Worst gain	Gain LCB	Align. LCB	Test cohorts	Verdict
Historical DFL	0.0000	−0.0384	−0.0306	5,000	abstain
Responsive DFL	0.0769	0.0384	0.1472	5,000	deploy
Causal SC-DFL	0.0875	0.0491	0.2313	5,000	deploy
Aligned DFL	0.0777	0.0393	**0.2769**	5,000	deploy
CS-DFL	**0.0880**	**0.0496**	0.2085	5,000	deploy

empirical test generator. Each evaluation draw formed an independent 20-person cohort and approved $k_{\text{cap}} = 6$. Two repayment actions lower the synthetic true default cost; debt reshuffling changes the visible score and can increase that cost. Four predeclared environments vary these effects and effort costs. Here, deployment gain is measured by the reduction in raw decision cost relative to the status quo. Table 1 reports 5,000 paired test cohorts. CS-DFL achieved the largest observed worst-environment gain and gain LCB; Causal SC-DFL was close in point gain, while Aligned DFL traded gain for the largest alignment LCB. These certificates are conditional on the empirical covariate generator and the explicit structural equations in Appendix H.

The ablation changes the calibration criterion: Responsive DFL minimizes predicted post-response decision cost; Causal SC-DFL maximizes nominal paired gain; Aligned DFL prioritizes worst-environment alignment; CS-DFL maximizes least-favorable gain subject to alignment.

Additional semi-synthetic experiments on three credit-risk datasets are reported in Appendix H. Across all three datasets, CS-DFL satisfies both deployment criteria, while Responsive DFL fails the alignment certificate on two datasets despite positive gain lower bounds. Equations, coefficients, audit tables, and provenance are also provided there, together with an additional frontier.

7 Discussion

Strategic behavior is neither inherently adversarial nor beneficial. Its sign is jointly determined by incentive geometry, actions' causal effects, and the downstream optimizer. An effort label supports a decision claim only when true-state improvement dominates prediction-induced distortion; coupled responses additionally require equilibrium perturbations smaller than the certified alignment margin.

Abstention is an inference result, not an optimization failure. The deployment decision therefore depends on the strength of the available evidence as well as the estimated improvement. A candidate with positive empirical gain may still retain the status quo when its lower confidence bound does not establish improvement.

ETHICS STATEMENT

Credit allocation raises fairness, privacy, and legal concerns. Because the response mechanism is semi-synthetic, the results do not establish transportability or fairness in operational lending. Any field use would require intervention-based validation, protected-group audits, and legal review.

AI USE STATEMENT

Generative AI tools assisted with language editing, literature triage, consistency checks, figure and experiment code, and formalization of selected algebraic proof cores. All numerical results were regenerated by the disclosed executable workflow; AI output is not treated as independent evidence. AI-assisted content was checked against cited sources, executable results, and the included proof artifacts; the authors take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

We provide the source code, experimental scripts, configuration files, and supplementary materials necessary to reproduce the empirical results reported in this paper. The main paper and appendix specify the experimental settings, data preprocessing procedures, evaluation metrics, and implementation details. All experiments use fixed random seeds where applicable. The datasets used in our experiments are publicly available, and the corresponding data preparation procedures are documented in the supplementary materials. For the theoretical results, complete proofs are provided in the appendix, and formal verification files for selected mathematical claims are included in the supplementary materials. These materials are provided to facilitate independent reproduction and verification of our results.

REFERENCES

Yatong Chen, Jialu Wang, and Yang Liu. Learning to incentivize improvements from strategic agents. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=W98AEKQ38Y.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5484–5494. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3fc2c60b5782f641f76bcefc39fb2392-Abstract.html.

Valia Efthymiou, Chara Podimata, Diptangshu Sen, and Juba Ziani. Incentivizing desirable effort profiles in strategic classification: The role of causality and uncertainty. In *Advances in Neural Information Processing Systems*, volume 38, pp. 146550–146586. Curran Associates, Inc., 2025. doi: 10.52202/085713-4902. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/d7ad6951ce4f6063469adbc7f54cd711-Abstract-Conference.html.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2003. doi: 10.1007/b97543.

Nika Haghtalab, Nicole Immorlica, Brendan Lucier, and Jack Z. Wang. Maximizing welfare with incentive-aware evaluation mechanisms. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pp. 160–166. International Joint Conferences on Artificial Intelligence Organization, 2020. doi: 10.24963/ijcai.2020/23.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 111–122, 2016. doi: 10.1145/2840728.2840730.

Guy Horowitz and Nir Rosenfeld. Causal strategic classification: A tale of two shifts. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 13233–13253. PMLR, 2023. URL https://proceedings.mlr.press/v202/horowitz23a.html.

Kaggle. Give me some credit. Kaggle Competition, 2011.

Kaggle. Home credit default risk. Kaggle Competition, 2018.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation*, 8(4):1–23, 2020. doi: 10.1145/3417742.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4929–4939. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/33e75ff09dd601bbe69f351039152189-Abstract.html.

John Miller, Smitha Milli, and Moritz Hardt. Strategic classification is causal modeling in disguise. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6917–6926, 2020. URL https://proceedings.mlr.press/v119/miller20b.html.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Yonadav Shavit, Benjamin Edelman, and Brian Axelrod. Causal strategic linear regression. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8676–8686. PMLR, 2020. URL https://proceedings.mlr.press/v119/shavit20a.html.

Stratis Tsirtsis, Behzad Tabibian, Moein Khajehnejad, Adish Singla, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. Optimal decision making under strategic behavior. *Management Science*, 70 (12):8506–8519, 2024. doi: 10.1287/mnsc.2021.02567.

UCI Machine Learning Repository. South german credit. UCI Machine Learning Repository, 2020.

Xinyu Wang, Jinxiao Du, Yiyang Peng, and Wei Ma. From sequential to recursive: Enhancing decision-focused learning with bidirectional feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26579–26587, 2026. doi: 10.1609/aaai.v40i31.39866.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658.

Shoki Yamao, Ken Kobayashi, Ryo Matsui, Shota Nagai, Naoki Nishimura, and Kazuhide Nakata. Robust decision-focused learning via worst-case regret minimization. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 7701–7736. PMLR, 2026. URL https://proceedings.mlr.press/v337/yamao26a.html.

Wenjing Yan and Xuanyu Cao. Zero-regret performative prediction under inequality constraints. In *Advances in Neural Information Processing Systems*, volume 36, pp. 1298–1308. Curran Associates, Inc., 2023. doi: 10.52202/075280-0063. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/047397849f63b4fcfced4ff720159f3d-Abstract-Conference.html.

I-Cheng Yeh. Default of credit card clients. UCI Machine Learning Repository, 2009.

APPENDIX CONTENTS

A Detailed Problem Setup . . . 13
A.1 Deployment Units, Causal States, and Decisions . . . 13
A.2 Strategic Response and the Endogenous Law . . . 14
A.3 Behavioral Semantics and Robust Improvement . . . 15
A.4 Design Objective and Diagnosable Assumptions . . . 15
B Detailed Theorem Statements . . . 15
B.1 Why Exogenous Evaluation Is Insufficient . . . 16
B.2 Strongly Monotone Strategic Equilibria . . . 16
B.3 Robust Causal Incentive Alignment . . . 16
B.4 The Alignment-to-Decision Bridge . . . 17
B.5 Finite-Sample and Asymptotic Certificates . . . 19
C Notation and Assumption Audit . . . 21
D Extended Literature Comparison . . . 23
E Proofs of Structural Results . . . 23
E.1 Historical Equivalence and Deployment Nonidentification . . . 23
E.2 Strongly Monotone Variational Inequalities . . . 23
E.3 Downstream-Aware Returns . . . 23
E.4 From Effort Direction to Decision Regret . . . 25
F Proofs of Statistical Certificates . . . 27
G Hypergradient Error for a Fixed Smooth Objective . . . 29
H Additional Experimental Evidence . . . 31
H.1 Five Controlled Audit Blocks . . . 31
H.2 UCI Semi-Synthetic Protocol . . . 32
H.3 Generalization Across Credit Datasets . . . 33
H.4 Artifacts and Seeded Regeneration . . . 35
I Reproducibility and Falsification Checklist . . . 35
J Machine-Checked Proof Scope . . . 36

A DETAILED PROBLEM SETUP

We separate three objects that are conflated by a fixed-distribution analysis: the historical deployment unit, the environment governing strategic response, and the post-response state on which a joint decision is made. This separation lets the model cover an individual decision as well as a capacity-constrained cohort allocation.

A.1 DEPLOYMENT UNITS, CAUSAL STATES, AND DECISIONS

Let $O \sim P_0$ be an i.i.d. pre-deployment unit. For an individual decision, O contains one agent's covariates and exogenous state. For a budgeted allocation, O is an entire cohort; its feature component is a matrix $X \in \mathbb{R}^{N \times p}$ and one observation for statistical analysis is one independently sampled cohort. This convention avoids treating agents coupled by a quota as independent observations.

An environment $\eta \in \mathfrak{U}$ specifies the causal propagation graph, effort technology, response model, and any exogenous response noise considered plausible before the evaluation set is inspected. For an action profile $a \in \mathcal{A}_\eta$, its structural map is

$$T_\eta(O, a) = \big(X_\eta(O, a), C_\eta(O, a)\big), \tag{2}$$

where $X_\eta(O, a)$ is the post-response feature array and $C_\eta(O, a) \in \mathbb{R}^q$ is the *true* cost vector relevant to the principal. Allowing C_η to depend on a is essential. An action that changes only X_η is gaming relative to the downstream state; an action that changes C_η may constitute substantive improvement. The set $\mathfrak{U}$ can be a finite library of graphs and response models, as in our finite-sample certificates, or a continuous ambiguity set approximated by a declared net.

The predictor f_θ maps post-response features to predicted costs. Given $\widehat{c}$, the principal chooses

$$z_\tau(\widehat{c}) \in \operatorname*{arg\,min}_{z \in \mathcal{S}} \left\{ \widehat{c}^\top z + \tau \Omega(z) \right\}, \qquad \tau \geq 0, \tag{3}$$

where $\mathcal{S} \subset \mathbb{R}^q$ is nonempty and compact. Whenever uniqueness and differentiability are inferred from a strongly convex Ω with $\tau > 0$, we additionally assume that $\mathcal{S}$ is convex; otherwise we explicitly require a unique locally differentiable selected branch. The unregularized map z_0 includes linear and combinatorial decisions after a fixed tie-breaking rule. In a top-k_{cap} allocation, for example, $q = N$, $k_{\text{cap}} \in \{1, \ldots, N-1\}$ is an integer, and $\mathcal{S} = \{z \in [0,1]^N : \mathbf{1}^\top z = k_{\text{cap}}\}$ couples all members of a cohort; the hard tie-breaking rule selects a binary extreme-point minimizer. Thus the mathematical decision matches the quota used in the credit experiments; it is not a collection of independent per-borrower optimizations.

For true cost c and prediction $\widehat{c}$, define the raw decision cost and regret by

$$d_\tau(c, \widehat{c}) := c^\top z_\tau(\widehat{c}), \qquad r_\tau(c, \widehat{c}) := d_\tau(c, \widehat{c}) - \min_{z \in \mathcal{S}} c^\top z \geq 0. \tag{4}$$

Regularization is a device for a unique, differentiable response channel; regret is always evaluated against the true unregularized oracle.

A.2 Strategic Response and the Endogenous Law

The deployed rule changes action incentives through the final decision. Assume $\mathcal{A}_\eta = \prod_i \mathcal{A}_{i,\eta}$. Agent i values its component of the joint decision according to $u_{i,\eta}(z, X, O)$ and pays effort cost $g_{i,\eta}(a_i; O)$. For a unilateral deviation $\widetilde{a}_i$, write

$$w_{i,\theta,\eta}(\widetilde{a}_i, a_{-i}; O) := u_{i,\eta}(z_\tau(f_\theta(X_\eta(O, (\widetilde{a}_i, a_{-i})))), X_\eta(O, (\widetilde{a}_i, a_{-i})), O) - g_{i,\eta}(\widetilde{a}_i; O).$$

A deterministic response profile is a selected Nash equilibrium

$$a^*_{\theta,\eta}(O) \in \mathrm{BR}_\eta(\theta, O) := \left\{ a \in \mathcal{A}_\eta : a_i \in \operatorname*{arg\,max}_{\widetilde{a}_i \in \mathcal{A}_{i,\eta}} w_{i,\theta,\eta}(\widetilde{a}_i, a_{-i}; O) \text{ for every } i \right\}. \tag{5}$$

For a single representative agent, BR_η is an ordinary argmax. For a cohort, it is a Nash equilibrium selection; the structural alignment theorem in Section 4 applies when this game admits the stated differentiable potential representation. The experiments use a price-taking response approximation and a smooth training relaxation, then evaluate the hard cohort decision.

To represent heterogeneous costs, bounded rationality, or incomplete knowledge, let $A_{\theta,\eta}(O) \sim Q_{\theta,\eta}(\cdot \mid O)$ be a random response with mean $\mu_{\theta,\eta}(O)$ and covariance $\Sigma_{\theta,\eta}(O)$. The deterministic model is the point-mass special case. The deployment law is

$$P_{\theta,\eta} := \mathcal{L}(X_\eta(O, A_{\theta,\eta}), C_\eta(O, A_{\theta,\eta})), \tag{6}$$

which depends on the very predictor evaluated under it.

The environment-specific deployment loss is

$$\ell_{\theta,\eta}(O, a) := r_\tau(C_\eta(O, a), f_\theta(X_\eta(O, a))), \qquad L_\eta(\theta) := \mathbb{E}\, \ell_{\theta,\eta}(O, A_{\theta,\eta}). \tag{7}$$

We reserve $D_\eta(\theta) := \mathbb{E}\, d_\tau(C_\eta(O, A_{\theta,\eta}), f_\theta(X_\eta(O, A_{\theta,\eta})))$ for raw decision cost; L_η is its regret counterpart after subtracting the action-specific oracle cost. The expectations include cohort sampling and, when present, response randomness. For the finite-sample certificate, one held-out observation must additionally contain, or admit an unbiased estimate of, the candidate and baseline potential losses under every predeclared environment. A structural simulator can provide paired evaluations by sharing exogenous response seeds; observational outcomes alone generally cannot.

A.3 Behavioral Semantics and Robust Improvement

Let $\mathcal{D}$ and $\mathcal{U}$ partition manipulable action coordinates into policy-designated beneficial and undesirable directions. This label is normative input, not a conclusion inferred from correlation. In lending, an institution might designate verifiable repayment as beneficial and balance reshuffling as undesirable, but the designation must be justified by domain policy and audited against $C_\eta(O, a)$. The theory deliberately keeps these steps separate: Section B.3 certifies where effort goes, while Section B.4 tests whether that effort improves the true decision state.

For $\beta \in [0, 1]$, $E_{\mathcal{D}}(a) = \sum_{j\in\mathcal{D}} a_j$, $E_{\mathcal{U}}(a) = \sum_{k\in\mathcal{U}} a_k$, and $E(a) = E_{\mathcal{D}}(a) + E_{\mathcal{U}}(a)$, define

$$h_\beta(a) := E_{\mathcal{D}}(a) - \beta E(a). \tag{8}$$

An action is β-desirable exactly when $h_\beta(a) \geq 0$. A weighted version replaces the binary partition by declared weights in $[-1, 1]$; we use the partition because it yields an interpretable threshold.

Fix a status-quo predictor θ_0 before evaluation. For every environment, define the paired deployment gain

$$g_{\theta,\eta}(O) := \ell_{\theta_0,\eta}(O, A_{\theta_0,\eta}) - \ell_{\theta,\eta}(O, A_{\theta,\eta}), \tag{9}$$

and the worst-case gain and alignment margins

$$\mathcal{G}_{\mathfrak{U}}(\theta; \theta_0) := \inf_{\eta\in\mathfrak{U}} \mathbb{E} g_{\theta,\eta}(O), \tag{10}$$

$$\mathcal{A}_{\beta,\mathfrak{U}}(\theta) := \inf_{\eta\in\mathfrak{U}} \mathbb{E} h_\beta(A_{\theta,\eta}). \tag{11}$$

Only $\mathcal{G}_{\mathfrak{U}} > 0$ warrants the phrase *robust downstream improvement*. A positive alignment margin alone does not.

A.4 Design Objective and Diagnosable Assumptions

At population level, the principal seeks a safe improvement over θ_0 subject to behavioral and response-uncertainty budgets:

$$\max_{\theta\in\Theta} \mathcal{G}_{\mathfrak{U}}(\theta; \theta_0) \quad \text{s.t.} \quad \mathcal{A}_{\beta,\mathfrak{U}}(\theta) \geq 0, \qquad \mathcal{V}_{\mathfrak{U}}(\theta) \leq \rho, \tag{12}$$

where

$$\mathcal{V}_{\mathfrak{U}}(\theta) := \sup_{\eta\in\mathfrak{U}} \frac{1}{2} \mathbb{E}\mathrm{tr}\{H_{r,\eta}\Sigma_{\theta,\eta}(O)\} \tag{13}$$

is the curvature certificate developed in Theorem B.10. The covariance term upper-bounds the increase in expected regret relative to the regret at the mean response. Training uses the Lagrangian of (12), while model selection uses the held-out lower confidence bound in Section B.5.

The guarantees are conditional, falsifiable, and allowed to abstain. The ambiguity set must contain the deployment environment; the response model must approximate observed or elicited behavior; the beneficial-action labels must withstand policy review; and the differentiable relaxation must be checked against the hard decision. Appendix C records, for every assumption, its role, a diagnostic, and the fallback when it fails.

Standard formulations are recovered as special cases. A singleton $\mathfrak{U}$ gives nominal strategic DFL; $A_{\theta,\eta} \equiv 0$ gives ordinary DFL; $C_\eta(O, a) = C_\eta(O, 0)$ permits gaming but not substantive outcome change; and $\tau = 0$ recovers an exact discrete optimizer wherever the selected response map is well defined.

B Detailed Theorem Statements

The results answer five questions in order. Can historical loss determine deployment loss? When is strategic equilibrium unique and stable? Which primitive causal-decision quantities sign regret? How do smooth and hard decisions consume the margin? Can deployment be certified from finite data? The first answer is negative. Each positive result states a diagnostic and a fallback.

B.1 Why Exogenous Evaluation Is Insufficient

Theorem B.1 (Historical equivalence, arbitrary deployment gap)**.** *For every $B > 0$, there is a strategic DFL problem and two predictors f_0, f_1 such that $f_0(X) = f_1(X)$ P_0-almost surely and both have zero historical decision regret, while their post-response deployment losses are respectively* 0 *and* B*.*

The construction uses one historical feature value, a binary action, and a binary downstream decision; Appendix E.1 gives it explicitly. The result is an off-support impossibility, not a claim that every deployment is unstable. It shows that no validation statistic based only on historical predictions can certify decision quality without restricting how rules induce actions and how actions change the state.

The same example can label the unique action direction as desirable, making every nonzero response 1-desirable while increasing regret by B. Thus alignment labels cannot substitute for an outcome model. This motivates the two-stage certificate below: first certify effort direction, then independently certify its decision effect.

B.2 Strongly Monotone Strategic Equilibria

Let $\mathcal{A} = \prod_i \mathcal{A}_i \subset \mathbb{R}^{p_M}$ be a nonempty compact convex action set. If each agent has differentiable loss $L_i(a_i, a_{-i})$ convex in its own action, define the game pseudogradient

$$F(a) = \big(\nabla_{a_1} L_1(a), \ldots, \nabla_{a_N} L_N(a)\big).$$

A Nash equilibrium solves $\mathrm{VI}(F, \mathcal{A})$; the converse holds by own-action convexity. A one-agent response and a potential game are special cases.

Theorem B.2 (VI uniqueness and operator stability)**.** *Suppose $F, \widehat{F} : \mathcal{A} \to \mathbb{R}^{p_M}$ are continuous and each is μ-strongly monotone in Euclidean norm. Then their variational inequalities have unique solutions a^* and $\widehat{a}^*$. If*

$$\sup_{a \in \mathcal{A}} \|F(a) - \widehat{F}(a)\|_2 \le \delta,$$

then

$$\|a^* - \widehat{a}^*\|_2 \le \delta/\mu, \tag{14}$$

$$h_\beta(a^*) \ge h_\beta(\widehat{a}^*) - \|w_\beta\|_2 \delta/\mu, \tag{15}$$

where $w_\beta = (1-\beta)\mathbf{1}_{\mathcal{D}} - \beta \mathbf{1}_{\mathcal{U}}$. Also $E(a^*) \ge E(\widehat{a}^*) - \sqrt{p_M}\delta/\mu$. The constants $1/\mu$ and $\|w_\beta\|_2/\mu$ are attained in one-dimensional affine examples.

The same feasible action set is essential; a moving feasible set also perturbs its normal cone. If strong monotonicity is not credible, the paper makes no uniqueness claim and instead ranges the certificate over a predeclared equilibrium library.

B.3 Robust Causal Incentive Alignment

We now specialize the response game to a representative agent or an exact potential game with the same homogeneous separable marginal-cost function across the coordinates and environments covered by the certificate. For fixed (θ, η, O), suppose the selected response maximizes

$$\Psi_{\theta,\eta}(a; O) := V_{\theta,\eta}(a; O) - \sum_{m=1}^{p_M} \psi(a_m) \quad \text{over } a \in \mathbb{R}_+^{p_M}, \tag{16}$$

where $V_{\theta,\eta}$ is the downstream utility induced by (5). Assume $V_{\theta,\eta}$ is differentiable at every selected optimum and ψ is continuously differentiable and strictly convex, with $\psi'(0) = 0$ and $\psi'(t) \to \infty$. Define

$$R^\theta_{m,\eta}(a; O) := \frac{\partial}{\partial a_m} V_{\theta,\eta}(a; O), \qquad \phi(r) := (\psi')^{-1}([r]_+). \tag{17}$$

Unlike a raw score derivative, $R^\theta_{m,\eta}$ retains the downstream decision channel.

Proposition B.3 (Downstream-aware causal factorization)**.** *Suppose* $V_{\theta,\eta}(a;O) = u_\eta(z_\tau(f_\theta(X_\eta(O,a))), f_\theta(X_\eta(O,a)), X_\eta(O,a), O)$ *and every displayed map is differentiable at* a*. Then*

$$R^\theta_{m,\eta} = \left(J_{z_\tau}^\top \nabla_z u_\eta + \nabla_{\widehat{c}} u_\eta\right)^\top J_{f_\theta} J_{X_\eta,a_m} + (\nabla_X u_\eta)^\top J_{X_\eta,a_m}, \tag{18}$$

where $\nabla_{\widehat{c}} u_\eta = 0$ *when utility has no direct dependence on predicted cost, and all Jacobians are evaluated along the deployed path. Equivalently, when utility depends on* a *only through the decision, the return is the product of decision preference, optimizer sensitivity, predictor sensitivity, and the action-to-feature Jacobian. Marginal effort cost enters only afterward, through* (16)*.*

The explicit direct-prediction term in (18) is included for models in which an announced score has value beyond the implemented decision; it vanishes in our canonical specification. The factorization identifies what must be bounded to obtain a robust margin. Define, over all selected equilibria in the ambiguity set,

$$\underline{R}_{\mathcal{D}} := \inf_{\eta,O,j\in\mathcal{D}} R^\theta_{j,\eta}(a^*_{\theta,\eta};O), \qquad \overline{R}_{\mathcal{U}} := \sup_{\eta,O,k\in\mathcal{U}} R^\theta_{k,\eta}(a^*_{\theta,\eta};O). \tag{19}$$

Theorem B.4 (Uniform effort-share certificate)**.** *Assume* (16)*, let* $\mathcal{D}$ *and* $\mathcal{U}$ *be nonempty, and suppose every selected response has positive total effort. If* $\underline{R}_{\mathcal{D}} > \overline{R}_{\mathcal{U}}$*, then every response in* $\mathfrak{U}$ *is* β_{cert}*-desirable, where*

$$\beta_{\mathrm{cert}} = \begin{cases} 1, & \overline{R}_{\mathcal{U}} \le 0, \\ \dfrac{|\mathcal{D}|\phi(\underline{R}_{\mathcal{D}})}{|\mathcal{D}|\phi(\underline{R}_{\mathcal{D}}) + |\mathcal{U}|\phi(\overline{R}_{\mathcal{U}})}, & \overline{R}_{\mathcal{U}} > 0. \end{cases} \tag{20}$$

The certificate increases with the worst-case desirable return and decreases with the worst-case undesirable return.

The common ψ assumption is what makes (20) a two-endpoint formula. Heterogeneous separable costs require coordinate-specific inverse marginal-cost envelopes; cross-action costs require the joint KKT geometry in Proposition B.6 and are not covered by Theorem B.4.

Corollary B.5 (Graph-error degradation)**.** *Suppose nominal returns* $\widehat{R}_m$ *satisfy* $|R^\theta_{m,\eta} - \widehat{R}^\theta_m| \le \varepsilon_G$ *uniformly over the certificate region. If the nominal gap is* $\widehat{\Gamma} = \min_{j\in\mathcal{D}} \widehat{R}^\theta_j - \max_{k\in\mathcal{U}} \widehat{R}^\theta_k$*, then the true gap is at least* $\widehat{\Gamma} - 2\varepsilon_G$*. Hence Theorem B.4 applies whenever* $\widehat{\Gamma} > 2\varepsilon_G$*. A computable conservative certificate is obtained by substituting* $\underline{R}^{\mathrm{cons}}_{\mathcal{D}} := \min_j \widehat{R}^\theta_j - \varepsilon_G$ *and* $\overline{R}^{\mathrm{cons}}_{\mathcal{U}} := \max_k \widehat{R}^\theta_k + \varepsilon_G$ *for the unknown endpoints in* (20)*.*

Proposition B.6 (Correct linear-quadratic response)**.** *Let* $V(a) = b^\top a$ *and let the effort cost be* $\frac{1}{2} a^\top Q a$ *for* $Q \succ 0$*. The unique response is*

$$a^* = \Pi^Q_{\mathbb{R}^{p_M}_+}(Q^{-1}b) := \arg\min_{a\ge 0} \tfrac{1}{2}\|a - Q^{-1}b\|^2_Q. \tag{21}$$

Only when Q *is diagonal does this reduce to the coordinatewise formula* $a^* = [Q^{-1}b]_+$*.*

B.4 The Alignment-to-Decision Bridge

Fix (θ,η,O) and abbreviate $r(a) = \ell_{\theta,\eta}(O,a)$. The next condition is substantive rather than merely smoothness: along the realized response path, beneficial actions must improve the true cost-decision pair.

Let $v_{\mathcal{S}}(c) = \min_{z\in\mathcal{S}} c^\top z$ and $Z^*(c) = \arg\min_{z\in\mathcal{S}} c^\top z$.

Proposition B.7 (Primitive decision-benefit bounds)**.** *Suppose* C, X, f_θ, z_τ *are differentiable at* a*. If* $Z^*(C(a)) = \{z^*(C(a))\}$*, then the ordinary derivative is*

$$\partial_m r(a) = \{z_\theta(a) - z^*(C(a))\}^\top \partial_m C(a) + C(a)^\top Dz_\tau \, Df_\theta \, \partial_m X(a). \tag{22}$$

At an oracle tie, the right directional derivative in coordinate e_m *is*

$$Dr(a;e_m) = \max_{z^*\in Z^*(C(a))} \{z_\theta(a) - z^*\}^\top \partial_m C(a) + C(a)^\top Dz_\tau \, Df_\theta \, \partial_m X(a). \tag{23}$$

Let $\gamma_{\mathcal{D}}, \gamma_{\mathcal{U}}, \delta_{\mathcal{D}}, \delta_{\mathcal{U}} \geq 0$. Suppose every certified nonnegative regret path is absolutely continuous and, uniformly at its almost-everywhere directional differentiability points and across environments, the first term (maximized over oracle minimizers at ties) is at most $-\gamma_{\mathcal{D}}$ for $j \in \mathcal{D}$ and at most $\gamma_{\mathcal{U}}$ for $k \in \mathcal{U}$, while the absolute second term is bounded by $\delta_{\mathcal{D}}$ and $\delta_{\mathcal{U}}$ respectively. If $\gamma_{\mathcal{D}} > \delta_{\mathcal{D}}$, Condition B.8 holds with

$$\kappa_{\mathcal{D}} = \gamma_{\mathcal{D}} - \delta_{\mathcal{D}} > 0, \qquad \kappa_{\mathcal{U}} = \gamma_{\mathcal{U}} + \delta_{\mathcal{U}}. \tag{24}$$

Condition B.8 (Decision-benefit envelope)**.** *The sets $\mathcal{D}$ and $\mathcal{U}$ are nonempty. For every certified nonnegative response a, the path $t \mapsto r(ta)$ is absolutely continuous and, for almost every $t \in [0,1]$,*

$$\frac{d}{dt} r(ta) \leq -\kappa_{\mathcal{D}} E_{\mathcal{D}}(a) + \kappa_{\mathcal{U}} E_{\mathcal{U}}(a), \tag{25}$$

for constants $\kappa_{\mathcal{D}} > 0$ and $\kappa_{\mathcal{U}} \geq 0$ that hold uniformly over $\mathfrak{U}$.

This condition is testable in a structural simulator and falsifiable with intervention or policy-variation data. If only interval bounds are credible, their worst endpoints define $\kappa_{\mathcal{D}}$ and $\kappa_{\mathcal{U}}$. If $\kappa_{\mathcal{D}} \leq 0$, the method must abstain from calling the designated actions beneficial.

Theorem B.9 (Decision-improvement threshold)**.** *Under Condition B.8, every nonzero β-desirable response satisfies*

$$r(a^*) - r(0) \leq -\left[\kappa_{\mathcal{D}}\beta - \kappa_{\mathcal{U}}(1-\beta)\right] E(a^*). \tag{26}$$

Consequently, strategic adaptation strictly reduces regret whenever

$$\beta > \beta_{\mathrm{crit}} := \frac{\kappa_{\mathcal{U}}}{\kappa_{\mathcal{D}} + \kappa_{\mathcal{U}}}. \tag{27}$$

Combining with Theorem B.4 yields a uniform same-policy response-benefit certificate whenever $\beta_{\mathrm{cert}} > \beta_{\mathrm{crit}}$. The threshold is sharp under Condition B.8 alone: for every $\beta \leq \beta_{\mathrm{crit}}$ and every prescribed positive effort level, a linear regret surface and a response with beneficial share exactly β attain a nonnegative regret change.

The theorem clarifies both the contribution and its boundary. A larger causal return gap changes where effort is allocated; it does not by itself sign the effect on regret. The additional decision-benefit envelope involves the action-dependent true cost $C_\eta(O, a)$ and is exactly what the previous score-only argument omitted. At this stage, "improvement" compares response with zero action under the same candidate rule; comparison with a different status quo requires the offset in Corollary B.11.

Theorem B.10 (Anisotropic survival under response uncertainty)**.** *Let A be a random response supported on the nonnegative action domain, with finite second moment and mean μ. Suppose $a \mapsto r(a)$ is twice continuously differentiable on every segment joining μ to the support of A, and a fixed matrix $H_r \succeq 0$ satisfies $\nabla^2 r(a) \preceq H_r$ throughout those segments. If $\mathrm{Cov}(A) = \Sigma$,*

$$\mathbb{E} r(A) \leq r(\mu) + \frac{1}{2} tr(H_r \Sigma). \tag{28}$$

If μ is β-desirable and satisfies Condition B.8, then

$$\mathbb{E} r(A) - r(0) \leq -m(\beta) E(\mu) + \frac{1}{2} tr(H_r \Sigma), \quad m(\beta) := \kappa_{\mathcal{D}}\beta - \kappa_{\mathcal{U}}(1-\beta). \tag{29}$$

Thus the bound certifies survival whenever $m(\beta)E(\mu) > \frac{1}{2} tr(H_r\Sigma)$. The earlier isotropic result is the corollary $H_r = L_r I$. Equality holds for quadratic regret with Hessian H_r.

Corollary B.11 (Status-quo deployment certificate)**.** *For each environment and deployment unit, define the no-response cross-policy offset*

$$b_{\theta,\eta}(O) := \mathbb{E}[\ell_{\theta_0,\eta}(O, A_{\theta_0,\eta}) \mid O] - \ell_{\theta,\eta}(O, 0). \tag{30}$$

Suppose the candidate response has conditional mean $\mu_{\theta,\eta}(O)$ and covariance $\Sigma_{\theta,\eta}(O)$, and that the hypotheses of Theorem B.10 hold conditionally on O, uniformly over $\mathfrak{U}$, with Hessian upper bound $H_{r,\eta}$ and $\mu_{\theta,\eta}(O)$ β-desirable. Then

$$\mathcal{G}_{\mathfrak{U}}(\theta; \theta_0) \geq \inf_{\eta \in \mathfrak{U}} \mathbb{E}\left[b_{\theta,\eta}(O) + m(\beta) E(\mu_{\theta,\eta}(O)) - \frac{1}{2} tr\{H_{r,\eta} \Sigma_{\theta,\eta}(O)\} \right]. \tag{31}$$

A positive right-hand side certifies improvement over the deployed status quo. The held-out certificate in Theorem B.14 can instead estimate the full paired gain directly, without separately estimating this decomposition.

Proposition B.12 (General polyhedral exchange). *Let $a \in \mathbb{R}_+^{p_M}$, let $\mathcal{S}$ be a fixed compact polytope, and let the hard linear optimizer use a vertex-valued fixed tie-breaker. Along $a(t) = ta$, suppose $c(t) = C_\eta(O, ta)$ and $\widehat{c}(t) = f_\theta(X_\eta(O, ta))$ are continuously differentiable. Assume the candidate and oracle selections are piecewise constant with one-sided limits and finitely many actual switches, and that the candidate has no selection-boundary tie at $t = 0, 1$. Let $\mathcal{T}_\theta \subset (0, 1)$ contain the candidate switch times. Then*

$$r(1) - r(0) = \int_0^1 \{z_\theta(t) - z^*(t)\}^\top \dot{c}(t)\, dt + \sum_{s \in \mathcal{T}_\theta} c(s)^\top \{z_\theta(s^+) - z_\theta(s^-)\}. \tag{32}$$

Oracle switches contribute zero to the value jump. If $q_j(t) \le -\kappa_{\mathcal{D}}$ for $j \in \mathcal{D}$ and $q_k(t) \le \kappa_{\mathcal{U}}$ for $k \in \mathcal{U}$ almost everywhere, where $q_m(t) = \{z_\theta(t) - z^*(t)\}^\top \partial_m C_\eta(O, ta)$, and a is β-desirable, then

$$r(1) - r(0) \le -m(\beta)E(a) + J_{\mathrm{fan}}, \qquad J_{\mathrm{fan}} := \sum_{s \in \mathcal{T}_\theta} [c(s)^\top \{z_\theta(s^+) - z_\theta(s^-)\}]_+. \tag{33}$$

The signed jump term cannot generally be deleted. Hard top-k_{cap} is recovered by the capped-simplex polytope; finite assignment, matching, and flow problems are covered through their exact convex hulls.

B.5 Finite-Sample and Asymptotic Certificates

Let Θ_0 be a finite set of N_Θ candidate policies fixed before test access; it may be generated using independent training and calibration data. Let $\mathfrak{U} = \{\eta_1, \dots, \eta_{N_\mathfrak{U}}\}$ be a finite environment library also fixed before test access. On n held-out independent deployment units, write $\widehat{G}(\theta) = \min_\eta P_n g_{\theta,\eta}$ and $G(\theta) = \min_\eta P g_{\theta,\eta}$. Also let $v_{\theta,\eta}(O) := h_\beta(A_{\theta,\eta}(O))$, $\widehat{A}(\theta) := \min_\eta P_n v_{\theta,\eta}$, and $A(\theta) := \min_\eta P v_{\theta,\eta}$.

The following result is the certificate used after model selection. Let $\mathcal{F}_{\mathrm{pre}}$ contain all information in the training and calibration samples, including the fitted response model, the selected policy, and the predeclared environment library. The test sample is not included in $\mathcal{F}_{\mathrm{pre}}$. In this result, P denotes conditional expectation given $\mathcal{F}_{\mathrm{pre}}$ and P_n denotes the held-out empirical mean.

Theorem B.13 (Independent-test certificate for the selected policy). *Let $\widehat{\theta}$ and a nonempty finite library $\mathfrak{U}$ be $\mathcal{F}_{\mathrm{pre}}$-measurable. Conditional on $\mathcal{F}_{\mathrm{pre}}$, suppose the held-out units $O_1, \dots, O_n$ are i.i.d. and independent of the data used to construct $\widehat{\theta}$ and $\mathfrak{U}$. Suppose, almost surely, the conditional ranges of $g_{\widehat{\theta},\eta}(O_i)$ and $v_{\widehat{\theta},\eta}(O_i)$ have lengths at most $2B$ and $2B_h$, respectively, for every $\eta \in \mathfrak{U}$. For $\alpha \in (0, 1)$, define*

$$\epsilon_n^{\mathrm{sel}} := B\sqrt{\frac{2\log(2/\alpha)}{n}}, \qquad \delta_n^{\mathrm{sel}} := B_h\sqrt{\frac{2\log(2/\alpha)}{n}}. \tag{34}$$

Then, conditional on $\mathcal{F}_{\mathrm{pre}}$, with probability at least $1 - \alpha$,

$$G(\widehat{\theta}) \ge \widehat{G}(\widehat{\theta}) - \epsilon_n^{\mathrm{sel}}, \qquad A(\widehat{\theta}) \ge \widehat{A}(\widehat{\theta}) - \delta_n^{\mathrm{sel}}. \tag{35}$$

Consequently, deployment only when both lower confidence bounds in (35) *are nonnegative certifies nonnegative worst-environment gain and alignment margin for the selected policy with conditional probability at least $1 - \alpha$. Neither radius contains N_Θ or $N_\mathfrak{U}$.*

To see why no environment union bound is needed, condition on $\mathcal{F}_{\mathrm{pre}}$ and fix a true least-favorable environment $\eta_G^* \in \arg\min_\eta P g_{\widehat{\theta},\eta}$. Since $\widehat{G}(\widehat{\theta}) \le P_n g_{\widehat{\theta},\eta_G^*}$, the one-sided Hoeffding event for this single, fixed environment implies the first inequality in (35). The same argument with $\eta_A^* \in \arg\min_\eta P v_{\widehat{\theta},\eta}$ yields the second. Allocating failure probability $\alpha/2$ to each event gives (34). This reasoning would fail if the test sample were used to select $\widehat{\theta}$ or to modify $\mathfrak{U}$.

For comparison, the next theorem supplies simultaneous validity over every candidate. That stronger reusable guarantee pays the corresponding multiplicity cost.

Theorem B.14 (Simultaneous robust gain certificate). *Suppose the almost-sure ranges of $g_{\theta,\eta}(O)$ and $v_{\theta,\eta}(O)$ have lengths at most $2B$ and $2B_h$ for every $(\theta, \eta) \in \Theta_0 \times \mathfrak{U}$. For $\alpha \in (0, 1)$, set*

$$\epsilon_n := B\sqrt{\frac{2\log(4N_\Theta N_\mathfrak{U}/\alpha)}{n}}, \qquad \delta_n := B_h\sqrt{\frac{2\log(4N_\Theta N_\mathfrak{U}/\alpha)}{n}}. \tag{36}$$

With probability at least $1-\alpha$*, simultaneously for all* $\theta \in \Theta_0$*,*

$$G(\theta) \geq \widehat{G}(\theta) - \epsilon_n. \tag{37}$$

$$A(\theta) \geq \widehat{A}(\theta) - \delta_n. \tag{38}$$

Therefore a policy deployed only when both lower bounds are nonnegative is, with the same probability, no worse than θ_0 *and* β*-aligned in expectation in every* $\eta \in \mathfrak{U}$*. Moreover, if* $\widehat{\theta} \in \arg\max_{\theta\in\Theta_0} \widehat{G}(\theta)$*, then*

$$G(\widehat{\theta}) \geq \max_{\theta\in\Theta_0} G(\theta) - 2\epsilon_n. \tag{39}$$

The certificate is honest only conditional on coverage of the structural library and independence of the held-out units. Enlarging $\mathfrak{U}$ makes the claim safer but potentially vacuous; shrinking it after inspecting test results invalidates the guarantee.

Proposition B.15 (Unbiasedness, variance, and the nonregular limit)**.** *For fixed* (θ, η)*,* $P_n g_{\theta,\eta}$ *is unbiased for* $Pg_{\theta,\eta}$ *and has variance* $\mathrm{Var}(g_{\theta,\eta})/n \leq B^2/n$*. If the* $N_{\mathfrak{U}}$*-vector of gains has finite covariance matrix* Ξ_θ*, then*

$$\sqrt{n}\{\widehat{G}(\theta) - G(\theta)\} \Rightarrow \min_{\eta\in I_\theta} Z_\eta, \qquad Z \sim \mathcal{N}(0, \Xi_\theta), \tag{40}$$

where $I_\theta = \arg\min_\eta Pg_{\theta,\eta}$*. The limit is normal when the least-favorable environment is unique and is generally nonnormal when several environments tie.*

Theorem B.16 (Minimax rate cannot beat $n^{-1/2}$)**.** *Even with one fixed policy and one environment, for the class of all gain distributions supported on an interval of length* $2B$*,*

$$\inf_{\widetilde{G}} \sup_{P} \mathbb{E}_P(\widetilde{G} - Pg)^2 \geq \frac{B^2}{64n}. \tag{41}$$

Thus the $n^{-1/2}$ *width in Theorem B.14 is unavoidable up to logarithmic factors and constants.*

The selected-policy lower confidence bound controls unsafe deployment without an environment multiplicity penalty. Requiring both that unsafe policies are rejected and that uniformly safe policies are accepted is harder: power over all environment coordinates necessarily costs $\log M$.

Theorem B.17 (Minimax testing cost of safe deployment)**.** *Let* $Y_1, \ldots, Y_n$ *be i.i.d. random vectors supported on* $[-B, B]^M$*, where* $B > 0$*,* $M \geq 2$*, and* $\mu_j = \mathbb{E}Y_{ij}$*. For a safety margin* $0 < \Delta \leq B/2$*, consider*

$$\begin{aligned} \mathcal{P}_0 &:= \left\{ P : \min_{1\leq j\leq M} \mu_j(P) \leq 0 \right\}, \\ \mathcal{P}_1(\Delta) &:= \left\{ P : \min_{1\leq j\leq M} \mu_j(P) \geq \Delta \right\}, \\ H_0 &: P \in \mathcal{P}_0, \qquad\qquad H_1 : P \in \mathcal{P}_1(\Delta). \end{aligned} \tag{42}$$

A possibly randomized test $\varphi_n \in \{0, 1\}$ *takes value one when it deploys. Define its worst-case type-I and type-II errors by*

$$\mathcal{E}_{\mathrm{I}}(\varphi_n) := \sup_{P\in\mathcal{P}_0} P(\varphi_n = 1), \qquad \mathcal{E}_{\mathrm{II}}(\varphi_n) := \sup_{P\in\mathcal{P}_1(\Delta)} P(\varphi_n = 0),$$

where each supremum ranges over all distributions on $[-B, B]^M$ *satisfying the indicated mean restriction. Let* $n_\star(B, \Delta, M, \alpha)$ *be the smallest* n *for which some test has both errors at most* α*. There are universal constants* $0 < c < C < \infty$ *such that, for* $\alpha \in (0, 1/4]$*,*

$$c\frac{B^2}{\Delta^2}\left\{\log M + \log\frac{1}{\alpha}\right\} \leq n_\star(B, \Delta, M, \alpha) \leq C\frac{B^2}{\Delta^2}\left\{\log M + \log\frac{1}{\alpha}\right\}. \tag{43}$$

In particular, the minimax sample complexity is $\Theta\big(B^2\Delta^{-2}\{\log M + \log(1/\alpha)\}\big)$*.*

For the upper bound, deploy exactly when $\min_j \overline{Y}_j > \Delta/2$. Under H_0, one coordinate has mean at most zero, so one-sided Hoeffding bounds false deployment by $\exp\{-n\Delta^2/(8B^2)\}$. Under H_1, a union bound over the M coordinates bounds false abstention by $M\exp\{-n\Delta^2/(8B^2)\}$. For the

lower bound, restrict attention to product distributions on $\{-B, B\}^M$ and write $t = \Delta/B$. Take the safe distribution P to have mean Δ in every coordinate, and let Q_j agree with P except that coordinate j has mean zero. The two-point pair (P, Q_1) and the Bretagnolle–Huber inequality give the $t^{-2}\log(1/\alpha)$ term. For $\overline{Q}_n = M^{-1}\sum_{j=1}^{M} Q_j^{\otimes n}$, direct calculation gives

$$\chi^2(\overline{Q}_n \| P^{\otimes n}) = \frac{\{1 + t^2/(1-t^2)\}^n - 1}{M}.$$

If n is smaller than a universal multiple of $t^{-2}\log M$, this divergence, and hence total variation, is bounded away from one; no test can then have both errors at most $1/4$. Combining the two lower bounds proves (43); Appendix E supplies the constants and all testing inequalities.

Appendix E proves every result, extends (37) to response-solver and Monte Carlo error, and records the exact boundary between smooth training and hard-decision evaluation.

C Notation and Assumption Audit

Table 2: Core notation. One statistical observation is a complete deployment unit, which may be a cohort.

Symbol	Meaning
$O \sim P_0$	Pre-deployment unit containing observed context and exogenous state
$\eta \in \mathfrak{U}$	Candidate causal, behavioral, and response-noise environment
$T_\eta(O, a)$	Structural map returning post-action features X_η and true costs C_η
f_θ	Predictor whose output parameterizes the downstream decision
$z_\tau(\widehat{c})$	Regularized or exact feasible decision induced by predicted cost
$A_{\theta,\eta}$	Possibly random strategic response under (θ, η)
$\ell_{\theta,\eta}$	Post-response decision regret evaluated under action-dependent true cost
$\mathcal{D}, \mathcal{U}$	Policy-designated beneficial and undesirable action coordinates
$\mathcal{G}_{\mathfrak{U}}$	Worst-case expected gain relative to predeclared status quo θ_0
$\mathcal{A}_{\beta,\mathfrak{U}}$	Worst-case expected β-alignment margin
$\mathcal{V}_{\mathfrak{U}}$	Curvature-weighted response-covariance certificate
$N, p, q, p_M, k_{\mathrm{cap}}$	Cohort size, feature dimension, cost/decision dimension, action dimension, and quota
E, h_β	Total effort and the beneficial-effort margin at target share β
R_m, ϕ	Downstream-aware causal return and inverse marginal-cost response map
$F, \widehat{F}, \mu, \delta$	True/nominal VI operators, strong-monotonicity modulus, and uniform operator error
$\kappa_{\mathcal{D}}, \kappa_{\mathcal{U}}, H_r$	Beneficial/harmful regret slopes and an anisotropic Hessian upper bound
J_{fan}	Sum of positive regret jumps at candidate normal-fan switches
$N_\Theta, N_{\mathfrak{U}}, n, B, B_h, \alpha$	Candidate count, environment count, test units, gain/alignment ranges, and error level
$K_{\mathrm{in}}, T_{\mathrm{out}}, \mu, L_a, L_\theta$	Inner/outer iterations, strong-concavity, and inner/outer smoothness constants

Proposition C.1 (Existence of a robust design)**.** *Suppose Θ and $\mathcal{A}_\eta$ are compact and $\mathfrak{U}$ is finite. For every η, assume the selected response is measurable in O and continuous in θ, and the pointwise loss, behavioral margin, and conditional covariance trace in* (13) *are continuous in θ almost surely and dominated by integrable envelopes. If the feasible set of* (12) *is nonempty, then its maximum is attained.*

Proof. For fixed η, continuity of the structural, response, predictor, and selected decision maps implies pointwise continuity in θ of the loss, gain, behavioral margin, and covariance certificate. The integrable envelope permits dominated convergence, so their expectations are continuous on Θ. A finite pointwise minimum or maximum of continuous functions is continuous. Hence $\mathcal{G}_{\mathfrak{U}}$, $\mathcal{A}_{\beta,\mathfrak{U}}$, and $\mathcal{V}_{\mathfrak{U}}$ are continuous. The two inverse images defining the constraints are closed; their intersection with compact Θ is therefore compact. It is nonempty by assumption. The continuous objective $\mathcal{G}_{\mathfrak{U}}$ attains its maximum on this feasible set by the extreme-value theorem. □

The proposition is intentionally finite-library. For a compact infinite ambiguity set, joint continuity in (θ, η) and a common integrable envelope yield the same conclusion by Berge's maximum theorem.

Table 3: Every substantive assumption has an empirical interpretation, a diagnostic, and a fallback.

Assumption	Why it is used	Diagnostic or falsification	Fallback if violated
$\eta_{\mathrm{true}} \in \mathfrak{U}$	Gives meaning to a worst-case certificate	Predeclare graph, cost, and response ranges; challenge them with policy variation	Enlarge $\mathfrak{U}$; if the lower bound becomes nonpositive, abstain
Potential response (16)	Converts returns into coordinatewise effort	Test best-response residuals and equilibrium multiplicity	Put alternative equilibria in $\mathfrak{U}$; do not use Theorem B.4
Strongly monotone VI	Makes a coupled equilibrium unique and stable	Lower-bound the symmetric-part Jacobian eigenvalue and audit VI residuals	Certify the worst equilibrium in a declared response library
Separable strictly convex cost	Makes marginal cost invertible	Elicit or fit marginal effort costs; inspect cross-action substitution	Solve the full KKT system and derive problem-specific bounds
Primitive causal-derivative margins	Links labeled actions to true regret	Estimate intervention effects and optimizer sensitivities, with sensitivity intervals	Report alignment only; do not claim decision improvement
Anisotropic Hessian bound	Controls the directional response-noise premium	Audit local spectral bounds over the response region	Use $L_r I$, a Lipschitz bound, or direct worst-case evaluation
i.i.d. deployment units	Enables finite-sample concentration	Time, cluster, and interference diagnostics	Treat a cohort/cluster as one unit or use sequential/cluster-robust inference
Predeclared finite classes	Makes multiplicity correction honest	Freeze policies, environments, and baseline before test access	Use sample splitting and covering-number bounds
Convex smooth decision set	Strong convexity gives a unique smooth branch	Check convexity and KKT regularity of $\mathcal{S}$	Require a unique local branch or use hard-rule evaluation
Finite polyhedral switches	Gives an exact piecewise-smooth exchange decomposition	Record active sets, ranking margins, endpoint ties, and every switch	Bound total variation or use the direct paired gain
Smooth interior response	Enables implicit differentiation	KKT slackness and Hessian eigenvalue checks	Use active-set derivatives or a response oracle
Finite uniform inner gap	Makes the constant C_0 in Theorem G.1 finite	Audit initial potential gaps over the declared library	State local, instance-dependent error only

Table 4: The distinguishing object is a certificate for endogenous downstream decision quality. "Explicit" means the object is part of the mathematical target, not merely a possible application.

Paradigm	Downstream optimizer	Strategic response	Causal state change	Desirable effort	Decision-gain certificate
Classical DFL	explicit	no	no	no	no
Performative prediction	generic	implicit	generic	no	no
Strategic classification	threshold	explicit	optional	optional	no
Causal strategic learning	prediction/rule	explicit	explicit	optional	no
Efthymiou et al. (2025)	binary rule	explicit	feature graph	explicit	no
This work	explicit	explicit	explicit	explicit	explicit

D Extended Literature Comparison

E Proofs of Structural Results

E.1 Historical Equivalence and Deployment Nonidentification

Proof of Theorem B.1. Fix $B > 0$. Let the historical feature be deterministic, $X = 0$, let the action set be $\mathcal{A} = \{0, 1\}$, and let $X(a) = a$. Let the feasible decision set be $\mathcal{S} = \{0, 1\}$ and use the fixed tie-breaking rule

$$z_0(\widehat{c}) = \begin{cases} 1, & \widehat{c} < 0, \\ 0, & \widehat{c} \geq 0. \end{cases}$$

This is an optimizer of $\min_{z \in \{0,1\}} \widehat{c}z$. Let the true cost be $C(a) = B$ for both actions, so the true oracle is $z = 0$ and the regret of decision z is Bz.

Consider predictors $f_0(x) = 1$ and $f_1(x) = 1 - 2x$. They agree at the only historical feature value: $f_0(0) = f_1(0) = 1$. Both historical decisions are therefore zero, and both historical regrets are zero.

The agent's utility is $z_0(f_j(X(a))) - \kappa a$, where $0 < \kappa < 1$. Under f_0, the decision is zero for both actions, so action 0 uniquely maximizes utility. The deployed regret is again zero. Under f_1, action 0 gives utility zero, whereas action 1 produces $f_1(1) = -1$, decision 1, and utility $1 - \kappa > 0$. Thus action 1 is the unique response and the deployed regret equals B. This proves all claims.

If the unique action coordinate is placed in $\mathcal{D}$, the response under f_1 is 1-desirable while increasing regret by B. Therefore no nonnegative decision-gain bound can depend on the effort share alone. □

Proof of Theorem 4.1. The preceding construction proves the main-text statement verbatim: the two predictors agree on the support of P_0, both historical regrets vanish, and their deployment losses are 0 and the arbitrary prescribed value B. □

E.2 Strongly Monotone Variational Inequalities

Proof of Theorem B.2. Continuity on the nonempty compact convex set $\mathcal{A}$ gives existence of both VI solutions. Strong monotonicity gives uniqueness: if x and y solve the same VI, testing the two inequalities with y and x and adding them yields $\langle F(x) - F(y), x - y\rangle \leq 0$, whereas strong monotonicity lower-bounds the left-hand side by $\mu\|x - y\|_2^2$.

Let $d = a^* - \widehat{a}^*$. The two VI inequalities imply $\langle F(a^*), d\rangle \leq 0$ and $\langle \widehat{F}(\widehat{a}^*), d\rangle \geq 0$. Hence

$$\begin{aligned} \mu\|d\|_2^2 &\leq \langle F(a^*) - F(\widehat{a}^*), d\rangle \\ &\leq \langle \widehat{F}(\widehat{a}^*) - F(\widehat{a}^*), d\rangle \leq \delta\|d\|_2. \end{aligned}$$

The first claimed bound follows after division when $d \neq 0$ and is immediate when $d = 0$. Cauchy–Schwarz then gives $h_\beta(a^*) - h_\beta(\widehat{a}^*) = w_\beta^\top d \geq -\|w_\beta\|_2\delta/\mu$. Likewise, $E(a^*) - E(\widehat{a}^*) = \mathbf{1}^\top d \geq -\sqrt{p_M}\delta/\mu$.

For sharpness, take a one-dimensional interval containing 0 and δ/μ, with $F(x) = \mu x$ and $\widehat{F}(x) = \mu x - \delta$. Their solutions are 0 and δ/μ, so the displacement bound is an equality; multiplying the signed displacement by the one-dimensional w_β attains the alignment constant as well. □

Proof of Theorem 4.2. This is Theorem B.2 with the notation used in the main text. The Nash interpretation follows only under the product-set, differentiability, and own-action convexity conditions stated before that theorem. □

E.3 Downstream-Aware Returns

Proof of Proposition B.3. Suppress (θ, η, O) and vectorize X when it is a matrix. Write $X(a) = X_\eta(O, a)$, $\widehat{c}(a) = f_\theta(X(a))$, and $z(a) = z_\tau(\widehat{c}(a))$. The multivariate chain rule gives

$$\frac{\partial V}{\partial a_m} = (\nabla_z u)^\top J_{z_\tau} J_{f_\theta} J_{X,a_m} + (\nabla_{\widehat{c}} u)^\top J_{f_\theta} J_{X,a_m} + (\nabla_X u)^\top J_{X,a_m}.$$

Factoring the first two terms produces (18). If utility has no direct predicted-cost argument, $\nabla_{\hat{c}} u = 0$; if it has no direct feature argument either, only the five-factor decision channel remains. □

Lemma E.1 (Inverse marginal cost)**.** *Under the assumptions following* (16)*, ψ' is a continuous strictly increasing bijection from $[0, \infty)$ onto $[0, \infty)$, and $\phi(r) = (\psi')^{-1}([r]_+)$ is nondecreasing.*

Proof. Strict convexity and continuous differentiability make ψ' continuous and strictly increasing. Its value at zero is zero and its limit is infinite, so the intermediate-value theorem gives range $[0, \infty)$. A continuous strictly monotone bijection has a continuous strictly increasing inverse. Composing it with the nondecreasing positive-part map proves the final claim. □

Lemma E.2 (Coordinatewise response identity)**.** *At every selected global optimum a^* of* (16)*, $a^*_m = \phi(R^\theta_{m,\eta}(a^*; O))$ for each coordinate m.*

Proof. Hold all coordinates except m fixed at their optimal values and consider the one-dimensional feasible slice $t \mapsto \Psi(a^* + (t - a^*_m)e_m)$ on $[0, \infty)$. If $a^*_m > 0$, Fermat's rule gives $R_m(a^*) - \psi'(a^*_m) = 0$, hence $a^*_m = (\psi')^{-1}(R_m(a^*))$ and $R_m(a^*) > 0$. If $a^*_m = 0$, optimality at the boundary gives the right derivative $R_m(a^*) - \psi'(0) \le 0$. Thus $R_m(a^*) \le 0$ and $\phi(R_m(a^*)) = (\psi')^{-1}(0) = 0$. Both cases give the identity. □

Proof of Theorem B.4. Fix an arbitrary (η, O) and write $a^* = a^*_{\theta,\eta}(O)$. For each $j \in \mathcal{D}$, the definitions in (19) and Lemma E.2 give

$$a^*_j = \phi(R^\theta_{j,\eta}(a^*; O)) \ge \phi(\underline{R}_{\mathcal{D}}).$$

For every $k \in \mathcal{U}$, the same argument gives

$$a^*_k = \phi(R^\theta_{k,\eta}(a^*; O)) \le \phi(\overline{R}_{\mathcal{U}}).$$

Consequently,

$$E_{\mathcal{D}}(a^*) \ge |\mathcal{D}|\phi(\underline{R}_{\mathcal{D}}), \qquad E_{\mathcal{U}}(a^*) \le |\mathcal{U}|\phi(\overline{R}_{\mathcal{U}}). \tag{44}$$

If $\overline{R}_{\mathcal{U}} \le 0$, Lemma E.2 yields $a^*_k = 0$ for every $k \in \mathcal{U}$. Positive total effort then implies $E_{\mathcal{D}}(a^*) = E(a^*)$, so the response is 1-desirable.

Suppose now that $\overline{R}_{\mathcal{U}} > 0$. By (44),

$$\frac{E_{\mathcal{D}}(a^*)}{E(a^*)} = \frac{E_{\mathcal{D}}(a^*)}{E_{\mathcal{D}}(a^*) + E_{\mathcal{U}}(a^*)} \ge \frac{|\mathcal{D}|\phi(\underline{R}_{\mathcal{D}})}{|\mathcal{D}|\phi(\underline{R}_{\mathcal{D}}) + |\mathcal{U}|\phi(\overline{R}_{\mathcal{U}})}.$$

This is (20). Since $a \mapsto a/(a + b)$ increases in a and decreases in b for positive a, b, and ϕ is nondecreasing, the asserted comparative statics follow. The choice of (η, O) was arbitrary, so the result is uniform over $\mathfrak{U}$. □

Proof of Corollary B.5. For every desirable coordinate, $R_j \ge \widehat{R}_j - \varepsilon_G$, hence

$$\underline{R}_{\mathcal{D}} \ge \min_{j \in \mathcal{D}} \widehat{R}_j - \varepsilon_G.$$

For every undesirable coordinate, $R_k \le \widehat{R}_k + \varepsilon_G$, hence

$$\overline{R}_{\mathcal{U}} \le \max_{k \in \mathcal{U}} \widehat{R}_k + \varepsilon_G.$$

Subtracting the latter bound from the former gives $\underline{R}_{\mathcal{D}} - \overline{R}_{\mathcal{U}} \ge \widehat{\Gamma} - 2\varepsilon_G$. The constant is attained: with one coordinate in each set, choose $R_{\mathcal{D}} = \widehat{R}_{\mathcal{D}} - \varepsilon_G$ and $R_{\mathcal{U}} = \widehat{R}_{\mathcal{U}} + \varepsilon_G$; both errors equal the budget and the true gap is exactly $\widehat{\Gamma} - 2\varepsilon_G$. If $\widehat{\Gamma} > 2\varepsilon_G$, the premise of Theorem B.4 holds. Moreover, the true lower endpoint is no smaller than $\underline{R}^{\mathrm{cons}}_{\mathcal{D}}$ and the true upper endpoint is no larger than $\overline{R}^{\mathrm{cons}}_{\mathcal{U}}$. Monotonicity of (20) therefore makes the certificate computed from these conservative endpoints a valid lower bound on the true certified effort share. □

Proof of Proposition B.6. Completing the square gives

$$b^\top a - \frac{1}{2}a^\top Q a = -\frac{1}{2}(a - Q^{-1}b)^\top Q(a - Q^{-1}b) + \frac{1}{2}b^\top Q^{-1}b.$$

The additive final term is constant in a. Maximizing the left-hand side over $a \geq 0$ is therefore equivalent to minimizing the Q-norm distance in (21). Strict positive definiteness makes the objective strictly convex, so the minimizer is unique.

If $Q = \mathrm{diag}(q_1, \ldots, q_{p_M})$, the problem separates as $\min_{a_m \geq 0} \frac{1}{2}q_m(a_m - b_m/q_m)^2$, whose solution is $a_m = [b_m/q_m]_+$. Hence $a^* = [Q^{-1}b]_+$. For non-diagonal Q, cross terms couple the coordinates, so Euclidean coordinatewise clipping need not solve the Q-metric projection. □

For completeness, the general response also has the linear-complementarity characterization

$$a^* \geq 0, \qquad Qa^* - b \geq 0, \qquad (a^*)^\top(Qa^* - b) = 0, \tag{45}$$

which follows directly from the KKT conditions and is often more convenient computationally.

E.4 From Effort Direction to Decision Regret

Proof of Proposition B.7. Write $c(a) = C(a)$ and $\widehat{c}(a) = f_\theta(X(a))$. The candidate-cost term is differentiable and the chain rule gives

$$\partial_m\{c(a)^\top z_\tau(\widehat{c}(a))\} = z_\theta(a)^\top \partial_m c(a) + c(a)^\top Dz_\tau \, Df_\theta \, \partial_m X(a).$$

If the linear oracle is unique, Danskin's theorem gives $\partial_m v_{\mathcal{S}}\{c(a)\} = z^*\{c(a)\}^\top \partial_m c(a)$; subtraction proves (22). At a tie, directional Danskin gives $Dv_{\mathcal{S}}\{c(a); \partial_m c(a)\} = \min_{z^* \in Z^*(c(a))}(z^*)^\top \partial_m c(a)$. Subtracting this minimum is the maximum in (23).

For $j \in \mathcal{D}$, the two primitive bounds imply $Dr(a; e_j) \leq -\gamma_{\mathcal{D}} + \delta_{\mathcal{D}} = -\kappa_{\mathcal{D}}$. For $k \in \mathcal{U}$, they imply $Dr(a; e_k) \leq \gamma_{\mathcal{U}} + \delta_{\mathcal{U}} = \kappa_{\mathcal{U}}$. At an oracle tie, directional Danskin in a nonnegative direction d and $\max_z \sum_m d_m q_m(z) \leq \sum_m d_m \max_z q_m(z)$ give the same weighted sum of coordinate bounds. Taking $d = a$ at ta, using absolute continuity, and applying these bounds almost everywhere yields $\frac{d}{dt}r(ta) \leq -\kappa_{\mathcal{D}}E_{\mathcal{D}}(a) + \kappa_{\mathcal{U}}E_{\mathcal{U}}(a)$. Thus Condition B.8 holds; nonnegativity of the four primitive constants gives $\kappa_{\mathcal{U}} \geq 0$, and $\gamma_{\mathcal{D}} > \delta_{\mathcal{D}}$ gives $\kappa_{\mathcal{D}} > 0$. □

Proof of Proposition 4.3. The unique-oracle identity and the primitive-to-envelope implication are the corresponding clauses of Proposition B.7. Its directional statement supplies the advertised fallback at oracle ties. □

Proof of Theorem B.9. Absolute continuity and the almost-everywhere path-derivative bound give

$$r(a^*) - r(0) = \int_0^1 \frac{d}{dt} r(ta^*)\, dt.$$

Condition B.8 therefore implies

$$\begin{aligned} r(a^*) - r(0) &\leq \int_0^1 [-\kappa_{\mathcal{D}}E_{\mathcal{D}}(a^*) + \kappa_{\mathcal{U}}E_{\mathcal{U}}(a^*)]\, dt \\ &= -\kappa_{\mathcal{D}}E_{\mathcal{D}}(a^*) + \kappa_{\mathcal{U}}E_{\mathcal{U}}(a^*). \end{aligned}$$

If a^* is β-desirable, then $E_{\mathcal{D}}(a^*) \geq \beta E(a^*)$ and $E_{\mathcal{U}}(a^*) = E(a^*) - E_{\mathcal{D}}(a^*) \leq (1-\beta)E(a^*)$. Substitution yields (26). Its bracket is positive exactly when $\beta(\kappa_{\mathcal{D}} + \kappa_{\mathcal{U}}) > \kappa_{\mathcal{U}}$, which is (27). Since total effort is positive, the resulting inequality is strict. The uniform conclusion follows by substituting β_{cert} from Theorem B.4.

To prove sharpness relative to the stated envelope, fix $e_0 > 0$, choose one coordinate in each nonempty set $\mathcal{D}$ and $\mathcal{U}$, and take a response with desirable effort βe_0, undesirable effort $(1-\beta)e_0$, and all other coordinates zero. Restrict the response region to $\{a \geq 0 : E(a) \leq e_0\}$ and define $r(a) = \kappa_{\mathcal{D}}e_0 - \kappa_{\mathcal{D}}E_{\mathcal{D}}(a) + \kappa_{\mathcal{U}}E_{\mathcal{U}}(a)$. This nonnegative linear surface attains both derivative bounds. Its regret change at the constructed response is $e_0\{\kappa_{\mathcal{U}}(1-\beta) - \kappa_{\mathcal{D}}\beta\} \geq 0$ exactly when $\beta \leq \beta_{\mathrm{crit}}$. It is realizable as a DFL regret on this bounded region: take $\mathcal{S} = \{0, 1\}$, make the predictor select $z = 1$, set the true scalar cost to $c(a) = r(a)$, and note that the true oracle selects $z = 0$. Thus no smaller universal threshold follows from Condition B.8 alone. □

Proof of Theorem 4.4. The inequality and its strict-improvement condition are Theorem B.9. The counterexample in its final paragraph proves that no smaller threshold yields a uniform guarantee over the stated envelope; it does not assert that every below-threshold instance fails. □

Proof of Theorem B.10. Let A have mean μ. Taylor's theorem with integral remainder yields, for every response realization a,

$$r(a) = r(\mu) + \nabla r(\mu)^\top (a-\mu) + \int_0^1 (1-t)(a-\mu)^\top \nabla^2 r(\mu + t(a-\mu))(a-\mu)\, dt.$$

The Hessian upper bound implies

$$r(a) \le r(\mu) + \nabla r(\mu)^\top (a-\mu) + \frac{1}{2}(a-\mu)^\top H_r (a-\mu).$$

Taking expectations eliminates the linear term, and the trace identity $\mathbb{E}[(A-\mu)^\top H_r (A-\mu)] = \mathrm{tr}(H_r \Sigma)$ proves (28). Applying Theorem B.9 to the mean response μ gives $r(\mu) - r(0) \le -m(\beta)E(\mu)$. Adding the covariance premium proves (29). The right-hand side is negative under the displayed survival condition, which proves strict improvement. For $r(a) = r(\mu) + q^\top (a-\mu) + \frac{1}{2}(a-\mu)^\top H_r (a-\mu)$, Taylor's identity is exact, proving the equality claim. □

Proof of Proposition 4.5. Apply Theorem B.10; its second clause follows by combining the exact anisotropic premium with Theorem B.9. Setting $H_r = L_r I$ gives $\mathrm{tr}(H_r \Sigma) = L_r \mathrm{tr}(\Sigma)$. □

Proof of Corollary B.11. Condition on O. Theorem B.10, applied to the candidate response, gives

$$\mathbb{E}[\ell_{\theta,\eta}(O, A_{\theta,\eta}) \mid O] \le \ell_{\theta,\eta}(O,0) - m(\beta)E(\mu_{\theta,\eta}(O)) + \frac{1}{2}\mathrm{tr}\{H_{r,\eta}\Sigma_{\theta,\eta}(O)\}.$$

Subtract this inequality from the conditional expected status-quo loss and use (30). Taking expectations over O yields the displayed lower bound for each η; taking the infimum over $\mathfrak{U}$ proves (31). Positivity of its right-hand side is the definition of positive worst-case paired gain in (10). □

Proof of Proposition B.12. Write $r(t) = c(t)^\top \{z_\theta(t) - z^*(t)\}$. The finite-switch assumption partitions $[0,1]$ into finitely many open intervals on which both selected vertices are constant. The no-boundary-tie condition makes the candidate selection constant in one-sided neighborhoods of both endpoints; an oracle endpoint tie has zero true-cost jump because every tied vertex is optimal for $c(0)$ or $c(1)$. Hence there is no unrecorded endpoint contribution. On each open interval the ordinary chain rule gives

$$r'(t) = \dot{c}(t)^\top \{z_\theta(t) - z^*(t)\} = \sum_m a_m \{z_\theta(t) - z^*(t)\}^\top \partial_m C_\eta(O, ta) = \sum_m a_m q_m(t).$$

At a switch of the true-cost oracle, the left and right selections are both optimal for $c(s)$; hence $c(s)^\top \{z^*(s^+) - z^*(s^-)\} = 0$. At a candidate switch, its contribution to the jump in r is exactly $c(s)^\top \{z_\theta(s^+) - z_\theta(s^-)\}$; this remains true at a simultaneous oracle switch because the oracle contribution is zero. Applying the fundamental theorem of calculus on every constant-selection interval and summing the jumps proves (32).

Under the displayed coordinate bounds, its integral term is at most

$$-\kappa_{\mathcal{D}} E_{\mathcal{D}}(a) + \kappa_{\mathcal{U}} E_{\mathcal{U}}(a) \le -\{\kappa_{\mathcal{D}}\beta - \kappa_{\mathcal{U}}(1-\beta)\}E(a) = -m(\beta)E(a).$$

Finally, the signed jump sum is at most its positive-part sum J_{fan}, which proves (33). To see that the signed term cannot be removed, let $\mathcal{S} = \mathrm{conv}\{e_1, e_2\}$, keep $c(t) = (0,1)$ fixed, and set $\widehat{c}(t) = (t, 1-t)$. The candidate changes from e_1 to e_2 at $t = 1/2$, while the oracle remains e_1. The continuous integral is zero and the regret rises from 0 to 1, entirely through the candidate jump. □

Proof of Proposition 4.6. This is Proposition B.12 with environment and context suppressed. The top-k_{cap}, assignment, matching, and flow cases follow whenever the stated feasible polytope is the exact convex hull of their feasible discrete decisions and the fixed tie-breaker returns a vertex. □

F PROOFS OF STATISTICAL CERTIFICATES

Proof of Theorem B.13. Condition throughout on $\mathcal{F}_{\mathrm{pre}}$. The policy, baseline, and environment library are then fixed, and the test units remain i.i.d. Let $\eta_G^* \in \arg\min_{\eta\in\mathfrak{U}} Pg_{\widehat{\theta},\eta}$. Because the empirical minimum is no larger than any one empirical mean,

$$\widehat{G}(\widehat{\theta}) \le P_n g_{\widehat{\theta},\eta_G^*}.$$

The variable $g_{\widehat{\theta},\eta_G^*}$ has range length $2B$, so one-sided Hoeffding gives

$$\Pr\left\{ P_n g_{\widehat{\theta},\eta_G^*} - Pg_{\widehat{\theta},\eta_G^*} > B\sqrt{\frac{2\log(2/\alpha)}{n}} \,\middle|\, \mathcal{F}_{\mathrm{pre}} \right\} \le \alpha/2.$$

On the complementary event, $G(\widehat{\theta}) = Pg_{\widehat{\theta},\eta_G^*} \ge \widehat{G}(\widehat{\theta}) - \epsilon_n^{\mathrm{sel}}$. Repeating the argument with a fixed $\eta_A^* \in \arg\min_\eta Pv_{\widehat{\theta},\eta}$ and range $2B_h$ proves the alignment bound with failure probability at most $\alpha/2$. A union bound proves (35). Only the two fixed population minimizers enter, so neither candidate nor environment multiplicity appears. □

Proof of Theorem 4.7. The main-text empirical minima and radii are exactly those of Theorem B.13. Applying that theorem after conditioning on training and calibration proves both bounds and the stated deploy-or-abstain interpretation. □

Proof of Theorem B.14. Fix (θ, η). Since the range of $g_{\theta,\eta}$ has length at most $2B$, Hoeffding's inequality gives

$$\Pr(|(P_n - P)g_{\theta,\eta}| > t) \le 2\exp\left(-\frac{nt^2}{2B^2}\right).$$

The same calculation for an alignment range of length at most $2B_h$ replaces B by B_h. With $t = \epsilon_n$ for gains and $t = \delta_n$ for alignment margins, the corresponding tail probability is $\alpha/(2N_\Theta N_\mathfrak{U})$ for each quantity. A union bound over the $N_\Theta N_\mathfrak{U}$ gain means and the same number of alignment means shows that, with probability at least $1-\alpha$,

$$\begin{aligned} &|(P_n - P)g_{\theta,\eta}| \le \epsilon_n, \\ &|(P_n - P)v_{\theta,\eta}| \le \delta_n \end{aligned} \quad \text{for every } (\theta,\eta). \tag{46}$$

Work on this event. For every θ and every η, $Pg_{\theta,\eta} \ge P_n g_{\theta,\eta} - \epsilon_n$. Taking the minimum over η proves (37). If the lower bound is positive, then every environment-specific expected gain is positive, so deployment is no worse than θ_0 throughout $\mathfrak{U}$.

Likewise, $Pv_{\theta,\eta} \ge P_n v_{\theta,\eta} - \delta_n$. Taking the minimum over η proves (38). A nonnegative alignment lower bound implies $\mathbb{E}h_\beta(A_{\theta,\eta}) \ge 0$ for every environment. This is an expectation-level alignment statement; it does not imply that every individual response is β-desirable.

For the oracle inequality, let $\theta^* \in \arg\max_{\theta\in\Theta_0} G(\theta)$. The two-sided event (46) implies both $G(\theta) \ge \widehat{G}(\theta) - \epsilon_n$ and $\widehat{G}(\theta) \ge G(\theta) - \epsilon_n$. Therefore

$$G(\widehat{\theta}) \ge \widehat{G}(\widehat{\theta}) - \epsilon_n \ge \widehat{G}(\theta^*) - \epsilon_n \ge G(\theta^*) - 2\epsilon_n,$$

which is (39). □

Proof of Proposition B.15. For fixed (θ, η), linearity of expectation gives $\mathbb{E}P_n g_{\theta,\eta} = Pg_{\theta,\eta}$. Independence gives

$$\mathrm{Var}(P_n g_{\theta,\eta}) = \frac{\mathrm{Var}(g_{\theta,\eta})}{n}.$$

Any variable supported on an interval of length $2B$ has variance at most B^2 by Popoviciu's inequality, proving the variance bound.

Let $\mu_\eta = Pg_{\theta,\eta}$ and $\widehat{\mu}_\eta = P_n g_{\theta,\eta}$. The multivariate central limit theorem yields

$$\sqrt{n}(\widehat{\mu} - \mu) \Rightarrow Z, \qquad Z \sim \mathcal{N}(0, \Xi_\theta).$$

Let $G = \min_\eta \mu_\eta$ and $I_\theta = \{\eta : \mu_\eta = G\}$. Because the library is finite, every inactive coordinate has a fixed positive gap from G. With probability tending to one, no inactive coordinate minimizes $\widehat{\mu}$. Hence

$$\sqrt{n}(\widehat{G} - G) = \min_{\eta \in I_\theta} \sqrt{n}(\widehat{\mu}_\eta - \mu_\eta) + o_p(1)$$
$$\Rightarrow \min_{\eta \in I_\theta} Z_\eta$$

by the continuous mapping theorem. If I_θ is a singleton, the minimum is one Gaussian coordinate. If it contains multiple indices, the minimum of correlated Gaussian coordinates is generally not Gaussian. □

Proof of Theorem B.16. It suffices to restrict the supremum to two distributions. Let $n \geq 1$, set $h = (8\sqrt{n})^{-1}$, and let P_- and P_+ place probabilities $1/2 - h$ and $1/2 + h$, respectively, on the value B, with the remaining mass on $-B$. Their means are $\mu_- = -2Bh$ and $\mu_+ = 2Bh$, separated by $\Delta = 4Bh = B/(2\sqrt{n})$.

For Bernoulli probabilities $p = 1/2 - h$ and $q = 1/2 + h$,

$$D_{\mathrm{KL}}(P_- \| P_+) = 2h \log\left(\frac{1/2 + h}{1/2 - h}\right).$$

Because $2h \leq 1/4$ and $\log\{(1+x)/(1-x)\} \leq 2x/(1-x)$ for $x \in [0, 1)$,

$$nD_{\mathrm{KL}}(P_- \| P_+) \leq \frac{32}{3} nh^2 = \frac{1}{6}.$$

Pinsker's inequality therefore gives $\|P_-^n - P_+^n\|_{\mathrm{TV}} \leq \sqrt{1/12} < 1/2$.

For any estimator $\widetilde{G}$, turn it into a test by selecting the closer of μ_- and μ_+. Whenever that test errs, the estimation error is at least $\Delta/2$. Markov's inequality, followed by the usual testing lower bound, yields

$$\max_{s \in \{-,+\}} \mathbb{E}_{P_s}(\widetilde{G} - \mu_s)^2 \geq \frac{\Delta^2}{8}\left(1 - \|P_-^n - P_+^n\|_{\mathrm{TV}}\right) \geq \frac{\Delta^2}{16} = \frac{B^2}{64n}.$$

Taking the infimum over estimators proves the result. □

Proof of Theorem B.17. For the upper bound, let $\varphi_n = 1$ exactly when $\min_j \overline{Y}_j > \Delta/2$. Under $P \in \mathcal{P}_0$, choose a coordinate j_0 with $\mu_{j_0} \leq 0$. One-sided Hoeffding gives

$$P(\varphi_n = 1) \leq P(\overline{Y}_{j_0} - \mu_{j_0} > \Delta/2) \leq \exp\left(-\frac{n\Delta^2}{8B^2}\right).$$

Under $P \in \mathcal{P}_1(\Delta)$, another one-sided bound and a union bound give

$$P(\varphi_n = 0) \leq \sum_{j=1}^{M} P(\overline{Y}_j - \mu_j \leq -\Delta/2) \leq M \exp\left(-\frac{n\Delta^2}{8B^2}\right).$$

Thus $n \geq 8B^2\Delta^{-2}\log(M/\alpha)$ suffices.

For the lower bound, put $t = \Delta/B \leq 1/2$ and restrict the model to product laws on $\{-B, B\}^M$. Let P have independent coordinates with $P(Y_j = B) = (1+t)/2$, so $P \in \mathcal{P}_1(\Delta)$. For each j, let Q_j agree with P except that $Q_j(Y_j = B) = 1/2$; then $Q_j \in \mathcal{P}_0$. First compare $P^{\otimes n}$ with $Q_1^{\otimes n}$. Their one-observation divergence is

$$D_{\mathrm{KL}}(Q_1 \| P) = -\frac{1}{2}\log(1 - t^2) \leq \frac{2}{3}t^2.$$

Bretagnolle–Huber therefore implies, for every test,

$$Q_1^{\otimes n}(\varphi_n = 1) + P^{\otimes n}(\varphi_n = 0) \geq \frac{1}{2}\exp(-2nt^2/3).$$

If both errors are at most α, then $n \geq \frac{3}{2}t^{-2}\log\{1/(4\alpha)\}$.

It remains to recover the environment cost. Let $\overline{Q}_n = M^{-1}\sum_{j=1}^{M} Q_j^{\otimes n}$. Independence across coordinates makes cross likelihood-ratio moments equal to one, while a diagonal moment equals $\{1 + t^2/(1-t^2)\}^n$. Consequently,

$$\chi^2(\overline{Q}_n \| P^{\otimes n}) = \frac{\{1 + t^2/(1-t^2)\}^n - 1}{M}.$$

Since $t^2/(1-t^2) \le 4t^2/3$, this divergence is at most $1/4$ whenever $n \le \frac{3}{4t^2}\log(1 + M/4)$. Then $\|\overline{Q}_n - P^{\otimes n}\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{\chi^2} \le 1/4$. Any test with both worst-case errors at most $1/4$ would instead have average null error plus safe-alternative error at most $1/2$, which requires total variation at least $1/2$, a contradiction. Hence $n \ge \frac{3}{4t^2}\log(1 + M/4)$.

For $M \ge 2$ and $\alpha \le 1/4$, these two lower bounds jointly dominate $ct^{-2}\{\log M + \log(1/\alpha)\}$ for a universal $c > 0$. Restoring $t = \Delta/B$ and combining with the upper bound proves (43). □

Corollary F.1 (Continuous ambiguity sets)**.** *Suppose $\mathfrak{U}$ is totally bounded under metric d, and $|Pg_{\theta,\eta} - Pg_{\theta,\eta'}| \le L_\eta d(\eta, \eta')$. For a fixed δ-net of size $N(\delta)$ and N_Θ policies, with probability at least $1 - \alpha$,*

$$\mathcal{G}_{\mathfrak{U}}(\theta; \theta_0) \ge \min_{\eta \text{ in the net}} P_n g_{\theta,\eta} - B\sqrt{\frac{2\log(2N_\Theta N(\delta)/\alpha)}{n}} - L_\eta \delta$$

simultaneously for all candidate policies.

Proof. Apply Theorem B.14 to the net. For any $\eta \in \mathfrak{U}$, choose a net point η' with $d(\eta, \eta') \le \delta$. Lipschitz continuity gives $Pg_{\theta,\eta} \ge Pg_{\theta,\eta'} - L_\eta\delta$. Taking the infimum over η and then using the finite-net certificate proves the claim. □

Proposition F.2 (Numerical and Monte Carlo error budget)**.** *Suppose the computed gain $\widetilde{g}_{\theta,\eta}$ satisfies $|\widetilde{g}_{\theta,\eta}(O) - g_{\theta,\eta}(O)| \le \delta_{\mathrm{num}}$ uniformly. Suppose also that a Monte Carlo approximation $\widetilde{P}_n\widetilde{g}$ satisfies $|\widetilde{P}_n\widetilde{g} - P_n\widetilde{g}| \le \epsilon_{\mathrm{MC}}$ simultaneously with probability at least $1 - \alpha_{\mathrm{MC}}$. Then, with probability at least $1 - \alpha - \alpha_{\mathrm{MC}}$,*

$$G(\theta) \ge \min_{\eta} \widetilde{P}_n\widetilde{g}_{\theta,\eta} - \epsilon_n - \delta_{\mathrm{num}} - \epsilon_{\mathrm{MC}}$$

simultaneously over the candidate set. If each loss is L_a-Lipschitz in response and the candidate and baseline response errors are δ_θ and δ_0, one may take $\delta_{\mathrm{num}} = L_a(\delta_\theta + \delta_0)$.

Proof. The uniform numerical bound gives $P_n g_{\theta,\eta} \ge P_n\widetilde{g}_{\theta,\eta} - \delta_{\mathrm{num}}$. The Monte Carlo event gives $P_n\widetilde{g}_{\theta,\eta} \ge \widetilde{P}_n\widetilde{g}_{\theta,\eta} - \epsilon_{\mathrm{MC}}$. Intersect these events with (46), combine the three inequalities, and take the minimum over η. A union bound yields the stated probability. The Lipschitz specialization follows by applying the triangle inequality separately to the candidate and baseline losses. □

G Hypergradient Error for a Fixed Smooth Objective

For an environment with differentiable potential, the inner iteration is

$$a^{(k+1)} = \Pi_{\mathcal{A}_\eta}\left(a^{(k)} + \rho_k \nabla_a \Psi_{\theta,\eta}(a^{(k)}; O)\right). \tag{47}$$

Let $F(\theta, a)$ be a fixed smooth outer objective with multipliers held fixed, set $\mathsf{H} = -\nabla^2_{aa}\Psi \succeq \mu I$ and $\mathsf{K} = \nabla^2_{a\theta}\Psi$, and suppose $\|\mathsf{K}\| \le C_K$. Assume H and K are Lipschitz in a with constants L_H, L_K; $\|\nabla_a F\| \le G_a$; and $\nabla_\theta F$ and $\nabla_a F$ are Lipschitz in a with constants $L_{\theta F}, L_{aF}$. Define

$$C_{\mathrm{imp}} := L_{\theta F} + G_a\left(\frac{L_K}{\mu} + \frac{C_K L_H}{\mu^2}\right) + \frac{C_K L_{aF}}{\mu}. \tag{48}$$

Theorem G.1 (Stationarity for a fixed smooth outer objective)**.** *Assume the inner potential is μ-strongly concave and L_a-smooth, its response is interior, the induced outer objective is bounded below with L_θ-Lipschitz gradient, and $C_0^2 := \sup_{\theta,O,\eta} \frac{2}{\mu}\{\Psi(a^*) - \Psi(a^{(0)})\} < \infty$. With inner stepsize $1/L_a$ and outer stepsize $\eta \le 1/L_\theta$, K_{in} inner iterations and T_{out} outer iterations satisfy*

$$\frac{1}{T_{\mathrm{out}}}\sum_{t=0}^{T_{\mathrm{out}}-1} \|\nabla F(\theta^{(t)})\|^2 \le \frac{2\{F(\theta^{(0)}) - F_{\mathrm{inf}}\}}{\eta T_{\mathrm{out}}} + C_{\mathrm{imp}}^2 C_0^2\left(1 - \frac{\mu}{L_a}\right)^{K_{\mathrm{in}}}, \tag{49}$$

where C_0 is a uniform scaled initial inner-error constant. Thus $T_{\mathrm{out}} = O(\varepsilon^{-2})$ and $K_{\mathrm{in}} = O((L_a/\mu)\log(1/\varepsilon))$ suffice for an ε-stationary iterate. This statement neither proves convergence of the stochastic primal–dual method nor covers hard top-k_{cap} decisions.

Lemma G.2 (Implicit response derivative and its perturbation)**.** *Under the assumptions preceding Theorem G.1, the exact interior response is differentiable and*

$$D_\theta a^*(\theta) = \mathsf{H}(\theta, a^*)^{-1}\mathsf{K}(\theta, a^*). \tag{50}$$

If $\widetilde{a}$ lies in the regularity region and $\delta = \|\widetilde{a} - a^*\|$, the hypergradient obtained by evaluating the implicit formula at $\widetilde{a}$ differs from the exact induced gradient by at most $C_{\mathrm{imp}}\delta$.

Proof. The interior first-order condition is $\nabla_a \Psi(\theta, a^*(\theta)) = 0$. Differentiating it gives

$$\nabla^2_{aa}\Psi \, D_\theta a^* + \nabla^2_{a\theta}\Psi = 0.$$

Since $\mathsf{H} = -\nabla^2_{aa}\Psi \succeq \mu I$, the implicit function theorem applies and rearrangement yields (50).

Write $D(a) = \mathsf{H}(a)^{-1}\mathsf{K}(a)$. Uniform strong concavity gives $\|\mathsf{H}(a)^{-1}\| \le 1/\mu$. The resolvent identity yields

$$\begin{aligned}\|D(\widetilde{a}) - D(a^*)\| &\le \|\mathsf{H}(\widetilde{a})^{-1}\|\|\mathsf{K}(\widetilde{a}) - \mathsf{K}(a^*)\| \\ &\quad + \|\mathsf{H}(\widetilde{a})^{-1} - \mathsf{H}(a^*)^{-1}\|\|\mathsf{K}(a^*)\| \\ &\le \left(\frac{L_K}{\mu} + \frac{C_K L_H}{\mu^2}\right)\delta.\end{aligned}$$

The exact hypergradient is $h(a^*) = \nabla_\theta F(\theta, a^*) + D(a^*)^\top \nabla_a F(\theta, a^*)$; the computed one is the same expression at $\widetilde{a}$. Adding and subtracting $D(\widetilde{a})^\top \nabla_a F(\theta, a^*)$, using $\|D(\widetilde{a})\| \le C_K/\mu$, and applying the assumed Lipschitz bounds gives

$$\begin{aligned}\|h(\widetilde{a}) - h(a^*)\| &\le L_{\theta F}\delta + \frac{C_K}{\mu}L_{aF}\delta + G_a\left(\frac{L_K}{\mu} + \frac{C_K L_H}{\mu^2}\right)\delta \\ &= C_{\mathrm{imp}}\delta.\end{aligned}$$

□

Lemma G.3 (Inner geometric error)**.** *Let Ψ be μ-strongly concave and L_a-smooth on a closed convex action set. Projected ascent with stepsize $1/L_a$ satisfies*

$$\|a^{(K_{\mathrm{in}})} - a^*\| \le C_0\left(1 - \frac{\mu}{L_a}\right)^{K_{\mathrm{in}}/2}$$

for a uniform scaled initial-error constant C_0.

Proof. Apply the standard projected-gradient descent inequality to the μ-strongly convex, L_a-smooth function $-\Psi$. With $q = 1 - \mu/L_a$,

$$\Psi(a^*) - \Psi(a^{(K_{\mathrm{in}})}) \le q^{K_{\mathrm{in}}}\{\Psi(a^*) - \Psi(a^{(0)})\}.$$

Strong concavity gives $\Psi(a^*) - \Psi(a^{(K_{\mathrm{in}})}) \ge (\mu/2)\|a^{(K_{\mathrm{in}})} - a^*\|^2$. Thus the claim holds with

$$C_0^2 := \sup_{\theta, O, \eta} \frac{2}{\mu}\{\Psi(a^*) - \Psi(a^{(0)})\},$$

which is finite by the stated uniform initialization assumption. In the unconstrained case, L_a-smoothness further bounds this constant by a scaled initial squared distance. □

Lemma G.4 (Smooth descent with deterministic gradient error)**.** *If F has L_θ-Lipschitz gradient, $\eta \le 1/L_\theta$, and $\theta^+ = \theta - \eta\{\nabla F(\theta) + e\}$, then*

$$F(\theta^+) \le F(\theta) - \frac{\eta}{2}\|\nabla F(\theta)\|^2 + \frac{\eta}{2}\|e\|^2.$$

Proof. The descent lemma gives

$$F(\theta^+) \le F(\theta) - \eta \nabla F(\theta)^\top (\nabla F(\theta) + e) + \frac{L_\theta \eta^2}{2} \|\nabla F(\theta) + e\|^2.$$

Using $L_\theta \eta \le 1$ and the identity $-u^\top(u+e) + \frac{1}{2}\|u+e\|^2 = -\frac{1}{2}\|u\|^2 + \frac{1}{2}\|e\|^2$ proves the claim. □

Proof of Theorem G.1. Let $q = 1 - \mu/L_a$. Lemma G.3 and Lemma G.2 imply that the outer hypergradient error at every iteration is bounded by

$$\|e_t\| \le C_{\mathrm{imp}} C_0 q^{K_{\mathrm{in}}/2}.$$

Apply Lemma G.4 and sum from $t = 0$ to $T_{\mathrm{out}} - 1$:

$$\frac{\eta}{2} \sum_{t=0}^{T_{\mathrm{out}}-1} \|\nabla F(\theta^{(t)})\|^2 \le F(\theta^{(0)}) - F(\theta^{(T_{\mathrm{out}})}) + \frac{\eta T_{\mathrm{out}}}{2} C_{\mathrm{imp}}^2 C_0^2 q^{K_{\mathrm{in}}}.$$

Use $F(\theta^{(T_{\mathrm{out}})}) \ge F_{\mathrm{inf}}$, divide by $\eta T_{\mathrm{out}}/2$, and obtain (49).

To make the first term $O(\varepsilon^2)$, take $T_{\mathrm{out}} = O(\varepsilon^{-2})$. To make the second term $O(\varepsilon^2)$, it is enough that $q^{K_{\mathrm{in}}} = O(\varepsilon^2)$. Since $-\log q \ge \mu/L_a$ for $q \in (0,1)$, it suffices to take $K_{\mathrm{in}} = O((L_a/\mu)\log(1/\varepsilon))$. This yields an iterate with gradient norm at most ε. If responses are exact, $e_t = 0$ and no inner iterations are required. □

H Additional Experimental Evidence

H.1 Five Controlled Audit Blocks

The controlled suite is theorem-directed: every panel has a numerical condition whose violation would indicate an implementation error or a missing assumption. It contains five audit blocks rather than a single aggregate benchmark.

Table 5: Evidence-to-theory map for the regenerated controlled suite. Counts are independent parameter draws, except that the uncertainty audit uses independent Monte Carlo responses and the deployment audit repeats each sample-size experiment.

Audit block	Formal target	Falsification statistic	Regenerated evidence
Response and threshold	Theorems B.4 and B.9	$\beta_{\mathrm{actual}} < \beta_{\mathrm{cert}}$ or $\Delta r \ge 0$ above β_{crit}	5,000 and 4,500 draws; zero observed failures
Primitive geometry	Proposition B.7	Analytic derivative differs from a central difference	1,800 draws; maximum error 8.66×10^{-11}
Operator stability	Theorem B.2	$\|a^* - \widehat{a}^*\|_2 > \delta/\mu$	1,800 draws; maximum ratio 0.993
Smooth and hard robustness	Theorem B.10 and Proposition B.12	Trace premium or path decomposition fails	3×10^5 responses per covariance; decomposition error 2.78×10^{-16}
Safe deployment	Gain-only case of Theorem B.13	Boundary-null policy passes the one-sided LCB	5,000 repetitions per setting; maximum observed rate 0.10%

Audit 1: response share and decision threshold. For each response instance, three beneficial returns are drawn independently from $\mathrm{Unif}[0.12, 1.35]$ and two undesirable returns from $\mathrm{Unif}[-0.28, 1.05]$. Unit quadratic costs give the exact actions $a_j = [r_j]_+$. We compare the resulting beneficial-effort share with the endpoint certificate

$$\beta_{\mathrm{cert}} = \begin{cases} 1, & \overline{r}_{\mathcal{U}} \le 0, \\ \dfrac{3\underline{r}_{\mathcal{D}}}{3\underline{r}_{\mathcal{D}} + 2\overline{r}_{\mathcal{U}}}, & \overline{r}_{\mathcal{U}} > 0. \end{cases}$$

The observed violation rate is zero over 5,000 draws; the median and fifth-percentile slacks are 0.2090 and 0.0118. A separate set of 4,500 draws samples $\kappa_D \in [0.35, 1.8]$, $\kappa_U \in [0.15, 1.4]$, and $\beta \in [0, 1]$. The realized coordinate effects are at least as favorable as the theorem envelope. No draw with $\beta > \kappa_U/(\kappa_D + \kappa_U)$ has nonnegative regret change. These are numerical checks under the stated generator, not distribution-free evidence that the assumptions hold in a target population.

Audit 2: primitive causal–decision geometry. For four-person soft allocation, the candidate chooses $z_j \propto \exp(-\widehat{C}_j/\tau)$ with $\tau \sim \mathrm{Unif}[0.18, 0.55]$, while the oracle chooses the minimum true cost. Independent linear paths perturb both C and $\widehat{C}$. We evaluate the two analytic terms in Proposition B.7 and compare their sum with a central difference at step 2×10^{-6}. The maximum and median absolute errors over 1,800 draws are 8.66×10^{-11} and 1.13×10^{-11}. The oracle minimizer is unique in these continuously sampled instances, so this audit does not test the directional derivative at an exact tie.

Audit 3: strongly monotone equilibrium error. We draw a random orthogonal matrix U and $Q = U\,\mathrm{diag}(\lambda)U^\top$ with $\lambda_j \sim \mathrm{Unif}[0.3, 2.5]$. For linear operators $F(a) = Qa - b$ and $\widehat{F}(a) = Qa - \widehat{b}$, the exact equilibria are $Q^{-1}b$ and $Q^{-1}\widehat{b}$. Perturbations satisfy $\|b - \widehat{b}\|_2 = \delta$ with $\delta \sim \mathrm{Unif}[0.005, 0.55]$, and the strong-monotonicity constant is $\mu = \lambda_{\min}(Q)$. Every one of 1,800 draws obeys the bound δ/μ; the largest and median displacement-to-bound ratios are 0.9929 and 0.6332.

Audit 4: anisotropic uncertainty and polyhedral exchange. For the smooth calculation, the Hessian is $H = \mathrm{diag}(8, 0.35)$ and the two covariances are $\Sigma_c = \mathrm{diag}(0.30, 0.02)$ and $\Sigma_f = \mathrm{diag}(0.02, 0.30)$. Both have trace 0.32, but $\frac{1}{2}\,\mathrm{tr}(H\Sigma_c) = 1.2035$ and $\frac{1}{2}\,\mathrm{tr}(H\Sigma_f) = 0.1325$; Monte Carlo estimates from 300,000 responses per covariance are 1.20324 and 0.13256. The hard-decision audit enumerates all $4!$ assignments along a linear cost path. Three candidate normal-fan switches and two oracle switches yield

$$\Delta r = 0.729464, \qquad \int_0^1 q(t)\,dt = 0.890946, \qquad \sum_s \Delta_s = -0.161482,$$

with absolute decomposition residual 2.78×10^{-16}. This path is a deterministic stress case selected by the recorded master seed.

Audit 5: selected-policy deployment. For $M = 8$ predeclared environments, cohort gains take values in $\{-1, 1\}$. The least-favorable mean is either 0 (boundary null) or 0.32 (safe alternative), with the other means spaced up to 0.28 higher. At $n \in \{40, 70, 120, 200, 350, 600, 1000, 1700\}$, we repeat the complete experiment 5,000 times. Conditional on selection before test access, the one-sided radius is $\sqrt{2\log(1/0.05)/n}$; the simultaneously reported library width $\sqrt{2\log(8/0.05)/n}$ is included only for comparison. The largest observed false-certification rate under the boundary null is 0.001. Under the safe alternative, abstention falls from 0.9602 at $n = 40$ to 0 at $n \ge 600$. Wilson intervals and all setting-level summaries are stored in the accompanying result file. These finite repetitions illustrate power and conservatism; the theorem, rather than a zero count in simulation, supplies the error guarantee.

H.2 UCI Semi-Synthetic Protocol

We use the UCI Default of Credit Card Clients data (Yeh, 2009): 30,000 records, 23 input variables, and the recorded default label. The repository identifies the data license as Creative Commons Attribution 4.0; the persistent identifier is `https://doi.org/10.24432/C55S3H`. The split is

fixed by master seed 20260914: 15,000 records for fitting, 6,000 for calibration and policy selection, and 9,000 held out from both for the evaluation generator. Standardized logistic regression ($C = 0.7$, `lbfgs`, at most 1,500 iterations) produces a base predicted default risk p_i and converges in 26 iterations.

The empirical component ends there. Actions, responses, effort costs, and causal effects are declared by the experiment rather than learned from UCI. Let

$$q_i = 1 + 0.28\,\sigma\left(\frac{\log(1 + [\text{BILL_AMT1}_i]_+) - m}{s}\right), \qquad \theta = (s_D, s_U),$$

where m and s are fitted using only the 15,000 training records and then frozen, and σ is the logistic map. The synthetic response is

$$a_{1i} = \Pi_{[0,1.35]}(s_D/q_i^\eta), \qquad a_{2i} = \Pi_{[0,1.35]}(0.78 s_D/q_i^\eta), \qquad a_{Ui} = \Pi_{[0,1.35]}(s_U/q_i^\eta),$$

with $q_i^\eta = q_i q_\eta$. In environment η, true and predicted post-response costs are

$$C_{i\eta}(\theta) = \Pi_{[0,1]}\{p_i - g_{1\eta}a_{1i} - g_{2\eta}a_{2i} + g_{U\eta}a_{Ui}\},$$
$$\widehat{C}_i(\theta) = \Pi_{[0,1]}\{p_i - 0.11 s_D a_{1i} - 0.07 s_D a_{2i} - 0.12 s_U a_{Ui}\}.$$

For each 20-person cohort, the decision approves the six smallest predicted costs and the reported decision cost is the mean true cost among those six; ties use the cohort's fixed row order. Held-out default labels are not test outcomes: only the covariates and the training-fitted p_i enter the declared synthetic cost. The alignment margin is $h_{0.82} = a_1 + a_2 - 0.82(a_1 + a_2 + a_U)$.

Table 6: The four predeclared semi-synthetic environments. The coefficients are assumptions used to challenge the certificate; they are not estimates of borrower behavior.

Environment	$g_{1\eta}$	$g_{2\eta}$	$g_{U\eta}$	q_η
Nominal	0.16	0.10	0.08	1.00
Weak benefit	0.11	0.07	0.09	1.05
Costly effort	0.14	0.09	0.08	1.25
Harmful gaming	0.14	0.09	0.13	1.00

The no-action Historical DFL rule is the fallback. The other methods select on calibration from the predeclared 12×7 grid $s_D \in \{0.25, 0.35, \ldots, 1.35\}$ and $s_U \in \{0, 0.091\overline{6}, \ldots, 0.55\}$: Responsive DFL minimizes nominal predicted post-response decision cost without reading $C_{i\eta}$; Causal SC-DFL maximizes nominal paired gain; Aligned DFL maximizes least-favorable alignment subject to nonnegative nominal gain; and CS-DFL maximizes least-favorable gain subject to nonnegative least-favorable alignment. Ties choose the smallest Euclidean action scale, then smaller (s_D, s_U). These rules select $(1.25, 0.183\overline{3})$, $(1.15, 0)$, $(1.35, 0)$, and $(1.05, 0)$, respectively. The responsive objective is flat at zero for 12 candidates and the nominal causal gain is tied for nine; the declared tie-break therefore matters and is part of the hard-decision sensitivity boundary. No choice is revised after test access.

For evaluation, we sample with replacement from the held-out 9,000-record pool to form 5,000 independent 20-person cohorts. Reusing a source record does not couple the cohort draws because each draw is independent conditional on the finite empirical generator. Gain and alignment are evaluated on the same paired cohorts in all four environments. The gain range length is 2 and the alignment range length is 1.593; allocating $\alpha/2$ to each certificate gives radii 0.038413 and 0.030596 from Theorem B.13. Each joint gain–alignment guarantee is marginal for its predeclared method, not simultaneous over the five displayed methods, and the test data are never used to choose among them. The figure below shows the resulting two-certificate frontier.

The scope is deliberately narrow: the study combines *real covariates and labels* with *synthetic actions, responses, effort costs, and causal effects*. It shows that the proposed pipeline can fit, select, and certify a policy under an auditable semi-synthetic generator.

H.3 Generalization Across Credit Datasets

We repeat the semi-synthetic evaluation on three additional credit-risk datasets: Give Me Some Credit (Kaggle, 2011), Home Credit Default Risk (Kaggle, 2018), and South German Credit (UCI Machine Learning Repository, 2020).

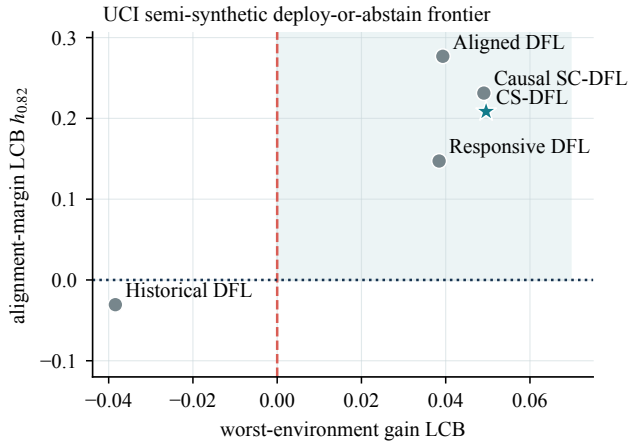

Figure 5: **Semi-synthetic UCI deploy-or-abstain frontier.** Positive coordinates are required on both axes. Each method is fixed using calibration data before the independent cohort stream is drawn; CS-DFL has the largest observed gain LCB, while Aligned DFL has the largest alignment LCB.

Give Me Some Credit is obtained from the Kaggle competition, containing 150,000 records, 10 input variables, and the recorded default label. Home Credit Default Risk is also obtained from Kaggle. We use the `application_train.csv` file, containing 307,511 records and the recorded default label. South German Credit is obtained from the UCI Machine Learning Repository, containing 1,000 records, 20 input variables, and the recorded default label. The dataset is licensed under Creative Commons Attribution 4.0, with persistent identifier `https://doi.org/10.24432/C5QG88`.

We follow the semi-synthetic protocol, with dataset-specific adaptations to the burden proxy used to construct heterogeneous effort costs. Specifically, we replace BILL_AMT1 with DebtRatio for Give Me Some Credit, the ratio of AMT_ANNUITY to AMT_INCOME_TOTAL for Home Credit Default Risk, and rate for South German Credit.

And to reduce Monte Carlo error, we increase the number of evaluation cohorts from 5,000 to 10,000 for each additional dataset.

Table 7: Additional semi-synthetic experiments across three credit-risk datasets.

Dataset	Method	Worst gain	Gain LCB	Alignment LCB	Verdict
Give Me Some Credit	Historical DFL	0.0000	-0.0272	-0.0216	Abstain
	Responsive DFL	0.0283	0.0012	-0.1013	Abstain
	Causal SC-DFL	0.0289	0.0017	0.0557	Deploy
	Aligned DFL	0.0286	0.0014	0.2768	Deploy
	CS-DFL	**0.0290**	**0.0018**	0.0999	**Deploy**
Home Credit Default Risk	Historical DFL	0.0000	-0.0272	-0.0216	Abstain
	Responsive DFL	0.0288	0.0016	0.0413	Deploy
	Causal SC-DFL	0.0288	0.0016	0.0792	Deploy
	Aligned DFL	0.0274	0.0002	0.2808	Deploy
	CS-DFL	**0.0288**	**0.0016**	0.1240	**Deploy**
South German Credit	Historical DFL	0.0000	-0.0272	-0.0216	Abstain
	Responsive DFL	0.0647	0.0376	-0.0281	Abstain
	Causal SC-DFL	0.0774	0.0503	0.2155	Deploy
	Aligned DFL	0.0758	0.0486	0.2832	Deploy
	CS-DFL	**0.0774**	**0.0503**	0.2155	**Deploy**

Table 7 reports the results across the three additional datasets. Across all three datasets, CS-DFL satisfies both deployment criteria, achieving positive lower confidence bounds on gain and alignment. In contrast, Responsive DFL abstains on Give Me Some Credit and South German Credit due to

negative alignment certificates, despite achieving positive gain lower confidence bounds. These experiments assess the pipeline across empirical covariate distributions under the same prespecified response mechanism.

H.4 Artifacts and Seeded Regeneration

After downloading the public UCI file, running

```
python scripts/run_certificate_experiments.py --uci DATAFILE
```

uses master seed 20260914 and 5,000 safety repetitions by default to regenerate the four vector figures and the following machine-readable outputs:

`results/effort_certificate.csv`,
`results/decision_threshold.csv`,
`results/primitive_geometry.csv`,
`results/operator_stability.csv`, `results/polyhedral_path.csv`,
`results/safe_deployment.csv`, `results/uci_semisynthetic.csv`,
`results/uci_calibration_trace.csv`,
`results/uci_test_by_environment.csv`, and
`results/experiment_summary.json`.

The script records trial counts, sample sizes, confidence level, per-method test time, split sizes, selected hyperparameters, environment coefficients, and the master seed. Random draws and reported estimates are seed-reproducible under the pinned software stack; last-bit floating-point residuals, runtimes, and PDF metadata remain machine-dependent. The script computes linear VI solutions analytically, enumerates all 24 assignments in the hard-decision audit, and writes vector PDF figures from the same data used for the reported numbers. `scripts/requirements.txt` pins the tested Python dependencies. The public UCI file is not redistributed; it should be obtained through the DOI above and supplied with the documented command-line option.

The additional semi-synthetic experiments can be reproduced by running

```
python scripts/run_certificate_generalization.py
```

after downloading the three credit-risk datasets. The script uses the same master seed 20260914 to generate the following machine-readable outputs:

`results/give_me_some_credit_semisynthetic.csv`,
`results/give_me_some_credit_calibration_trace.csv`,
`results/give_me_some_credit_test_by_environment.csv`,
`results/give_me_some_credit_metadata.json`,
`results/home_credit_semisynthetic.csv`,
`results/home_credit_calibration_trace.csv`,
`results/home_credit_test_by_environment.csv`,
`results/home_credit_metadata.json`,
`results/south_german_semisynthetic.csv`,
`results/south_german_calibration_trace.csv`,
`results/south_german_test_by_environment.csv`,
`results/south_german_metadata.json`, and
`results/generalization_summary.csv`.

The additional experiments use the same pinned software dependencies as the original UCI experiment. Dataset-specific metadata record the source-file hashes, split sizes, selected policies, burden-proxy normalization statistics, and environment coefficients.

I Reproducibility and Falsification Checklist

The decisive falsification tests are prospective. First, changing the ambiguity library after observing test outcomes invalidates the selected-policy certificate. Second, omission of the true environment invalidates its coverage guarantee and should be stress-tested by moving the generator outside the

Table 8: Reproducibility status of the regenerated evidence. "Boundary" identifies information that the semi-synthetic design cannot supply.

Item	Status	Audit trail or remaining action
Generator and theorem mapping	Available	Equations and falsification statistics are stated above and implemented in the included script
Numerical audit tables	Available	Nine CSV files and one strict JSON summary retain unrounded values
Randomness and repetitions	Available	Master seed 20260914; 5,000 repetitions per deployment setting and 5,000 UCI cohorts
UCI provenance and license	Available	DOI 10.24432/C55S3H; repository metadata states CC BY 4.0
Train/calibration/test separation	Available	Fixed 15,000/6,000/9,000 record split; policy selection uses no evaluation draw
Evaluation dependence	Documented	Cohorts are independent draws with replacement conditional on the untouched empirical pool
Metric and decision rule	Available	Mean true cost among six lowest predicted costs; stable cohort-position tie-break
Environment library	Available	Four environments are fixed before selection; coefficients appear in Table 6
Confidence procedure	Available	One-sided selected-policy LCBs; Wilson intervals summarize Monte Carlo frequencies
Solver diagnostics	Available	Closed-form responses and enumerated assignments; UCI `lbfgs` converged in 26 iterations
Runtime and software lock	Partial	Package versions and per-setting times are recorded; publish the hardware specification
Behavioral and causal validity	Boundary	Requires intervention data or a separate identification design; UCI observation alone cannot establish it

predeclared library. Third, changing the action semantics or replacing the synthetic response by an intervention estimate may reverse the alignment ordering. Fourth, if every defensible library yields a nonpositive gain or alignment LCB, the correct output is abstention, not post hoc deletion of adverse environments. Finally, the cohort stream must be regenerated from a genuinely independent source before any field-deployment claim.

J Machine-Checked Proof Scope

The accompanying `LeanProofs/` project kernel-checks only the algebraic theorem cores summarized in `LeanProofs/README.md` and audited in `LeanProofs/AxiomAudit.lean`; their Lean declarations are defined in `LeanProofs/LeanProofs/StrategicDFL.lean`. Its checker rejects `sorry`, `admit`, custom axioms, and unsafe declarations. Measure-theoretic concentration, analytic regularity, and polyhedral path assumptions remain conventional proofs unless the audit says otherwise. Lean does not verify the Python generator, fitting, Monte Carlo sampling, figures, or dataset and license metadata. Formalization checks implications under encoded hypotheses; numerical audits check the stated generators. Neither establishes causal identification in the UCI application.