GHOSTS WITHOUT TARGET SHIFT: EXACT ADMISSIBILITY FOR STOCHASTIC CANDIDATE REUSE

Anonymous authors
Paper under double-blind review

ABSTRACT

Stochastic language-model evaluation often judges only the response selected from a candidate batch. Unselected responses offer control information, but useful reuse must preserve the selected policy's target, respect score support, and repay its full cost. We introduce Target-Preserving Ghost Qualification (TPGQ): a policy-centered correction is projected into an external support envelope and qualified on independent calibration data. We prove target and support preservation and, with independently supplied conditional slot means, variance optimality in the certified class. Exact boundaries account for range-sensitive workload and componentwise amortized cost. In a controlled clustered instance, projection lowers variance by 2.857% versus feasible Global Coefficient Shrinkage. An independent-split study accepts active projection with a disjoint variance ratio of .6224; a 2,304-cell audit identifies violations missed by simpler rules. Across three generator families and two judges, one exploratory Llama–Gemini setting shows a 4.99% variance reduction versus raw and $.70\%$ versus feasible shrinkage, yet none qualifies for fresh evaluation. Feasible variance gains and qualified cost savings are therefore distinct outcomes.

1 INTRODUCTION

Stochastic language-model evaluation can score several sampled responses cheaply, yet obtains a costly judgment only for the deployed one. The unselected candidates, or ghosts, offer control information. Reusing it is worthwhile only if the estimator preserves the policy target, respects the certified score envelope, and reduces full evaluation cost.

Proposal recycling, leave-one-out baselines, and action-informed controls motivate reuse (Delmas & Jourdain, 2009; Kool et al., 2019; Burch et al., 2018). Reranking or conditioning on favorable judgments, however, changes the policy being evaluated. Even a policy-centered correction may widen the score range enough to erase its variance gain under bounded inference (Howard et al., 2020; 2021; Waudby-Smith & Ramdas, 2024). Methods that reduce label demand (Berrada et al., 2025; Kossen et al., 2021; Maia Polo et al., 2024; Angelopoulos et al., 2023a; Fisch et al., 2024) also incur proxy, calibration, failure, and resource costs. The unresolved question is whether candidate reuse satisfies these requirements jointly.

Target-Preserving Ghost Qualification (TPGQ) projects a policy-centered correction into the certified feasible set before response selection. An independent calibration split then tests alignment, activity, workload, and componentwise cost under a sealed policy, proxy, contract, planning rule, and ledger. Only ACCEPT opens disjoint evaluation. Calibration can be shared across unchanged registered campaigns; we examine $S = 20$ such units.

Our analysis establishes target and support preservation and recovers unprojected PPI-style reuse when the envelope does not bind. With independently supplied conditional slot means, Support Projection minimizes variance in the certified instrument-measurable class. A clustered family shows a strict gain over feasible Global Coefficient Shrinkage because projection retains interior contrasts. Fitted proxies retain target and support preservation, without the oracle optimum.

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.

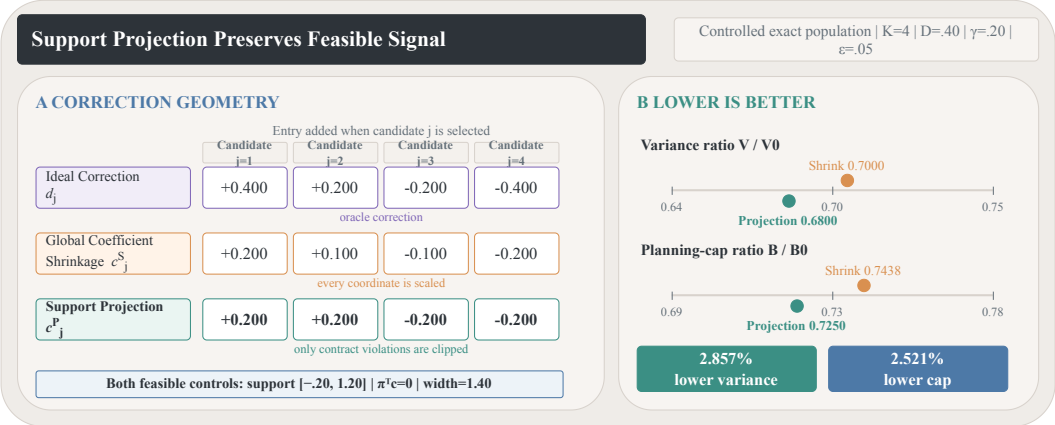


Figure 1: Exact controlled instance ($K = 4$, $D = .40$, $\gamma = .20$, $\epsilon = .05$). (A) Projection preserves feasible interior coordinates; global shrinkage attenuates all coordinates. (B) Under the same width-1.40 contract, projection attains the feasible variance optimum and a lower planning cap.

We derive the exact range-sensitive workload boundary and the amortized componentwise cost condition $\kappa_k(\phi + q) < \tau$. Controlled and independent-split studies establish an admissible region with active projection. Reward-model records reconstruct headroom rejections, and a 2,304-cell audit isolates constraints missed by simpler rules. Diagnostics across three generator families and two judges find a local fitted-proxy variance gain that still fails the workload gate. The evidence thus separates feasible correction from qualified reuse.

2 Problem Setting and Related Work

Each round produces K responses to one input. A cheap proxy scores all candidates; an expensive instrument judges only the response selected by the declared stochastic policy. Ghosts supply scores, not observed outcomes. We use them to estimate that policy's value under an externally supplied corrected-score envelope.

Target and timing. At round t, $\mathcal{G}_t$ contains the sealed candidate batch, policy masses $\pi_{t1:K}$, proxy values $v_{t1:K}$, slot outcome intervals $[L_{tj}, U_{tj}]$, and method parameters. The system then draws $J_t \mid \mathcal{G}_t \sim \pi_t$ and observes only Y_{tJ_t}. Define

$$\mu_{tj} = \mathbb{E}[Y_{tj} \mid \mathcal{G}_t, J_t = j], \qquad \bar{\mu}_t = \sum_{j=1}^{K} \pi_{tj}\mu_{tj}, \qquad \mu = \mathbb{E}[\bar{\mu}_t]. \tag{1}$$

Here $\bar{\mu}_t$ is the selected-policy conditional mean, and μ averages over sealed rounds. The other slot outcomes remain unobserved. Sequential inference around a fixed target additionally requires $\mathbb{E}[\bar{\mu}_t \mid \mathcal{F}_{t-1}] = \mu$; round-wise centering does not imply this condition. Ghost information may choose a correction but cannot change π_t, selection, the instrument, or post-outcome failure handling.

Qualification requirements. Effective support, $n_{\text{eff},t} = 1/\sum_j \pi_{tj}^2$, measures inverse policy concentration. A route contains a sealed batch, policy, selected index, and outcome; a nonzero selected correction makes it nonidentity. Registered route fraction or policy mass defines structural activity. Independent calibration must give a positive lower bound on proxy alignment, the centered proxy's moment with the outcome. Qualification must improve the registered variance-and-width cap and componentwise full cost, including setup, calibration, attempts, failures, and projection. The first failed condition returns a typed REJECT.

2.1 Related Work

Candidate reuse, constrained calibration, and cost-aware evaluation meet at the fixed-policy, external-envelope requirement.

Candidate reuse and control variates. Policy-gradient baselines reduce gradient variance; VIMCO and RLOO use other samples to form learning signals or baselines (Greensmith et al., 2004; Mnih & Rezende, 2016; Kool et al., 2019). AIVAT provides unbiased action-informed evaluation, and AV-AIVAT adds predictable updates and sequential range control (Burch et al., 2018; Li et al., 2026a). Waste recycling reuses rejected proposals but may increase variance (Delmas & Jourdain, 2009).

Prediction and constrained calibration. PPI and StratPPI combine predictions with costly labels; PPI++ tunes their weights, and cross-prediction reuses labels through cross-fitting (Angelopoulos et al., 2023a; Fisch et al., 2024; Angelopoulos et al., 2023b; Zrnic & Candès, 2024b). Active testing and inference guide label acquisition (Kossen et al., 2021; Zrnic & Candès, 2024a). Supplied constraints arise in survey and range-restricted calibration and constrained control variates (Deville & Särndal, 1992; Espuny-Pujol et al., 2018; Maskell et al., 2022). Appendix A derives our unprojected RLOO/PPI forms.

Cost and evaluation targets. Multifidelity Monte Carlo, weak/strong rating allocation, and PEMC price aspects of evaluation cost (Peherstorfer et al., 2016; Angelopoulos et al., 2025; Li et al., 2026b). Our gate also prices shared calibration and each physical resource. Off-policy inference estimates other policies (Karampatziakis et al., 2021; Kallus et al., 2021); here selection follows the declared policy. LLM-judge bias motivates adjusted preference metrics (Zheng et al., 2023; Dubois et al., 2025), while prediction-powered ranking and selective evaluation target preference order or agreement (Chatzi et al., 2024; Jung et al., 2025). Our estimand is a fixed instrument's policy mean.

3 Target-Preserving Estimation

3.1 Policy-matched support projection

Let $\bar{v}_t = \sum_j \pi_{tj} v_{tj}$ and define the ideal correction

$$d_{tj} = b(\bar{v}_t - v_{tj}). \tag{2}$$

Policy centering preserves the target, but the correction may violate the preselection support contract. Let $[L, U]$ contain every declared slot interval. Before proxy selection, the benchmark owner fixes the maximum corrected width $W_{\max} \geq U - L$, giving $\gamma = (W_{\max} - (U - L))/2$ and feasible set

$$\mathcal{C}_{t,\gamma} = \left\{ c : \sum_j \pi_{tj} c_j = 0, \quad L - \gamma - L_{tj} \leq c_j \leq U + \gamma - U_{tj} \right\}. \tag{3}$$

The zero vector is feasible. TPGQ computes

$$c_t^* = \arg\min_{c \in \mathcal{C}_{t,\gamma}} \sum_j \pi_{tj} (c_j - d_{tj})^2, \qquad Z_t = Y_{tJ_t} + c_{tJ_t}^*. \tag{4}$$

Proposition 1 (Target and support). *If $\mathcal{G}_t$ is fixed before J_t, $J_t \mid \mathcal{G}_t \sim \pi_t$, and each declared interval contains its slot outcome and lies in $[L, U]$, then*

$$\mathbb{E}[Z_t \mid \mathcal{G}_t] = \bar{\mu}_t, \qquad Z_t \in [L - \gamma, U + \gamma]. \tag{5}$$

Proof and computation. Conditioning on $\mathcal{G}_t$ gives $\mathbb{E}[Z_t \mid \mathcal{G}_t] = \sum_j \pi_{tj}(\mu_{tj} + c_{tj}^*) = \bar{\mu}_t$ by weighted centering. The coordinate bounds ensure $L - \gamma \leq Y_{tj} + c_{tj}^* \leq U + \gamma$. The KKT solution (Boyd & Vandenberghe, 2004, Sec. 5.5.3) clips $d_{tj} - \lambda$ to those bounds; monotone bisection finds the centering multiplier λ. Zero-mass coordinates use a canonical zero extension.

Proposition 2 (Scope boundary). *If $d_t \in \mathcal{C}_{t,\gamma}$, then $c_t^* = d_t$. Without an external support or certificate contract, choosing γ to contain every d_{tj} makes projection inactive, with no projection-specific statistical benefit.*

Proof. The nonnegative objective in Equation 4 is zero at feasible $c = d_t$. □

Definition 1 (Instrument-measurable certified estimator). *The class $\mathcal{Z}_{t,\gamma}$ consists of estimators $Y_{tJ_t} + c_{tJ_t}$ with $c \in \mathcal{C}_{t,\gamma}$ fixed before selection. Each uses the outcome instrument executed on the deployed response: calibrated conditional means may choose c but cannot replace the observed Y_{tJ_t}.*

Table 1: TPGQ procedure. Sealed quantities and numerical tolerance ξ are fixed before the selected outcome.

Stage	Operation
Seal	Fix the target wrapper, π, proxy, slot bounds, external envelope and γ, splits, cap, thresholds, tolerance, and full-cost map.
Fit	Estimate b on the fit split only; form $d_j = b(\bar{v} - v_j)$.
Project	Let $\ell_j = L - \gamma - L_j$, $u_j = U + \gamma - U_j$. Bisect λ and set $c_j = \mathrm{clip}(d_j - \lambda, \ell_j, u_j)$ until $\lvert\sum_j \pi_j c_j\rvert \le \xi$.
Qualify	Check six predicates in order on the calibration split; return the first typed REJECT.
Evaluate	After ACCEPT, freeze b and the gate; use $Z = Y_J + c_J$ on a fresh disjoint split.

This class requires the selected outcome: a calibrated mean substituted for it would certify the calibration model rather than the deployed response.

Proposition 3 (Contract-feasible variance optimum)**.** *With sealed conditional-mean oracle $d_{tj} = -(\mu_{tj} - \bar{\mu}_t)$, c_t^* in Equation 4 minimizes $\mathrm{Var}(Y_{tJ_t} + c_{tJ_t} \mid \mathcal{G}_t)$ over $\mathcal{Z}_{t,\gamma}$, uniquely on positive-policy-mass coordinates.*

Proof. For feasible c, conditional variance is

$$\mathrm{Var}(Y_{tJ_t} + c_{tJ_t} \mid \mathcal{G}_t) = \sum_j \pi_{tj} \mathrm{Var}(Y_{tj} \mid \mathcal{G}_t, J_t = j) + \sum_j \pi_{tj}(c_j - d_{tj})^2. \tag{6}$$

The first term is fixed; the second is strictly convex on positive-mass coordinates and equals the projection objective. Independent conditional-mean calibration supplies the oracle; fitted proxies need not attain its optimum. □

Corollary 1 (Clustered strict gap)**.** *Let $K \ge 4$ be even, $\pi_j = 1/K$, and d contain one extreme pair $\{+D, -D\}$ and $K - 2$ coordinates equally split between $+a$ and $-a$, where $0 < a < D < 1/2$. For $0 < \gamma < D$ and $Y \mid J = j \sim \mathrm{Bernoulli}(1/2 - d_j)$, symmetry makes d_j the conditional-mean oracle with raw support $[0, 1]$. The controls $c_j^P = \mathrm{clip}(d_j, -\gamma, \gamma)$ and $c_j^S = (\gamma/D)d_j$ are both policy centered, with corrected width $1 + 2\gamma$, and*

$$V(c^S) - V(c^P) = \frac{K-2}{K}\left[a^2\left(1 - \frac{\gamma}{D}\right)^2 - (a - \gamma)_+^2\right] > 0. \tag{7}$$

Thus projection strictly improves every planning functional that is strictly increasing in variance at fixed width, including Equation 12.

The gap follows from the interior residuals; Appendix A.1 gives the calculation.

Bisection takes $O(\log(R_\lambda/\xi))$ iterations with $O(K)$ time and storage each. The theorem requires an exact root: residual $r = \sum_j \pi_j c_j$ causes conditional bias r. Deployment must register $|r| \le \xi$ or find an exact representable root; the controlled construction has $r = 0$. Projection work enters every resource cost.

3.2 RANGE-SENSITIVE PLANNING

Cap formulas and inputs are sealed before evaluated outcomes. One controlled experiment uses $x_n = \log(2n(n+1)/\alpha)$ and the alpha-spent Bernstein radius

$$r(n; V, W) = \frac{W x_n}{3n} + \sqrt{\frac{2V x_n}{n} + \left(\frac{W x_n}{3n}\right)^2},$$
$$N(V, W) = \min\{n : r(n; V, W) \le \epsilon\}. \tag{8}$$

With $N_0 = N(V_0, W_0)$ and $N_c = N(V_c, W_0 + 2\gamma)$, the net-headroom gate at ratio τ is

$$H_N(\tau) = \tau N_0 - N_c > 0. \tag{9}$$

The wider certified range can erase a variance gain.

For a fixed-design i.i.d. gate, fit b independently and compute unbiased sample variances S_0^2 and S_c^2 on m gate observations. With width W and atomic failure probability η, Maurer & Pontil (2009, Theorem 10) give

$$q = \sqrt{\frac{2\log(1/\eta)}{m-1}}, \qquad \underline{V}(S,W) = \left(\sqrt{S^2} - Wq\right)_+^2,$$
$$\overline{V}(S,W) = \min\left\{\frac{W^2}{4}, \left(\sqrt{S^2} + Wq\right)^2\right\}. \tag{10}$$

For policy-centered proxy contrast X and $Y \in [0,1]$, seal the deterministic product width W_{XY} and use

$$L_A = \overline{XY} - W_{XY}\sqrt{\frac{\log(1/\eta)}{2m}} \tag{11}$$

as an alignment lower bound. A union bound allocates η across all frozen predicates and regimes.

Proposition 4 (Finite-sample qualification)**.** *On the joint confidence event, $L_A > 0$, an active upper cap below τ times the raw lower cap, and all full-cost ratios using these conservative caps below τ imply the corresponding population fixed-design inequalities.*

Appendix A.1 gives the joint-confidence argument.

To quantify the signal–range tradeoff, fix $\delta \in (0,1)$ and $x = \log(2/\delta)$, and consider the continuous fixed-design workload

$$B(V,W;x) = \frac{2x\{V + W\epsilon/3\}}{\epsilon^2}. \tag{12}$$

Corollary 2 (Signal pays the range price)**.** *For $V_0 > 0$, define the observed variance-removal fraction $g(c) = 1 - V_c/V_0$. Then $B(V_c, W_c; x) < \tau B(V_0, W_0; x)$ if and only if*

$$g(c) > g_{\min} := 1 - \tau + \frac{(W_c - \tau W_0)\epsilon}{3V_0}. \tag{13}$$

For the optimal unprojected linear control with centered proxy X, $g(c) = \rho^2$, where $\rho = \mathrm{Corr}(Y,X)$.

This boundary is exact for B and interpretive for other registered caps. Projected or fitted controls require $g(c)$; ρ^2 applies only to the optimal unprojected case.

Appendix A.1 gives the algebra.

Equation 8 supports population-planning sensitivity, with n-dependent error allocation across horizons. Prediction-powered e-values address sequential inference (Csillag et al., 2025). Proposition 4 and the projection-active sample use fixed-design Equation 16; martingale extensions require the additional condition in Equation 1.

4 The Qualification Gate

Let P_k^c and P_k^0 denote active and raw setup costs in resource coordinate k, and let a_k and r_k denote per-attempt costs. Componentwise full-cost qualification requires

$$R_k = \frac{P_k^c + a_k N_c}{P_k^0 + r_k N_0} < \tau \quad \text{for every registered } k. \tag{14}$$

The resource coordinates are accelerator reservation, wall time, host CPU, and durable bytes. The constructed ratio vector $(.84, .84, 1.04, .84)$ averages $.89 < .90$ yet fails on CPU.

Proposition 5 (Economic admissibility under sealed reuse)**.** *Let calibration cost M active-attempt equivalents shared across S campaigns with the same frozen policy, model revision, proxy, instrument, population, and support contract. Define $\kappa_k = a_k/r_k$, $q = N_c/N_0$, and $\phi = M/(SN_0)$. With calibration as the only setup cost, coordinate k qualifies exactly when*

$$\kappa_k(\phi + q) < \tau. \tag{15}$$

Equivalently, $\phi < \tau/\kappa_k - q$. With additional setup costs, defining $P_k^c = Q_k^c + a_k M/S$, $p_k^c = Q_k^c/(r_k N_0)$, and $p_k^0 = P_k^0/(r_k N_0)$ gives the exact extension $\{p_k^c + \kappa_k(\phi + q)\}/(1 + p_k^0) < \tau$.

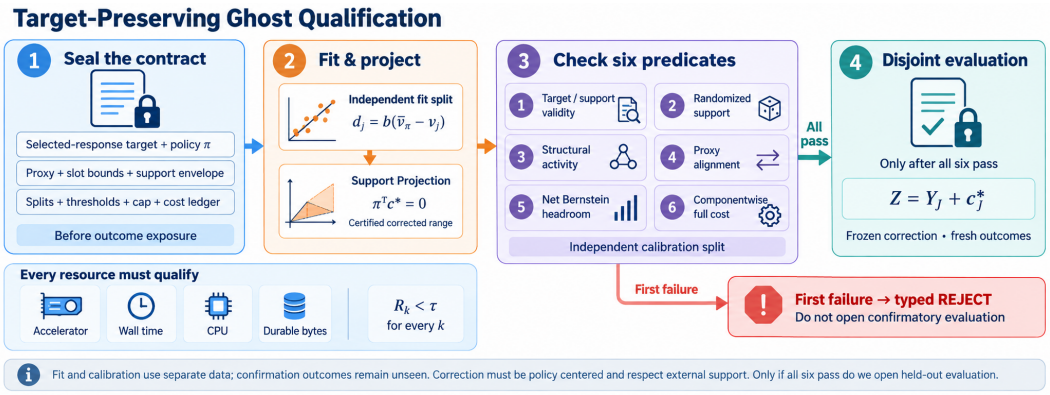


Figure 2: TPGQ workflow. (1) Seal the contract. (2) Fit independently and project. (3) Check six predicates on separate calibration data; the first failure returns a typed REJECT. (4) Only ACCEPT opens disjoint evaluation with unseen confirmation outcomes.

Table 2: Controlled registry: $\delta = .05$, $\tau = .90$, and realized $n_{\text{eff}} = 4.0$ throughout; the finite-sample support floor is 1.5. S counts invariant units sharing calibration. The amortized point prospectively reuses frozen bounds.

Study	Cap formula	ϵ	γ	S	Activity floor
Clustered exact instance	Equation 12	.05	.20	–	n/a
Projection-active sample	Equation 16	.001	.20	1	.80
Amortized operating point	Equation 16	.01	.20	20	.80

Dividing active by raw cost gives this condition; the derivation is in Appendix A.1. Precision enters through the registered caps in q and ϕ.

Stage 3 of Figure 2 checks six predicates with floors fixed before calibration (Table 2). Target / Support Validity verifies the wrapper and intervals. Randomized Support and Structural Activity require $n_{\text{eff}} \geq 1.5$ and nonidentity mass at least .80 in the controlled study. Proxy Alignment requires a positive one-sided lower bound with frozen coefficient. Net Bernstein Headroom checks conservative caps in Equation 9; Componentwise Full Cost checks every ratio in Equation 14. The first failure returns a typed REJECT; only ACCEPT opens disjoint evaluation.

Proposition 6 (No universal positive-cost dominance)**.** *Suppose the admissible class contains a constant-outcome no-signal law with sealed ghosts independent of Y. Couple ghost and target-matched raw evaluation through shared selected outcomes and stopping. Positive expected ghost cost in any registered coordinate rules out simultaneously no larger variance and no larger expected componentwise cost for every law in the class.*

Constant outcomes supply the counterexample (Appendix A.1).

5 ADMISSIBILITY EVIDENCE

The evidence proceeds from feasible optimality and sensitivity to statistical and economic qualification, then tests real-pipeline boundaries and gate coverage. Controlled and independent-split results are distinct from the exploratory model diagnostics.

5.1 RQ1: DOES PROJECTION IMPROVE FEASIBLE CONTROLS?

The clustered instance uses $K = 4$, uniform π, $D = .40$, and $a = \gamma = .20$, with external support $[-.20, 1.20]$. Figure 1 gives the corrections: projection changes two ideal coordinates, while global shrinkage changes all four. Both remain centered at the same width.

Table 3: Exact clustered instance ($K = 4$, uniform π, $D = .40$, $\gamma = .20$, $\epsilon = .05$). $R(c) = \mathbb{E}_\pi[(d-c)^2]$ and B/B_0 uses Equation 12. All rows satisfy width 1.40; global shrinkage is Global Coefficient Shrinkage.

Method	$R(c)$	Variance V	V/V_0	Width	B/B_0
Raw	.100000	.250000	1.0000	1.00	1.0000
Global shrinkage	.025000	.175000	.7000	1.40	.7438
Support Projection	**.020000**	**.170000**	**.6800**	1.40	**.7250**

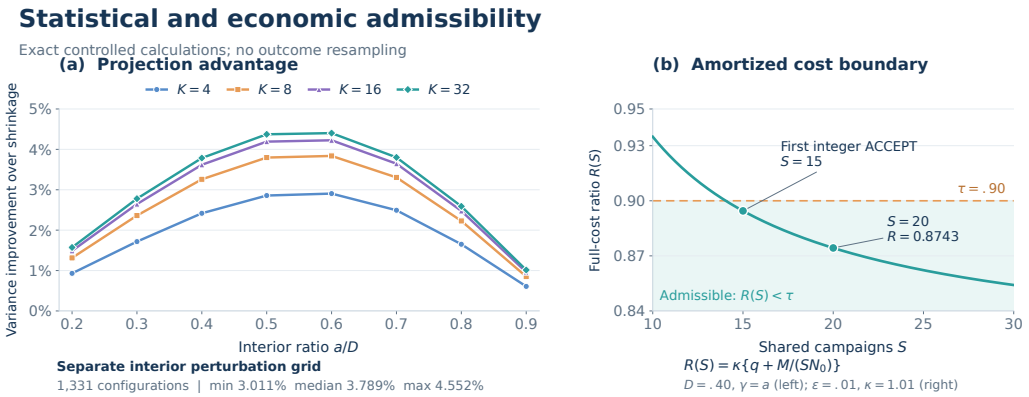


Figure 3: Controlled sensitivity and amortization. (a) Exact variance improvement across a/D and K ($D = .40$, $\gamma = a$); the separate summary gives the minimum, median, and maximum over 1,331 perturbations. (b) Constructed full-cost ratio under shared calibration: $S_{\min} = 15$, with ratio .8743 at $S = 20$ ($\epsilon = .01$).

Projection reduces squared correction error from .025000 to .020000, variance by 2.857%, and the continuous planning cap by 2.521% (Table 3). It adjusts the extreme pair while retaining feasible interior corrections. Proposition 3 extends the comparison to the full certified class under oracle calibration.

5.2 RQ2: Which features sustain the advantage?

All 32 exact cells with $a/D \in \{.2, .3, \ldots, .9\}$ and $K \in \{4, 8, 16, 32\}$ favor projection (Figure 3a). Gains range from .608% to 2.905% for $K = 4$ and 1.014% to 4.403% for $K = 32$. The separate summary covers 1,331 symmetric interior perturbations: exact gains span 3.011%–4.552%, with median 3.789%. For nonuniform $\pi = (.4, .3, .2, .1)$, the nonzero KKT multiplier $\lambda = .0833$ yields a 1.606% gain, with both controls still centered. All 91 multiplicative conditional-mean errors in $[-.40, .50]$ retain a positive advantage, with minimum .820%.

Holding the proxy marginal and ρ^2 fixed leaves the cap ratio at .7644 for $K \in \{2, 4, 8, 16\}$. The gain depends on predictive contrast retained within the envelope, rather than candidate count alone. Appendix A specifies the perturbations and error family.

5.3 RQ3: When do statistical and cost gains coexist?

The fixed-design planning quantity is

$$N_{\text{plan}}(\widehat{V}, W) = \left\lceil \frac{\log(2/\delta)\{2\widehat{V} + 2W\epsilon/3\}}{\epsilon^2} \right\rceil. \tag{16}$$

The projection-active study uses independent fit/gate/evaluation splits of 4,096/16,384/32,768, familywise error .05, $\gamma = .20$, and $\tau = .90$. It clips 50% of coordinates with full structural activity. At $\epsilon = .001$, conservative variance bounds give cap ratio .7979 and maximum componentwise cost ratio .8181. Both disjoint target intervals contain .5; the projected variance ratio is .6224.

Table 4: All four sealed settings return REJECT. Direct checks test registered predicates. Reward-Bench 2 (RB2) rows reconstruct integer-cap decisions from g and Equation 13; they are plug-in diagnostics, not finite-sample certificates.

Setting	Required boundary	Sealed value
Direct branch coverage		
Target unseen	Alignment: $L_A > 0$	-9.56×10^{-6}
Target policy	Activity $\geq 103/128$	$17/128$
Plug-in headroom reconstruction		
RB2 DeBERTa proxy	$g > .3768$	.0479
RB2 Skywork prefix proxy	$g > .2554$	.1228

Table 5: Exploratory pairwise-proxy Gate results ($n = 60$ per combination; through-origin Fit rule). V_0, V_S, V_P are raw, feasible-shrinkage, and projected sample variances; $r_P = N_P/N_0$ and U_{90} are the plug-in cap ratio and its paired-bootstrap 90th percentile. All routes return HOLD. Bold marks a row minimum, not significance.

Generator	Judge	V_0	V_S	V_P	r_P	U_{90}
Qwen	Gemini	.205862	.205862	.205862	1.0381	1.0447
Qwen	Claude	.172599	.172599	.172599	1.0447	1.0579
Llama	Gemini	.232768	.222704	**.221152**	.9864	1.0350
Llama	Claude	.245480	.245480	.245480	1.0323	1.0339
DeepSeek	Gemini	**.156144**	.156538	.156372	1.0510	1.1062
DeepSeek	Claude	.152260	**.151625**	**.151625**	1.0461	1.0682

At $\epsilon = .01$, the same bounds give $N_0 = 17{,}075$, $N_c = 13{,}757$, and $q = .8057$ (Appendix A). For $\kappa = 1.01$, $M = 20{,}480$, and $S = 20$ invariant units, $\phi = .0600$ and every constructed cost ratio is $.8743 < .90$. The first admissible integer is $S_{\min} = 15$ (Figure 3b). Calibration is repaid across sufficiently demanding or invariant repeated evaluations.

5.4 RQ4: WHAT DO REAL-PIPELINE BOUNDARIES EXPLAIN?

The first two rows of Table 4 test alignment and activity. The RB2 rows (Malik et al., 2026) reconstruct integer-cap decisions from sealed variances and certified widths, without post-outcome tuning. Both reward models fail headroom. For Skywork, the minimum threshold over $\gamma \geq 0$ is $.125893 > .122815$, a margin of .00308 that grows with γ.

The range price explains why positive variance reduction can fail workload qualification. Table 4 separates these plug-in diagnostics from certification.

5.5 CROSS-FAMILY PROXY DIAGNOSTICS

We generate four candidates per prompt with Qwen3.8-27B, Llama-3.3-70B-Instruct, and DeepSeek-V3.2, and use Gemini-2.5-Flash and Claude-Haiku-4.5 as separate selected-response judges. Each generator uses the same 60 Fit and 60 Gate RewardBench 2 prompts, spanning Factuality, Focus, and Safety. Two paired proxy changes preserve policy, selection, and target outcome: full-context Skywork replaces its 256-token input, and Qwen predicts each candidate's pairwise win-or-tie score against the reference. Both reuse inspected Gate data and are exploratory rather than independent confirmation. Appendix B reports all combinations and both prespecified Fit rules.

In the pairwise-proxy Llama–Gemini setting, the through-origin Fit rule gives raw, shrinkage, and projected variances .232768, .222704, and .221152. Projection lowers variance by 4.99% versus raw and $.70\%$ versus feasible shrinkage, clipping $13/60$ routes. The centered Fit rule yields a $.29\%$ projection gain over shrinkage. Full-context Skywork gives gains of $.67\%$ and $.54\%$ on Llama–Gemini and Llama–Claude, though both remain worse than raw. These fitted-proxy gains do not invoke the oracle guarantee.

The pairwise Llama–Gemini alignment lower estimate is $.00930 > 0$, but $40/60$ selected corrections are nonzero, below the required $48/60$. Its cap ratio $.9864$ exceeds $\tau = .90$, and its paired-bootstrap 90th-percentile upper value is 1.0350. Positive alignment and lower empirical variance therefore do not qualify the route. All combinations return HOLD; no fresh evaluation runs. Across six combinations and two Fit rules, all 720 Gate certificates satisfy exact policy centering and the $[-.25, .25]$ correction bounds. These are algebraic checks on recorded inputs, not independent statistical samples.

5.6 RQ5: Which constraints do simpler rules miss?

The $4 \times 6 \times 4 \times 4 \times 6 = 2{,}304$ audit varies policy concentration, signal, γ, ϵ, and cost profiles. The full gate admits 75 cells without violations. Appendix A gives the rule counts and visualization (Figure 4; Table 6).

Each of 132 scalar/componentwise disagreements has one resource crossing τ despite a passing mean, with 33 per resource. No simpler rule rejects a full-gate ACCEPT on this grid. Alignment does not certify headroom, and scalar cost misses individual bottlenecks. The counts establish constraint coverage in the declared family, not deployment frequency or universal optimality.

6 Discussion and Scope

An external envelope determines whether projection matters. Without a binding contract, Proposition 2 recovers unprojected PPI-style reuse. With one, projection retains interior contrasts attenuated by Global Coefficient Shrinkage. Its variance optimum assumes independently supplied conditional slot means; fitted proxies preserve the target and support but need not attain that optimum. The certification authority fixes the envelope.

A variance gain must repay the certified range (Equation 13) and full cost in every resource coordinate (Proposition 5). A single bottleneck defeats a favorable average. Calibration sharing applies only across unchanged registered campaigns; typed rejection identifies the failed condition while retaining raw evaluation.

The amortized boundary separates two routes to a lower full cost. A stronger control signal can reduce the active planning ratio q, subject to the fixed range contract. More invariant campaigns reduce the calibration fraction ϕ without changing that contract or the target. The constructed ledger quantifies these thresholds; realized savings still require measured costs for the same resource coordinates.

The studies separate the geometry of the correction from the decision to use it. The clustered example isolates the benefit of projection inside a fixed envelope; the independent-split study tests qualification after fitting and calibration. In the cross-family diagnostic, a fitted proxy improves local variance while activity and planned workload still fail. A favorable variance comparison therefore informs the control signal but does not establish cost-saving qualification.

The guarantees concern fixed-design qualification with independent splits. Controlled studies establish admissibility in specified families, while the 91-point error study covers one calibration family. Cross-family fitted-proxy gains reuse Gate data and are not independent confirmation. Every real-model setting remains unqualified, with no measured accepted four-resource ledger. Hosted API charges do not measure provider-side accelerator and CPU use. Adaptive evaluation, human-judge validity, and empirical superiority over WR, PEMC, or RLOO remain untested.

7 Conclusion

TPGQ connects policy centering, Support Projection, and independent qualification for fixed-policy evaluation. Exact boundaries show when proxy signal and shared calibration can repay the certified range and full cost. Controlled results establish an admissible region; real-model diagnostics show a local projection advantage without qualification. Independent evaluation with a measured resource ledger is needed to establish realized savings. A typed rejection also identifies whether support, signal, workload, or cost prevents reuse under the registered contract.

STATEMENTS

Reproducibility. The main text states the estimator, assumptions, proofs, qualification predicates, and primary experimental settings. Appendix A gives the sensitivity families, operating-point calculations, independent-split protocol, and exact rule counts. Controlled cost profiles and real-pipeline measurements are identified separately.

Use of AI tools. Generative AI was used solely to assist with writing and language polishing in this paper. All research methods, code implementation, theoretical development and proof derivations, literature retrieval and synthesis, and the preparation of figures, tables, and other research outputs were independently completed by the authors.

Scope of evaluation. The method evaluates a declared response-selection policy using a fixed outcome instrument. Target preservation concerns that instrument's mean; it does not establish human-judge validity or justify changing the treatment of failed attempts.

REFERENCES

Anastasios N. Angelopoulos, Stephen Bates, Clara Fannjiang, Michael I. Jordan, and Tijana Zrnic. Prediction-powered inference. *Science*, 382(6671):669–674, 2023a. doi: 10.1126/science.adi6000.

Anastasios N. Angelopoulos, John C. Duchi, and Tijana Zrnic. PPI++: Efficient prediction-powered inference, 2023b. URL https://arxiv.org/abs/2311.01453. Revised March 2024.

Anastasios N. Angelopoulos, Jacob Eisenstein, Jonathan Berant, Alekh Agarwal, and Adam Fisch. Cost-optimal active AI model evaluation, 2025. URL https://arxiv.org/abs/2506.07949.

Gabrielle Berrada, Jannik Kossen, Freddie Bickford Smith, Muhammed Razzak, Yarin Gal, and Tom Rainforth. Scaling up active testing to large language models. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2508.09093.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004. ISBN 978-0-521-83378-3. doi: 10.1017/CBO9780511804441. URL https://web.stanford.edu/~boyd/cvxbook/.

Neil Burch, Martin Schmid, Matej Moravčík, Dustin Morill, and Michael Bowling. AIVAT: A new variance reduction technique for agent evaluation in imperfect information games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11481. URL https://ojs.aaai.org/index.php/AAAI/article/view/11481.

Ivi Chatzi, Eleni Straitouri, Suhas Thejaswi, and Manuel Gomez Rodriguez. Prediction-powered ranking of large language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3594. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/cd47cd67caa87f5b1944e00f6781598f-Abstract-Conference.html.

Daniel Csillag, Claudio José Struchiner, and Guilherme Tegoni Goedert. Prediction-powered E-values. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 11493–11514. PMLR, 2025. URL https://proceedings.mlr.press/v267/csillag25a.html.

Jean-François Delmas and Benjamin Jourdain. Does waste recycling really improve the multi-proposal Metropolis–Hastings algorithm? an analysis based on control variates. *Journal of Applied Probability*, 46(4):938–959, 2009. doi: 10.1239/jap/1261670681.

Jean-Claude Deville and Carl-Erik Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382, 1992. doi: 10.1080/01621459.1992.10475217.

Yann Dubois, Balázs Galambosi, Percy Liang, and Tatsunori B. Hashimoto. Length-controlled AlpacaEval: A simple way to debias automatic evaluators, 2025. URL https://arxiv.org/abs/2404.04475v2. Version 2, revised 10 March 2025; first posted in 2024.

Ferran Espuny-Pujol, Karyn Morrissey, and Paul Williamson. A global optimisation approach to range-restricted survey calibration. *Statistics and Computing*, 28(2):427–439, 2018. doi: 10.1007/s11222-017-9739-5.

Adam Fisch, Joshua Maynez, R. Alex Hofer, Bhuwan Dhingra, Amir Globerson, and William W. Cohen. Stratified prediction-powered inference for effective hybrid evaluation of language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c9fcd02e6445c7dfbad6986abee53d0d-Abstract-Conference.html.

Evan Greensmith, Peter L. Bartlett, and Jonathan Baxter. Variance reduction techniques for gradient estimates in reinforcement learning. *Journal of Machine Learning Research*, 5:1471–1530, November 2004. URL https://jmlr.org/papers/v5/greensmith04a.html.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020. doi: 10.1214/18-PS321.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991.

Jaehun Jung, Faeze Brahman, and Yejin Choi. Trust or escalate: LLM judges with provable guarantees for human agreement. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/08dabd5345b37fffcbe335bd578b15a0-Abstract-Conference.html.

Nathan Kallus, Yuta Saito, and Masatoshi Uehara. Optimal off-policy evaluation from multiple logging policies. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5247–5256, 2021. URL https://proceedings.mlr.press/v139/kallus21a.html.

Nikos Karampatziakis, Paul Mineiro, and Aaditya Ramdas. Off-policy confidence sequences. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5301–5310, 2021. URL https://proceedings.mlr.press/v139/karampatziakis21a.html.

Krzysztof C. Kiwiel. Breakpoint searching algorithms for the continuous quadratic knapsack problem. *Mathematical Programming*, 112(2):473–491, 2008. doi: 10.1007/s10107-006-0050-z. URL https://link.springer.com/article/10.1007/s10107-006-0050-z.

Wouter Kool, Herke van Hoof, and Max Welling. Buy 4 REINFORCE samples, get a baseline for free! In *Deep Reinforcement Learning Meets Structured Prediction Workshop at ICLR*, 2019. URL https://openreview.net/forum?id=r1lgTGL5DE.

Jannik Kossen, Sebastian Farquhar, Yarin Gal, and Tom Rainforth. Active testing: Sample-efficient model evaluation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5753–5763, 2021. URL https://proceedings.mlr.press/v139/kossen21a.html.

Boning Li, Yu Chen, and Longbo Huang. AV-AIVAT: 74x cheaper agent evaluation with certified anytime-valid stopping in imperfect-information games, 2026a. URL https://arxiv.org/abs/2608.06362.

Fengpei Li, Haoxian Chen, Jiahe Lin, Arkin Gupta, Xiaowei Tan, Honglei Zhao, Gang Xu, Yuriy Nevmyvaka, Agostino Capponi, and Henry Lam. Prediction-enhanced Monte Carlo: A machine learning view on control variate, 2026b. URL https://arxiv.org/abs/2412.11257v4. Version 4, revised 4 August 2026; first posted in 2024.

Chris Yuhao Liu, Liang Zeng, Yuzhen Xiao, Jujie He, Jiacai Liu, Chaojie Wang, Rui Yan, Wei Shen, Fuxiang Zhang, Jiacheng Xu, Yang Liu, and Yahui Zhou. Skywork-Reward-V2: Scaling preference data curation via human-AI synergy. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2507.01352v3.

Felipe Maia Polo, Lucas Weber, Leshem Choshen, Yuekai Sun, Gongjun Xu, and Mikhail Yurochkin. tinyBenchmarks: Evaluating LLMs with fewer examples. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34303–34326, 2024. URL https://proceedings.mlr.press/v235/maia-polo24a.html.

Saumya Malik, Valentina Pyatkin, Sander Land, Jacob Morrison, Noah A. Smith, Hannaneh Hajishirzi, and Nathan Lambert. RewardBench 2: Advancing reward model evaluation. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2506.01937v2.

Simon Maskell, Yifan Zhou, and Antonietta Mira. Control variates for constrained variables. *IEEE Signal Processing Letters*, 29:2333–2337, 2022. doi: 10.1109/LSP.2022.3221347.

Andreas Maurer and Massimiliano Pontil. Empirical Bernstein bounds and sample variance penalization. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009. URL https://www.cs.mcgill.ca/~colt2009/papers/012.pdf.

Andriy Mnih and Danilo J. Rezende. Variational inference for Monte Carlo objectives. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2188–2196, New York, New York, USA, June 2016. PMLR. URL https://proceedings.mlr.press/v48/mnihb16.html.

OpenAssistant. OpenAssistant reward-model-deberta-v3-large-v2: Model card. Hugging Face, n.d. URL https://huggingface.co/OpenAssistant/reward-model-deberta-v3-large-v2. Accessed 11 September 2026.

Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016. doi: 10.1137/15M1046472.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2024. doi: 10.1093/jrsssb/qkad009.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/91f18a1287b398d378ef22505bf41832-Abstract-Datasets_and_Benchmarks.html.

Tijana Zrnic and Emmanuel J. Candès. Active statistical inference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 62993–63010. PMLR, 2024a. URL https://proceedings.mlr.press/v235/zrnic24a.html.

Tijana Zrnic and Emmanuel J. Candès. Cross-prediction-powered inference. *Proceedings of the National Academy of Sciences*, 121(15):e2322083121, 2024b. doi: 10.1073/pnas.2322083121. URL https://www.pnas.org/doi/10.1073/pnas.2322083121.

APPENDIX

A SUPPLEMENTARY SETTINGS AND CALCULATIONS

Reproducibility statement. Sections 3–4 state the results and prove target and support preservation and the variance optimum. The remaining short proofs appear below. Section 5 reports the split sizes, error allocation, conservative caps, cost coordinates, and exact sensitivity families used in the controlled evaluation. The figures present exact controlled calculations or summarize the sealed qualification protocol.

A.1 PROOF DETAILS

Clustered strict gap (Corollary 1). Both controls map the extreme pair to $\{+\gamma, -\gamma\}$, so its variance contribution cancels. Interior residual magnitudes are $a(1-\gamma/D)$ under shrinkage and $(a-\gamma)_+$ under projection. The former is larger when $\gamma \geq a$ and, for $\gamma < a$, because $a(1-\gamma/D) > a-\gamma$ when $a < D$. Squaring these residuals and applying Equation 6 yields Equation 7.

Finite-sample qualification (Proposition 4). Condition on the independent fit. The registered union bound jointly covers the alignment lower bound and variance bounds. The cap is monotone in variance and width, giving an upper bound for the active count and a lower bound for the raw count. Nonnegative cost coefficients preserve this ordering, so the strict sample cap and cost inequalities imply the corresponding population fixed-design inequalities on the joint confidence event.

Range price (Corollary 2). Substitute $V_c = V_0\{1-g(c)\}$ into $B(V_c, W_c; x) < \tau B(V_0, W_0; x)$, cancel $2x/\epsilon^2$, and rearrange. For the optimal unprojected linear coefficient, $V_c = V_0(1-\rho^2)$, giving its stated special case.

Economic admissibility (Proposition 5). Dividing $a_k(M/S + N_c)$ by $r_k N_0$ gives $\kappa_k(\phi + q)$. Normalizing the additional setup costs by $r_k N_0$ gives $\{p_k^c + \kappa_k(\phi+q)\}/(1+p_k^0)$.

No universal positive-cost dominance (Proposition 6). Under a constant-outcome no-signal law, raw variance is zero. A nonzero centered correction has positive variance; a zero correction retains raw stopping but adds ghost cost. Either prevents universal weak dominance. The six-predicate gate need not be uniquely optimal.

Unprojected estimator connections. For $\bar{v}_\pi = \sum_i \pi_i v_i$ and $\bar{v}_{-j,\pi} = \sum_{i \neq j} \pi_i v_i/(1-\pi_j)$, the RLOO identity is

$$(1-\pi_j)(\bar{v}_{-j,\pi} - v_j) = \bar{v}_\pi - v_j, \qquad \pi_j < 1.$$

Thus the coefficient-matched peer baseline agrees with the unprojected policy-centered correction. For frozen $f(v) = bv$, the prediction-powered form

$$Y_{tJ_t} + \sum_j \pi_{tj} f(v_{tj}) - f(v_{tJ_t})$$

is the unprojected form in Equation 2, with its coefficient absorbed into f. A nonlinear f instead defines a transformed proxy $f(v)$, to which the same identity applies. These are estimator identities for the selected-response target; they do not transfer a gradient-estimation guarantee or remove the external support and cost requirements.

Exact clustered sensitivity. The main $K = 4$ instance has $d = (.40, .20, -.20, -.40)$, equivalently $X = (-1, -.5, .5, 1)$ with $\Pr(Y = 1 \mid X) = 1/2 + .40X$. Under $[-.20, 1.20]$, projection gives $c^P = (.20, .20, -.20, -.20)$ and global shrinkage gives $c^S = (.20, .10, -.10, -.20)$. These change two and four ideal coordinates, respectively.

For each even K, one correction pair is $\{+D, -D\}$ and the remaining coordinates form matched interior pairs. At $\gamma = a$, Equation 7 reduces to

$$\Delta_K(D, a) = \frac{K-2}{K} a^2 (1 - a/D)^2.$$

The reported curves divide Δ_K by the exact shrinkage variance, which includes the Bernoulli conditional-noise term. The perturbation calculation keeps $K = 8$, $D = .40$, and $\gamma = .20$, enumerates three positive interior magnitudes on the eleven-point grid $[.18, .22]$, and includes their negative counterparts. All 1,331 configurations are evaluated by exact finite sums.

Economic operating point. The frozen variance bounds are $\underline{V}_0 = .2281001609$ and $\overline{V}_c = .1817982201$. Substitution into Equation 16 at $\epsilon = .01$ gives $N_0 = 17{,}075$ and $N_c = 13{,}757$. With $M = 20{,}480$, $S = 20$, and $\kappa = 1.01$, the normalized terms are

$$q = .8056808, \qquad \phi = .0599707, \qquad \kappa(q + \phi) = .8743081.$$

The strict threshold is $\phi < \tau/\kappa - q = .0854083$. Solving the integer inequality gives $S_{\min} = 15$ for these frozen inputs.

Reward-model sources and outcomes. The two reconstructions use frozen subsets of RewardBench 2 (Malik et al., 2026). Both use the Skywork-Reward-V2 Llama-3.1-8B scorer (Liu et al., 2026), with different outcome definitions. The DeBERTa setting uses OpenAssistant's DeBERTa-v3-large-v2 reward model as its proxy (OpenAssistant, n.d.); its target is a clipped linear margin between the selected and reference responses under the full Skywork scorer. The prefix setting uses a 256-token Skywork proxy and a full-response pairwise-win outcome, with ties assigned one-half. These frozen instrument-defined outcomes are distinct from the benchmark's aggregate accuracy and from newly collected human judgments.

Real removal-boundary reconstruction. The sealed DeBERTa route records give $V_0 = .01806424$, $V_c = .01719883$, $W_c = 1.20$, $\epsilon = .05$, and $\tau = .90$. Hence $g(c) = .0479070$ and Equation 13 gives $g_{\min} = .3767900$. The sealed Skywork prefix records give $V_0 = .06436782$, $V_c = .05646246$, and $W_c = 1.50$, which yield $g(c) = .1228154$ and $g_{\min} = .2553571$. The resulting continuous cap ratios are 1.07106 and 1.00528; the recorded integer ratios are $110/103$ and $241/240$. Centering residuals are below 7×10^{-18} in both reconstructions.

For Skywork, minimizing Equation 13 over every $\gamma \geq 0$ sets $W_c = W_0 = 1$ and gives $g_{\min}(0) = .125893$. The sealed removal $.122815$ remains below this value; the gap is $.003078$, or 2.45% of the minimum threshold. Since $g_{\min}$ increases linearly with γ, the recorded headroom rejection does not depend on the registered support budget.

Asymmetric policy and calibration error. For $\pi = (.4, .3, .2, .1)$, $d = (.2, .05, -.3, -.35)$, and $\gamma = .10$, the exact projection multiplier is $1/12$. Projection gives $(.1, -.0333, -.1, -.1)$ and variance $.2233333$; feasible global shrinkage gives $(.0571429, .0142857, -.0857143, -.1)$ and variance $.2269796$. Both weighted centering residuals are zero. In the multiplicative calibration error grid $[-.40, .50]$ with step $.01$, all 91 projected controls have lower variance than feasible Global Coefficient Shrinkage; the relative advantage ranges from $.8198\%$ to 2.2772%.

Decision coverage and vector cost. The exact 60-cell gate sweep records decision counts $(4, 30, 12, 9, 5)$ for ACCEPT, randomized-support, alignment, headroom, and cost. The finite-sample six-regime study records one ACCEPT and one rejection each for randomized support, structural activity, alignment, headroom, and cost. The accepted cost ratios $.89466$, $.89555$, $.89734$, and $.89510$ also pass scalarization and therefore do not establish decision disagreement. The separate scalar counterexample has component ratios $(.84, .84, 1.04, .84)$: their equal-weight average is $.89$, but the componentwise rule rejects the CPU regression.

Alternative-rule factorial audit. The exact audit crosses four policy vectors, six signal levels, four support budgets, four precision levels, and six per-attempt cost-ratio profiles, for 2,304 cells. The six profiles contain balanced and efficient ledgers plus one accelerator, wall, CPU, or durable-byte bottleneck. The calibration fraction is $\phi = .02$ and every component ratio is computed as $\kappa_k(\phi + q)$. Alignment-only, randomized-support-plus-alignment, variance-ratio-only, range-aware-without-cost, and scalarized-cost rules admit 1,845, 1,365, 237, 141, and 132 cells, respectively, that violate another registered constraint. The 132 scalar/componentwise disagreements split equally across the four bottleneck sources. These are exact constraint-coverage counts, not samples from a deployment distribution.

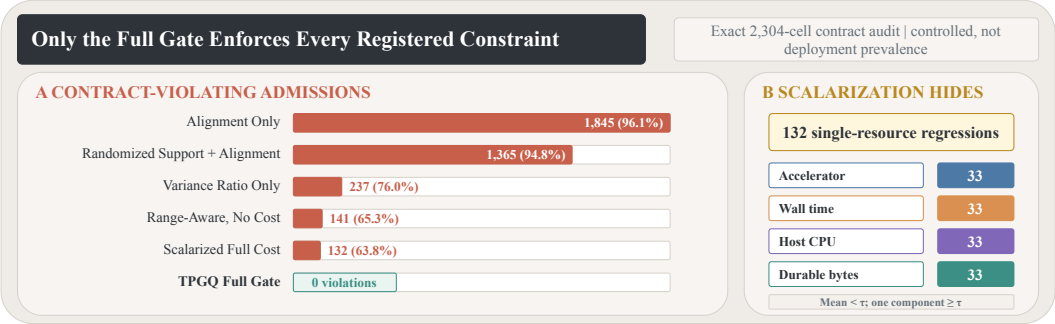


Figure 4: Exact 2,304-cell contract audit. (A) Admissions violating another registered constraint; TPGQ admits 75 cells with no violations. (B) The 132 scalar/componentwise disagreements split equally across accelerator, wall-time, CPU, and durable-byte bottlenecks.

Table 6: Exact rule comparison over 2,304 controlled cells. Violating admissions fail another registered contract constraint; the last column divides that count by all admissions. The full gate is the conjunction being audited, not a learned gold classifier.

Rule	Admissions	Violating admissions	Violation rate
Alignment only	1,920	1,845	96.1%
Randomized support + alignment	1,440	1,365	94.8%
Variance ratio only	312	237	76.0%
Range-aware, no cost	216	141	65.3%
Scalarized full cost	207	132	63.8%
TPGQ full gate	75	0	0

Projection solver. The box-constrained quadratic objective with one equality is a continuous quadratic knapsack problem (Kiwiel, 2008). A valid initial multiplier bracket is $[\min_j(d_j - u_j), \max_j(d_j - \ell_j)]$, where $\ell_j = L - \gamma - L_j$ and $u_j = U + \gamma - U_j$ on positive-mass coordinates. The weighted clipped sum has opposite weak signs at these endpoints and is monotone with Lipschitz constant at most one. Thus the iteration bound depends on this multiplier-bracket width, which includes the ideal correction d, rather than only the coordinate-box width. A stopped residual $r = \sum_j \pi_j c_j$ is the conditional target bias; the declared tolerance must account for it.

Independent evaluation. The projection-active sample uses separate seed namespaces for its 4,096 fit, 16,384 gate, and 32,768 evaluation observations. The gate computes $\hat{b} = .5030883$, clips two of four correction coordinates, and obtains alignment lower bound $.2554148$. On the disjoint split, the raw and projected target intervals both contain .5, and their variance ratio is $.6223664$.

Anonymous artifact. The controlled calculations distinguish supplied support contracts, conditional-mean calibration, and constructed cost ledgers from measured real-pipeline quantities. Fit, gate, and evaluation splits use separate seed namespaces; shared calibration is valid only while the policy, model revision, proxy, instrument, population, and support contract remain unchanged.

B Cross-Family Development Diagnostics

Scope and shared inputs. These diagnostics compare frozen controls on model-generated responses, not on the controlled Bernoulli populations used for the oracle theorem. The original registry contains 360 RewardBench 2 prompts, with 20 Fit, 20 Gate, and 80 conditional-evaluation prompts from each of Factuality, Focus, and Safety. Each generator uses the same 60 Fit and 60 Gate prompts. The 240 conditional-evaluation prompts are not executed because no setting qualifies. Full-context and pairwise proxy development reuse these previously inspected Fit/Gate inputs and selected-response outcomes. Consequently, their comparisons are paired exploratory diagnostics, not fresh hypothesis tests.

Models and fixed target. The generator identifiers are `qwen/qwen3.8-27b`, `meta-llama/llama-3.3-70b-instruct`, and `deepseek/deepseek-v3.2`. The recorded providers are DeepInfra (`deepinfra/bf16`), DeepInfra (`deepinfra/turbo`), and Novita (`novita/fp8`), respectively. Each prompt yields four candidates at temperature .8, top-p = .95, and maximum completion length 768 tokens. The selected-response instruments are `google/gemini-2.5-flash` through `google-vertex/global` and `anthropic/claude-haiku-4.5` through `anthropic`. The target is the selected candidate's win/tie/loss score against the public reference, encoded as 1, .5, or 0, with deterministic display permutations and exact-identity ties. These instruments define separate targets; neither is treated as human ground truth.

The original policy uses the frozen Skywork prefix proxy at temperature .1 and margin scale 37.71875. Both later diagnostics retain that policy, its selected slot J, and the recorded target outcome. They do not rerank candidates using the new proxy. Hosted model identifiers and recorded provider metadata specify the observed run, rather than a guarantee that a mutable endpoint remains identical in the future.

Proxy changes. The full-context ablation replaces the 256-token Skywork input with the complete prompt and response, including reference scoring. Of its 1,800 candidate/reference sequences, 1,241 exceed 256 tokens, and the maximum length is 1,156, below the 8,192-token safety limit. This documents substantial truncation in the prefix input, not an automatic benefit from removing it. Tables 9 and 10 show the resulting paired comparisons.

The pairwise proxy uses `qwen/qwen3.8-27b` through `deepinfra/bf16`. A single request reads the full prompt, public reference, and four candidates, then returns four scores in $[0, 1]$. Each score estimates $P(\text{win}) + \frac{1}{2}P(\text{tie})$ against the reference; the four scores are not required to sum to one. Candidate positions are deterministically permuted and mapped back. Exact candidate/reference identity sets the score to .5. The request uses temperature zero, a four-score JSON schema, a 128-token completion limit, disabled reasoning, and no provider fallback. No new target-judge request is made.

Frozen coefficient rules and feasible comparator. Let $x_t = v_{tJ_t} - \sum_j \pi_{tj} v_{tj}$. Both rules estimate their coefficient from the 60 Fit observations only:

$$\hat{b}_{\mathrm{O}} = \mathrm{clip}_{[0,16]} \frac{\sum_t x_t Y_t}{\sum_t x_t^2}, \qquad \hat{b}_{\mathrm{C}} = \mathrm{clip}_{[0,16]} \frac{\sum_t (x_t - \bar{x})(Y_t - \bar{Y})}{\sum_t (x_t - \bar{x})^2}.$$

A zero denominator gives coefficient zero. The pairwise study specifies both rules before execution and reports both, rather than selecting a rule using Gate outcomes. The full-context study uses the through-origin rule. For $d_{tj} = \hat{b}(\bar{v}_{\pi,t} - v_{tj})$, the matched feasible shrinkage control is $c^S_{tj} = s_t d_{tj}$ with $s_t = \min\{1, \gamma / \max_j |d_{tj}|\}$, taking $c^S = 0$ if $d = 0$. Projection uses the same policy and correction budget. The registered $\gamma = .25$ is a diagnostic budget, not evidence of an externally mandated deployment contract. The theorem's external-contract assumption is separate.

Screening quantities and uncertainty. The diagnostic settings are $\epsilon = .05$, $\delta = .05$, $\tau = .90$, $W_0 = 1$, and $W_S = W_P = 1.5$. Sample variances enter Equation 16; they are not simultaneous conservative variance bounds. The fixed width 1.5 remains in the recorded active cap even when $\hat{b} = 0$. Thus equal raw and corrected variances can have unequal caps; this is a registered range charge, not a measured variance increase.

The randomized-support check requires effective support at least 1.5 on 48 of 60 routes. The activity counter requires a nonzero selected correction on at least 48 routes; it is not an estimate of the probability that an arbitrary slot is active. The remaining checks require $\hat{b} > 0$, a positive alignment lower estimate, and a planning ratio below .90. With $m_t = x_t Y_t$, the alignment estimate is $\bar{m} - 1.2815515655\, s_m/\sqrt{60}$, a normal-approximation diagnostic. The reported U_{90} is the 90th percentile of cap ratios over 10,000 paired route-index bootstrap resamples with the coefficient fixed. It is not multiplicity-adjusted and does not replace Proposition 4. All six settings under both pairwise coefficient rules remain HOLD. Tables 7 and 8 complement the main through-origin table.

Table 7: Pairwise-proxy results with the centered Fit coefficient. V_0, V_S, and V_P are raw, feasible-shrinkage, and projected sample variances; $N_0/N_S/N_P$ lists their integer planning caps. All six settings return HOLD.

Generator–judge	V_0	V_S	V_P	$N_0/N_S/N_P$	U_{90}
Qwen–Gemini	0.205862	0.203321	0.203321	657/674/674	1.0467
Qwen–Claude	0.172599	0.171200	0.171200	559/580/580	1.0784
Llama–Gemini	0.232768	0.226713	0.226046	737/743/741	1.0446
Llama–Claude	0.245480	0.245480	0.245480	774/799/799	1.0339
DeepSeek–Gemini	0.156144	0.155469	0.155469	510/533/533	1.0721
DeepSeek–Claude	0.152260	0.152260	0.152260	499/524/524	1.0696

Table 8: Pairwise-proxy diagnostics for both prespecified Fit rules: O (through-origin) and C (centered). LCB is the normal-approximation alignment lower estimate; $\Delta\bar{Y} = \bar{Z} - \bar{Y}$. Active counts selected nonzero corrections and Clip counts routes on which projection changes an ideal coordinate; each denominator is 60. Neither a small mean difference nor an algebraic certificate is an independent statistical certification.

Generator–judge	Rule	$\hat{b}$	LCB	$\Delta\bar{Y}$	Active	Clip
Qwen–Gemini	O	0.0000	-0.01414	0.00000	0	0
Qwen–Gemini	C	0.1325	-0.01414	0.00199	33	0
Qwen–Claude	O	0.0000	-0.02241	0.00000	0	0
Qwen–Claude	C	0.2263	-0.02241	0.00341	33	0
Llama–Gemini	O	1.1293	0.00930	-0.01289	40	13
Llama–Gemini	C	0.6837	0.00930	-0.01064	40	10
Llama–Claude	O	0.0000	0.01349	0.00000	0	0
Llama–Claude	C	0.0000	0.01349	0.00000	0	0
DeepSeek–Gemini	O	0.8602	-0.02506	0.00437	28	7
DeepSeek–Gemini	C	0.2463	-0.02506	0.00274	28	0
DeepSeek–Claude	O	0.1408	-0.02546	0.00157	28	0
DeepSeek–Claude	C	0.0000	-0.02546	0.00000	0	0

Algebraic checks and interpretation. Exact rational checks on all 720 pairwise Gate correction certificates (six generator–judge combinations, two rules, and 60 routes) verify $\sum_j \pi_j c_j = 0$ and $|c_j| \leq .25$ using the recorded categorical masses. Repeated certificates share routes and outcomes. They establish these constraints on the recorded inputs, not 720 independent trials. Likewise, the observed $\bar{Z} - \bar{Y}$ is a descriptive paired mean difference, not a proof of unbiasedness. Target preservation follows from policy centering under the pre-outcome measurability assumptions in the method, not from a small empirical mean difference on reused data.

Measured resources and provenance. The full-context proxy stage, including cold model loading, records 103.23 seconds of wall time, 113.49 seconds of host CPU time, and 15,547,594,752 bytes of peak allocated CUDA memory, with no API expenditure. Peak memory is not accelerator time or durable storage. The pairwise study records 360 new proxy requests, zero new target requests, zero transport retries, and $0.170436075 of new API charges. Known cumulative API charges are $1.383042397; the conservative cumulative charge including earlier uncertainty and reserves is $1.642686797. These are acquisition costs for the development studies, not matched lifecycle savings. No independent four-resource comparison is measured at the hosted providers.

The sealed local campaigns are identified in the accompanying source package's README. Their result summaries are copied without changing measured values; the protocol settings are documented above. The original cross-family run, full-context ablation, and pairwise development all report every generator–judge combination. Their local projection gains support a mechanism comparison; they do not establish a real accepted deployment, fresh confirmation, or empirical superiority to WR, PEMC, or RLOO.

Table 9: Full-context Skywork ablation on the same development inputs and outcomes. The prefix and full columns give projected integer cap ratios to raw; V_S and V_P refer to the full-context proxy. Raw variances are the same as in Table 5. Every setting returns HOLD.

Generator–judge	V_S	V_P	Prefix ratio	Full ratio	U_{90}
Qwen–Gemini	0.205862	0.205862	1.0381	1.0381	1.0447
Qwen–Claude	0.171993	0.171993	1.0447	1.0411	1.0578
Llama–Gemini	0.242223	0.240598	1.0516	1.0638	1.1503
Llama–Claude	0.252969	0.251593	1.1227	1.0556	1.1158
DeepSeek–Gemini	0.156144	0.156144	1.0490	1.0490	1.0633
DeepSeek–Claude	0.152215	0.152215	1.0501	1.0481	1.0682

Table 10: Full-context Skywork coefficients and diagnostics. Definitions match Table 8; all alignment lower estimates are nonpositive.

Generator–judge	$\hat{b}$	LCB	$\Delta\bar{Y}$	Active	Clip
Qwen–Gemini	0.0000	-0.00598	0.00000	0	0
Qwen–Claude	0.1818	-0.00573	0.00011	60	0
Llama–Gemini	4.4783	-0.00333	0.00716	58	26
Llama–Claude	2.9031	-0.00452	0.00411	58	12
DeepSeek–Gemini	0.0000	-0.00664	0.00000	0	0
DeepSeek–Claude	0.1404	-0.00678	0.00000	60	0