GEOMETRIC IDENTIFICATION IN PREDICT-THEN-OPTIMIZE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Decision-focused surrogates can recover downstream decisions without identifying the quotient report. We characterize the equality set of the convex Smart Predict-then-Optimize surrogate (SPO+) population risk. Under central symmetry, the centered mean class is the unique Bayes minimizer exactly when every nonzero effective displacement makes the old optimizer leave the shifted optimal face with positive probability. This condition separates face crossing from selected-oracle disagreement and gives quantitative local coercivity. Without symmetry, strict crossing alone need not identify the mean; selection balance with reflected crossing restores quotient-report identification, and conditional versions extend the result to measurable predictors. These are population statements, without finite-sample report-recovery or generic transfer-regret guarantees. Closed-form mechanisms reproduce the analytic identities and rates. Portfolio, complete-matrix KuaiRec, and Energy/Storage studies measure predictive fidelity, shifted regret, and fitted-report geometry. A known data-generating process (DGP) companion retains their application geometries while isolating conditional-mean recovery and crossing, without testing the original observational assumptions.

1 INTRODUCTION

Predict-then-optimize learning uses a predictive report as the cost vector of a downstream optimization problem, often trained by decision-focused methods (Donti et al., 2017; Mandi et al., 2022). The Smart Predict-then-Optimize (SPO) loss and its convex SPO+ surrogate (Elmachtoub & Grigas, 2022) are commonly assessed through decision quality, with calibration and generalization analyses addressing when small surrogate risk yields good decisions (Liu & Grigas, 2021; Balghiti et al., 2023). Decision quality alone does not determine the predictive report. If multiple reports induce the same downstream action or remain in the same optimality region, they can have indistinguishable training-task decisions while encoding different cost predictions. Decision consistency is therefore weaker than quotient-report identification: small task regret does not identify the predictive report, even modulo invisible directions.

The distinction becomes operational when a fitted report is reused. A cost vector that is decision-equivalent for one feasible set need not remain equivalent after an objective, constraint, or feasible-set change. The resulting transfer and reuse implication is only a testable hypothesis: similar training-task regret need not imply similar quotient-report fidelity or shifted-task performance. This paper establishes the population geometry; it does not prove a transfer benefit. The real-data studies evaluate a pre-specified, application-specific transfer hypothesis.

A one-dimensional example already shows the nonuniqueness. In Example 4.1, $S = [-1/2, 1/2]$ and a symmetric absolutely continuous cost law supported on $[1, 2]$ yield $\arg\min_{\hat{c}} R_{\mathrm{SPO}+}(\hat{c}) = [1, \infty)$, although every point in this ray induces the Bayes-optimal decision. Thus neither a density nor correct downstream decisions identify the quotient report. Our question is: under which geometric conditions does SPO+ identify the effective component of a predictive cost, namely the quotient report $[c]$, rather than only a downstream-optimal action?

For a deterministic measurable oracle $w^*(c) \in W^*(c)$ and a report displacement Δ, consider the old decision evaluated at the shifted cost,

$$G_\Delta(c) = (c + 2\Delta)^\top w^*(c) - z^*(c + 2\Delta) \geq 0.$$

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.

The gap is positive exactly when the old selected decision leaves the entire shifted optimal face. Prior SPO+ analyses give the centered excess-risk identity and decision consistency or mean minimization under stronger conditions (Liu & Grigas, 2021). We do not claim another SPO+ consistency theorem: Theorem 5.2 identifies optimal-face strict crossing as the exact condition for unique quotient-report identification, characterizing the zero-excess-risk geometry after invisible directions are quotiented out. Subsequent results separate face departure from oracle disagreement and give conditions for local coercivity.

Central symmetry centers the Bayes target at $\mathbb{E}[C] = \bar{c}$ and yields the nonnegative centered excess-risk identity; it is not required to define optimal-face strict crossing, establish geometric crossing arguments, or compare crossing with oracle disagreement under absolute continuity. Without symmetry, optimal-face strict crossing (SC) alone does not guarantee mean identification: Theorem F.1 constructs a full-support, mean-zero asymmetric law for which SC holds but the mean is not SPO+-optimal. Theorem 8.8 gives exact quotient-report identification when the oracle balances under reflection and reflected crossing holds; approximate symmetry instead localizes the Bayes set under a coercivity condition.

The fixed-covariate theory also has a contextual lift. Theorem F.2 treats $C \mid X = x$ as a Borel probability kernel and optimizes over a measurable predictor class. Conditional central symmetry and strict crossing give uniqueness of the conditional-mean predictor after quotient projection, up to almost-sure equality; the quantitative corollary integrates a uniform conditional coercivity modulus. These are population Bayes statements and do not imply automatic finite-sample quotient-report recovery, neural-network optimization convergence, or improved shifted-task regret without additional assumptions.

The contributions are:

- **Exact zero-excess-risk geometry.** Theorem 5.2 characterizes the exact condition under which the complete population minimizer set is one quotient-report class: every nonzero effective displacement must produce optimal-face departure with positive probability. Theorem 6.1 gives a geometric sufficient condition.
- **Oracle/crossing separation and quantitative coercivity.** Selected-oracle disagreement can occur within a shared optimal face; absolute continuity removes this ambiguity for each fixed displacement (Theorem 7.3). Quantitative crossing yields local parameter coercivity, including quadratic growth for direction-covering polyhedral boundary mass and sharp one-dimensional exponents (Section 8).
- **A calibrated symmetry boundary.** Theorem F.1 rules out symmetry-free SC identification of the mean. Theorem 8.8 recovers exact quotient-report identification under oracle selection balance and reflected crossing; total-variation asymmetry yields Bayes-set localization.
- **Contextual population identification.** Theorem F.2 and Corollary G.2 lift the fixed-x result to measurable predictors under explicit conditional assumptions.

The next section distinguishes this equality-set question from regret analyses, decision-sufficiency methods, and local analyses near decision boundaries.

Controlled synthetic experiments reproduce crossing, oracle separation, and coercivity. Portfolio, complete-matrix KuaiRec, and Energy/Storage measure predictive fidelity, pre-specified-shift regret, and fitted-report crossing. Their observational outcomes do not identify a conditional Bayes report or establish population strict crossing. We therefore add a known-DGP layer that keeps each application's feature rows and optimization geometry fixed while setting the conditional cost law by construction. This layer tests whether conditional-mean recovery, crossing mass, and finite-family transfer regret follow the predicted mechanism; it does not establish a universal transfer guarantee. The visualizations are reported in Appendix J.

2 Related work

Decision-focused learning with SPO+. Elmachtoub & Grigas (2022) introduced SPO and its convex surrogate SPO+. Existing SPO+ theory gives the centered excess-risk identity and establishes decision consistency or unique mean minimization under stronger symmetry, density, and

full-dimensionality conditions; later calibration, margin, and generalization results relate excess risk to task regret (Liu & Grigas, 2021; Balghiti et al., 2023; Ho-Nguyen & Kılınç-Karzan, 2022); Im et al. (2025) extends this target to robust constraints. Wan & Liu (2026) proposes a solver-free surrogate for predict-then-optimize training and proves Fisher consistency and excess-risk bounds for that loss. These results concern decision consistency or sufficient conditions for a target; they do not characterize the complete zero-excess-risk equality set of SPO+. Our contribution is not another SPO+ consistency theorem: under centered symmetry, we characterize the population minimizer geometry after quotienting invisible directions, with optimal-face strict crossing as the necessary and sufficient condition for a single quotient-report class.

What optimal decisions identify. Recent work characterizes data or deterministic proxies sufficient for an optimal decision, sometimes allowing nonunique optimal point forecasts, while inverse optimization characterizes objective parameters compatible with observed decisions (Bennouna et al., 2025; Ye et al., 2026; Schutte et al., 2026; de Mello et al., 2026; Ahmadi et al., 2026). Sun et al. (2023) builds a max-margin objective from optimality conditions and trains from observed optimal solutions without objective coefficients; it does not study the population SPO+ equality set. Related work on elicitable functionals and calibrated structured surrogates distinguishes the statistical target, the report, and the regret induced by a surrogate (Gneiting, 2011; Fissler et al., 2021; Frongillo & Waggoner, 2021; Finocchiaro et al., 2024; Osokin et al., 2017; Nowak-Vila et al., 2019). Those are information-, proxy-, or observation-identification targets. Ours is the Bayes equality set of a fixed surrogate: when does zero SPO+ excess risk identify one effective component, i.e. one quotient report, rather than merely preserve a downstream decision?

Geometry near decision boundaries. Normal-fan margins and contextual fast-rate analyses examine behavior near decision boundaries (Balghiti et al., 2023; Hu et al., 2022). In contrast, we use optimal-face crossing to characterize the global equality set of the population SPO+ risk, and use direction-covering boundary mass to obtain local coercivity. This separates disagreement of a selected oracle from departure from the full shifted optimal face.

3 Problem Setup and Population Notions

Sections 5 through 8 work at a fixed covariate value and suppress conditioning notation; Appendix F states the measurable contextual lift. Let $S \subset \mathbb{R}^d$ be nonempty, compact, and convex. For a cost vector $q \in \mathbb{R}^d$, define

$$z^*(q) := \min_{w \in S} q^\top w, \qquad W^*(q) := \arg\min_{w \in S} q^\top w. \tag{1}$$

The optimizer correspondence W^* is set-valued. Throughout, w^* is a deterministic measurable selection satisfying $w^*(q) \in W^*(q)$ for every q. The support function of S is $\xi_S(q) := \max_{w \in S} q^\top w = -z^*(-q)$.

For a prediction $\hat{c}$ and observed cost c, the SPO+ loss is

$$\ell_{\mathrm{SPO+}}(\hat{c}, c) := \xi_S(c - 2\hat{c}) + 2\hat{c}^\top w^*(c) - z^*(c). \tag{2}$$

Write $R_{\mathrm{SPO+}}(\hat{c}) := \mathbb{E}[\ell_{\mathrm{SPO+}}(\hat{c}, c)]$ whenever the expectation is finite, and let P_c denote the law of c. To make the downstream target explicit, we deploy the same fixed selection and define

$$\ell_{\mathrm{SPO}}(\hat{c}, c) := c^\top w^*(\hat{c}) - z^*(c), \qquad R_{\mathrm{SPO}}(\hat{c}) := \mathbb{E}[\ell_{\mathrm{SPO}}(\hat{c}, c)]. \tag{3}$$

The target parameter is $\bar{c} \in \mathbb{R}^d$. Let

$$L := \mathrm{span}(S - S), \qquad N := L^\perp, \qquad [q] := q + L^\perp \in \mathbb{R}^d / L^\perp.$$

Let P_L denote the orthogonal projection onto L. Write $k := \dim L$ and let λ_L denote k-dimensional Lebesgue measure on L. Since every $n \in N = L^\perp$ is constant on S, the SPO+ risk is invariant under $\hat{c} \mapsto \hat{c} + n$. Accordingly, every identification statement below is a statement about the quotient report $[\hat{c}] = \hat{c} + L^\perp$, equivalently the effective projected report $P_L \hat{c}$. When S is full-dimensional, $N = \{0\}$ and the quotient coincides with the ambient parameter space.

Assumption 3.1 (Centered distribution and integrability)**.** The cost law is centrally symmetric about $\bar{c}$, namely $c \overset{d}{=} 2\bar{c} - c$, and $\mathbb{E}\|c\| < \infty$. The deterministic measurable selection w^* is fixed before population risks are formed.

Compactness makes $w^*(c)$ bounded, so Assumption 3.1 is sufficient for the integrability used below. It also implies $\bar{c} = \mathbb{E}[c]$. For coordinatewise nonnegative costs, symmetry further forces $0 \le c_j \le 2\bar{c}_j$ almost surely for each coordinate j. In particular,

$$R_{\mathrm{SPO}}(\hat{c}) = \bar{c}^\top w^*(\hat{c}) - \mathbb{E}[z^*(c)], \qquad \arg\min_{\hat{c}} R_{\mathrm{SPO}}(\hat{c}) = \{\hat{c} : w^*(\hat{c}) \in W^*(\bar{c})\}. \tag{4}$$

Thus this task-risk notion is explicit about the deployed tie-breaking rule.

Proposition 3.2 (Centered Bayes minimizer)**.** *Under Assumption 3.1, $R_{\mathrm{SPO}+}$ is convex and $\bar{c} \in \arg\min_{\hat{c}} R_{\mathrm{SPO}+}(\hat{c})$.*

We distinguish the following population notions. They are often conflated in decision-focused learning, and the examples in this paper separate them.

1. **Bayes existence**: $\bar{c} \in \arg\min R_{\mathrm{SPO}+}$.
2. **Unique quotient-report identification**:
$$\arg\min R_{\mathrm{SPO}+} = \bar{c} + N \iff \arg\min_{[\hat{c}]} R_{\mathrm{SPO}+}([\hat{c}]) = \{[\bar{c}]\}.$$
3. **Fixed-rule Fisher consistency**: every Bayes report induces a Bayes-optimal decision for the deployed selection: $\arg\min R_{\mathrm{SPO}+} \subseteq \arg\min R_{\mathrm{SPO}}$, with the latter set given exactly by equation 4.
4. **Selection-independent decision consistency**:
$$W^*(\hat{c}) \subseteq W^*(\bar{c}) \quad \text{for every} \quad \hat{c} \in \arg\min R_{\mathrm{SPO}+}.$$
5. **Quantitative quotient-report identification**: there is a modulus ψ such that, for effective displacements $\Delta \in L$, $R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) \ge \psi(\|\Delta\|)$ locally. This is quotient-parameter coercivity, not task-regret calibration.
6. **Estimator recovery**: an empirical minimizer near $\bar{c}$ has effective-component error controlled by an empirical uniform-deviation and the local inverse modulus. This requires an additional localization argument and is not implied by the population coercivity alone.

Remark 3.3 (Terminology). We use "unique quotient-report identification" to mean uniqueness of the quotient Bayes report $[\hat{c}]$, never uniqueness of an ambient representative. In the terminology of scoring-function theory this is closely related to strict consistency and should not be confused with identification-function terminology for statistical functionals (Gneiting, 2011; Fissler et al., 2021). The present paper is not a scoring-rule survey; the distinction matters because a strictly consistent loss can have a unique quotient Bayes report without implying any particular downstream decision rule, and conversely decision consistency can hold without unique identification.

4 Absolute Continuity Does Not Ensure Identification

Example 4.1 (Compact support yields a flat SPO+ risk)**.** Absolute continuity does not by itself make a decision-focused surrogate identify the quotient report. Let $S = [-1/2, 1/2]$ and $C \sim \mathrm{Unif}[1, 2]$, so $\bar{c} = 3/2$. Direct evaluation gives

$$\arg\min_{\hat{c}} R_{\mathrm{SPO}+}(\hat{c}) = [1, \infty), \tag{5}$$

although every minimizer induces the same singleton decision as the mean. Indeed, for $c > 0$,

$$\ell_{\mathrm{SPO}+}(\hat{c}, c) = \tfrac{1}{2}|c - 2\hat{c}| - \hat{c} + \tfrac{c}{2}, \tag{6}$$

so the loss vanishes almost surely for $\hat{c} \ge 1$ and is positive with positive probability for $\hat{c} < 1$. The exact piecewise risk and the same construction with a smooth compactly supported density are given in Appendix D.

The obstruction is mass missing from the optimization-transition regions, not the lack of a density. Under the centered-symmetry setup, unique quotient-report identification instead requires positive mass in every nonzero effective displacement's optimal-face crossing event, formalized by optimal-face strict crossing in Section 5 and quantified in Section 8.

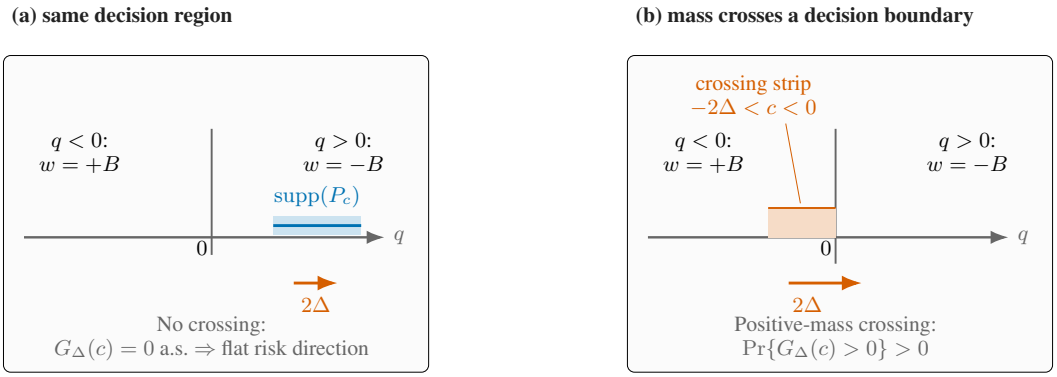


Figure 1: Missing transition-region mass leaves a flat surrogate-risk direction; mass in a crossed boundary strip yields a positive gap.

5 Strict Crossing and Exact Equality-Set Geometry

For a displacement $\Delta \in \mathbb{R}^d$, define the *optimization crossing gap*

$$G_\Delta(c) := (c + 2\Delta)^\top w^*(c) - z^*(c + 2\Delta) \geq 0. \tag{7}$$

It is the suboptimality of the old selected decision at the shifted cost. The nonnegativity in equation 7 is pointwise and the equality event has a set-valued meaning:

$$G_\Delta(c) > 0 \quad \Longleftrightarrow \quad w^*(c) \notin W^*(c + 2\Delta). \tag{8}$$

Definition 5.1 (Optimal-face strict crossing)**.** The triple (P_c, S, w^*) satisfies *optimal-face strict crossing* (SC) when, for every effective displacement $\Delta \in L \setminus \{0\}$,

$$\Pr\{w^*(c) \notin W^*(c + 2\Delta)\} > 0. \qquad \text{(SC)}$$

Theorem 5.2 (Identification if and only if strict crossing)**.** *Under Assumption 3.1,*

$$\bar{c} + N \textit{ is the unique minimizer set of } R_{\mathrm{SPO}+} \quad \Longleftrightarrow \quad \textit{SC holds.} \tag{9}$$

Proof idea. The centered excess-risk identity is

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) = \mathbb{E}\left[G_\Delta(c)\right]. \tag{10}$$

By equation 8, the expectation in equation 10 is strictly positive exactly when the SC event has positive probability. Proposition 3.2 then gives both directions of equation 9. Appendix A records the centered algebra and the full quantifier argument. □

Remark 5.3 (What is known and what is new here). The centered representation equation 10 and uniqueness under stronger full-support conditions are known (Liu & Grigas, 2021). Theorem 5.2 characterizes their exact equality case after quotienting invisible directions: zero gap means that the old decision remains in the shifted optimal face. The distinct geometric and quantitative consequences are developed below.

Theorem 5.2 treats a decision transition as the primitive object and identifies only the effective quotient report. It does not compare the two selected oracle outputs; that distinction is essential on a positive-probability tie set.

6 Geometry of Strict Crossing

The support function makes the set-valued geometry behind Theorem 5.2 explicit. All geometry in this section is intrinsic to L: translating S by any $s_0 \in S$ gives the full-dimensional body $S - s_0 \subset L$ without changing its optimizers. Let

$$\mathcal{T}_L := \{q \in L : W^*(q) \text{ is not a singleton}\}. \tag{11}$$

The restricted support function is finite convex on L, so $\mathcal{T}_L$ is λ_L-null. For a polytope, $q \mapsto W^*(q)$ is constant on the relative interiors of the minimization normal cones in L.

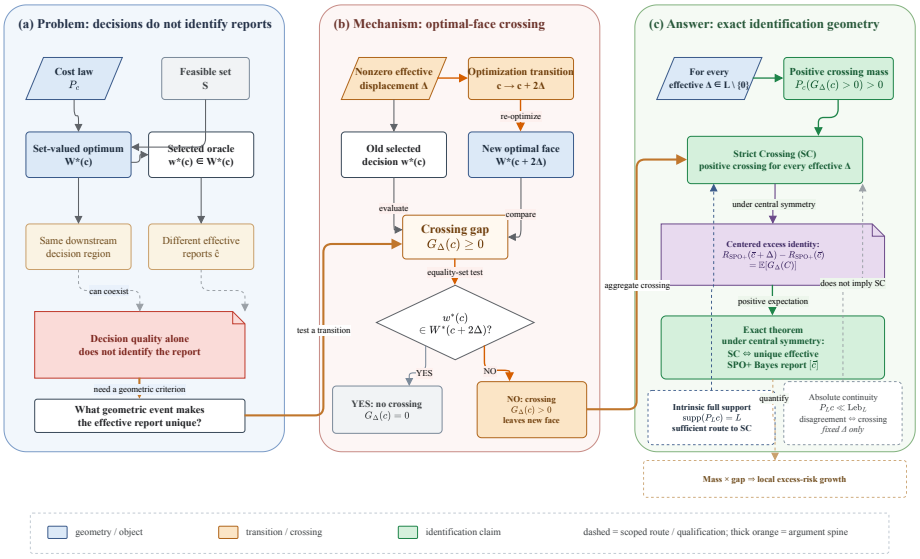


Figure 2: Optimal-face strict crossing links optimal-face departure to exact quotient identification; the side conditions and quantitative refinement are shown in the diagram.

Theorem 6.1 (Intrinsic full support implies strict crossing)**.** *Suppose* $\mathrm{supp}(P_L c) = L$*. Then SC in Definition 5.1 holds.*

Proof idea. Fix $\Delta \in L \setminus \{0\}$ and set $q_0 = -\Delta$. Every $w \in W^*(q_0)$ maximizes $\Delta^\top w$, whereas $z^*(q_0 + 2\Delta) = z^*(\Delta)$. Hence

$$\Delta^\top w - z^*(\Delta) = \max_{u \in S} \Delta^\top u - \min_{u \in S} \Delta^\top u > 0. \tag{12}$$

The width is positive because a vector in L that is orthogonal to every element of $S - S$ must vanish. Put

$$g_\Delta(q, w) := (q + 2\Delta)^\top w - z^*(q + 2\Delta).$$

Since z^* is continuous, so is g_Δ. At $q = q_0$, uniformly over $w \in W^*(q_0)$,

$$g_\Delta(q_0, w) = \max_{u \in S} \Delta^\top u - \min_{u \in S} \Delta^\top u =: \delta_\Delta > 0.$$

Compactness and continuity give relative neighborhoods U_1 of q_0 in L and V of $W^*(q_0)$ on which $g_\Delta > \delta_\Delta/2$. By the maximum theorem, $q \mapsto W^*(q)$ is upper hemicontinuous, so some relative neighborhood U_2 of q_0 satisfies $W^*(q) \subset V$. Thus, for $U = U_1 \cap U_2$,

$$W^*(q) \subset V, \qquad g_\Delta(q, w) > \delta_\Delta/2 \quad \forall w \in W^*(q),$$

which implies $w^*(q) \notin W^*(q + 2\Delta)$. Hence the SC event contains the preimage of the nonempty relatively open set U, and intrinsic full support gives $\Pr\{P_L c \in U\} > 0$. The appendix gives the reduction and full argument. □

The proof constructs a crossing neighborhood before invoking the distribution. It explains why intrinsic full support is sufficient and why a density inside one decision region, as in Example 4.1, is not.

Remark 6.2 (Quotient-report identification is intrinsic). Let $L := \mathrm{span}(S - S)$ and $L^\perp$ be its orthogonal complement. For any $\Delta \in L^\perp$, the linear form $\Delta^\top w$ is constant on S, and a pointwise calculation gives $\ell_{\mathrm{SPO}+}(\hat{c} + \Delta, c) = \ell_{\mathrm{SPO}+}(\hat{c}, c)$ for every $\hat{c}, c$ (Appendix C). Hence the ambient SPO+ risk is constant along $L^\perp$ for *every* distribution, including asymmetric laws. The identification target is therefore the quotient report $[\bar{c}] = \bar{c} + L^\perp$, or equivalently the effective component $P_L \bar{c}$. Theorem 5.2 accordingly quantifies SC over $L \setminus \{0\}$ and identifies the quotient report in $\mathbb{R}^d / L^\perp$; no ambient full-dimensionality is required.

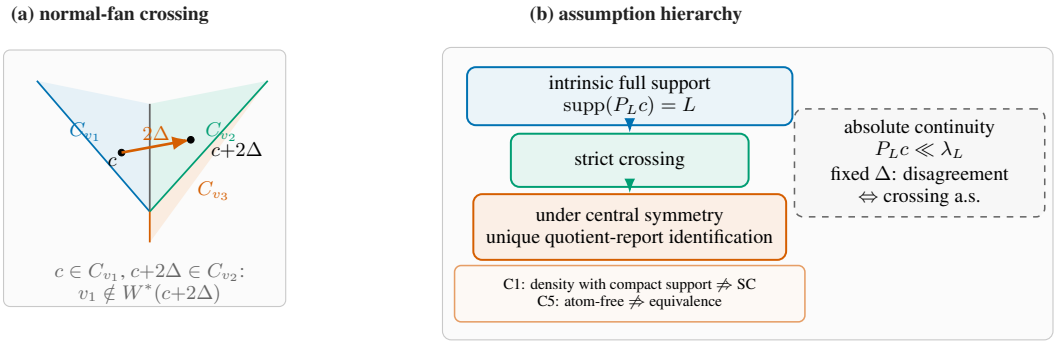


Figure 3: Normal-fan crossing and the assumption hierarchy. Intrinsic full support supplies crossing mass; under central symmetry, strict crossing gives unique quotient-report identification. Absolute continuity only resolves oracle ambiguity.

7 Oracle Disagreement and Optimal-Face Crossing

For a fixed displacement define

$$A_\Delta := \{w^*(c) \neq w^*(c+2\Delta)\}, \qquad B_\Delta := \{w^*(c) \notin W^*(c+2\Delta)\}. \tag{13}$$

Always $B_\Delta \subseteq A_\Delta$, but the converse can fail on a non-singleton shifted optimal face.

Example 7.1 (Atomic oracle disagreement without crossing)**.** The atomic construction in Appendix D has $\Pr(A_\Delta) = 1/2$ and $\Pr(B_\Delta) = 0$.

Example 7.2 (Atom-free disagreement without crossing)**.** The singular Cantor construction in Appendix D is atom-free but still has $\Pr(A_\Delta) = 1/2$ and $\Pr(B_\Delta) = 0$.

In both cases the oracle moves tangentially within a common optimal face while the SPO+ crossing gap remains zero. Thus atom-freeness is insufficient.

Absolute continuity removes this ambiguity displacement-wise:

Theorem 7.3 (Absolute continuity removes output ambiguity)**.** *Suppose $P_L c \ll \lambda_L$. For every fixed $\Delta \in \mathbb{R}^d$,*

$$\Pr(A_\Delta \triangle B_\Delta) = 0. \tag{14}$$

Proof idea. The projected tie locus in L is λ_L-null because the restricted support function is finite convex. Hence the source and shifted projected costs avoid it almost surely, and selected-output disagreement is equivalent to departure from the shifted singleton optimal face. The statement is for each fixed Δ; the full argument is in Appendix C. □

8 Quantitative Crossing and Excess-Risk Growth

Strict crossing certifies positivity but does not supply a modulus. The selected gap always obeys a linear envelope: if $D_S := \sup_{u,v \in S} \|u - v\|$, then

$$0 \leq G_\Delta(c) \leq 2D_S\|\Delta\|. \tag{15}$$

We therefore control both the mass of a crossing event and the magnitude of the gap on that event.

Definition 8.1 (Quantitative strict crossing (QSC))**.** For $r, \gamma, \kappa > 0$, $\alpha \geq 1$, and $\beta \geq 0$, $\mathrm{QSC}(\alpha, \beta; \gamma, \kappa, r)$ holds if, for every effective displacement $\Delta \in L$ with $0 < \|\Delta\| \leq r$,

$$\Pr\{G_\Delta(c) \geq \gamma\|\Delta\|^\alpha\} \geq \kappa\|\Delta\|^\beta. \tag{16}$$

Theorem 8.2 (Quantitative crossing transfer)**.** *Under Assumption 3.1, QSC$(\alpha, \beta; \gamma, \kappa, r)$ implies*

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) \geq \gamma\kappa\|\Delta\|^{\alpha+\beta} \quad \textit{for } \Delta \in L,\ 0 < \|\Delta\| \leq r. \tag{17}$$

Proposition 8.3 (Local polynomial coercivity implies quantitative crossing)**.** *Under Assumption 3.1, assume the effective space is nontrivial,* $\dim L \geq 1$ *(equivalently* $D_S > 0$*), and suppose there are* $C, r > 0$ *and* $p \geq 1$ *such that*

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) \geq C\|\Delta\|^p \quad \textit{for every } \Delta \in L,\ 0 < \|\Delta\| \leq r. \tag{18}$$

Then QSC$(p, p-1; \frac{C}{2}, \frac{C}{4D_S}, r)$ *holds.*

Proof. See Appendix E. □

Remark 8.4 (Quantitative crossing and coercivity are two views of one local phenomenon). Theorem 8.2 and Proposition 8.3 form a two-way relationship at the level of local polynomial moduli, up to transformed exponents and constants: QSC(α, β) yields growth exponent $p = \alpha + \beta$, while a growth exponent $p \geq 1$ yields QSC$(p, p-1)$. The universal linear envelope equation 15 forces $p \geq 1$ for any nontrivial polynomial lower bound. Strict crossing is the qualitative limit: it asserts $\mathbb{E}[G_\Delta] > 0$ for every nonzero Δ, but no modulus. QSC strengthens this by jointly controlling crossing mass and crossing gap on the effective quotient directions.

For polytopes, an intrinsic normal-fan condition makes the two exponents explicit. Let S be a compact polytope with affine hull $s_0 + L$. For a vertex v, let $C_v := \{q \in L : v \in W^*(q)\}$ be its minimization cone. For each edge $e = [v_i, v_j]$, let F_e be the codimension-one face shared by C_{v_i} and C_{v_j}, and set $a_e = v_i - v_j$ and $L_e := \mathrm{span}(F_e) = L \cap a_e^\perp$. Here L_e is the linear span of the fan facet, not its lineality space.

Lemma 8.5 (Edge directions span the effective space)**.** *The directions of the edges of* S *span* L*. Consequently, for any finite family* $\mathcal{E}$ *whose directions span* L*,*

$$\tau_{\mathcal{E}} := \min_{u \in L : \|u\| = 1} \max_{e \in \mathcal{E}} |a_e^\top u| > 0. \tag{19}$$

Proof. See Appendix E. □

The following assumption strengthens Theorem 6.1 into a quantitative statement. It does *not* require every edge of the polytope to have a density-positive patch: only a finite family whose directions cover all unit directions is needed. This is the direction-covering form of the boundary-mass condition.

Assumption 8.6 (Intrinsic polyhedral boundary mass)**.** Let $k = \dim L \geq 1$. Let $\mathcal{E}_0$ be a finite family of edges of S whose directions span L; equivalently $\tau_{\mathcal{E}_0} > 0$ in equation 19. The law of $P_L c$ has density p with respect to λ_L. For every $e \in \mathcal{E}_0$, there are a compact patch $K_e \Subset \mathrm{relint}(F_e)$ of positive $(k-1)$-dimensional measure, a radius $\rho_e > 0$, and a common constant $m > 0$. With

$$K_e^{[\rho_e]} := \{y + h : y \in K_e,\ h \in L_e,\ \|h\| \leq \rho_e\}, \tag{20}$$

assume $K_e^{[\rho_e]} \subset \mathrm{relint}(F_e)$ and, for every $y \in K_e^{[\rho_e]}$,

$$y + \frac{s}{\|a_e\|^2} a_e \in \mathrm{int}(C_{v_i}) \qquad \text{if } -\rho_e < s < 0, \tag{21}$$

$$y + \frac{s}{\|a_e\|^2} a_e \in \mathrm{int}(C_{v_j}) \qquad \text{if } 0 < s < \rho_e. \tag{22}$$

Finally, p admits a Borel version satisfying, for $(\mathcal{H}^{k-1} \lfloor_{K_e}) \otimes \lambda_1$-almost every $(y, s) \in K_e \times [-\rho_e, \rho_e]$,

$$p\left(y + \frac{s}{\|a_e\|^2} a_e\right) \geq m. \tag{23}$$

Theorem 8.7 (Direction-covering boundary mass yields quadratic growth)**.** *Let* S *be a compact polytope. Under Assumption 3.1 and Assumption 8.6, there exist* $C, r > 0$ *such that*

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) \geq C\|\Delta\|^2 \quad \textit{for every } \Delta \in L,\ 0 < \|\Delta\| \leq r. \tag{24}$$

All cones, interiors, densities, and Hausdorff measures in this condition are taken in L. Thus the result applies directly to equality-constrained polytopes without imposing an ambient density on the cost law.

Mechanism. By Lemma 8.5, the finite family in Assumption 8.6 has a positive covering constant $\tau_{\mathcal{E}_0}$. For any unit direction, at least one family edge has uniformly positive inner product with it. A displacement of size t then sweeps a prism of width proportional to t across the corresponding fan facet. The density floor assigns this prism probability proportional to t, and its trimmed interior has crossing gap proportional to t. The product proves equation 24. Appendix E gives the uniform finite-edge argument with explicit constants. □

The growth bound is an inverse quotient-report identification modulus, not task-regret transfer; Appendix E gives the conditional approximate-ERM consequence. It is population-only and does not automatically imply finite-sample quotient-report recovery, neural-network optimization convergence, or improved transfer regret without additional assumptions.

For $S = [-B, B]$ and even $p(c) \asymp |c|^\nu$ near the boundary ($\nu > -1$), the exact rate is

$$R_{\mathrm{SPO+}}(t) - R_{\mathrm{SPO+}}(0) = \Theta\big(|t|^{\nu+2}\big). \tag{25}$$

A positive boundary density gives a quadratic rate; a vanishing density gives a flatter polynomial rate, and SC alone need not give a polynomial modulus.

Beyond central symmetry. For asymmetric laws, SC alone can fail to identify the mean; selection balance with reflected crossing restores exact quotient-report identification (Appendix G).

Theorem 8.8 (Selection-balanced exact identification)**.** *Let $\mu = \mathbb{E}[C]$, $C^\circ = 2\mu - C$, and $\mathbb{E}\|C\| < \infty$. If $\mathbb{E}[w^*(C)] = \mathbb{E}[w^*(C^\circ)]$, then $\arg\min_\theta R_{P,\mathrm{SPO+}}(\theta) = \mu + N$ if and only if $\Pr\{w^*(C^\circ) \notin W^*(C^\circ + 2\Delta)\} > 0$ for every $\Delta \in L \setminus \{0\}$.*

The selection-balance condition compares the average deployed oracle under the observed cost law with its reflected-cost counterpart. It is automatic under central symmetry and can be plausible when the cost law and tie-breaking rule have a reflection-compatible aggregate, but it is not generic for asymmetric laws. If the cost law, its mean, and the oracle are known, the two bounded expectations can be estimated or computed; for ordinary observational data, the reflected counterfactual and exact equality are not directly observable, so this remains a population condition rather than a routine empirical test. Under central symmetry this is Theorem 5.2; reflection preserves the same crossing geometry.

9 Real-Data Diagnostics

Observational data cannot test Theorem 5.2: they do not reveal conditional means, central symmetry, or population strict crossing. Appendix J retains application features and geometries under a known cost law. RMSE and fixed-shift regret are not theorem or transfer tests. Repeated outcomes assess mean recovery under centered noise; a deterministic law controls flat equality. Finite-grid curves illustrate, but do not establish, assumptions or universal transfer.

Applications and protocol. Portfolio raises a 48-industry cap (0.10 to 0.12); KuaiRec uses user-disjoint complete matrices ($k \in \{5, 10, 15\}$); Energy/Storage uses 48 prices, an intercept, and tilts $s \in \{0, 0.5, 1\}$. Appendix I gives five seeds and full shift matrices (Tables 3, 4, 6).

Observed patterns. On KuaiRec, Linear SPO+ leads formal methods at all tested budgets; on Energy, SPO+ leads only at the largest tilt. Portfolio shifted regrets are close despite up to sevenfold report-RMSE differences. Fitted crossing is descriptive. Appendix J shows improved mean recovery and crossing on tested geometries, not population strict crossing or universal transfer.

10 Discussion, Limitations, and Conclusion

Under symmetry, $\mathbb{E}[C]$ is the Bayes target and the centered identity holds. Without symmetry, selection balance with reflected crossing identifies the quotient report; approximate symmetry localizes the Bayes set, but SC alone may fail. These population results imply neither finite-sample recovery, neural convergence, nor transfer regret; lower-dimensional feasible sets identify only a quotient. Empirical checks (fixed shifts, known DGPs, no KuaiRec online exposure, paired Energy results, finite neural grid) do not verify the universal quantifiers in the theorems.

Reproducibility Statement

The main text and appendices state the assumptions and quantifiers for every result. The accompanying materials contain the full proofs, counterexamples, closed-form illustrations, experiment protocols, complete result cells, and the mathematical checks used in this paper.

AI Use Statement

We used generative AI tools to assist with research planning, literature synthesis, checking mathematical derivations and proof presentation, reviewing experimental design and implementation, interpreting experiment outputs, and revising manuscript prose. We did not use generative AI to generate synthetic datasets, clean or reformat datasets, or conduct qualitative or thematic data analysis. The authors manually checked AI-assisted mathematical material against the stated arguments, verified software changes and numerical outputs with the reported tests, and take responsibility for the final text, claims, proofs, and software artifacts.

References

Farzin Ahmadi, Fardin Ganjkhanloo, and Kimia Ghobadi. From non-identifiability to goal-integrated decision-making in parametric inverse optimization, 2026. URL https://arxiv.org/abs/2603.17033.

Othman El Balghiti, Adam N. Elmachtoub, Paul Grigas, and Ambuj Tewari. Generalization bounds in the predict-then-optimize framework. *Mathematics of Operations Research*, 48(4):2043–2065, 2023. doi: 10.1287/moor.2022.1330.

Omar Bennouna, Amine Bennouna, Saurabh Amin, and Asuman Ozdaglar. What data enables optimal decisions? an exact characterization for linear optimization. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Tito Homem de Mello, Juan Valencia, Felipe Lagos, and Guido Lagos. Forecasting outside the box: Application-driven optimal pointwise forecasts for a class of two-stage stochastic programs. *Mathematical Programming*, 2026. doi: 10.1007/s10107-026-02359-9.

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, 2017. URL https://arxiv.org/abs/1703.04529.

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

Jessie Finocchiaro, Rafael M. Frongillo, and Bo Waggoner. An embedding framework for the design and analysis of consistent polyhedral surrogates. *Journal of Machine Learning Research*, 25(63): 1–60, 2024. URL https://arxiv.org/abs/2206.14707.

Tobias Fissler, Jana Hlavinová, and Birgit Rudloff. Elicitability and identifiability of set-valued measures of systemic risk. *Finance and Stochastics*, 25:133–165, 2021. doi: 10.1007/s00780-020-00446-z.

Rafael M. Frongillo and Bo Waggoner. Surrogate regret bounds for polyhedral losses. In *Advances in Neural Information Processing Systems*, volume 34, pp. 21569–21580, 2021. URL https://arxiv.org/abs/2110.14031.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011. doi: 10.1198/jasa.2011.r10138.

Nam Ho-Nguyen and Fatma Kılınç-Karzan. Risk guarantees for end-to-end prediction and optimization processes. *Management Science*, 68(12):8680–8698, 2022. doi: 10.1287/mnsc.2022.4321.

Yichun Hu, Nathan Kallus, and Xiaojie Mao. Fast rates for contextual linear optimization. *Management Science*, 68(6):4236–4245, 2022. doi: 10.1287/mnsc.2022.4383.

Hyungki Im, Wyame Benslimane, and Paul Grigas. Smart surrogate losses for contextual stochastic linear optimization with robust constraints. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Heyuan Liu and Paul Grigas. Risk bounds and calibration for a smart predict-then-optimize method. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://arxiv.org/abs/2108.08887.

Jayanta Mandi, Víctor Bucarey, Maxime Mulamba Ke Tchomba, and Tias Guns. Decision-focused learning: Through the lens of learning to rank. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 14935–14947, 2022. URL https://arxiv.org/abs/2112.03609.

Alex Nowak-Vila, Francis Bach, and Alessandro Rudi. A general theory for structured prediction with smooth convex surrogates, 2019.

Anton Osokin, Francis Bach, and Simon Lacoste-Julien. On structured prediction theory with calibrated convex surrogate losses. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1703.02403.

Noah Schutte, Grigorii Veviurko, Krzysztof Postek, and Neil Yorke-Smith. Sufficient decision proxies for decision-focused learning. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2337–2345, 2026. doi: 10.24963/ijcai.2026/260. URL https://arxiv.org/abs/2505.03953.

Chunlin Sun, Shang Liu, and Xiaocheng Li. Maximum optimality margin: A unified approach for contextual linear programming and inverse linear programming. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32886–32912, 2023. URL https://proceedings.mlr.press/v202/sun23e.html.

Beichen Wan and Mo Liu. A solver-free training method for predict-then-optimize, 2026. Accepted at ICML 2026.

Yuhan Ye, Saurabh Amin, and Asuman Ozdaglar. Learning decision-sufficient representations for linear optimization, 2026. URL https://arxiv.org/abs/2603.18551. Accepted at COLT 2026.

A Proofs for Centering and Strict Crossing

Proof of Proposition 3.2. For fixed c, equation 2 is convex in $\hat{c}$, and

$$g_c := 2\big(w^*(c) - w^*(2\bar{c} - c)\big) \in \partial_{\hat{c}} \ell_{\mathrm{SPO}+}(\bar{c}, c). \tag{26}$$

Indeed, $\partial_{\hat{c}} \xi_S(c - 2\hat{c}) = -2W^*(2\hat{c} - c)$, so g_c is the sum of $-2w^*(2\bar{c} - c)$ and the derivative $2w^*(c)$ of the report-dependent term. By convexity,

$$\ell_{\mathrm{SPO}+}(\hat{c}, c) \geq \ell_{\mathrm{SPO}+}(\bar{c}, c) + g_c^\top(\hat{c} - \bar{c}) \quad \text{for every } \hat{c}, c. \tag{27}$$

Both sides are integrable: the loss is bounded by a constant multiple of $\|c\| + \|\hat{c}\| + 1$ because S is compact, and g_c is bounded for the same reason. Taking expectations in equation 27 gives

$$R_{\mathrm{SPO}+}(\hat{c}) \geq R_{\mathrm{SPO}+}(\bar{c}) + \mathbb{E}[g_c]^\top(\hat{c} - \bar{c}). \tag{28}$$

Under Assumption 3.1, $c \overset{d}{=} 2\bar{c} - c$, so the deterministic selections satisfy $\mathbb{E}[w^*(c)] = \mathbb{E}[w^*(2\bar{c} - c)]$; hence $\mathbb{E}[g_c] = 0$. Therefore equation 28 shows $R_{\mathrm{SPO}+}(\hat{c}) \geq R_{\mathrm{SPO}+}(\bar{c})$ for every $\hat{c}$, proving $\bar{c} \in \arg\min_{\hat{c}} R_{\mathrm{SPO}+}(\hat{c})$. Convexity of $R_{\mathrm{SPO}+}$ follows from convexity of equation 2 in $\hat{c}$ and linearity of the expectation. □

Lemma A.1 (Centered excess-risk identity)**.** *Under Assumption 3.1, for every $\Delta \in \mathbb{R}^d$,*

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) = \mathbb{E}[G_\Delta(c)]. \tag{29}$$

Proof. Subtract equation 2 at $\bar{c}$ from its value at $\bar{c} + \Delta$. Before using symmetry, the difference is

$$\mathbb{E}\big[\xi_S(c - 2\bar{c} - 2\Delta) - \xi_S(c - 2\bar{c}) + 2\Delta^\top w^*(c)\big]. \tag{30}$$

Since $c - 2\bar{c} \overset{d}{=} -c$, symmetry changes the expectation of the support-function difference into the corresponding expression at $-c$; the linear term is unchanged. Thus

$$R_{\mathrm{SPO}+}(\bar{c} + \Delta) - R_{\mathrm{SPO}+}(\bar{c}) = \mathbb{E}\big[\xi_S(-c - 2\Delta) - \xi_S(-c) + 2\Delta^\top w^*(c)\big] \tag{31}$$

$$= \mathbb{E}\big[(c + 2\Delta)^\top w^*(c) - z^*(c + 2\Delta)\big]. \tag{32}$$

For the second equality, use $\xi_S(-q) = -z^*(q)$ and $c^\top w^*(c) = z^*(c)$. This is equation 29. □

Proof of Theorem 5.2. Lemma A.1 writes the excess risk as the expectation of a nonnegative random variable. By optimality at $c + 2\Delta$, the random variable vanishes exactly when $w^*(c) \in W^*(c + 2\Delta)$. Therefore the excess is positive for every effective displacement $\Delta \in L \setminus \{0\}$ exactly when SC holds. Proposition 3.2 shows that $\bar{c}$ is already a global minimizer. If SC holds, every report outside the coset $\bar{c} + N$ has strictly larger risk; if SC fails for some nonzero $\Delta \in L$, the excess is zero and $\bar{c} + \Delta$ is a second quotient minimizer. Reports differing only by an element of N are equivalent by the quotient invariance in Section 3. □

B Optimal-Set Containment for General Compact Convex Sets

This section records a foundational result for the set-valued SPO loss. The containment statement appears in Elmachtoub and Grigas (Elmachtoub & Grigas, 2022); we do not claim it as a new main theorem. We include a self-contained compact-convex proof because the set-valued distinction is useful for our identification analysis, and we explicitly avoid perturbing arbitrary face-interior points. It is separate from the fixed-selection deployed loss in equation 3: ties are evaluated by the worst decision in the predicted optimal face. Define

$$\ell^{\mathrm{set}}_{\mathrm{SPO}}(\hat{c}, c) := \max_{w \in W^*(\hat{c})} c^\top w - z^*(c), \qquad R^{\mathrm{set}}_{\mathrm{SPO}}(\hat{c}) := \mathbb{E}\big[\ell^{\mathrm{set}}_{\mathrm{SPO}}(\hat{c}, c)\big]. \tag{33}$$

Theorem B.1 (Optimal-set containment)**.** *Let $S \subset \mathbb{R}^d$ be nonempty, compact, and convex, assume $\mathbb{E}\|c\| < \infty$, and write $\bar{c} := \mathbb{E}[c]$.*

1. *If $c^* \in \arg\min_{\hat{c}} R^{\mathrm{set}}_{\mathrm{SPO}}(\hat{c})$, then $W^*(c^*) \subseteq W^*(\bar{c})$.*

2. *Conversely, if* $W^*(c^*)$ *is a singleton and* $W^*(c^*) \subseteq W^*(\bar{c})$*, then* $c^* \in \arg\min_{\hat{c}} R^{\text{set}}_{\text{SPO}}(\hat{c})$.

Proof. Let $F := W^*(\bar{c})$. We first prove the containment. If F is a singleton, write $F = \{e\}$ and take $\tilde{c}_n = \bar{c}$. Otherwise choose an exposed point e of the compact convex set F relative to its affine hull. Thus there is $d \in \mathbb{R}^d$ such that

$$\arg\min_{w \in F} d^\top w = \{e\}. \tag{34}$$

For $\epsilon_n \downarrow 0$, put $\tilde{c}_n = \bar{c} + \epsilon_n d$.

We claim that $W^*(\tilde{c}_n)$ converges to $\{e\}$ in Hausdorff distance. Take any sequence $w_n \in W^*(\tilde{c}_n)$. Optimality against e gives

$$\epsilon_n(d^\top e - d^\top w_n) \geq \bar{c}^\top w_n - z^*(\bar{c}) \geq 0, \tag{35}$$

so $d^\top w_n \leq d^\top e$. Every convergent subsequence has a limit $w \in W^*(\bar{c}) = F$: this follows by passing to the limit in $\tilde{c}_n^\top w_n \leq \tilde{c}_n^\top u$ for every $u \in S$. Equation equation 34 and the preceding inequality force $w = e$. Compactness then implies the asserted Hausdorff convergence.

Let c^* minimize $R^{\text{set}}_{\text{SPO}}$. Since the support functions of $W^*(\tilde{c}_n)$ converge pointwise to that of $\{e\}$ and are dominated by $M\|c\|$, where $M := \sup_{w \in S} \|w\|$, dominated convergence yields

$$0 \leq \lim_{n \to \infty} \left(R^{\text{set}}_{\text{SPO}}(\tilde{c}_n) - R^{\text{set}}_{\text{SPO}}(c^*) \right) \tag{36}$$

$$= \bar{c}^\top e - \mathbb{E}\left[\max_{w \in W^*(c^*)} c^\top w \right] \tag{37}$$

$$\leq \bar{c}^\top e - \max_{w \in W^*(c^*)} \bar{c}^\top w \leq 0. \tag{38}$$

The penultimate inequality is Jensen's inequality for the support function of $W^*(c^*)$; the last uses $\bar{c}^\top e = z^*(\bar{c}) \leq \bar{c}^\top w$ for every $w \in S$. Equality throughout equation 38 implies $\bar{c}^\top w = z^*(\bar{c})$ for every $w \in W^*(c^*)$, proving $W^*(c^*) \subseteq W^*(\bar{c})$.

For the converse, write $W^*(c^*) = \{w^*\} \subseteq W^*(\bar{c})$. Then

$$R^{\text{set}}_{\text{SPO}}(c^*) = z^*(\bar{c}) - \mathbb{E}[z^*(c)]. \tag{39}$$

For arbitrary $\hat{c}$ and any $w \in W^*(\hat{c})$,

$$R^{\text{set}}_{\text{SPO}}(\hat{c}) \geq \mathbb{E}[c^\top w] - \mathbb{E}[z^*(c)] \tag{40}$$

$$= \bar{c}^\top w - \mathbb{E}[z^*(c)] \geq z^*(\bar{c}) - \mathbb{E}[z^*(c)], \tag{41}$$

which proves optimality of c^*. □

Example B.2 (A face-interior point need not be exposable)**.** Let $S = \text{conv}\{(0,0),(1,0),(0,1)\}$ and $\bar{c} = 0$, so $W^*(\bar{c}) = S$. The point $\bar{w} = (0.4, 0.3)$ lies in $\text{relint}(S)$, and no cost vector q satisfies $W^*(q) = \{\bar{w}\}$. Indeed, a singleton exposed face of a polytope is a vertex. This example rules out proofs that perturb the cost to make an arbitrary point of an optimal face uniquely optimal. Theorem B.1 instead exposes one suitable point of the mean-optimal face and lets the entire perturbed optimal set converge to it.

Remark B.3 (Why compactness and the set-valued loss matter). Compactness supplies both optimal solutions and the integrable envelope used in dominated convergence. The maximum in equation 33 is also essential to this theorem's interface. For the fixed-selection risk in equation 3, the Bayes set is already characterized directly by equation 4; the two risks should not be identified on tie predictions.

C Geometry of Tie Loci and Full Support

Proposition C.1 (Affine-hull obstruction and quotient invariance)**.** *Let* $L := \text{span}(S - S)$*. If* $L^\perp$ *contains a nonzero vector* Δ*, then for every* $\hat{c}, c \in \mathbb{R}^d$*,*

$$\ell_{\text{SPO+}}(\hat{c} + \Delta, c) = \ell_{\text{SPO+}}(\hat{c}, c). \tag{42}$$

Consequently, $R_{\mathrm{SPO}+}(\hat{c}+\Delta) = R_{\mathrm{SPO}+}(\hat{c})$ *for* every *distribution satisfying the integrability needed to define the risk, and*

$$W^*(c+2\Delta) = W^*(c) \quad \textit{and} \quad G_\Delta(c) = 0 \qquad \textit{for every } c \in \mathbb{R}^d. \tag{43}$$

Thus SC and ambient-space unique identification are impossible for every distribution. In particular, full dimensionality is necessary for a distribution-free sufficient condition that covers all nonzero ambient directions.

Proof. For $\Delta \in L^\perp$, there is a constant $k = \Delta^\top w$ independent of $w \in S$. Hence

$$\xi_S(q - 2\Delta) = \xi_S(q) - 2k \quad \text{for every } q \in \mathbb{R}^d. \tag{44}$$

Applying this with $q = c - 2\hat{c}$ and adding the linear report shift,

$$\ell_{\mathrm{SPO}+}(\hat{c}+\Delta, c) = \xi_S(c - 2\hat{c} - 2\Delta) + 2(\hat{c}+\Delta)^\top w^*(c) - z^*(c) \tag{45}$$

$$= \xi_S(c - 2\hat{c}) - 2k + 2\hat{c}^\top w^*(c) + 2k - z^*(c) \tag{46}$$

$$= \ell_{\mathrm{SPO}+}(\hat{c}, c). \tag{47}$$

This proves equation 42.

To prove optimizer invariance, fix $c \in \mathbb{R}^d$. For any $w, u \in S$, $\Delta \in L^\perp$ gives $\Delta^\top(w-u) = 0$, hence

$$(c+2\Delta)^\top(w-u) = c^\top(w-u). \tag{48}$$

Thus c and $c+2\Delta$ induce exactly the same pairwise objective differences on S, so $W^*(c+2\Delta) = W^*(c)$. Since $w^*(c) \in W^*(c) = W^*(c+2\Delta)$, the crossing gap vanishes by equation 8. □

Remark C.2 (Identification modulo the invisible subspace). Let $L = \mathrm{span}(S - S)$. Proposition C.1 shows that the SPO+ population risk is constant on every coset of $L^\perp$, so the natural parameter space is the quotient $\mathbb{R}^d/L^\perp$, represented by the projected cost $P_L c$. Under Assumption 3.1, $\bar{c}$ is the unique minimizer of $R_{\mathrm{SPO}+}$ on the affine subspace $\bar{c} + L$ if and only if strict crossing holds for every nonzero $\Delta \in L$. This is exactly Theorem 5.2 applied after restricting the displacement set to the directions that affect W^*; no additional full-dimensionality is needed for the quotient statement.

The next two proofs verify the intrinsic versions stated in the main text. They use only that $W^*(c) = W^*(P_L c)$ and that every effective shift is represented by its projection onto L.

For a compact polytope with affine hull $s_0 + L$, strict crossing admits an intrinsic normal-fan representation. Let V be the vertex set and

$$C_v := \{q \in L : v \in W^*(q)\}, \qquad v \in V, \tag{49}$$

be the closed cones of the minimization normal fan.

Proposition C.3 (Normal-fan representation of crossing)**.** *Suppose S is a compact polytope and $P_L c \ll \lambda_L$. For every fixed $\Delta \in \mathbb{R}^d$,*

$$\Pr(B_\Delta) = \Pr\left(\bigcup_{v \in V} \{P_L c \in \mathrm{int}_L(C_v) : P_L c + 2P_L\Delta \notin C_v\} \right). \tag{50}$$

The union is disjoint up to a null set. Thus, away from tie strata, strict crossing is exactly positive probability mass leaving the closed normal cone of the formerly optimal vertex; landing on a shared boundary does not count when that vertex remains optimal.

Proof. The union of the lower-dimensional faces of the normal fan is λ_L-null. Absolute continuity therefore implies that $P_L c$ belongs almost surely to the relative interior of exactly one vertex cone, say C_v, and hence $W^*(c) = \{v\}$ and $w^*(c) = v$. By the definition of C_v,

$$w^*(c) \notin W^*(c+2\Delta) \iff v \notin W^*(c+2\Delta) \iff P_L c + 2P_L\Delta \notin C_v. \tag{51}$$

Taking the union over the almost-surely unique source cone proves equation 50. The interiors of distinct full-dimensional-in-L vertex cones are disjoint. □

Remark C.4 (Tie boundaries and oracle independence). Proposition C.3 removes the source oracle from the event because absolute continuity makes the source optimizer unique almost surely. At the shifted cost, the closed cone C_v retains the set-valued test: if $c + 2\Delta$ lands on a shared fan face that still contains v, then the old decision remains optimal and no crossing occurs. This is the geometric distinction between optimal-face departure and an arbitrary change of selected output.

Proof of Theorem 7.3. Fix $s_0 \in S$ and restrict the support function of $S - s_0$ to L. It is finite convex on L, so its nondifferentiability locus, namely $\mathcal{T}_L$, is λ_L-null. Let A_Δ, B_Δ be as in equation 13. The inclusion $B_\Delta \subseteq A_\Delta$ is immediate. Outside $\{P_L c + 2P_L\Delta \in \mathcal{T}_L\}$, the shifted optimal face is a singleton; hence $A_\Delta \setminus B_\Delta$ is contained in that null event. Since a translation of an absolutely continuous law on L is absolutely continuous, equation 14 follows for the fixed displacement Δ. □

Proof of Theorem 6.1. Fix $\Delta \in L \setminus \{0\}$ and put $q_0 = -\Delta$. The directional width in equation 12 is positive because $\Delta \notin L^\perp$. Define

$$h(q, w) := (q + 2\Delta)^\top w - z^*(q + 2\Delta), \qquad q \in L. \tag{52}$$

At q_0, $h(q_0, w)$ equals the positive width for every $w \in W^*(q_0)$. If no relative neighborhood with the asserted uniform positivity existed, there would be $q_n \to q_0$ in L and $w_n \in W^*(q_n)$ with $h(q_n, w_n) \le 0$. Compactness yields a convergent subsequence. Berge upper hemicontinuity places its limit in $W^*(q_0)$, while continuity of z^* makes the limiting value of h positive, a contradiction. Hence a nonempty relatively open $U \subset L$ satisfies

$$\forall q \in U,\ \forall w \in W^*(q): \quad (q + 2\Delta)^\top w > z^*(q + 2\Delta). \tag{53}$$

Every legal selection therefore crosses on $\{P_L c \in U\}$. Intrinsic full support implies $\Pr\{P_L c \in U\} > 0$, and the argument applies to every nonzero $\Delta \in L$. □

D Counterexample Details

Exact compact-support risk for Example 4.1. Integrating equation 6 gives

$$R_{\mathrm{SPO}+}(\hat{c}) = \begin{cases} \frac{3}{2} - 2\hat{c}, & \hat{c} \le \frac{1}{2}, \\ 2(1-\hat{c})^2, & \frac{1}{2} < \hat{c} < 1, \\ 0, & \hat{c} \ge 1. \end{cases} \tag{54}$$

Thus the loss is strictly decreasing before the plateau and vanishes on the entire ray $[1, \infty)$. This calculation verifies that the displayed Bayes set is exact rather than a partial list of additional minimizers.

The same plateau holds for a C^∞ density p that is symmetric about $3/2$, strictly positive on $(1, 2)$, and zero outside $[1, 2]$: for $\hat{c} \ge 1$ the event $\{c > 2\hat{c}\}$ has probability zero, so the formula in equation 6 gives zero loss almost surely; for $\hat{c} < 1$ the event $\{c > 2\hat{c}\}$ has positive probability because p is strictly positive on $\big(\max\{1, 2\hat{c}\}, 2\big)$, which is nonempty. Thus the phenomenon is not an artifact of the uniform law.

Atomic construction for Example 7.1. In this and the following singular construction, let $S = [-1, 1]^2$. Let P_c put mass $1/2$ at each of $(1, 0)$ and $(-1, 0)$, so the law is centrally symmetric about zero. Define the global Borel measurable selection

$$w^*(q_1, q_2) = \begin{cases} (-\operatorname{sgn}(q_1), -\operatorname{sgn}(q_2)), & q_2 \ne 0, \\ (-\operatorname{sgn}(q_1), 1), & q_2 = 0,\ q_1 \le 1, \\ (-1, -1), & q_2 = 0,\ q_1 > 1, \end{cases} \tag{55}$$

where $\operatorname{sgn}(0) = 0$ and $\operatorname{sgn}(x) = x/|x|$ for $x \ne 0$. This is Borel measurable and $w^*(q) \in W^*(q)$ for every q: the first coordinate is the unique minimizer whenever $q_1 \ne 0$, and when $q_1 = 0$ the value 0 is also optimal; the second coordinate is chosen from the optimal interval $[-1, 1]$ on the tie slice $q_2 = 0$, and is the unique minimizer when $q_2 \ne 0$.

Fix $\Delta = (\delta, 0)$ with $0 < \delta < 1/2$. At $c = (1, 0)$, the shift to $(1 + 2\delta, 0)$ changes the selected second coordinate from 1 to -1, but both selections lie in the shifted face $\{-1\} \times [-1, 1]$. At $c = (-1, 0)$, both c and $c + 2\Delta$ have negative first coordinate and the selected output is unchanged. Hence

$$\Pr(A_\Delta) = \frac{1}{2}, \qquad \Pr(B_\Delta) = 0, \qquad G_\Delta(c) = 0 \quad \text{a.s.} \tag{56}$$

The disagreement is entirely tangential to a common optimal face.

Singular construction for Example 7.2. Let X have the standard Cantor distribution. It is atom-free, singular, and satisfies $X \overset{d}{=} 1 - X$. Define the global Borel measurable selection

$$w^*(q_1, q_2) = \begin{cases} (-\operatorname{sgn}(q_1), -\operatorname{sgn}(q_2)), & q_2 \neq 0, \\ (-\operatorname{sgn}(q_1), 1), & q_2 = 0,\ q_1 \leq 2, \\ (-1, -1), & q_2 = 0,\ q_1 > 2. \end{cases} \tag{57}$$

The same legality check as in the atomic example applies. The cost $c = (1 + X, 0)$ is centrally symmetric about $(3/2, 0)$ and is supported on the tie slice. For $\Delta = (1/4, 0)$, the output changes exactly on $\{X > 1/2\}$, which has probability $1/2$ by symmetry and atom-freeness. The old point $(-1, 1)$ remains in the shifted face $\{-1\} \times [-1, 1]$ throughout. Hence

$$\Pr(A_\Delta) = \frac{1}{2}, \qquad \Pr(B_\Delta) = 0, \qquad G_\Delta(c) = 0 \quad \text{a.s.} \tag{58}$$

This closes the gap between atom-freeness and absolute continuity: the former does not prevent a distribution from living on a Lebesgue-null tie locus.

The SPO+ loss is not uniformly bounded under first moments. Even when S is compact, integrability does not give a finite uniform loss envelope. Let $S = [-1/2, 1/2]$, $\hat{c} = 0$, and $c = M > 0$. Then

$$\ell_{\mathrm{SPO}+}(0, M) = \xi_S(M) - z^*(M) = M \longrightarrow \infty. \tag{59}$$

Consequently, any finite-sample argument that invokes $\sup_{\hat{c},c} \ell_{\mathrm{SPO}+}(\hat{c}, c) < \infty$ requires an additional bounded support or tail assumption. The population identities in this paper instead use the stated first-moment and report-wise integrability conditions.

No power law from SC alone. In the one-dimensional problem $S = [-1/2, 1/2]$, take an even density that is proportional near zero to $\exp(-1/c^2)$. Every nonzero displacement crosses a positive-mass interval, so SC holds. Nevertheless the excess at a positive displacement t is bounded by a constant times $t^2 \exp(-1/(4t^2))$, which is smaller than every polynomial in t.

E Quantitative Proofs and Sharpness

Proof of Lemma 8.5. If the edge directions did not span L, there would be a nonzero $u \in L$ orthogonal to every edge direction. Along every edge, the linear function $u^\top v$ is therefore constant. The vertex-edge graph of a polytope is connected, so $u^\top v$ is constant on all vertices. Every point of the polytope is a convex combination of vertices, hence $u^\top v$ is constant on S; because $u \in L$, this forces $u \in L \cap L^\perp = \{0\}$, a contradiction. Since the unit sphere of L is compact and the maximum of finitely many continuous functions is continuous and strictly positive, equation 19 follows. □

Proof of Proposition 8.3. Fix Δ with $0 < \|\Delta\| \leq r$ and write $t = \|\Delta\|$. Let

$$E_\Delta := \left\{ G_\Delta(c) \geq \tfrac{C}{2} t^p \right\}. \tag{60}$$

By Lemma A.1 and equation 15,

$$\begin{aligned} C t^p &\leq \mathbb{E}[G_\Delta(c)] && (61)\\ &= \mathbb{E}[G_\Delta 1_{E_\Delta}] + \mathbb{E}[G_\Delta 1_{E_\Delta^c}] && (62)\\ &\leq 2 D_S t \, \Pr(E_\Delta) + \tfrac{C}{2} t^p. && (63) \end{aligned}$$

Rearranging gives $\Pr(E_\Delta) \geq \frac{C}{4D_S} t^{p-1}$, which is equation 16 with $\alpha = p$, $\beta = p - 1$, $\gamma = \frac{C}{2}$, and $\kappa = \frac{C}{4D_S}$. □

Proof of equation 15. Let $w_0 = w^*(c)$ and $w_1 = w^*(c+2\Delta)$. Optimality of w_0 at c gives $c^\top(w_0 - w_1) \le 0$. Therefore

$$G_\Delta(c) = (c+2\Delta)^\top(w_0 - w_1) \le 2\|\Delta\|\,\|w_0 - w_1\| \le 2D_S\|\Delta\|. \tag{64}$$

□

Proof of Theorem 8.2. Restrict the expectation in equation 10 to the event in equation 16. The gap is at least $\gamma\|\Delta\|^\alpha$ on an event of probability at least $\kappa\|\Delta\|^\beta$, proving equation 17. □

Proof of Theorem 8.7. Write $\Delta = tu$ with $u \in L$ and $\|u\| = 1$. Let $\mathcal{E}_0$ be the finite family in Assumption 8.6. By Lemma 8.5,

$$\tau := \tau_{\mathcal{E}_0} = \min_{u\in L:\|u\|=1} \max_{e\in\mathcal{E}_0} |a_e^\top u| > 0. \tag{65}$$

For the direction u, choose an edge in $\mathcal{E}_0$ and orient it as $e = [v_i, v_j]$ (relabeling its endpoints if necessary) so that $a_e^\top u \ge \tau$, and set $\delta := 2ta_e^\top u$. The patch property in Assumption 8.6 is symmetric under this relabeling. Let $\rho_{\min} := \min_{e\in\mathcal{E}_0}\rho_e$ and choose

$$0 < t \le r_0 := \frac{1}{2}\min\left\{\rho_{\min}, \frac{\rho_{\min}}{\max_{e\in\mathcal{E}_0}\|a_e\|}\right\}. \tag{66}$$

Then $2t\|u - (a_e^\top u)a_e/\|a_e\|^2\| \le \rho_e$ and $0 < \delta \le \rho_e$. The prism

$$\mathcal{P}_{e,t,u} := \left\{y + \frac{s}{\|a_e\|^2}a_e : y \in K_e,\ -\frac{3\delta}{4} \le s \le -\frac{\delta}{2}\right\} \tag{67}$$

lies in the first vertex cone while its shift by $2tu$ lies in the second. Indeed, write $u = u_T + (a_e^\top u)a_e/\|a_e\|^2$ with $u_T \in L_e$. The tangential component $2tu_T$ stays within the neighborhood in Assumption 8.6, whereas the normal coordinate changes from s to $s+\delta$. On the prism, $s \in [-3\delta/4, -\delta/2]$, so $s+\delta \in [\delta/4, \delta/2]$. Hence the optimizer is unique on the prism. Since the optimizer and gap depend only on $P_L c$ for $\Delta \in L$, on $\{P_L c \in \mathcal{P}_{e,t,u}\}$,

$$G_{tu}(c) = s + \delta \ge \frac{\delta}{4} \ge \frac{t\tau}{2}. \tag{68}$$

The map $T_e(y,s) := y + sa_e/\|a_e\|^2$ is an injective product coordinate map on $K_e \times [-\rho_e, \rho_e]$. Its Jacobian with respect to λ_L is $\|a_e\|^{-1}$. Hence, by Fubini's theorem and the almost-everywhere density lower bound in Assumption 8.6, if $A_{\min} := \min_{e\in\mathcal{E}_0}\mathcal{H}^{k-1}(K_e)$ and $a_{\max} := \max_{e\in\mathcal{E}_0}\|a_e\|$,

$$\begin{aligned}\Pr\{P_L c \in \mathcal{P}_{e,t,u}\} &= \int_{K_e}\int_{-3\delta/4}^{-\delta/2} p\big(T_e(y,s)\big)\,\frac{ds\,d\mathcal{H}^{k-1}(y)}{\|a_e\|} \\ &\ge \frac{m\,\mathcal{H}^{k-1}(K_e)}{\|a_e\|}\,\frac{\delta}{4} \ge \frac{mA_{\min}\tau}{2a_{\max}}t.\end{aligned} \tag{69}$$

Multiplying equation 68 and equation 69 in the centered excess identity proves equation 24, for example with $C = mA_{\min}\tau^2/(4a_{\max})$ and $r = r_0$. □

Corollary E.1 (Local inverse risk and conditional approximate-ERM recovery)**.** *Suppose that, for some $C, r, p > 0$,*

$$R_{\mathrm{SPO}+}(\bar{c}+\Delta) - R_{\mathrm{SPO}+}(\bar{c}) \ge C\|\Delta\|^p \quad \textit{for every } \Delta \in L \textit{ with } 0 < \|\Delta\| \le r. \tag{70}$$

Write $\overline{B}_L(\bar{c}, r) := \{\bar{c}+\Delta : \Delta \in L,\ \|\Delta\| \le r\}$ for the canonical effective slice. Then every $\hat{c} \in \overline{B}_L(\bar{c}, r)$ obeys

$$\|\hat{c} - \bar{c}\| \le C^{-1/p}\big(R_{\mathrm{SPO}+}(\hat{c}) - R_{\mathrm{SPO}+}(\bar{c})\big)^{1/p}. \tag{71}$$

Let $\widehat{R}_n$ be an empirical criterion and let a (possibly random) $\hat{c}_n \in \overline{B}_L(\bar{c}, r)$ satisfy

$$\widehat{R}_n(\hat{c}_n) \le \widehat{R}_n(\bar{c}) + \eta_n. \tag{72}$$

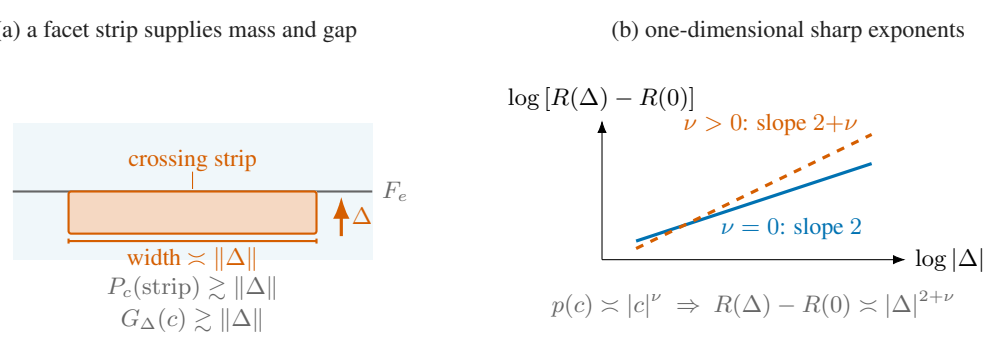


Figure 4: Quantitative crossing is a mass-gap product. A displacement sweeps a strip of width proportional to $\|\Delta\|$ across a fan facet. Under the stated density-floor and trimmed-interior conditions, the strip has probability proportional to $\|\Delta\|$ and its crossing gap is also proportional to $\|\Delta\|$, yielding quadratic parameter-excess growth. In one dimension a vanishing boundary density changes the exponent to $2 + \nu$.

On any event on which

$$\sup_{\theta \in \overline{B}_L(\bar{c}, r)} |\widehat{R}_n(\theta) - R_{\mathrm{SPO}+}(\theta)| \le \varepsilon_n, \tag{73}$$

one has

$$\|\hat{c}_n - \bar{c}\| \le \left(\frac{2\varepsilon_n + \eta_n}{C}\right)^{1/p}. \tag{74}$$

This statement is deterministic conditional on the local uniform-deviation event equation 73 and on local membership $\hat{c}_n \in \overline{B}_L(\bar{c}, r)$; it is not a standalone finite-sample rate theorem.

Proof of Corollary E.1. If $\hat{c} = \bar{c}$, equation 71 is immediate. Otherwise, apply equation 70 with $\Delta = \hat{c} - \bar{c}$ and take pth roots. On the event equation 73,

$$R_{\mathrm{SPO}+}(\hat{c}_n) - R_{\mathrm{SPO}+}(\bar{c}) \le \widehat{R}_n(\hat{c}_n) - \widehat{R}_n(\bar{c}) + 2\varepsilon_n \tag{75}$$

$$\le \eta_n + 2\varepsilon_n. \tag{76}$$

Apply equation 71 to obtain equation 74. □

Theorem E.2 (One-dimensional sharp exponent)**.** *Let $B > 0$, let $S = [-B, B]$, let $\bar{c} = 0$, and let c have an even density p with $\mathbb{E}|c| < \infty$. For $t > 0$ sufficiently small,*

$$R_{\mathrm{SPO}+}(t) - R_{\mathrm{SPO}+}(0) = 2B \int_0^{2t} (2t - x) p(-x)\, dx. \tag{77}$$

If $mx^\nu \le p(x) \le Mx^\nu$ for $0 < x < \varepsilon$ and $\nu > -1$, then for $0 < t < \varepsilon/2$,

$$\frac{2Bm(2t)^{\nu+2}}{(\nu+1)(\nu+2)} \le R_{\mathrm{SPO}+}(t) - R_{\mathrm{SPO}+}(0) \le \frac{2BM(2t)^{\nu+2}}{(\nu+1)(\nu+2)}. \tag{78}$$

Proof. Only $c \in (-2t, 0)$ crosses the unique decision boundary. On that interval, $w^*(c) = B$, $w^*(c + 2t) = -B$, and $G_t(c) = 2B(c + 2t)$. Substitute $x = -c$ to obtain equation 77, then integrate the density bounds. □

F Contextual Prediction and Further Extensions

For a cost law P governing a random cost C, we write

$$R_{P,\mathrm{SPO}+}(\theta) := \int \ell_{\mathrm{SPO}+}(\theta, c)\, P(dc). \tag{79}$$

This notation agrees with $R_{\mathrm{SPO}+}$ in the fixed-law setup and makes the law used by the asymmetric extensions explicit.

Theorem F.1 (Strict crossing alone does not identify the mean)**.** *For $S = [-1, 1]$ with $w^*(c) = -\operatorname{sign}(c)$ for $c \neq 0$ and $w^*(0) = 0$, there exists a full-support law on $\mathbb{R}$ with mean zero such that SC holds but $0 \notin \arg\min_\theta R_{P,\mathrm{SPO}+}(\theta)$.*

Appendix G gives the proof of Theorem 8.8, an asymmetric counterexample for SC alone, and the approximate-symmetry localization result. Because reflection preserves full support, Theorem 6.1 also gives exact quotient-report identification under selection balance and full support.

For the contextual result, let $(\mathcal{X}, \mathcal{A})$ be standard Borel, P_X a probability measure, and $K(x, \cdot)$ a Borel kernel for $C \mid X = x$. Set $\mu(x) = \int cK(x, dc)$ and $m(x) = \int \|c\|K(x, dc)$.

For the conditional law, define

$$R_{x,\mathrm{SPO}+}(\theta) := \int \ell_{\mathrm{SPO}+}(\theta, c)\, K(x, dc). \tag{80}$$

Theorem F.2 (Contextual SPO+ Bayes identification)**.** *Assume $m(x) < \infty$ P_X-a.s. and $\int m\, dP_X < \infty$. Suppose there is $E \in \mathcal{A}$ with $P_X(E) = 1$ such that, for every $x \in E$ and every Borel set $A \subseteq \mathbb{R}^d$ and every $\Delta \in L \setminus \{0\}$,*

$$\begin{aligned} K(x, A) &= K(x, 2\mu(x) - A), \\ K\left(x, \{c : w^*(c) \notin W^*(c + 2\Delta)\}\right) &> 0. \end{aligned} \tag{81}$$

For any measurable predictor class $\mathcal{F}$ containing μ and satisfying $\mathbb{E}\|f(X)\| < \infty$ for every $f \in \mathcal{F}$, $\mathcal{R}(f) = \mathbb{E}[\ell_{\mathrm{SPO}+}(f(X), C)]$ is well-defined and μ is its unique minimizer modulo the equivalence $P_L f(X) = P_L \mu(X)$ P_X-almost surely.

Appendix G gives the corresponding conditional quantitative bound. The completed empirical scope is stated in Section 9 and Appendix H. Those diagnostics do not turn these population Bayes results into finite-sample, neural-optimization, or shifted-task regret guarantees.

G Proofs for Symmetry Relaxation and Contextual Prediction

For approximate symmetry, let P denote the law of C and write $\mu = \mathbb{E}_P[C]$. Set $P^\circ(A) = P(2\mu - A)$, $P^{\mathrm{sym}} = (P + P^\circ)/2$, $B_S = \sup_{w \in S} \|w\|$, and $\eta = \sup_{h \text{ Borel}, \|h\|_\infty \leq 1} |\int h\, d(P - P^{\mathrm{sym}})|$.

Theorem G.1 (Approximate-symmetry Bayes-set localization)**.** *If $p > 1$ and, for every $\Delta \in L$ with $0 < \|\Delta\| \leq r$,*

$$R_{P^{\mathrm{sym}},\mathrm{SPO}+}(\mu + \Delta) - R_{P^{\mathrm{sym}},\mathrm{SPO}+}(\mu) \geq C\|\Delta\|^p, \tag{82}$$

and $4B_S\eta < Cr^{p-1}$, then the risk has a global minimizer and every quotient minimizer $[\theta_P]$ satisfies

$$\|P_L(\theta_P - \mu)\| \leq \left(\frac{4B_S\eta}{C}\right)^{1/(p-1)}. \tag{83}$$

This localizes the Bayes set but does not establish exact uniqueness.

Corollary G.2 (Contextual quantitative identification)**.** *Under Theorem F.2, suppose that for P_X-a.e. x and every $\Delta \in L$ with $0 < \|\Delta\| \leq r$,*

$$R_{x,\mathrm{SPO}+}(\mu(x) + \Delta) - R_{x,\mathrm{SPO}+}(\mu(x)) \geq C\|\Delta\|^p. \tag{84}$$

Then every $f \in \mathcal{F}$ with $\|P_L(f(X) - \mu(X))\| \leq r$ almost surely obeys

$$\mathcal{R}(f) - \mathcal{R}(\mu) \geq C\, \mathbb{E}\|P_L(f(X) - \mu(X))\|^p. \tag{85}$$

Proof of Theorem F.1. Let Y have asymmetric Laplace density

$$f_Y(y) = \frac{ab}{a+b} \begin{cases} e^{-ay}, & y \geq 0, \\ e^{by}, & y < 0, \end{cases} \qquad a = 1, \quad b = 2,$$

and set $C = Y - \mathbb{E}[Y] = Y - 1/2$. The law has full support and mean zero, so the geometric full-support argument gives SC. At $\theta = 0$, the SPO+ risk is differentiable almost surely and

$$R'_{P,\mathrm{SPO}+}(0) = 2\mathbb{E}[w^*(C) - w^*(-C)] = -4\mathbb{E}[\operatorname{sign}(C)].$$

But $\Pr(C > 0) = \Pr(Y > 1/2) = \frac{2}{3}e^{-1/2} \neq \frac{1}{2}$, so the derivative is nonzero. Hence zero is not a minimizer. □

Proof of Theorem 8.8. Set $C^\circ = 2\mu - C$. Subtracting equation 2 at μ from its value at $\mu + \Delta$, using $\xi_S(-q) = -z^*(q)$ and $(C^\circ)^\top w^*(C^\circ) = z^*(C^\circ)$, gives the pointwise identity

$$\ell_{\mathrm{SPO}+}(\mu + \Delta, C) - \ell_{\mathrm{SPO}+}(\mu, C) = G_\Delta(C^\circ) + 2\Delta^\top \big(w^*(C) - w^*(C^\circ)\big).$$

Compactness of S and integrability of C justify taking expectations. The balance assumption cancels the final term, so

$$R_{P,\mathrm{SPO}+}(\mu + \Delta) - R_{P,\mathrm{SPO}+}(\mu) = \mathbb{E}[G_\Delta(C^\circ)]. \tag{86}$$

The random variable on the right is nonnegative and vanishes exactly when $w^*(C^\circ) \in W^*(C^\circ + 2\Delta)$. Thus the reflected crossing condition in Theorem 8.8 makes the excess strictly positive for every nonzero $\Delta \in L$. Quotient invariance then gives $\arg\min R_{P,\mathrm{SPO}+} = \mu + N$. Conversely, failure of that condition makes the right-hand side of equation 86 zero for a nonzero effective displacement, so uniqueness in the quotient fails. □

Proof of Theorem G.1. For an effective displacement $\Delta \in L$, write $D_P(\Delta) = R_{P,\mathrm{SPO}+}(\mu + \Delta) - R_{P,\mathrm{SPO}+}(\mu)$. The report-dependent loss difference is bounded uniformly in c by $4B_S\|\Delta\|$: the support-function difference is at most $2B_S\|\Delta\|$ by Lipschitz continuity, and the linear term $2\Delta^\top w^*(c)$ contributes at most another $2B_S\|\Delta\|$. Therefore

$$|D_P(\Delta) - D_{P^{\mathrm{sym}}}(\Delta)| \le 4B_S\eta\|\Delta\|.$$

Combining this bound with equation 82 for $\Delta \in L$ yields

$$D_P(\Delta) \ge C\|\Delta\|^p - 4B_S\eta\|\Delta\| \quad (0 < \|\Delta\| \le r).$$

The small-asymmetry condition makes this expression positive on the boundary of the radius-r ball in L. Continuity and convexity give a minimizer in its interior in the quotient; convexity rules out a lower value outside the ball. Every report decomposes as $\mu + \Delta + n$ with $\Delta \in L$ and $n \in N$, and the risk is invariant in n. Thus a quotient minimizer has representative $\theta_P = \mu + \Delta_P$, and the displayed lower bound with $D_P(\Delta_P) \le 0$ implies equation 83. □

Proof of Theorem F.2. The loss is Borel measurable in (x, c, θ) and continuous in θ. Kernel integration therefore makes $(x, \theta) \mapsto R_{x,\mathrm{SPO}+}(\theta) = \int \ell_{\mathrm{SPO}+}(\theta, c)K(x, dc)$ jointly Borel measurable. For every $x \in E$, Theorem 5.2 applies to the conditional law and shows

$$R_{x,\mathrm{SPO}+}(\theta) = R_{x,\mathrm{SPO}+}(\mu(x)) \quad \Longleftrightarrow \quad P_L\theta = P_L\mu(x).$$

Integrability and Tonelli's theorem give

$$\mathcal{R}(f) - \mathcal{R}(\mu) = \mathbb{E}\big[R_{X,\mathrm{SPO}+}(f(X)) - R_{X,\mathrm{SPO}+}(\mu(X))\big] \ge 0.$$

If $P_L f(X) \ne P_L\mu(X)$ on a set of positive P_X measure, the integrand is strictly positive there, proving uniqueness in the quotient modulo P_X-almost-sure equality of projected reports. □

Proof of Corollary G.2. Apply the assumed conditional lower bound at the effective displacement $\Delta = P_L(f(X) - \mu(X)) \in L$ and use quotient invariance. The radius condition holds almost surely, so Tonelli's theorem yields the claimed inequality. □

H Real-Data Protocol and Evidence Scope

The falsifiable empirical hypothesis is that, at similar training-task decision quality, better predictive fidelity may coincide with less degradation under a pre-specified optimization shift. The theorems do not imply this hypothesis, and the observed outcomes do not reveal a latent conditional Bayes report.

All three activated applications are complete. Portfolio is evaluated as a continuous linear program (LP). Recommendation uses a complete common KuaiRec user-item matrix, so exact additive top-k regret is available rather than logged-only evaluation. Energy/Storage is evaluated both as an LP relaxation and as binary scheduling, with an appended constant feature in every formal decision-focused model. Appendix I records the full method-by-shift matrices and ablations.

Appendix J reports a separate mechanism run with a known data-generating process (DGP). It keeps the three application geometries fixed and generates costs with a known conditional mean. Its

conditional-mean, crossing, and transfer curves are not additional tests of the original observational hypothesis; they isolate the assumptions used by the population analysis.

For realized cost c_i, predicted report $\hat{c}_i$, shifted feasible set S_s, optimizer w_s, and optimum value z_s, we report

$$\widehat{R}_s = \frac{1}{n}\sum_{i=1}^{n}\left[c_i^\top w_s(\hat{c}_i) - z_s(c_i)\right]. \tag{87}$$

Stochastic methods use five seeds and seed-level 95% t intervals. Complete per-example rows are retained for bootstrap auditing, but examples are never pooled across seeds as independent replications. Deterministic baselines are reported as such. Root mean squared error (RMSE) is a predictive proxy, not a measurement of quotient recovery or conditional identification.

I Complete Real-Data Results and Ablations

This appendix reports every method and pre-specified shift used in the three real-data studies. Means are over seeds $0, \ldots, 4$. Error bars in the figures are seed-level 95% t intervals. Ordinary least squares (OLS) and ridge regression (Ridge), together with the deterministic linear KuaiRec fits produce identical rows when replayed under the five seed indices; we mark these by † and do not interpret the repeated rows as independent fits. Per-example bootstrap intervals, checkpoints, and the exact unrounded values remain in the accompanying CSV files.

Method definitions. All linear reports have the form $\hat{c}_B(x) = Bx$; Energy appends a constant feature, while Ridge uses an unpenalized intercept. With empirical average $\widehat{\mathbb{E}}$, the two SPO+ objectives are

$$\widehat{J}_{\mathrm{SPO+}}(B) = \widehat{\mathbb{E}}\,\ell_{\mathrm{SPO+}}(Bx, c), \qquad \widehat{J}_{\mathrm{hyb}}(B) = \widehat{J}_{\mathrm{SPO+}}(B) + \lambda\widehat{\mathbb{E}}\|Bx - c\|_2^2.$$

Here $\lambda = 0.1$ in the reported hybrid cells. Table 1 records the remaining training choices; T counts parameter-update steps, not epochs. T, batch size, and η_0 are listed as Portfolio/KuaiRec/Energy, respectively.

Table 1: Self-contained definitions of the empirical methods. The perturbed-oracle decision-gradient method (PG) uses Gaussian perturbations Z: σ is their report-space smoothing scale and η is the decayed step size $\eta_0/\sqrt{t+1}$ at update step t.

Method	Objective and model	Training and randomness
OLS / Ridge	Linear squared-report regression with ridge coefficient 10^{-3}; the intercept is unpenalized when used.	Deterministic.
Linear SPO+	$\widehat{J}_{\mathrm{SPO+}}(B)$, optimized by the SPO+ subgradient through the fixed oracle.	$T = 40/100/40$; batch $= 128$/full/32; $\eta_0 = 1/1/.01$. Stochastic except full-batch KuaiRec.
SPO+ + report	$\widehat{J}_{\mathrm{hyb}}(B)$; thus λ multiplies the squared report residual, not decision regret.	Same backbone and budget as Linear SPO+; $\lambda = .1$.
PG weak report	Linear report trained with the perturbed-oracle decision gradient $\widehat{\mathbb{E}}[(w^*(Bx + \sigma Z) - w^*(c))/\sigma]x^\top$. "Weak report" means that no squared report-fidelity term is included.	Stochastic: $n_{\mathrm{pert}} = 3/16/3$ Gaussian perturbations/example; $\sigma = .01/.01/10$, $\eta_0 = .001/.01/10^{-5}$, and the same T; batches $128/512/32$. Portfolio uses a low-learning-rate setting chosen for numerical stability after validation screening; KuaiRec uses validation regret as the primary selection criterion.
Neural SPO+	KuaiRec-only MLP report $x \mapsto \hat{c}(x)$ trained with the SPO+ chain-rule subgradient.	Stochastic Adam; hidden widths $(128, 64)$, 100 epochs, batch 128, learning rate 10^{-3}, weight decay 10^{-4}.

Table 2: Final PG weak-report configurations. Values are five-seed means ± sample SD; the reported settings were selected using validation screening and refit before the reported test evaluation.

Dataset	PG configuration $(\sigma, \eta_0, n_{\mathrm{pert}})$	Validation RMSE	Test RMSE	Test shifted regret
Portfolio	$(.01, .001, 3)$	0.945 ± 0.029	0.878 ± 0.027	0.01517 ± 0.00009
KuaiRec	$(.01, .01, 16)$	145.996 ± 1.274	150.60 ± 1.50	93.23 ± 1.11

I.1 PORTFOLIO

For PG, both screening grids fixed $\sigma = .01$, used five seeds, and crossed $\eta_0 \in \{1, .1, .01, .001\}$ with $n_{\text{pert}} \in \{3, 16\}$. Portfolio screened 40-step fits on a fixed 512-example validation prefix. Validation regret was nearly tied across cells, whereas RMSE and parameter norm increased sharply with η_0; we chose $(.001, 3)$ for numerical stability, not because it minimized validation regret. The reported Portfolio checkpoint was then refit for 40 steps and evaluated on the full validation split and test split. KuaiRec used 100-step fits and selected $(.01, 16)$ by mean validation top-k regret, with RMSE as a secondary criterion. The original $(1, 3)$ PG cell remains in both validation grids.

The chronological split is 70/15/15. Features contain only lagged returns, historical rolling volatility/momentum, and calendar variables. Table 3 and Figure 5 show that the pre-specified cap shift changes regret similarly for all methods although their report RMSE differs by up to roughly sevenfold.

Table 3: Complete Portfolio five-seed means ± sample SD. $R_{0.10}$ is training-task regret and $R_{0.12}$ is shifted regret. † denotes a deterministic fit.

Method	Report RMSE	$R_{0.10}$	$R_{0.12}$
OLS†	0.1324 ± 0.0000	0.013924 ± 0.000000	0.015078 ± 0.000000
SPO+	0.2036 ± 0.0146	0.013977 ± 0.000050	0.015160 ± 0.000058
SPO+ + report	0.2114 ± 0.0175	0.013955 ± 0.000088	0.015120 ± 0.000109
PG weak report	0.878 ± 0.027	0.013972 ± 0.000071	0.015170 ± 0.000086

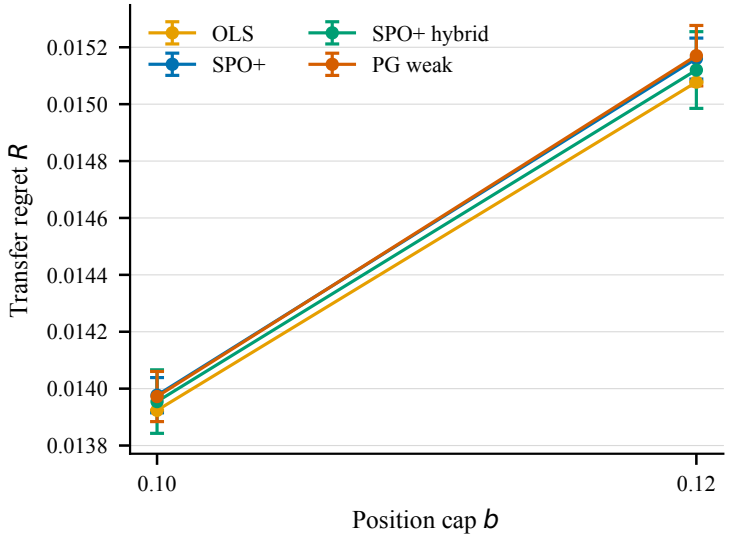


Figure 5: Portfolio regret under the training cap $b = 0.10$ and shifted cap $b = 0.12$. Error bars are seed-level 95% intervals; deterministic OLS has no seed variation.

The report-penalty ablation uses $\lambda \in \{0, 0.01, 0.1, 1\}$. At $b = 0.12$, $\lambda = 0.01$ gives 0.015164 ± 0.000067 regret (sample SD) and RMSE 0.207 ± 0.025. The $\lambda = 1$ endpoint has similar regret (0.015192 ± 0.000082) but RMSE $(3.14 \pm 2.74) \times 10^6$ and degenerate crossing slopes. Thus $\lambda \in \{0, 0.01, 0.1\}$ supplies a stable small-grid check. The $\lambda = 1$ result is numerically unstable under the fixed step-size schedule and does not establish a geometric failure boundary.

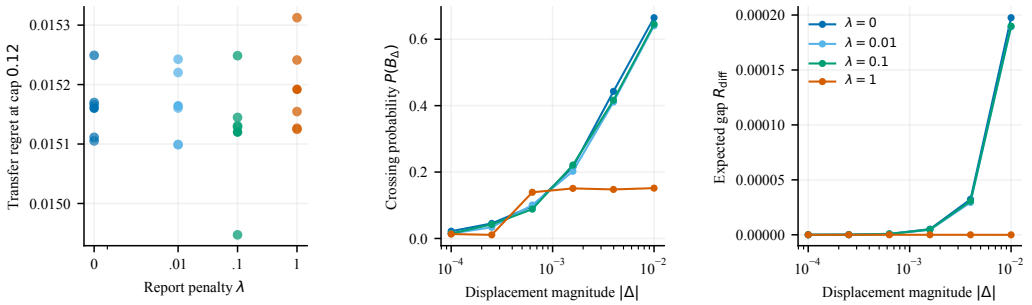


Figure 6: Portfolio report-penalty ablation. Left: five seed outcomes at the shifted cap. Middle and right: symmetrized crossing probability and unconditional expected transition gap versus displacement. The $\lambda = 1$ model has exploding RMSE; its crossing slopes are unreliable.



Figure 7: Portfolio local log-log slopes for raw and symmetrized perturbations. Mean, OLS, and SPO+ give the decomposition $P(B_\Delta)\mathbb{E}[G_\Delta \mid B_\Delta] = R_{\text{diff}}$ with an approximately quadratic product. PG is retained as a contrast. These finite-sample diagnostics do not verify symmetry or strict crossing of the population law.

I.2 Complete-matrix KuaiRec recommendation

Users are disjoint across train, validation, and test splits. Candidates with any missing feedback in the common matrix are removed rather than imputed. Costs are therefore observed for every candidate used in exact additive top-k regret. This is not a logged-bandit or off-policy evaluation (OPE) or position-counterfactual experiment.

Table 4: Complete KuaiRec five-seed means ± sample SD. The MLP report model is auxiliary; all other rows are pre-specified formal methods. † marks deterministic linear fits.

Method	Report RMSE	$R_{k=5}$	$R_{k=10}$	$R_{k=15}$
Ridge†	58.92 ± 0.00	53.96 ± 0.00	71.00 ± 0.00	82.68 ± 0.00
Linear SPO+†	71.73 ± 0.00	$\mathbf{50.63} \pm 0.00$	$\mathbf{66.25} \pm 0.00$	76.86 ± 0.00
SPO+ + report†	$\mathbf{58.04} \pm 0.00$	50.99 ± 0.00	66.71 ± 0.00	77.43 ± 0.00
PG weak report	150.60 ± 1.50	57.76 ± 0.56	78.77 ± 1.01	93.23 ± 1.11
Neural SPO+ (pre-specified)	100.53 ± 2.31	53.16 ± 0.20	70.72 ± 0.33	83.23 ± 0.25
MLP (MSE, auxiliary)	58.71 ± 0.02	52.03 ± 0.14	67.88 ± 0.28	78.67 ± 0.17

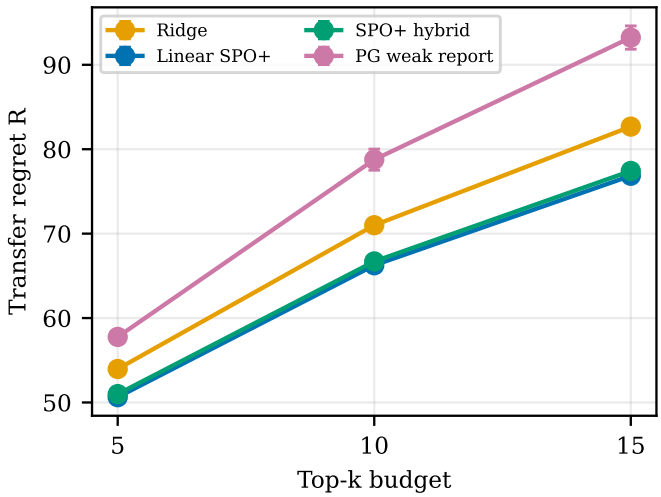


Figure 8: Complete KuaiRec regret across all pre-specified top-k budgets. Linear SPO+ has the lowest formal-method mean at each k; the hybrid is close, and validation-selected PG remains less competitive. Error bars are seed-level 95% intervals.

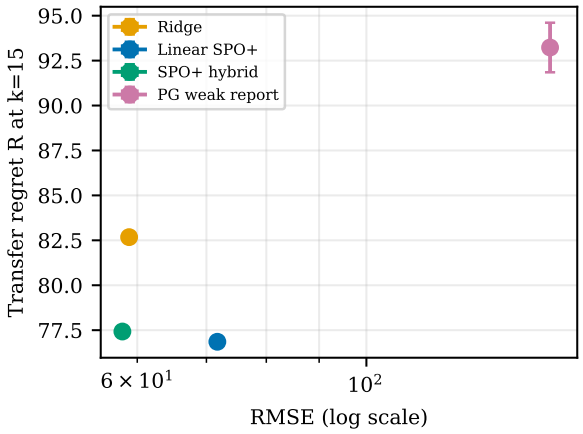


Figure 9: KuaiRec report RMSE versus $k = 15$ regret. Horizontal and vertical bars are seed-level 95% intervals. Linear SPO+ improves regret relative to Ridge, while the hybrid improves linear report RMSE; PG remains less competitive after validation-selected stabilization.

The Neural SPO+ ablation varies epochs in $\{50, 100\}$ and hidden widths in $\{64, (128, 64)\}$. Within this finite grid, the 50-epoch, width-64 model has the lowest observed regret, while the 50-epoch, $(128, 64)$ model has the lowest RMSE. This is an optimization/capacity sensitivity check, not a test of SC, quotient-report recovery, or neural-model-class consistency; it does not attribute either outcome to the architecture.

Table 5: Neural SPO+ capacity/epoch sensitivity check (five-seed means ± sample SD), not an identification or strict-crossing test.

Configuration	RMSE	R_5	R_{10}	R_{15}
50 epochs, 64	73.35 ± 0.18	$\mathbf{50.67} \pm 0.08$	$\mathbf{65.96} \pm 0.17$	$\mathbf{76.55} \pm 0.29$
50 epochs, 128×64	$\mathbf{72.32} \pm 0.27$	51.07 ± 0.04	66.58 ± 0.17	77.36 ± 0.23
100 epochs, 64	85.71 ± 0.73	51.99 ± 0.12	68.76 ± 0.29	80.36 ± 0.25
100 epochs, 128×64	100.53 ± 2.14	53.16 ± 0.20	70.72 ± 0.33	83.23 ± 0.25

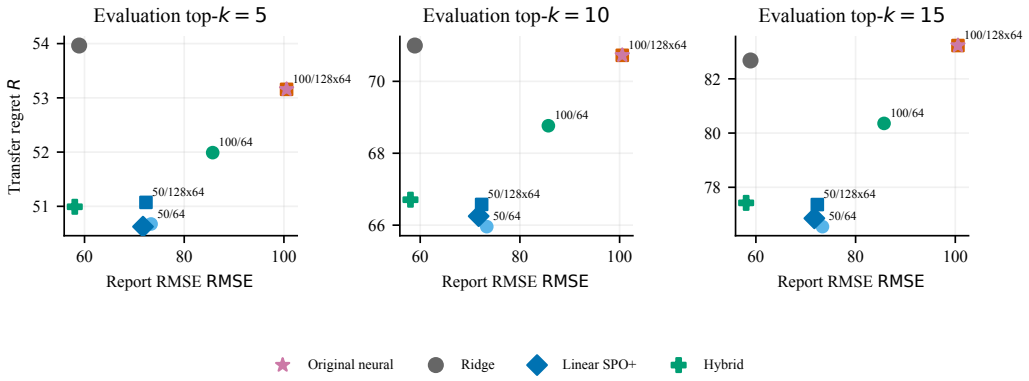


Figure 10: KuaiRec RMSE versus regret on the Pareto plane at $k \in \{5, 10, 15\}$. Labels give epochs/hidden widths. The originally pre-specified 100/128×64 network is shown separately from the ablation-selected 50-epoch models. This finite-grid sensitivity check does not measure quotient-report error or establish strict crossing.

Figure 11 reports exact top-k crossing diagnostics. Positive crossing and transition gaps occur for Ridge, Linear SPO+, and the hybrid at larger perturbations; the weak-report PG model is flat in the tested neighborhood. These curves diagnose the fitted reports under the finite candidate matrix and do not identify a population Bayes cost.

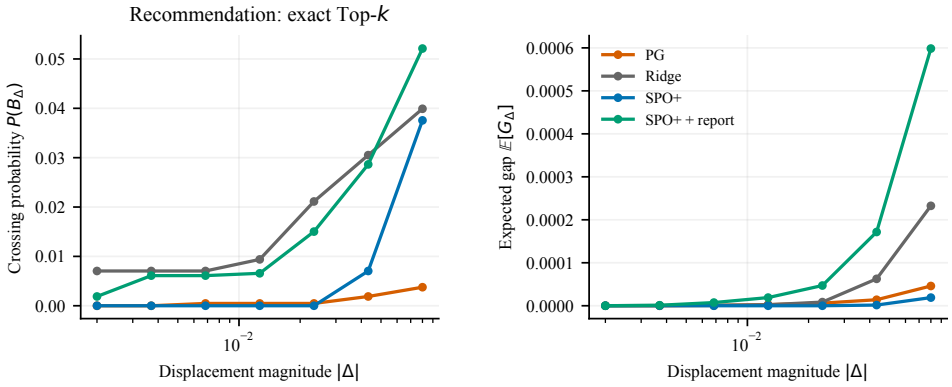


Figure 11: Complete-matrix KuaiRec top-k crossing probability and expected transition gap. The two panels are separated sufficiently for independent axis labels and share the displacement scale.

I.3 Energy/Storage

Energy uses complete 48-period days, a chronological 70/15/15 split, and only past or contemporaneously available features. All formal decision-focused fits standardize the decision features and

append a constant feature. The LP and binary optimizers share the same 48-period scheduling instance. Table 6 shows near-matching mean regrets on these paired evaluations; this does not establish integrality of the relaxation. In fact, the LP can return fractional schedules for this instance.

Table 6: Complete intercept-corrected Energy five-seed means ± sample SD. Values are in thousands. Ridge is deterministic (†).

		LP relaxation			Binary scheduling		
Method	RMSE	$s = 0$	$s = 0.5$	$s = 1$	$s = 0$	$s = 0.5$	$s = 1$
Ridge†	31.65 ± 0.00	71.10 ± 0.00	52.96 ± 0.00	51.10 ± 0.00	71.11 ± 0.00	52.97 ± 0.00	51.09 ± 0.00
SPO+	111.30 ± 7.56	73.14 ± 29.12	58.89 ± 26.36	**48.29** ± 12.23	73.12 ± 29.12	58.88 ± 26.36	**48.29** ± 12.22
SPO+ + report	107.02 ± 6.23	72.59 ± 23.63	56.76 ± 22.72	50.16 ± 10.97	72.59 ± 23.63	56.76 ± 22.72	50.16 ± 10.97
PG weak report	74.67 ± 0.00	573.07 ± 102.57	700.72 ± 151.36	783.80 ± 197.00	573.07 ± 102.57	700.72 ± 151.36	783.79 ± 197.00

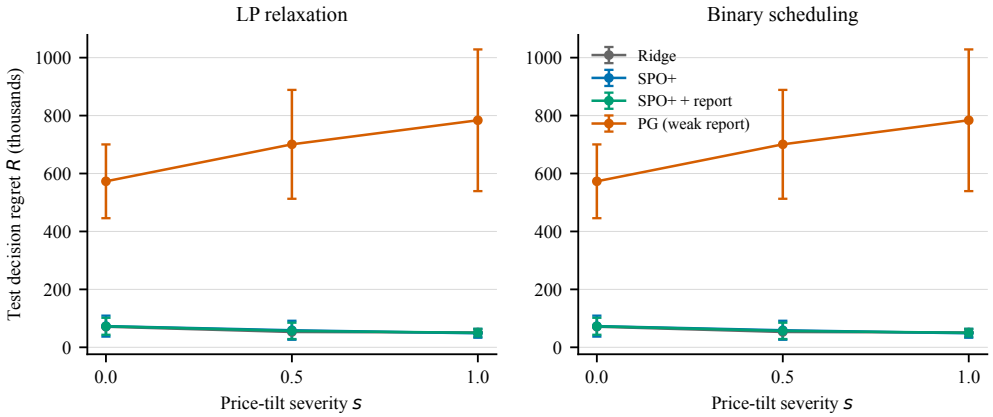


Figure 12: Intercept-corrected Energy regret for the LP relaxation and binary scheduler, shown in thousands. Ridge is deterministic; other bars are seed-level 95% intervals. SPO+ wins only at the largest pre-specified tilt, while PG remains unstable and substantially worse.

At $s = 1$, the report-penalty grid gives LP regrets 48.29, 49.73, 50.16, and 120.56 thousand for $\lambda = 0, .01, .1, 1$, respectively. The iteration grid gives 251.86, 48.29, and 42.43 thousand for $T = 10, 40, 80$. The best tested PG cell ($\sigma = 1$, $\eta = 10^{-5}$) still gives 759.87 thousand. Binary values are nearly identical. SPO+ improves with the tested training budget; the largest report penalty and tested PG settings are numerically unstable. The grid does not establish a universally optimal report penalty.

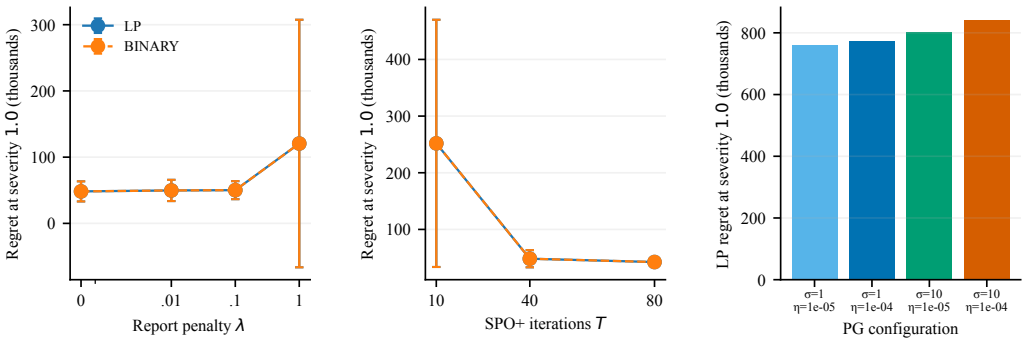


Figure 13: Intercept-corrected Energy ablations at $s = 1$, with regret in thousands. Left: report penalty; middle: SPO+ training budget; right: PG stability grid. LP and binary curves overlap almost exactly.

The method-averaged LP curves in Figure 14 increase with relative displacement for all fitted reports. Individual direction cells need not be monotone; local slopes vary by method and direction. This remains qualitative mechanism evidence, not a claim of the uniform boundary-mass condition or a universal quadratic rate.

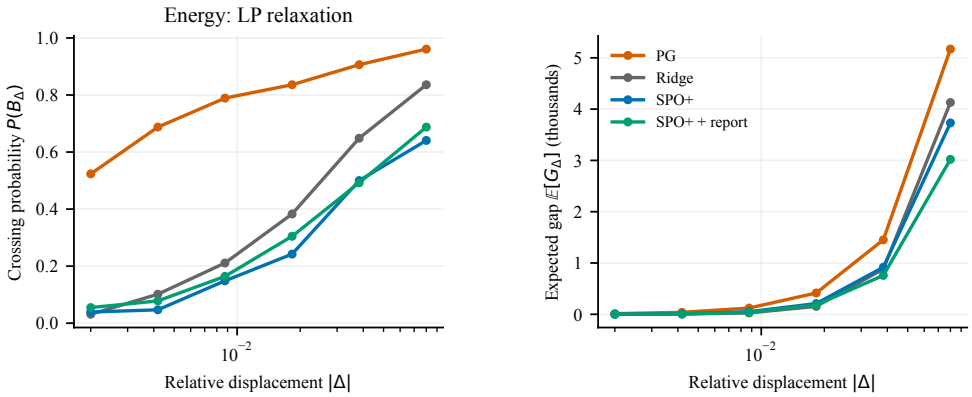


Figure 14: Energy LP crossing probability and expected transition gap (in thousands) versus relative displacement. Curves are finite-sample fitted-report diagnostics; no population symmetry or boundary-density assumption is inferred.

I.4 Cross-application synopsis

Figure 15 normalizes each method's regret by its own pre-specified baseline within the corresponding application. It summarizes sensitivity to the pre-specified shifts; it does not make objective scales comparable, supply a hypothesis test, or turn the three applications into independent replications of one estimand.

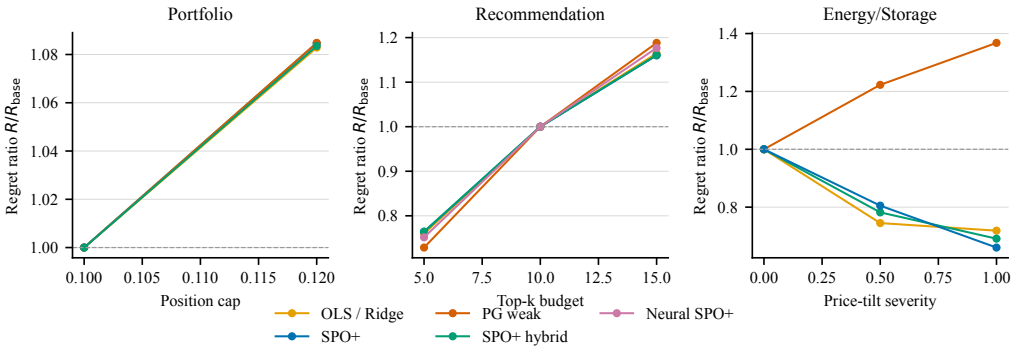


Figure 15: Cross-application shift profiles. Each curve reports regret relative to that method's application-specific pre-specified baseline.

I.5 Artifact provenance

The supplementary machine-readable materials contain the unrounded summary tables, per-example records, checkpoint metadata, ablation outputs, and the cross-application audit inventory. GPU Scheduling is not part of the final empirical scope because its replay protocol was not suitable for a formal claim.

J Experiments with Known Data-Generating Processes on Real Application Geometries

The observational experiments in Section 9 cannot reveal $\mu(x) = \mathbb{E}[C \mid X = x]$, conditional symmetry, or population strict crossing. We therefore add a semi-synthetic layer that preserves the feature rows and optimization geometry of each application while replacing the realized costs with a registered law whose conditional mean is known:

$$C = \mu(X) + \varepsilon.$$

The experiment is designed to isolate the mechanism in the theorems. It is not a reanalysis of the original observational outcomes.

Data-generating processes (DGPs) and evaluation. For Portfolio, KuaiRec, and Energy/Storage, the adapters reuse the existing data loaders and application-specific optimizers. The SC-ON law uses a feature-dependent mean and centered diagonal Gaussian noise, with each coordinate scaled by the corresponding training-cost dispersion and a fixed multiplier of 3. The SC-OFF control uses a deterministic conditional law at a locally stable reference cost. The conditional-mean diagnostic repeats outcomes at the same feature value and measures the RMSE of their sample mean relative to the supplied $\mu(x)$. It therefore measures Monte Carlo recovery under the known law; it does not fit or identify a regression function from the original data.

For each application, the crossing diagnostic evaluates the transition gap G_Δ on a fixed effective-direction and displacement grid. We report crossing mass, expected gap, and conditional mean recovery. The transfer diagnostic evaluates the known mean and controlled perturbations over the registered shift families: position caps for Portfolio, top-k values for KuaiRec, and price tilts for Energy/Storage. The shift family is finite by construction.

The current output is a mechanism and figure run rather than the preregistered high-budget confirmation. It used 8 feature rows and 16 Monte Carlo repetitions for Portfolio, 8 and 8 for KuaiRec, and 3 and 8 for Energy/Storage. The mean-recovery grid used $1, 4, 16, 64$ repeated outcomes. The run used three effective directions and five displacements for Portfolio and KuaiRec, and two directions and four displacements for Energy/Storage.

Results. The conditional-mean curves decrease with the number of repeated outcomes in all three applications under SC-ON; the deterministic SC-OFF control reaches numerical zero. The crossing curves show positive mass under SC-ON on part of the registered grid. Under SC-OFF, the smallest displacements remain locally flat in Portfolio and KuaiRec, while larger displacements leave the local region and produce crossing in some application geometries. In the current run, positive crossing occurred in $16/36$, $5/36$, and $10/20$ direction-scale rows for SC-ON in Portfolio, KuaiRec, and Energy/Storage, respectively; the corresponding SC-OFF counts were $2/36$, $0/36$, and $5/20$.

These results support a mechanism statement: the same application geometries can exhibit the predicted distinction between a full-support noisy law and a locally flat control when the conditional law is known. They do not support a claim that the original data satisfy strict crossing. The finite grid also does not establish the all-displacement quantifier in the population definition.

The transfer plot should be read together with the finite-shift restriction. It tests whether controlled report perturbations change regret under the specified application shifts. It does not quantify transfer over arbitrary objectives, constraints, feasible sets, or distributions. In particular, a known conditional mean in the semi-synthetic law removes one source of uncertainty; it does not supply the assumptions required for a universal transfer theorem.

K Deterministic Illustrations

Figure 18 visualizes exact one-dimensional formulae and the analytic probabilities in the two oracle-separation examples; it uses neither Monte Carlo estimates nor fitted rates. In the compact-support construction, the plotted risk is $\mathbb{E}[(c - 2\hat{c})_+]$ for $c \sim \mathrm{Unif}[1, 2]$, hence it is exactly zero for $\hat{c} \geq 1$. For the boundary-density family on $[-1, 1]$ with $p_\nu(c) = (\nu + 1)|c|^\nu/2$, the plotted excess

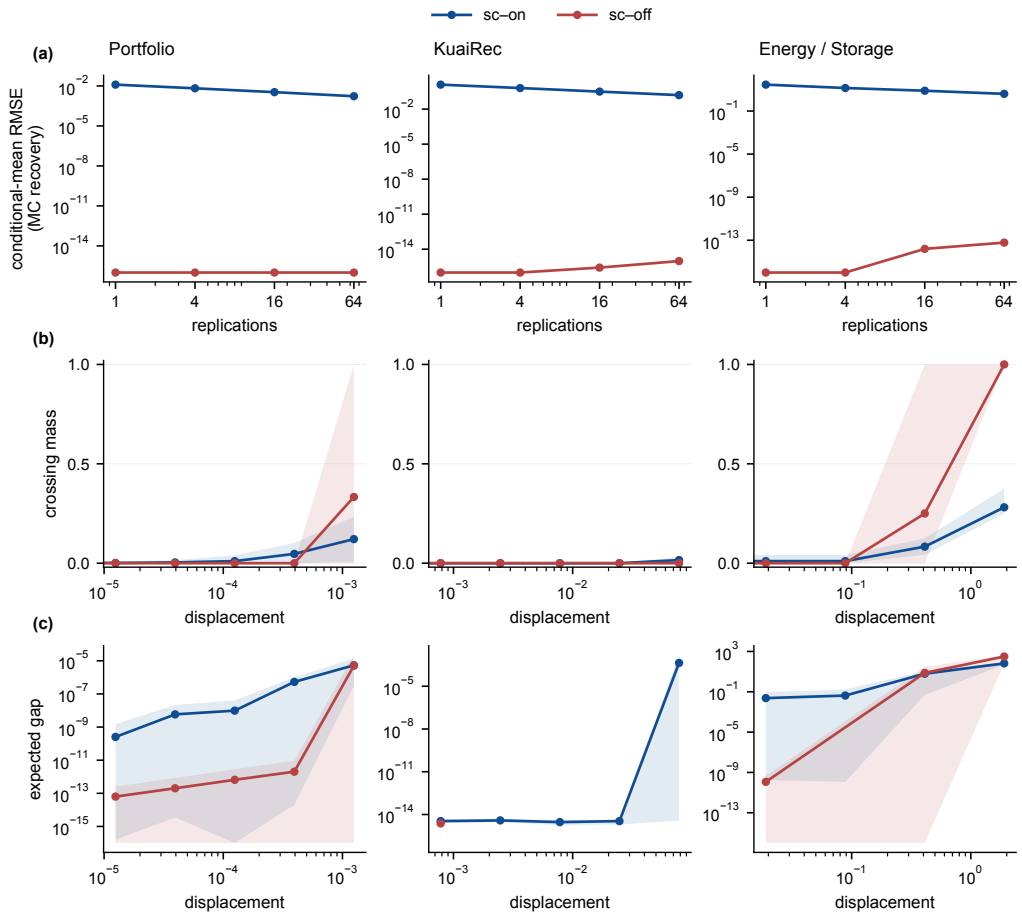


Figure 16: Known-DGP identification and crossing diagnostics on the real Portfolio, KuaiRec, and Energy/Storage geometries. The top row reports conditional-mean Monte Carlo recovery as the number of repeated outcomes increases. The middle row reports crossing mass over the tested effective directions and displacements. The bottom row reports expected transition gaps, with ribbons spanning the tested directions and signs. The curves are finite-grid diagnostics; they do not certify population strict crossing or conditional-mean identification from the original observational data.

is the closed form in Appendix L. The final panel records that C4 and C5 have positive selected-output disagreement but zero crossing probability and zero excess gap. The panels display these mechanisms but do not supply evidence for the theorems.

L Details for the Deterministic Illustrations

All curves in Figure 18 are evaluated from closed forms. The first panel uses the exact risk in Equation 54 for Example 4.1. For the centered one-dimensional model with $S = [-B, B]$ and $p_\nu(c) = (\nu+1)|c|^\nu/2$ on $[-1, 1]$, direct integration gives, for $0 < |t| < 1/2$,

$$R_{\mathrm{SPO}+}(t) - R_{\mathrm{SPO}+}(0) = \frac{B(2|t|)^{\nu+2}}{\nu+2}. \tag{88}$$

The third panel plots the exact probabilities in Examples 7.1 and 7.2: $\Pr(A_\Delta) = 1/2$ and $\Pr(B_\Delta) = 0$. The vector figure renders these closed forms and contains no random sampling.

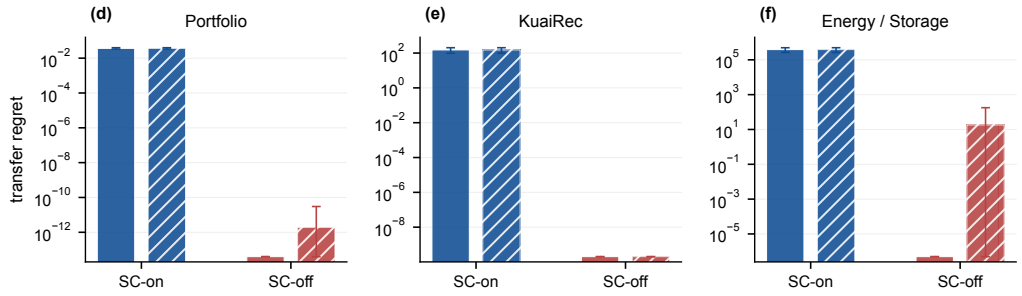


Figure 17: Transfer stress test under the registered finite shift families. Solid bars use the known conditional mean and hatched bars use controlled report perturbations. Error bars span the registered shifts and controlled perturbations within each application and regime. The logarithmic axis keeps the small SC-OFF regrets visible. The plot is finite-family evidence and does not establish a universal transfer guarantee.

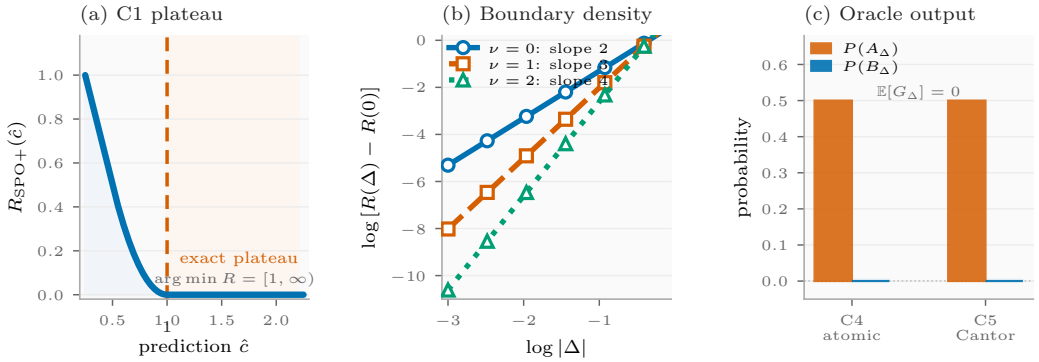


Figure 18: Closed-form illustrations: the C1 plateau, boundary-density slopes $2 + \nu$, and oracle disagreement without crossing. These are analytic constructions, not empirical evidence.

M FENCHEL TRANSITION EXTENSION (SUPPORTING)

The support-function identity has a precise Fenchel-transition extension, but this appendix is a supporting extension, not a general decision-focused surrogate identification theorem. The normal-fan geometry, decision consistency, and polyhedral crossing interpretation remain tied to the SPO+/support-function structure. In this appendix only, costs and reports lie in $\mathbb{R}^m$. Let Ω be proper, lower-semicontinuous, and convex, let $\Phi := \Omega^*$ be finite on $\mathbb{R}^m$, and fix a measurable selection $s(q) \in \partial\Phi(q)$. Define

$$\mathcal{W}_\Omega(c) := \partial\Phi(-c), \qquad w_\Omega(c) := s(-c), \tag{89}$$

and consider the Fenchel transition loss

$$\ell_\Phi(\theta, c) := \Phi(c - 2\theta) + 2\theta^\top w_\Omega(c) + \Phi(-c) + a(c), \tag{90}$$

where $a(c)$ is integrable and does not depend on θ.

Write

$$R_\Phi(\theta) := \mathbb{E}[\ell_\Phi(\theta, c)] \tag{91}$$

for the corresponding population risk.

Theorem M.1 (Fenchel transition identity)**.** *Suppose $c \overset{d}{=} 2\bar{c} - c$ and, for every $\theta \in \mathbb{R}^m$, each of $\Phi(c - 2\theta)$, $\theta^\top s(-c)$, $\Phi(-c)$, and $a(c)$ is integrable. Then*

$$R_\Phi(\bar{c} + \Delta) - R_\Phi(\bar{c}) = \mathbb{E}\left[D_\Phi(-c - 2\Delta, -c; s(-c))\right], \tag{92}$$

where $D_\Phi(x, y; s(y)) := \Phi(x) - \Phi(y) - s(y)^\top(x - y)$.

Proof of Theorem M.1. Subtract equation 90 at $\bar{c}$ from its value at $\bar{c} + \Delta$. The offset $a(c)$ cancels and central symmetry gives

$$R_\Phi(\bar{c} + \Delta) - R_\Phi(\bar{c}) = \mathbb{E}\big[\Phi(-c - 2\Delta) - \Phi(-c) + 2\Delta^\top s(-c)\big] \tag{93}$$

$$= \mathbb{E}\big[D_\Phi(-c - 2\Delta, -c; s(-c))\big]. \tag{94}$$

The last equality is the definition of the selected Bregman divergence. □

Because the selected Bregman divergence vanishes exactly when $s(y) \in \partial\Phi(x)$, the target is uniquely identified precisely when this selected divergence is positive with positive probability for every nonzero displacement.

Remark M.2 (Scope of the extension). For $\Omega = I_S$, $\Phi = \xi_S$, and the support-function selection must be aligned as $s(q) := w^*(-q)$. Then $w_\Omega(c) = w^*(c)$ and $D_\Phi(-c - 2\Delta, -c; w^*(c)) = G_\Delta(c)$, recovering Theorem 5.2. Quadratic potentials instead identify by pointwise curvature, while entropy identifies costs only modulo additive constants. No generic SPO task-consistency statement follows from Theorem M.1.

Theorem M.3 (Fenchel identification by strict Bregman crossing)**.** *Under the assumptions of Theorem M.1, $\bar{c}$ is the unique minimizer of R_Φ if and only if, for every $\Delta \neq 0$,*

$$\Pr\{w_\Omega(c) \notin \mathcal{W}_\Omega(c + 2\Delta)\} > 0. \tag{95}$$

Proof. For a finite convex Φ and $s(y) \in \partial\Phi(y)$,

$$D_\Phi(x, y; s(y)) = 0 \quad \Longleftrightarrow \quad s(y) \in \partial\Phi(x). \tag{96}$$

Apply equation 96 to the nonnegative expectation in equation 92. The same strict-positive-expectation argument as in Theorem 5.2 proves the result. □

Theorem M.4 (Fenchel absolute-continuity equivalence)**.** *Let $\mathcal{T}_\Phi := \{q : \partial\Phi(q)$ is non-singleton$\}$. If $P_c \ll \lambda_m$, then for every fixed Δ,*

$$\Pr\left(\{w_\Omega(c) \neq w_\Omega(c + 2\Delta)\} \triangle \{w_\Omega(c) \notin \mathcal{W}_\Omega(c + 2\Delta)\}\right) = 0. \tag{97}$$

Proof. A finite convex function is differentiable Lebesgue-almost everywhere, so $\lambda_m(\mathcal{T}_\Phi) = 0$. Outside the translated tie locus, the new subdifferential is a singleton and output disagreement forces crossing. The reverse event inclusion is always true. □

N MECHANISM EXPERIMENTS

The four mechanism experiments reproduce the predicted geometry on synthetic laws with closed-form or numerically exact objectives. Their curves are deterministic evaluations of these constructions.

Sharp exponent and absence of polynomial modulus (Exps. 1 and 2). For $S = [-1, 1]$, $\bar{c} = 0$, and $p_\nu(c) = (\nu + 1)|c|^\nu/2$ on $[-1, 1]$, Theorem E.2 gives the exact closed form $R_{\mathrm{SPO+}}(t) - R_{\mathrm{SPO+}}(0) = B(2|t|)^{\nu+2}/(\nu + 2)$. Figure 19(a) plots this exact curve for $\nu \in \{0, 0.5, 1, 2\}$; the log-log slopes are exactly $2 + \nu$. For the C^∞ flat density $p(c) \propto e^{-1/|c|}$, strict crossing holds for every $t > 0$, yet the excess risk falls below t^2, t^3, t^4; the local log-log slope grows without bound as $t \downarrow 0$ (Figure 19(b)). SC alone carries no polynomial modulus.

Oracle disagreement versus optimal-face crossing (Exp. 3). Figure 20 shows the atomic tie example (Example 7.1) and the Cantor example (Example 7.2) under increasing noise $c_\varepsilon = c + \varepsilon Z$. At $\varepsilon = 0$, $\Pr(A_\Delta) = 1/2$ while $\Pr(B_\Delta) = 0$ and $\mathbb{E}[G_\Delta] = 0$: the selected oracle moves while the optimal face is never departed. For $\varepsilon > 0$ the law is absolutely continuous and $\Pr(A_\Delta) = \Pr(B_\Delta)$ (Theorem 7.3); both grow with ε. The gap-based events are insensitive to the tolerance $\mathrm{tol} \in \{0, 10^{-8}, 10^{-6}\}$.

Normal-fan mass yields quadratic growth (Exp. 4). For the box $S = [-1, 1]^2$ with $\bar{c} = (2, 2)$ and $\mathrm{Unif}(0, 4)^2$ costs, the mechanism $P(B_{tu}) \asymp t$, $\mathbb{E}[G_{tu} \mid B_{tu}] \asymp t$, $R_{\mathrm{SPO+}}(\bar{c} + tu) - R_{\mathrm{SPO+}}(\bar{c}) \asymp t^2$ holds uniformly over the tested directions (Figure 21(a), slopes $1/1/2$); inward directions have no crossing and flat risk. Removing the boundary patch near one facet (symmetrically about $\bar{c}$) creates an exact dead zone $P(B_{tu}) = 0$ and flat risk for t below the gap, with growth resuming later, while the intact direction is unaffected (Figure 21(b)). Lowering the density floor keeps the slopes and shrinks the constants. The same mechanism holds for the non-symmetric triangle $\mathrm{conv}\{(0, 0), (2, 0), (0, 1)\}$ (Figure 21(c)), computed by vertex enumeration without solver-output comparisons.

Controlled neural quotient-report diagnostic (Exp. 5). This experiment uses the box $S = [-1, 1]^2$ and a known conditional mean $\mu(x) = 2 + 0.25x$ for $x \sim \mathrm{Unif}([-1, 1]^2)$. Conditional costs are $C = \mu(X) + \varepsilon$, where the coordinates of ε are independent uniform draws on $[-a, a]$. We use $a = 3$ (SC-on), $a = 2.25$ (low-crossing), and $a = 0.25$ (SC-off). Each law remains conditionally centrally symmetric. In the SC-off design, C lies in the positive orthant, so the positive displacement $(0.5, 0)$ is a flat direction of the selected box decision rule. For this same displacement, held-out crossing masses are 0.152, 0.056, and 0, respectively. These finite-sample diagnostics illustrate the designed regimes; they do not estimate a real-data population assumption.

We train Linear SPO+, MLP-MSE, MLP-SPO+, and MLP-SPO+ with a squared-report term ($\lambda = 0.1$) on five independent seeds. Table 7 reports held-out Monte Carlo surrogate excess relative to the known μ, decision regret, quotient-report error $\mathbb{E}\|P_L(\hat{c}(X) - \mu(X))\|_2$, and report variance across fitted seed predictions. Excess, regret, and Q-error are five-seed means with sample SD. Small signed excess estimates with magnitude below 5×10^{-9} are displayed as zero at table precision. In SC-off, all four methods obtain zero observed decision regret while their quotient errors and report variances differ. This is controlled evidence about equality geometry, not a finite-sample neural-convergence result.

Table 7: Controlled neural identification experiment. "Excess", "Regret", and "Q-error" are five-seed means ± sample SD; "Excess" is held-out SPO+ excess relative to the known conditional mean and "Q-error" is quotient-report error. "Var" is mean coordinatewise fitted report variance across seeds. The SC-off rows have equal observed decision regret but different fitted reports.

Regime	Method	Excess	Regret	Q-error	Var
SC-on	Linear SPO+	0.0146 ± 0.0064	0.3326 ± 0.0000	0.1217 ± 0.0330	0.0095
	MLP-MSE	0.0315 ± 0.0159	0.3326 ± 0.0000	0.1863 ± 0.0437	0.0261
	MLP-SPO+	0.0400 ± 0.0132	0.3326 ± 0.0000	0.2046 ± 0.0422	0.0290
	MLP-SPO+ + MSE	0.0388 ± 0.0119	0.3326 ± 0.0000	0.2039 ± 0.0446	0.0295
Low mass	Linear SPO+	0.0040 ± 0.0018	0.0360 ± 0.0000	0.0671 ± 0.0194	0.0026
	MLP-MSE	0.0205 ± 0.0051	0.0360 ± 0.0000	0.1481 ± 0.0262	0.0163
	MLP-SPO+	0.0111 ± 0.0035	0.0360 ± 0.0000	0.1164 ± 0.0323	0.0080
	MLP-SPO+ + MSE	0.0111 ± 0.0034	0.0360 ± 0.0000	0.1135 ± 0.0249	0.0067
SC-off	Linear SPO+	0.0000 ± 0.0000	0.0000 ± 0.0000	1.2192 ± 0.0027	0.0001
	MLP-MSE	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0305 ± 0.0118	0.0007
	MLP-SPO+	0.0000 ± 0.0000	0.0000 ± 0.0000	0.3384 ± 0.1645	0.0601
	MLP-SPO+ + MSE	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0742 ± 0.0271	0.0019

Table 8: Fitted local log-log slopes (theory in parentheses). BOX and TRI are the polytopes of Exp. 4; slopes are averaged over the crossing directions; the flat-density row reports a finite-grid slope from Exp. 2 and its asymptotic behavior.

Case	$P(B_{tu})$	$\mathbb{E}[G_{tu} \mid B_{tu}]$	Rdiff
Box $S = [-1, 1]^2$ (5 directions)	1.00 (1)	1.00 (1)	2.01 (2)
TRI (4 directions)	0.98 to 1.01 (1)	1.00 (1)	1.93 to 2.07 (2)
Lower density floor	0.99 (1)	0.99 (1)	2.00 (2)
Flat density $e^{-1/\lvert c\rvert}$ (Exp. 2)	finite-grid slope 175.30 at $t \approx 0.002$; diverges as $t \downarrow 0$		

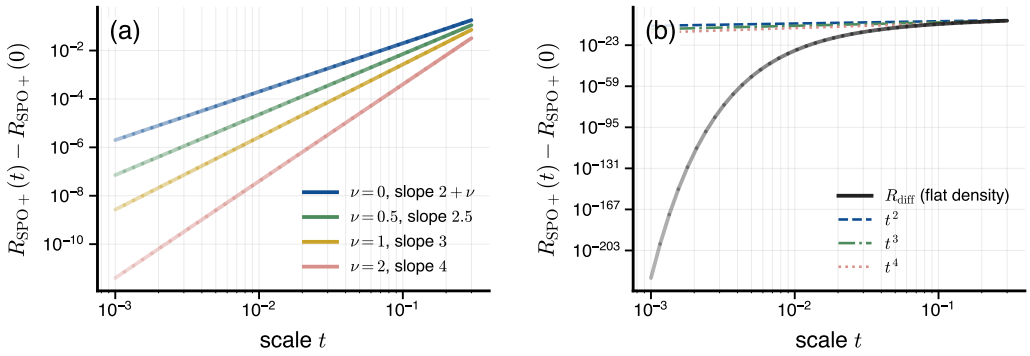


Figure 19: (a) Exact excess risk of the one-dimensional model for $\nu \in \{0, 0.5, 1, 2\}$ (log-log; slopes $2 + \nu$). (b) A flat C^∞ density satisfies strict crossing but yields no polynomial modulus: the excess risk falls below t^2, t^3, t^4.

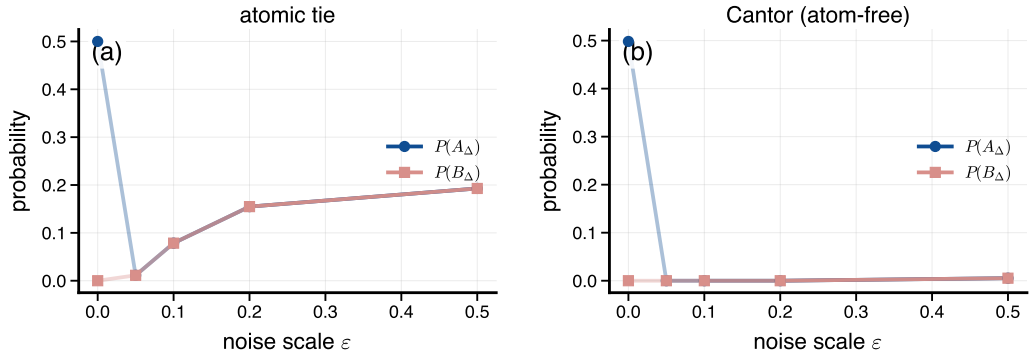


Figure 20: Selected-oracle disagreement $\Pr(A_\Delta)$ versus optimal-face departure $\Pr(B_\Delta)$ for the atomic and Cantor examples under noise $c_\varepsilon = c + \varepsilon Z$. At $\varepsilon = 0$, $\Pr(A_\Delta) = 1/2$ and $\Pr(B_\Delta) = 0$; for $\varepsilon > 0$ they coincide (Theorem 7.3).

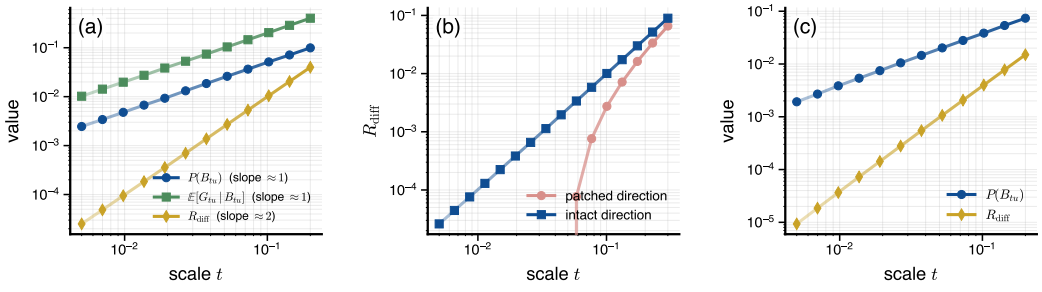


Figure 21: (a) Crossing mass, conditional gap, and excess risk versus scale for the box polytope (slopes $1/1/2$). (b) Removing the boundary patch near one facet kills growth below the gap (dead zone), while the intact direction is unchanged. (c) The same mechanism on a non-symmetric triangle.