

Highlights

Validation-Calibrated Risk-Function Ambiguity for Minimax-Regret Portfolio Optimization

Zheqi Fan

- Decision-relevant risk discrepancy ignores common additive level shifts.
- Validation calibrates ambiguity over convex risk-engine aggregations.
- Finite-sample audit validation extends to the full portfolio domain.
- Minimax regret admits Expected Shortfall certificates and model-class rejection.

Validation-Calibrated Risk-Function Ambiguity for Minimax-Regret Portfolio Optimization

Zheqi Fan^{a,b,*}

^a*Division of Emerging Interdisciplinary Areas, Hong Kong University of Science and Technology, Hong Kong SAR*

^b*Thrust of Financial Technology, Hong Kong University of Science and Technology (Guangzhou), China*

Abstract

Level-based Value-at-Risk and Expected Shortfall checks on benchmark portfolios need not control portfolio choice: common additive risk shifts leave rankings unchanged, whereas relative-risk distortions can alter the optimizer. We develop a validation-calibrated minimax-regret framework for Expected Shortfall portfolio optimization. Using a decision-relevant discrepancy that is invariant to common additive shifts, fixed candidate risk engines, an independent validation sample, and a pre-specified audit library, we form a polyhedral ambiguity set over convex aggregations of implemented risk functions. Finite-sample calibration under bounded returns, together with a Lipschitz covering argument from audits to the full feasible domain, yields a high-probability true Expected Shortfall regret certificate that separates model discrepancy, validation error, audit-coverage error, and optimization error. The procedure also permits model-class rejection and admits an exact linear-programming reformulation for moderate ambiguity systems.

Keywords: Expected Shortfall, portfolio optimization, model validation, minimax regret, finite-sample guarantee

1. Introduction

Portfolio optimization converts estimated return distributions into investment decisions. This conversion can amplify statistical and modeling errors because a small error in an estimated risk function may induce a substantial change in the optimizer. The sensitivity of portfolio choice to estimated inputs has motivated a large literature on robust portfolio selection; see, among others, Goldfarb and Iyengar (2003), Zhu and Fukushima (2009), and Huang et al. (2010).

Expected Shortfall (ES), also referred to as Conditional Value-at-Risk (CVaR) under the convention adopted in this paper, is particularly suitable for scenario-based portfolio optimization. It is a coherent tail-risk measure and admits a convex optimization representation that becomes linear under a finite scenario distribution (Artzner et al., 1999; Acerbi and Tasche, 2002; Rockafellar

*Corresponding author.

Email address: zheqi.fan@connect.ust.hk (Zheqi Fan)

URL: <https://orcid.org/0000-0003-4752-9019> (Zheqi Fan)

and Uryasev, 2002). Worst-case CVaR portfolio models exploit this tractability to protect decisions against distributional uncertainty (Zhu and Fukushima, 2009). Relative robust CVaR formulations further evaluate a portfolio relative to model-specific or distribution-specific benchmark portfolios (Huang et al., 2010). Minimax and minimax-regret criteria provide a complementary decision-theoretic language for robust combinatorial and continuous optimization (Aissi et al., 2009).

A separate line of research constructs uncertainty sets from statistical evidence and derives out-of-sample guarantees for the resulting decisions (Delage and Ye, 2010; Bertsimas et al., 2018; Mohajerin Esfahani and Kuhn, 2018; Van Parys et al., 2021). These approaches establish an important principle: an uncertainty set should be connected to observable data rather than selected solely as a deterministic robustness parameter. Parallel work on Expected Shortfall estimation provides asymptotics and concentration bounds that can support finite-sample validation radii (Chun et al., 2012; Prashanth et al., 2020).

Nevertheless, a gap remains between portfolio model validation and portfolio optimization. In practice, candidate tail-risk engines are often compared on a limited collection of portfolios or through forecast scoring rules. Although the pair consisting of Value-at-Risk and Expected Shortfall is jointly elicitable under appropriate conditions (Fissler and Ziegel, 2016), forecast accuracy does not automatically provide a uniform bound on the downstream portfolio regret. More generally, decision-focused learning emphasizes that the quality of a prediction should be evaluated in relation to the optimization problem in which the prediction is used (Elmachtoub and Grigas, 2022).

This paper studies the following complementary interface:

$$\textit{validation on finitely many audit portfolios} \implies \textit{regret control on a continuous portfolio domain}.$$

The key observation is that an error in a reported risk level need not be decision-relevant. Suppose two risk functions differ by a common constant on the entire feasible set. Their absolute risk forecasts are different, but they rank all portfolios identically and generate identical optimization regret. This motivates validation of *relative risk differences* across portfolios rather than validation of risk levels alone.

Our main contributions are as follows.

- (i) We introduce a decision-relevant discrepancy between portfolio risk functions that measures distortions in pairwise risk differences, is invariant to common additive level shifts, and transfers directly to portfolio regret.
- (ii) Given fixed candidate ES risk engines, an independent validation sample, and a pre-specified audit library, we construct a polyhedral ambiguity set over convex aggregations of implemented risk functions. The construction provides finite-sample containment under bounded independent validation observations and permits rejection of the entire candidate model class when no aggregation passes validation.

- (iii) We extend finite-audit validation to the full feasible portfolio domain through Lipschitz continuity and audit coverage. This yields an oracle-consistent minimax-regret portfolio and a high-probability certificate for true ES regret that separates accepted model discrepancy, validation estimation error, audit-coverage error, and optimization error.
- (iv) We clarify why aggregate-risk regret differs from weighted engine-specific regret through an oracle-incompatibility decomposition. For fixed finite scenario representations and moderate ambiguity systems, we also provide an exact linear-programming reformulation of the resulting minimax-regret problem.

Our framework is related to relative robust CVaR optimization, which evaluates a portfolio relative to model-specific or distribution-specific benchmark portfolios. However, the source and interpretation of ambiguity are different. Relative robust CVaR models typically begin with a prescribed uncertainty set of return distributions or scenarios (Huang et al., 2010). In contrast, we begin with a finite set of already trained risk engines and use independent validation evidence to determine which convex aggregations of their implemented risk functions are admissible. The resulting ambiguity set is therefore a statistically calibrated set of risk functions rather than a postulated family of return distributions. Moreover, the finite audit-to-domain analysis yields an explicit high-probability certificate for true Expected Shortfall regret over the full feasible portfolio set.

The paper likewise differs from worst-case CVaR optimization, which hedges against an a priori worst distribution in a given uncertainty set, and from distributionally robust optimization, which typically calibrates ambiguity over return laws rather than over validated risk-function aggregations. It also differs from decision-focused learning: rather than training a predictor end-to-end for an optimization loss, we treat candidate engines as fixed and ask which aggregations remain admissible under independent validation. We do not claim to replace these approaches. The contribution is complementary: connecting model validation to downstream portfolio regret.

The remainder of the paper is organized as follows. Section 2 defines the portfolio environment and candidate risk engines. Section 3 introduces the decision discrepancy. Section 4 constructs the validation-calibrated ambiguity set. Section 5 extends finite-audit validation to the full feasible portfolio domain. Section 6 develops the oracle-consistent minimax-regret portfolio and its regret certificate. Section 7 collects supporting analysis and computation: sensitivity to statistical expansion, an exact linear-programming reformulation, and a counterexample to vertex-only validation. Section 8 reports numerical experiments. Section 9 discusses extensions and limitations and concludes.

2. Portfolio environment and candidate risk engines

2.1. Feasible portfolio set

Let n denote the number of investable assets or portfolio sleeves. Portfolio weights belong to a nonempty compact polytope

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : x \geq 0, \mathbf{1}^\top x = 1, Gx \leq g \right\}, \quad (2.1)$$

where $Gx \leq g$ may encode upper bounds, group exposures, tracking constraints, and a linearized turnover limit.

For example,

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : x \geq 0, \mathbf{1}^\top x = 1, x_i \leq \bar{x}_i, \left\| x - x^{\text{old}} \right\|_1 \leq \tau \right\}. \quad (2.2)$$

The ℓ_1 -turnover constraint can be represented through standard auxiliary variables $s \geq 0$ satisfying

$$-s \leq x - x^{\text{old}} \leq s, \quad \mathbf{1}^\top s \leq \tau.$$

Hence, the feasible region remains a compact polytope.

When turnover is included, the current portfolio x^{old} is treated as known at the decision date. We impose turnover through the common feasible set so that every candidate decision and every model-specific oracle is evaluated under identical implementation constraints.

2.2. Holding-period loss

Let $R \in \mathbb{R}^n$ be the vector of simple asset returns over one deployment period. The portfolio is selected at the beginning of the period and held without rebalancing during that period. Its loss is

$$L(x, R) = -x^\top R. \quad (2.3)$$

If a deployment period contains H daily observations, the asset-level holding-period return is

$$R_i = \prod_{h=1}^H (1 + r_{i,h}) - 1. \quad (2.4)$$

This convention preserves linearity of the holding-period loss in the initial portfolio weights.

2.3. True Expected Shortfall

Let P_0 denote the unknown deployment distribution of R , and fix a confidence level $\alpha \in (0, 1)$. Write

$$p = 1 - \alpha.$$

For a loss random variable L , define

$$\text{ES}_\alpha(L) = \min_{\zeta \in \mathbb{R}} \left\{ \zeta + \frac{1}{p} \mathbb{E} [[L - \zeta]_+] \right\}. \quad (2.5)$$

This definition remains valid for discrete distributions and distributions with probability mass at the quantile threshold (Rockafellar and Uryasev, 2002; Acerbi and Tasche, 2002).

The true portfolio risk function is

$$f_0(x) = \text{ES}_\alpha^{P_0}(-x^\top R). \quad (2.6)$$

2.4. Fixed candidate risk engines

Suppose M risk engines have already been trained using data that are separate from the validation sample. Engine m produces scenarios

$$\{(R_{ms}, \pi_{ms})\}_{s=1}^{S_m}, \quad \pi_{ms} \geq 0, \quad \sum_{s=1}^{S_m} \pi_{ms} = 1.$$

Its implemented scenario risk function is

$$f_m(x) = \min_{\zeta_m \in \mathbb{R}} \left\{ \zeta_m + \frac{1}{p} \sum_{s=1}^{S_m} \pi_{ms} \left[-x^\top R_{ms} - \zeta_m \right]_+ \right\}. \quad (2.7)$$

Let $\mathcal{G}$ denote the σ -field generated by the training-stage data, the generated scenario sets, the fixed candidate risk engines, the pre-specified audit library $\mathcal{B}$, the tolerance δ , and the confidence level β . The analysis conditions on $\mathcal{G}$. Consequently, $f_1, \dots, f_M$ are treated as fixed functions during model validation and portfolio optimization. All probability statements below are understood conditionally on $\mathcal{G}$.

Remark 2.1. *The object in (2.7) is the risk function actually used by the optimization system. It is not identified with a hypothetical infinite-simulation limit. This distinction prevents training uncertainty, Monte Carlo approximation error, and validation uncertainty from being combined without an explicit decomposition.*

2.5. Assumptions

Assumption 2.1 (Independent validation sample). *Conditional on $\mathcal{G}$, the validation observations*

$$R_1^v, \dots, R_T^v \stackrel{\text{iid}}{\sim} P_0$$

are independent of the training-stage data used to construct the candidate risk engines and the audit library.

Assumption 2.2 (Uniform boundedness). *There exists a known constant $\bar{R} > 0$ such that*

$$\|R\|_\infty \leq \bar{R} \quad P_0\text{-a.s.}, \quad \|R_{ms}\|_\infty \leq \bar{R}, \quad m = 1, \dots, M, \quad s = 1, \dots, S_m. \quad (2.8)$$

Assumption 2.3 (Pre-specified validation design). *The audit library $\mathcal{B}$, validation tolerance δ , confidence level β , and candidate risk-engine library are specified before observing the validation sample. The audit library may depend on the feasible set, economic restrictions, and training-stage information, but not on validation outcomes.*

Assumption 2.1 provides a transparent starting point for finite-sample analysis. It is not satisfied merely because holding-period returns are non-overlapping. Serial dependence and distribution drift are discussed in Section 9.

3. Decision-relevant discrepancy between risk functions

3.1. Definition

For two continuous risk functions $f, g : \mathcal{X} \rightarrow \mathbb{R}$, define

$$d_{\mathcal{X}}(f, g) = \sup_{x, y \in \mathcal{X}} |[f(x) - f(y)] - [g(x) - g(y)]|. \quad (3.1)$$

Let $h = f - g$. Compactness of $\mathcal{X}$ gives

$$d_{\mathcal{X}}(f, g) = \max_{x \in \mathcal{X}} h(x) - \min_{x \in \mathcal{X}} h(x). \quad (3.2)$$

Thus, $d_{\mathcal{X}}(f, g)$ is the oscillation of the model error over the decision domain.

Proposition 3.1 (Equivalent centered representation). *For continuous $f, g : \mathcal{X} \rightarrow \mathbb{R}$,*

$$d_{\mathcal{X}}(f, g) = 2 \min_{c \in \mathbb{R}} \sup_{x \in \mathcal{X}} |f(x) - g(x) - c|. \quad (3.3)$$

Proof. Let

$$a = \min_{x \in \mathcal{X}} [f(x) - g(x)], \quad b = \max_{x \in \mathcal{X}} [f(x) - g(x)].$$

The smallest uniform error after subtracting a constant is obtained at $c = (a + b)/2$, and its value is $(b - a)/2$. Equation (3.2) completes the proof. $\square$

Equivalently, $d_{\mathcal{X}}$ is the quotient sup-norm distance obtained after identifying risk functions that differ only by an additive constant over the feasible domain.

Remark 3.1. *The discrepancy removes a common additive error but not a multiplicative error. A common additive shift does not alter portfolio ranking or regret. A multiplicative distortion changes the economic scale of regret and is therefore retained.*

3.2. Basic properties

The function $d_{\mathcal{X}}$ is nonnegative and symmetric. The triangle inequality follows immediately by applying the triangle inequality to the pairwise-difference representation in (3.1). Moreover,

$$d_{\mathcal{X}}(f, g) = 0 \iff f - g \text{ is constant on } \mathcal{X}. \quad (3.4)$$

It is therefore a semimetric on the original function space and a metric on equivalence classes modulo additive constants.

3.3. Regret transfer

For any continuous risk function f , define

$$v_f = \min_{y \in \mathcal{X}} f(y) \quad (3.5)$$

and

$$r_f(x) = f(x) - v_f. \quad (3.6)$$

Lemma 3.1 (Regret transfer). *For any continuous $f, g : \mathcal{X} \rightarrow \mathbb{R}$ and every $x \in \mathcal{X}$,*

$$|r_f(x) - r_g(x)| \leq d_{\mathcal{X}}(f, g). \quad (3.7)$$

Proof. Define

$$a = \min_{x \in \mathcal{X}} [f(x) - g(x)], \quad b = \max_{x \in \mathcal{X}} [f(x) - g(x)].$$

Then

$$g(x) + a \leq f(x) \leq g(x) + b.$$

Taking minima over x gives

$$a \leq v_f - v_g \leq b.$$

Because

$$r_f(x) - r_g(x) = [f(x) - g(x)] - [v_f - v_g],$$

both bracketed quantities belong to $[a, b]$. Their difference has absolute value at most $b - a = d_{\mathcal{X}}(f, g)$. $\square$

Corollary 3.1 (Optimizer transfer). *If*

$$x_g^* \in \arg \min_{x \in \mathcal{X}} g(x),$$

then

$$f(x_g^*) - \min_{x \in \mathcal{X}} f(x) \leq d_{\mathcal{X}}(f, g). \quad (3.8)$$

This result supplies the decision-theoretic motivation for the validation criterion developed below. Thus, a bound on decision-relevant risk distortion immediately yields a bound on the true regret incurred by optimizing an approximating risk function.

4. Validation-calibrated risk-function ambiguity

4.1. Risk-function aggregation

Let

$$\Delta_M = \left\{ q \in \mathbb{R}^M : q \geq 0, \mathbf{1}^\top q = 1 \right\}. \quad (4.1)$$

For $q \in \Delta_M$, define

$$f_q(x) = \sum_{m=1}^M q_m f_m(x). \quad (4.2)$$

The vector q aggregates risk evaluation functions. It is not interpreted as a posterior probability that a model is true.

Remark 4.1 (Risk aggregation versus distribution mixing). *In general,*

$$\sum_{m=1}^M q_m \text{ES}_\alpha^{P_m}(L) \neq \text{ES}_\alpha^{\sum_{m=1}^M q_m P_m}(L).$$

Accordingly, $\{f_q : q \in \Delta_M\}$ is a class of risk functions and need not coincide with the ES functions generated by a mixture family of return distributions.

4.2. Audit portfolios

Let

$$\mathcal{B} = \{b_1, \dots, b_J\} \subseteq \mathcal{X} \quad (4.3)$$

be the pre-specified audit library.

Define the true audit risks

$$\theta_j = f_0(b_j), \quad j = 1, \dots, J. \quad (4.4)$$

The validation estimator is

$$\hat{\theta}_j = \min_{\zeta \in \mathbb{R}} \left\{ \zeta + \frac{1}{pT} \sum_{t=1}^T \left[-b_j^\top R_t^v - \zeta \right]_+ \right\}. \quad (4.5)$$

Let the model audit matrix $A \in \mathbb{R}^{J \times M}$ be

$$A_{jm} = f_m(b_j). \quad (4.6)$$

For any vector $z \in \mathbb{R}^J$, define its range by

$$\text{rng}(z) = \max_{j \leq J} z_j - \min_{j \leq J} z_j. \quad (4.7)$$

The population audit discrepancy of q is

$$d_{\mathcal{B}}(q; \theta) = \text{rng}(Aq - \theta). \quad (4.8)$$

Equivalently,

$$d_{\mathcal{B}}(q; \theta) = \max_{j, k \leq J} |(A_j - A_k)q - (\theta_j - \theta_k)|. \quad (4.9)$$

4.3. Population acceptance set

For a pre-specified tolerance $\delta \geq 0$, define

$$\mathcal{Q}_{\delta} = \{q \in \Delta_M : d_{\mathcal{B}}(q; \theta) \leq \delta\}. \quad (4.10)$$

Membership in $\mathcal{Q}_{\delta}$ means that the aggregated risk function reproduces the true risk differences among the audit portfolios up to the accepted tolerance δ .

No assumption is made that a unique “true” aggregation vector exists. The set may contain multiple risk functions, a single risk function, or no acceptable risk function.

4.4. Simultaneous validation error

Define the event

$$\mathcal{E}_T = \left\{ \max_{j \leq J} |\hat{\theta}_j - \theta_j| \leq e_T \right\}. \quad (4.11)$$

Under Assumptions 2.1–2.3, one transparent choice is

$$e_T = \frac{2\bar{R}}{p} \sqrt{\frac{\log(2J/\beta)}{2T}}, \quad (4.12)$$

where $1 - \beta$ is the desired confidence level.

Lemma 4.1 (Simultaneous audit-risk concentration). *Under Assumptions 2.1–2.3,*

$$\mathbb{P}(\mathcal{E}_T \mid \mathcal{G}) \geq 1 - \beta \quad a.s., \quad (4.13)$$

where $\mathcal{E}_T$ and e_T are given by (4.11)–(4.12).

Proof. For every $b_j \in \mathcal{X}$, Assumption 2.2 implies

$$-b_j^{\top} R \in [-\bar{R}, \bar{R}].$$

Let F_j and $\widehat{F}_j$ denote the population and empirical loss distribution functions. Conditionally on $\mathcal{G}$, the Dvoretzky–Kiefer–Wolfowitz–Massart inequality (Massart, 1990) and a union bound imply

$$\mathbb{P}\left(\max_{j \leq J} \|\widehat{F}_j - F_j\|_\infty > \varepsilon \mid \mathcal{G}\right) \leq 2J \exp(-2T\varepsilon^2).$$

For any $\zeta \in [-\bar{R}, \bar{R}]$,

$$\mathbb{E}_F[(L - \zeta)_+] = \int_\zeta^{\bar{R}} [1 - F(u)] du,$$

and hence

$$|\mathbb{E}_{\widehat{F}}[(L - \zeta)_+] - \mathbb{E}_F[(L - \zeta)_+]| \leq 2\bar{R}\|\widehat{F} - F\|_\infty.$$

Define

$$\phi_F(\zeta) = \zeta + \frac{1}{p}\mathbb{E}_F[(L - \zeta)^+].$$

Because $L \in [-\bar{R}, \bar{R}]$ almost surely, every minimizer of the population or empirical Expected Shortfall representation can be chosen in $[-\bar{R}, \bar{R}]$. Hence it suffices to compare the objectives over this compact interval. In particular,

$$\sup_{\zeta \in [-\bar{R}, \bar{R}]} |\phi_{\widehat{F}}(\zeta) - \phi_F(\zeta)| \leq \frac{2\bar{R}}{p}\|\widehat{F} - F\|_\infty,$$

and therefore

$$\left| \min_{\zeta} \phi_{\widehat{F}}(\zeta) - \min_{\zeta} \phi_F(\zeta) \right| \leq \frac{2\bar{R}}{p}\|\widehat{F} - F\|_\infty.$$

Consequently,

$$|\widehat{\text{ES}}_\alpha - \text{ES}_\alpha| \leq \frac{2\bar{R}}{p}\|\widehat{F} - F\|_\infty.$$

Setting

$$\varepsilon = \sqrt{\frac{\log(2J/\beta)}{2T}}$$

proves the result. $\square$

Remark 4.2. *The radius in (4.12) is deliberately simple and distribution-free under bounded support. It may be conservative, particularly for high confidence levels and small tail probabilities. Sharper ES concentration inequalities or bootstrap procedures can replace e_T , provided that they deliver a simultaneous event of the form (4.11).*

4.5. Estimated acceptance set

Define

$$\widehat{\mathcal{Q}}_\delta = \left\{ q \in \Delta_M : \text{rng}(Aq - \widehat{\theta}) \leq \delta + 2e_T \right\}. \quad (4.14)$$

Theorem 4.1 (Finite-sample set containment). *On the event $\mathcal{E}_T$,*

$$\mathcal{Q}_\delta \subseteq \widehat{\mathcal{Q}}_\delta \subseteq \mathcal{Q}_{\delta+4e_T}. \quad (4.15)$$

Consequently, conditionally on $\mathcal{G}$, the containment holds with probability at least $1 - \beta$.

Proof. For a fixed $q \in \Delta_M$, define

$$u = Aq - \theta, \quad \widehat{u} = Aq - \widehat{\theta}.$$

On $\mathcal{E}_T$,

$$\|u - \widehat{u}\|_\infty \leq e_T.$$

The range functional satisfies

$$|\text{rng}(u) - \text{rng}(\widehat{u})| \leq 2e_T.$$

If $q \in \mathcal{Q}_\delta$, then

$$\text{rng}(Aq - \widehat{\theta}) \leq \delta + 2e_T,$$

so $q \in \widehat{\mathcal{Q}}_\delta$. Conversely, if $q \in \widehat{\mathcal{Q}}_\delta$, then

$$\text{rng}(Aq - \theta) \leq \delta + 4e_T,$$

so $q \in \mathcal{Q}_{\delta+4e_T}$. □

4.6. Model-class rejection

Define the minimum empirical discrepancy

$$\widehat{\delta}_{\min} = \min_{q \in \Delta_M} \text{rng}(Aq - \widehat{\theta}). \quad (4.16)$$

Then

$$\widehat{\mathcal{Q}}_\delta = \emptyset \iff \widehat{\delta}_{\min} > \delta + 2e_T. \quad (4.17)$$

On $\mathcal{E}_T$, emptiness of $\widehat{\mathcal{Q}}_\delta$ implies emptiness of $\mathcal{Q}_\delta$. Thus, the framework permits the candidate model class to fail validation rather than forcing a credibility vector to exist.

5. From audit portfolios to the full decision domain

5.1. Audit coverage radius

Define

$$h(\mathcal{B}, \mathcal{X}) = \sup_{x \in \mathcal{X}} \min_{b \in \mathcal{B}} \|x - b\|_1. \quad (5.1)$$

The quantity $h(\mathcal{B}, \mathcal{X})$ is the worst-case distance between a feasible portfolio and its nearest audited portfolio.

5.2. Lipschitz continuity of Expected Shortfall

Lemma 5.1 (Portfolio Lipschitz continuity). *Under Assumption 2.2, for every $m \in \{0, 1, \dots, M\}$ and every $x, y \in \mathcal{X}$,*

$$|f_m(x) - f_m(y)| \leq \bar{R} \|x - y\|_1. \quad (5.2)$$

Moreover, for every $q \in \Delta_M$,

$$|f_q(x) - f_q(y)| \leq \bar{R} \|x - y\|_1. \quad (5.3)$$

Proof. For any admissible return vector,

$$|L(x, R) - L(y, R)| = |(x - y)^\top R| \leq \bar{R} \|x - y\|_1.$$

Hence,

$$L(x, R) \leq L(y, R) + \bar{R} \|x - y\|_1$$

almost surely. Monotonicity and translation equivariance of ES imply

$$f_m(x) \leq f_m(y) + \bar{R} \|x - y\|_1.$$

Interchanging x and y gives (5.2). Equation (5.3) follows from convex aggregation. $\square$

Remark 5.1. *The Lipschitz constant in (5.2) does not contain the factor $1/p$. The factor $1/p$ appears in the statistical estimation radius because tail estimation becomes more difficult as p decreases. It is not required for deterministic sensitivity of ES to a uniformly bounded perturbation of the loss.*

5.3. Audit-to-domain transfer

Theorem 5.1 (Decision-domain discrepancy bound). *On the event $\mathcal{E}_T$, every $q \in \hat{\mathcal{Q}}_\delta$ satisfies*

$$d_{\mathcal{X}}(f_q, f_0) \leq \delta + 4e_T + 4\bar{R}h(\mathcal{B}, \mathcal{X}). \quad (5.4)$$

Consequently, conditionally on $\mathcal{G}$, the bound holds with probability at least $1 - \beta$.

Proof. Define

$$g_q(x) = f_q(x) - f_0(x).$$

By Lemma 5.1, f_q and f_0 are each $\bar{R}$ -Lipschitz with respect to $\|\cdot\|_1$, so

$$|g_q(x) - g_q(y)| \leq 2\bar{R} \|x - y\|_1.$$

Write $h = h(\mathcal{B}, \mathcal{X})$. For arbitrary $x, y \in \mathcal{X}$, choose $b_x, b_y \in \mathcal{B}$ such that

$$\|x - b_x\|_1 \leq h, \quad \|y - b_y\|_1 \leq h.$$

Then

$$\begin{aligned} g_q(x) - g_q(y) &= [g_q(x) - g_q(b_x)] + [g_q(b_x) - g_q(b_y)] + [g_q(b_y) - g_q(y)] \\ &\leq |g_q(x) - g_q(b_x)| + [g_q(b_x) - g_q(b_y)] + |g_q(b_y) - g_q(y)| \\ &\leq [g_q(b_x) - g_q(b_y)] + 4\bar{R}h. \end{aligned}$$

Taking the supremum over $x, y \in \mathcal{X}$ yields

$$d_{\mathcal{X}}(f_q, f_0) \leq d_{\mathcal{B}}(q; \theta) + 4\bar{R}h(\mathcal{B}, \mathcal{X}).$$

By Theorem 4.1, $q \in \hat{\mathcal{Q}}_\delta$ implies

$$d_{\mathcal{B}}(q; \theta) \leq \delta + 4e_T$$

on $\mathcal{E}_T$, proving the result. $\square$

Define the total validation-to-decision error

$$\Gamma_T = \underbrace{\delta}_{\text{accepted model discrepancy}} + \underbrace{4e_T}_{\text{validation estimation error}} + \underbrace{4\bar{R}h(\mathcal{B}, \mathcal{X})}_{\text{audit coverage error}}. \quad (5.5)$$

The decomposition in (5.5) distinguishes three conceptually different sources of uncertainty. A large value of Γ_T may result from a permissive governance tolerance, a short validation sample, or an audit library that poorly covers the feasible portfolio domain.

6. Oracle-consistent minimax-regret optimization

6.1. Regret under an aggregated risk function

For $q \in \Delta_M$, define

$$v(q) = \min_{y \in \mathcal{X}} f_q(y) \quad (6.1)$$

and

$$r_q(x) = f_q(x) - v(q). \quad (6.2)$$

This is the regret associated with using x instead of the optimal portfolio under the aggregated risk evaluation f_q .

6.2. Weighted engine-specific regret

An alternative expression frequently associated with multiple-model regret is

$$\tilde{r}_q(x) = \sum_{m=1}^M q_m [f_m(x) - v_m], \quad v_m = \min_{y \in \mathcal{X}} f_m(y). \quad (6.3)$$

The two regret definitions generally differ.

Proposition 6.1 (Oracle-incompatibility decomposition). *For every $q \in \Delta_M$ and $x \in \mathcal{X}$,*

$$\tilde{r}_q(x) = r_q(x) + I(q), \quad (6.4)$$

where

$$I(q) = v(q) - \sum_{m=1}^M q_m v_m \geq 0. \quad (6.5)$$

Under compactness, $I(q) = 0$ if and only if the positively weighted engines have a common optimal portfolio.

Proof. The equality follows from

$$\tilde{r}_q(x) = f_q(x) - \sum_m q_m v_m = f_q(x) - v(q) + v(q) - \sum_m q_m v_m.$$

Furthermore,

$$\min_y \sum_m q_m f_m(y) \geq \sum_m q_m \min_y f_m(y),$$

so $I(q) \geq 0$. Equality holds exactly when a portfolio simultaneously attains the model-specific minima for every engine with $q_m > 0$. $\square$

The weighted engine-specific regret compares the implemented portfolio with a hypothetical benchmark that is allowed to select a different oracle portfolio for each engine. The resulting incompatibility term $I(q)$ is independent of the implemented decision and cannot be reduced by portfolio optimization. Therefore, $r_q(x)$, rather than $\tilde{r}_q(x)$, is the appropriate regret measure when f_q is interpreted as the aggregated risk evaluation.

As a simple illustration, take $x \in [0, 1]$, $f_1(x) = x$, $f_2(x) = 1 - x$, and $q = (1/2, 1/2)$. Then $f_q \equiv 1/2$, so $r_q(x) = 0$, whereas

$$\tilde{r}_q(x) = \frac{1}{2}x + \frac{1}{2}(1 - x) = \frac{1}{2}$$

and $I(q) = 1/2$. The entire weighted engine-specific regret is attributable to oracle incompatibility.

6.3. Validation-calibrated minimax regret

Define

$$\widehat{F}_\delta(x) = \sup_{q \in \widehat{\mathcal{Q}}_\delta} r_q(x). \quad (6.6)$$

The proposed portfolio problem is

$$\widehat{V}_\delta = \min_{x \in \mathcal{X}} \sup_{q \in \widehat{\mathcal{Q}}_\delta} \left\{ \sum_{m=1}^M q_m f_m(x) - \min_{y \in \mathcal{X}} \sum_{m=1}^M q_m f_m(y) \right\}. \quad (6.7)$$

If $\widehat{\mathcal{Q}}_\delta = \Delta_M$, then convexity of $r_q(x)$ in q implies

$$\sup_{q \in \Delta_M} r_q(x) = \max_{m \leq M} \left\{ f_m(x) - \min_{y \in \mathcal{X}} f_m(y) \right\}. \quad (6.8)$$

Thus, the proposed model nests an engine-wise relative robust risk criterion, while allowing independent validation evidence to remove or combine unsupported evaluations.

6.4. True regret certificate

Define the true regret

$$r_0(x) = f_0(x) - \min_{y \in \mathcal{X}} f_0(y). \quad (6.9)$$

Theorem 6.1 (Uniform regret approximation). *Suppose $\widehat{\mathcal{Q}}_\delta \neq \emptyset$. On the event $\mathcal{E}_T$, for every $x \in \mathcal{X}$,*

$$|\widehat{F}_\delta(x) - r_0(x)| \leq \Gamma_T. \quad (6.10)$$

Consequently, conditionally on $\mathcal{G}$, the bound holds with probability at least $1 - \beta$.

Proof. By Theorem 5.1 and Lemma 3.1, every $q \in \widehat{\mathcal{Q}}_\delta$ satisfies

$$|r_q(x) - r_0(x)| \leq \Gamma_T.$$

Thus,

$$r_0(x) - \Gamma_T \leq r_q(x) \leq r_0(x) + \Gamma_T.$$

Taking the supremum over $q \in \widehat{\mathcal{Q}}_\delta$ proves the result. $\square$

Let $\widehat{x}$ be an approximately optimal solution satisfying

$$\widehat{F}_\delta(\widehat{x}) \leq \widehat{V}_\delta + \xi, \quad (6.11)$$

where $\xi \geq 0$ is the optimization error.

Corollary 6.2 (A posteriori regret certificate). *Let $\hat{x} \in \mathcal{X}$ satisfy (6.11), and suppose $\hat{\mathcal{Q}}_\delta \neq \emptyset$. On the event $\mathcal{E}_T$,*

$$r_0(\hat{x}) \leq \hat{V}_\delta + \xi + \Gamma_T. \quad (6.12)$$

Equivalently,

$$\boxed{f_0(\hat{x}) - \min_{x \in \mathcal{X}} f_0(x) \leq \hat{V}_\delta + \xi + \delta + 4e_T + 4\bar{R}h(\mathcal{B}, \mathcal{X})}.$$

Consequently, conditionally on $\mathcal{G}$, the certificate holds with probability at least $1 - \beta$. This is the paper's main operational bound: the true ES regret is controlled by the computed robust objective, optimization error, accepted model discrepancy, validation estimation error, and audit-coverage error.

Corollary 6.3 (A priori regret bound). *On the same event $\mathcal{E}_T$,*

$$r_0(\hat{x}) \leq 2\Gamma_T + \xi = 2\delta + 8e_T + 8\bar{R}h(\mathcal{B}, \mathcal{X}) + \xi. \quad (6.13)$$

Proof. Let $x_0^* \in \arg \min_x f_0(x)$. Theorem 6.1 gives

$$\hat{F}_\delta(x_0^*) \leq r_0(x_0^*) + \Gamma_T = \Gamma_T.$$

Hence $\hat{V}_\delta \leq \Gamma_T$. Combining this inequality with Corollary 6.2 proves the result. $\square$

The a posteriori bound (6.12) is preferable for reporting because it retains the computed robust objective $\hat{V}_\delta$ instead of replacing it by another worst-case bound.

Given a governance threshold ε_{gov} , the portfolio may be certified whenever

$$\hat{V}_\delta + \xi + \Gamma_T \leq \varepsilon_{\text{gov}}. \quad (6.14)$$

Failure of (6.14) means that the available evidence does not certify the requested ES-regret tolerance. It does not establish that the portfolio necessarily violates that tolerance.

7. Supporting analysis and computation

This section collects three supporting ingredients that complete the method but are not part of the core validation-to-certificate chain: sensitivity of the robust value to statistical expansion of the acceptance set, an exact LP reformulation for moderate ambiguity systems, and a warning counterexample showing why sparse benchmark validation can fail.

7.1. Sensitivity to statistical expansion

Define the population robust objective

$$F_\delta(x) = \sup_{q \in \mathcal{Q}_\delta} r_q(x) \quad (7.1)$$

and its optimal value

$$V_\delta = \min_{x \in \mathcal{X}} F_\delta(x). \quad (7.2)$$

On $\mathcal{E}_T$, Theorem 4.1 implies

$$F_\delta(x) \leq \widehat{F}_\delta(x) \leq F_{\delta+4e_T}(x). \quad (7.3)$$

Consequently,

$$F_\delta(\widehat{x}) - V_\delta \leq V_{\delta+4e_T} - V_\delta + \xi. \quad (7.4)$$

To derive an explicit sensitivity bound, suppose that the population acceptance set contains a strictly feasible aggregation.

Assumption 7.1 (Validation margin). *There exist $\bar{q} \in \Delta_M$ and $\sigma > 0$ such that*

$$d_{\mathcal{B}}(\bar{q}; \theta) \leq \delta - \sigma. \quad (7.5)$$

Define

$$D = \max_{m \leq M} \left\{ \max_{x \in \mathcal{X}} f_m(x) - \min_{x \in \mathcal{X}} f_m(x) \right\}. \quad (7.6)$$

Under Assumption 2.2, $D \leq 2\bar{R}$.

Theorem 7.1 (Expansion sensitivity). *Under Assumption 7.1, for every $t \geq 0$,*

$$0 \leq V_{\delta+t} - V_\delta \leq \frac{2Dt}{\sigma+t}. \quad (7.7)$$

Proof. Fix $q \in \mathcal{Q}_{\delta+t}$ and define

$$q' = \frac{\sigma}{\sigma+t}q + \frac{t}{\sigma+t}\bar{q}.$$

Convexity of $d_{\mathcal{B}}(\cdot; \theta)$ gives

$$\begin{aligned} d_{\mathcal{B}}(q'; \theta) &\leq \frac{\sigma}{\sigma+t}(\delta+t) + \frac{t}{\sigma+t}(\delta-\sigma) \\ &= \delta. \end{aligned}$$

Hence $q' \in \mathcal{Q}_\delta$.

For fixed x ,

$$r_q(x) = \max_{y \in \mathcal{X}} \sum_{m=1}^M q_m [f_m(x) - f_m(y)].$$

Since

$$|f_m(x) - f_m(y)| \leq D,$$

we obtain

$$|r_q(x) - r_{q'}(x)| \leq D \|q - q'\|_1.$$

Moreover,

$$\|q - q'\|_1 = \frac{t}{\sigma + t} \|q - \bar{q}\|_1 \leq \frac{2t}{\sigma + t}.$$

Taking suprema over q , followed by minima over x , proves (7.7). $\square$

Combining (7.4) and Theorem 7.1 yields

$$F_\delta(\hat{x}) - V_\delta \leq \frac{8De_T}{\sigma + 4e_T} + \xi. \quad (7.8)$$

The bound shows that statistical expansion is relatively inexpensive when the candidate risk-function class passes validation with a substantial margin. Near the feasibility boundary, however, small perturbations in the validation estimates may materially change the robust decision.

7.2. Polyhedral structure and exact optimization

7.2.1. Polyhedral representation

The estimated acceptance set can be written as

$$(A_j - A_k)q \leq \hat{\theta}_j - \hat{\theta}_k + \delta + 2e_T, \quad j, k = 1, \dots, J, \quad (7.9)$$

together with

$$q \geq 0, \quad \mathbf{1}^\top q = 1. \quad (7.10)$$

Thus, $\hat{Q}_\delta$ is a compact polytope whenever it is nonempty.

Let its extreme points be

$$q^{(1)}, \dots, q^{(K)}. \quad (7.11)$$

7.2.2. Extreme-point reduction

Proposition 7.1 (Extreme-point reduction). *For every fixed $x \in \mathcal{X}$,*

$$\sup_{q \in \hat{Q}_\delta} r_q(x) = \max_{k=1, \dots, K} r_{q^{(k)}}(x). \quad (7.12)$$

Proof. For fixed x ,

$$r_q(x) = \max_{y \in \mathcal{X}} \sum_m q_m [f_m(x) - f_m(y)]$$

is convex in q . If

$$q = \sum_{k=1}^K \lambda_k q^{(k)}, \quad \lambda \in \Delta_K,$$

then convexity implies

$$r_q(x) \leq \sum_k \lambda_k r_{q^{(k)}}(x) \leq \max_k r_{q^{(k)}}(x).$$

The reverse inequality follows because every extreme point is feasible. $\square$

7.2.3. Oracle problems

For each extreme point $q^{(k)}$, compute

$$v_k = \min_{y \in \mathcal{X}} \sum_{m=1}^M q_m^{(k)} f_m(y). \quad (7.13)$$

Using the finite-scenario ES representation, this is the linear program

$$\begin{aligned} \min_{y, \eta, u} \quad & \sum_{m=1}^M q_m^{(k)} \left(\eta_m + \frac{1}{p} \sum_{s=1}^{S_m} \pi_{ms} u_{ms} \right) \\ \text{s.t.} \quad & u_{ms} \geq -y^\top R_{ms} - \eta_m, \quad m = 1, \dots, M, \ s = 1, \dots, S_m, \\ & u_{ms} \geq 0, \quad m = 1, \dots, M, \ s = 1, \dots, S_m, \\ & y \in \mathcal{X}. \end{aligned} \quad (7.14)$$

7.2.4. Master problem

After computing $v_1, \dots, v_K$, solve

$$\begin{aligned} \min_{x, z, \zeta, u} \quad & z \\ \text{s.t.} \quad & z \geq \sum_{m=1}^M q_m^{(k)} \left(\zeta_m + \frac{1}{p} \sum_{s=1}^{S_m} \pi_{ms} u_{ms} \right) - v_k, \quad k = 1, \dots, K, \\ & u_{ms} \geq -x^\top R_{ms} - \zeta_m, \quad m = 1, \dots, M, \ s = 1, \dots, S_m, \\ & u_{ms} \geq 0, \quad m = 1, \dots, M, \ s = 1, \dots, S_m, \\ & x \in \mathcal{X}. \end{aligned} \quad (7.15)$$

The shared ES auxiliary variables in (7.15) are exact because all aggregation coefficients are nonnegative. For a fixed x , reducing any engine-specific ES epigraph value weakly reduces every extreme-point constraint.

Remark 7.1 (Computational scope). *For a fixed finite scenario representation, audit library, and finite ambiguity system, the minimax-regret problem admits an exact linear-programming reformulation. This formulation is practical for moderate problem sizes. Its computational burden*

nevertheless grows with the number of assets, scenarios, audit portfolios, candidate engines, and robust-regret constraints; the principal bottleneck is extreme-point enumeration of the validation-calibrated ambiguity polytope. Larger instances may require decomposition, cutting-plane methods, scenario reduction, or other scalable implementations, which are beyond the scope of the present paper.

Remark 7.2 (Implementation note). *In implementation, engine-specific ES auxiliary variables may be bounded using the return bound $\bar{R}$, and engines that receive zero weight at every relevant extreme point may be removed from the LP formulation without changing the optimal value.*

7.3. Why sparse benchmark validation can fail

Consider two assets and portfolios

$$x = (t, 1 - t), \quad t \in [0, 1].$$

Let $c > 0$, and consider two risk engines, each with two equally likely return scenarios.

Engine 1 has

$$R = \begin{cases} (3c, 0), & \text{with probability } 1/2, \\ (0, c), & \text{with probability } 1/2. \end{cases} \quad (7.16)$$

Engine 2 has

$$R = \begin{cases} (c, 0), & \text{with probability } 1/2, \\ (0, 3c), & \text{with probability } 1/2. \end{cases} \quad (7.17)$$

For $\alpha \geq 1/2$, ES equals the larger of the two scenario losses. Therefore,

$$f_1(t) = \max\{-3ct, -c(1 - t)\} \quad (7.18)$$

and

$$f_2(t) = \max\{-ct, -3c(1 - t)\}. \quad (7.19)$$

The two engines agree at both vertices:

$$f_1(0) = f_2(0) = 0, \quad f_1(1) = f_2(1) = 0. \quad (7.20)$$

Nevertheless, their optimal portfolios differ:

$$t_1^* = \frac{1}{4}, \quad t_2^* = \frac{3}{4}, \quad (7.21)$$

and their common optimal value is $-3c/4$.

If both engines are admissible, the minimax-regret solution is

$$t^* = \frac{1}{2},$$

with worst-case regret

$$\frac{c}{4}. \tag{7.22}$$

Figure 1 illustrates the two risk functions and the resulting interior optima.

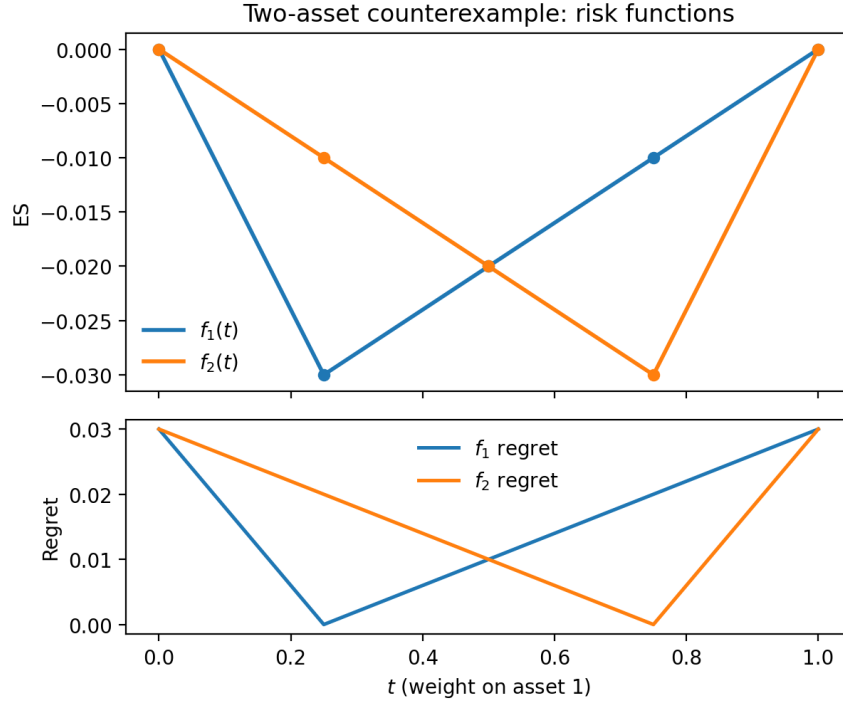


Figure 1: Two-asset counterexample ($c = 0.04$). Engine-specific ES risk functions agree at the pure-asset vertices but attain distinct interior optima at $t = 1/4$ and $t = 3/4$; the minimax-regret portfolio lies at $t = 1/2$.

Hence,

$$\text{agreement on all pure-asset portfolios} \not\Rightarrow \text{agreement on diversified portfolio decisions}.$$

The counterexample demonstrates why representative benchmark portfolios must be evaluated in relation to their geometric coverage of the feasible domain. Validation at the vertices alone does not control the behavior of convex piecewise-linear ES functions in the interior.

8. Numerical experiments

This section reports controlled simulations designed to illustrate the mechanisms implied by the theory under known data-generating processes. The experiments are not intended as a

Table I: Baseline simulation parameters for the controlled experiments.

Parameter	Value
Number of assets n	4
Feasible set	long-only simplex
ES level α	0.95
Tail probability $p = 1 - \alpha$	0.05
True discrete scenarios	200
Training sample size	400
Support / return bound $\bar{R}$	0.20
Governance tolerance δ	0.01
Confidence level $1 - \beta$	0.95
Gaussian engine scenarios	500
Sampled diagnostic portfolios $ \mathcal{G} $	800
Random seed	42

rolling empirical asset-management study and do not use realized market outcomes to verify the population certificate. Instead, the known simulation law permits exact evaluation of true Expected Shortfall and true portfolio regret. A full empirical application would require a sufficiently long sequence of non-overlapping deployment periods, explicit transaction-cost and rebalancing specifications, and certified audit-coverage bounds.

Throughout, Expected Shortfall is computed at level $\alpha = 0.95$ ($p = 0.05$), and the validation radius uses the distribution-free bound (4.12) unless stated otherwise. Portfolio losses and regrets are reported in simple-return units (so that 0.01 corresponds to one percentage point).

8.1. Experimental design

Table I summarizes the baseline simulation design used below. The true law P_0 is a discrete regime-mixture scenario model on the simplex

$$\mathcal{X} = \{x \in \mathbb{R}^n : x \geq 0, \mathbf{1}^\top x = 1\}$$

with $n = 4$ assets and 200 equally weighted scenarios whose coordinates are clipped to the known support bound $\bar{R} = 0.20$. True ES values are computed exactly from this discrete scenario representation. Candidate engines are constructed from an independent training sample of size 400 drawn from the same DGP, except for synthetic level-shift and scale engines that transform the true risk function directly.

8.2. Level error versus decision-relevant error

We compare candidate engines relative to the true ES risk function:

- common additive shifts of the true risk function;
- a dependence-misspecified shape distortion estimated from training data;

- multiplicative scale factors 0.8 and 1.25;
- a Gaussian scenario approximation fitted on the training sample;
- a historical empirical scenario engine.

Audit portfolios are the conventional vertex-plus-equal-weight library. Domain diagnostics are computed on a finite set of 800 random interior portfolios.

Let $\mathcal{G} \subset \mathcal{X}$ denote the finite sampled portfolio set used for the numerical domain diagnostic. For two risk functions f and g , define

$$\tilde{d}_{\mathcal{G}}(f, g) = \max_{x, y \in \mathcal{G}} |[f(x) - f(y)] - [g(x) - g(y)]|.$$

Because $\mathcal{G} \subseteq \mathcal{X}$, this quantity satisfies

$$\tilde{d}_{\mathcal{G}}(f, g) \leq d_{\mathcal{X}}(f, g).$$

Accordingly, it is a Monte Carlo lower diagnostic for the full-domain discrepancy rather than a certified estimate or upper bound for $d_{\mathcal{X}}$.

The additive-shift specifications are synthetic risk-function perturbations introduced solely to isolate the invariance property of the proposed discrepancy. They are not intended to represent separately trained scenario generators satisfying the scenario-engine construction in Section 2. Their purpose is to show that common risk-level errors can be large under absolute forecast comparisons while being irrelevant for portfolio ranking and regret.

Table II reports absolute audit ES error, the audit discrepancy, the sampled-domain diagnostic, and true regret of each engine-specific optimizer. Figure 2 visualizes the contrast between absolute level error and true regret. Numbers are single-realization outputs under the fixed seed in Table I; they are intended to exhibit the mechanism rather than to estimate sampling distributions.

The pattern is consistent with Proposition 3.1 and Lemma 3.1: absolute risk-level error can be large while regret remains unchanged under a common shift, whereas relative-risk distortions can affect the optimizer and true regret. In the reported simulations, observed true regret is no larger than the sampled-domain discrepancy diagnostic. This comparison is illustrative only: because $\tilde{d}_{\mathcal{G}}$ is computed on a finite random portfolio set, it does not numerically verify the full-domain regret-transfer inequality over the entire feasible set.

8.3. Audit-library geometry and coverage diagnostics

We compare audit designs on polytopes with $n \in \{3, 6\}$ assets: pure-asset (vertex) libraries, vertices plus the equal-weight portfolio, random diversified portfolios, and a greedy farthest-point design. Table III reports a Monte Carlo *lower diagnostic*

$$\hat{h}_{\text{diag}} = \max_{g \in \mathcal{G}_{\text{sampled}}} \min_{b \in \mathcal{B}} \|g - b\|_1 \leq h(\mathcal{B}, \mathcal{X}).$$

Table II: Level error versus decision-relevant discrepancy. Columns report absolute audit ES level error, audit discrepancy $d_{\mathcal{B}}$, the sampled-domain diagnostic $\tilde{d}_{\mathcal{G}}$, and true regret. A common additive shift inflates absolute audit ES error while leaving the sampled decision-discrepancy diagnostic and true regret unchanged; shape distortions raise both quantities. The quantity $\tilde{d}_{\mathcal{G}}$ is a Monte Carlo lower diagnostic for $d_{\mathcal{X}}$, not a certified full-domain upper bound. The oracle true-ES row is included only for reference and is not available to the validation-calibrated procedure. Units are simple returns.

Engine	Level err.	$d_{\mathcal{B}}$	$\tilde{d}_{\mathcal{G}}$	Regret
Oracle (true ES)	0.000	0.000	0.000	0.000
Historical	0.011	0.012	0.009	0.0004
Shift +0.005	0.005	0.000	0.000	0.000
Shift +0.010	0.010	0.000	0.000	0.000
Shift +0.020	0.020	0.000	0.000	0.000
Shape distortion	0.016	0.028	0.022	0.0010
Scale 0.8	0.019	0.006	0.005	0.000
Scale 1.25	0.023	0.007	0.006	0.000
Gaussian approx.	0.010	0.010	0.012	0.0022

Table III: Monte Carlo audit-coverage diagnostics for selected audit-library designs. The reported quantity $\hat{h}_{\text{diag}} = \max_{g \in \mathcal{G}_{\text{sampled}}} \min_{b \in \mathcal{B}} \|g - b\|_1$ is a lower diagnostic for $h(\mathcal{B}, \mathcal{X})$, not a certified upper bound and not a valid certificate input.

n	Design	J	$\hat{h}_{\text{diag}}$
3	Vertices	3	1.32
3	Conventional (vertices + EW)	4	0.67
3	Random diversified	12	0.79
3	Farthest-point	12	0.40
6	Vertices	6	0.74
6	Conventional	7	0.74
6	Random diversified	12	0.84
6	Farthest-point	12	0.77

This quantity is useful for comparing designs on a finite sampled portfolio set, but it is *not* a valid upper bound on $h(\mathcal{B}, \mathcal{X})$ and is therefore not inserted into the formal certificate Γ_T . Certified coverage requires a verified finite ρ -net upper bound as in Section 9.

Consistent with the counterexample in Section 7.3, vertex-only validation can leave a large coverage residual because agreement at pure-asset portfolios does not control risk-function behavior in the interior of the feasible set. Audit enrichment does not automatically improve coverage, however: randomly selected diversified portfolios may still leave substantial uncovered regions, particularly in higher-dimensional simplices. This motivates audit designs that explicitly target geometric coverage, such as verified nets or carefully constructed space-filling designs, rather than relying solely on unstructured random enrichment.

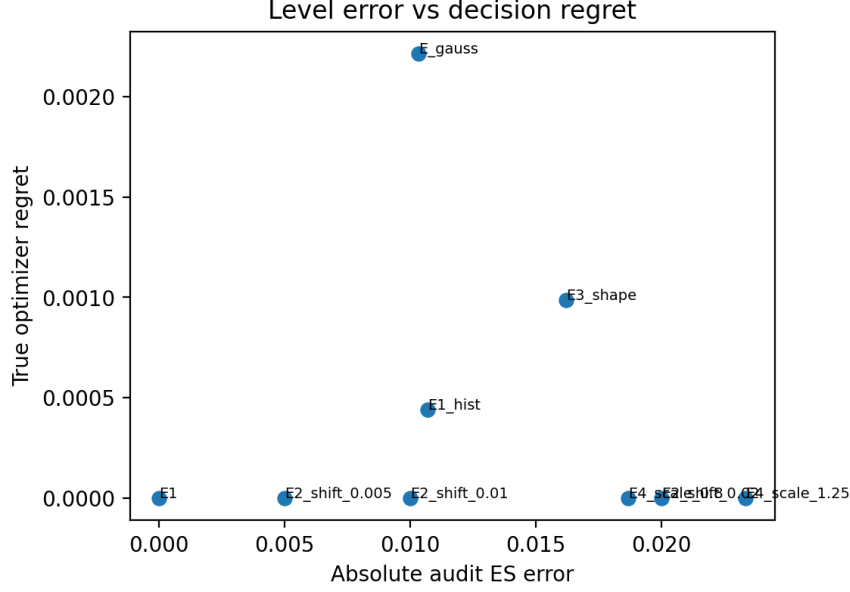


Figure 2: Absolute audit ES error versus true regret. Common additive shifts increase absolute error without changing regret; shape and Gaussian approximation errors induce positive regret.

8.4. Validation length and the statistical certificate component

For fixed $\alpha = 0.95$, $\beta = 0.05$, $\bar{R} = 0.20$, and a conventional vertex-plus-equal-weight audit library with $J = 5$, increasing the validation length T shrinks the statistical estimation radius e_T . Table IV reports the closed-form decay of e_T and its contribution $4e_T$. The full certificate $\Gamma_T = \delta + 4e_T + 4\bar{R}h(\mathcal{B}, \mathcal{X})$ additionally depends on the accepted tolerance and on a certified upper bound for the audit-coverage radius. Because no certified coverage upper bound is used here, the table does not report a complete Γ_T .

The distribution-free DKW calibration is valid but economically loose under the baseline parameters. Since portfolio losses lie in $[-\bar{R}, \bar{R}]$, true Expected Shortfall regret satisfies $0 \leq r_0(x) \leq 2\bar{R}$. With $\bar{R} = 0.20$, this trivial range bound equals 0.40. At every sample size reported in Table IV, the statistical component $4e_T$ alone exceeds this value. The reported calculation should therefore be interpreted as illustrating the finite-sample scaling and conservativeness of a simple distribution-free calibration, rather than demonstrating a practically tight certificate. Economically informative certificates would require substantially longer validation histories, sharper ES-specific confidence bounds, or additional distributional structure.

8.5. Model-class rejection

When the candidate class contains an aggregation that approximately reproduces the true audit risk differences, the estimated acceptance set $\hat{\mathcal{Q}}_\delta$ is nonempty and the robust optimization problem is well posed. Conversely, tightening δ , shortening the validation sample, or considering a candidate library that cannot reproduce the audit differences can yield $\hat{\mathcal{Q}}_\delta = \emptyset$. Such emptiness

Table IV: Decay of the validation-estimation component with validation length under the closed-form radius (4.12) ($\bar{R} = 0.20$, $p = 0.05$, $\beta = 0.05$, $J = 5$). The full certificate is $\Gamma_T = \delta + 4e_T + 4\bar{R}h(\mathcal{B}, \mathcal{X})$, which additionally requires a certified upper bound on the audit-coverage radius. The reference level $2\bar{R} = 0.40$ is the trivial upper bound on true ES regret under the bounded-support assumption. Units are simple returns.

T	e_T	$4e_T$	$4e_T/(2\bar{R})$
100	1.3021	5.2084	13.02
250	0.8235	3.2941	8.24
500	0.5823	2.3293	5.82
1000	0.4118	1.6470	4.12
2500	0.2604	1.0417	2.60
5000	0.1841	0.7366	1.84
10000	0.1302	0.5208	1.30

is not an optimization failure, a solver error, or a numerical bug. It is a diagnostic outcome indicating that, at the selected tolerance and confidence level, the candidate risk-function class is not supported by the validation exercise. In practice, emptiness should trigger expansion or retraining of the engine library, reassessment of the tolerance, or collection of additional validation data.

8.6. Protocol for end-to-end certificate coverage in repeated simulation

A full repeated-simulation diagnostic would require a certified upper bound on $h(\mathcal{B}, \mathcal{X})$. For each independent validation replication, one would construct $\hat{\mathcal{Q}}_\delta$, record model-class rejection, solve the minimax-regret problem when $\hat{\mathcal{Q}}_\delta \neq \emptyset$, and compare the exact simulated true regret with the certificate in (6.12). We do not report such coverage frequencies here because the present numerical study does not construct a certified audit-coverage upper bound.

8.7. Implementation note on empirical panels

The accompanying computational package also supports a strict rolling workflow on multi-asset ETF panels: training, validation, optimization, and subsequent deployment use non-overlapping holding-period returns. We do not report rolling empirical results because the available pilot history is insufficient for stable out-of-sample risk comparisons. A full application would need to specify the asset universe and providers, window lengths, rebalancing calendar, transaction costs, certified coverage bounds, and a sufficiently long sequence of deployment periods.

9. Extensions, limitations, and conclusion

9.1. Serial dependence

The finite-sample radius in (4.12) relies on independent validation observations. Financial holding-period returns may be serially dependent even when windows do not overlap. Under

mixing or other dependence conditions, the DKW-based radius should be replaced by a valid dependent-data bound or a carefully justified block-bootstrap procedure.

9.2. Distribution drift

If the deployment distribution P_{dep} differs from the validation distribution P_{val} , define

$$\Delta_{\text{drift}} = d_{\mathcal{X}}(f_0^{\text{dep}}, f_0^{\text{val}})$$

and write r_0^{dep} for Expected Shortfall regret under P_{dep} . The expanded certificate takes the form

$$r_0^{\text{dep}}(\hat{x}) \leq \hat{V}_\delta + \xi + \Gamma_T + \Delta_{\text{drift}}.$$

This extension requires an externally justified bound on Δ_{drift} . Validation data alone cannot certify performance under unrestricted distribution shift.

9.3. Unbounded returns

The bounded-support assumption produces transparent constants and separates the statistical and geometric arguments. For unbounded returns, one may use truncation, robust ES estimation, moment conditions, or light-tail concentration results. Any replacement must preserve a simultaneous bound for all audit portfolios.

9.4. Approximate audit coverage

Computing $h(\mathcal{B}, \mathcal{X})$ exactly may itself be difficult in high dimension. A dense numerical search yields a lower bound on h , not a valid upper bound. Formal certification requires either an exact computation or a verified upper bound obtained from a known finite ρ -net $\mathcal{G} \subseteq \mathcal{X}$:

$$h(\mathcal{B}, \mathcal{X}) \leq \max_{g \in \mathcal{G}} \min_{b \in \mathcal{B}} \|g - b\|_1 + \rho.$$

Only the right-hand side may enter Γ_T . Monte Carlo lower diagnostics such as $\hat{h}_{\text{diag}}$ in Section 8.3 must not be substituted for this upper bound.

9.5. Scope of the guarantee

The proposed guarantee concerns the difference between the true ES of the implemented portfolio and the minimum true ES over the same feasible set. It is not a guarantee on maximum drawdown, terminal wealth, or the realized loss in a single deployment period.

9.6. Conclusion

This paper develops a validation-calibrated interface between portfolio model validation and robust tail-risk optimization. The central object is a discrepancy between risk functions that removes common additive level errors and controls the transfer of portfolio regret.

Independent validation observations on a finite audit library are used to construct a statistically expanded polyhedral set of admissible risk-function aggregations. A geometric covering argument extends audit-level validation to the continuous feasible portfolio domain. The resulting validation-calibrated minimax-regret portfolio admits an explicit high-probability ES-regret certificate that separates accepted model error, statistical estimation error, and audit-coverage error under the stated assumptions.

The framework also permits the complete candidate model class to fail validation, distinguishes aggregate-risk regret from incompatible engine-specific oracle comparisons, and remains exactly solvable by linear programming when the number of engines is moderate.

The analysis suggests a broader governance principle: model validation for optimization should certify the quantities that determine decisions rather than rely exclusively on predictive risk levels at a small number of benchmarks. Future work may develop sharper ES confidence regions, dependence-robust validation procedures, adaptive audit design, and large-scale algorithms that avoid explicit extreme-point enumeration.

Declaration of competing interests

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author thanks colleagues and seminar participants for helpful comments. Any remaining errors are the author's own.

Appendix A. Linear formulation of the minimum audit discrepancy

The quantity $\widehat{\delta}_{\min}$ in (4.16) can be computed from

$$\begin{aligned}
 \min_{q, \ell, u} \quad & u - \ell \\
 \text{s.t.} \quad & \ell \leq A_j q - \widehat{\theta}_j \leq u, \quad j = 1, \dots, J, \\
 & q \geq 0, \\
 & \mathbf{1}^\top q = 1.
 \end{aligned} \tag{A.1}$$

The same representation can be used to construct the acceptance polytope with two additional variables:

$$\begin{aligned} \ell &\leq A_j q - \widehat{\theta}_j \leq u, \quad j = 1, \dots, J, \\ u - \ell &\leq \delta + 2e_T, \\ q &\in \Delta_M. \end{aligned} \tag{A.2}$$

This formulation uses $O(J)$ constraints instead of all $O(J^2)$ pairwise inequalities, while defining the same set after projection onto q .

References

- Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 1487–1503. doi:[10.1016/S0378-4266\(02\)00283-2](https://doi.org/10.1016/S0378-4266(02)00283-2).
- Aissi, H., Bazgan, C., Vanderpooten, D., 2009. Min-max and min-max regret versions of combinatorial optimization problems: A survey. *European Journal of Operational Research* 197, 427–438. doi:[10.1016/j.ejor.2008.09.012](https://doi.org/10.1016/j.ejor.2008.09.012).
- Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228. doi:[10.1111/1467-9965.00068](https://doi.org/10.1111/1467-9965.00068).
- Bertsimas, D., Gupta, V., Kallus, N., 2018. Data-driven robust optimization. *Mathematical Programming* 167, 235–292. doi:[10.1007/s10107-017-1125-8](https://doi.org/10.1007/s10107-017-1125-8).
- Chun, S.Y., Shapiro, A., Uryasev, S., 2012. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research* 60, 739–756. doi:[10.1287/opre.1120.1072](https://doi.org/10.1287/opre.1120.1072).
- Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58, 595–612. doi:[10.1287/opre.1090.0741](https://doi.org/10.1287/opre.1090.0741).
- Elmachtoub, A.N., Grigas, P., 2022. Smart “predict, then optimize”. *Management Science* 68, 9–26. doi:[10.1287/mnsc.2020.3922](https://doi.org/10.1287/mnsc.2020.3922).
- Fissler, T., Ziegel, J.F., 2016. Higher order elicibility and Osband’s principle. *The Annals of Statistics* 44, 1680–1707. doi:[10.1214/16-AOS1439](https://doi.org/10.1214/16-AOS1439).
- Goldfarb, D., Iyengar, G., 2003. Robust portfolio selection problems. *Mathematics of Operations Research* 28, 1–38. doi:[10.1287/moor.28.1.1.14260](https://doi.org/10.1287/moor.28.1.1.14260).
- Huang, D., Zhu, S., Fabozzi, F.J., Fukushima, M., 2010. Portfolio selection under distributional uncertainty: A relative robust CVaR approach. *European Journal of Operational Research* 203, 185–194. doi:[10.1016/j.ejor.2009.07.010](https://doi.org/10.1016/j.ejor.2009.07.010).

- Massart, P., 1990. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability* 18, 1269–1283. doi:[10.1214/aop/1176990746](https://doi.org/10.1214/aop/1176990746).
- Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171, 115–166. doi:[10.1007/s10107-017-1172-1](https://doi.org/10.1007/s10107-017-1172-1).
- Prashanth, L.A., Jagannathan, K., Kolla, R.K., 2020. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions, in: *Proceedings of the 37th International Conference on Machine Learning*, PMLR. pp. 5577–5586.
- Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, 1443–1471. doi:[10.1016/S0378-4266\(02\)00271-6](https://doi.org/10.1016/S0378-4266(02)00271-6).
- Van Parys, B.P.G., Mohajerin Esfahani, P., Kuhn, D., 2021. From data to decisions: Distributionally robust optimization is optimal. *Management Science* 67, 3387–3402. doi:[10.1287/mnsc.2020.3678](https://doi.org/10.1287/mnsc.2020.3678).
- Zhu, S., Fukushima, M., 2009. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research* 57, 1155–1168. doi:[10.1287/opre.1080.0684](https://doi.org/10.1287/opre.1080.0684).