MINIMAX ACQUISITION UNDER A SHARED BOUNDED BIAS: CRITICAL WINDOWS AND SPARSE PORTFOLIOS

Anonymous authors
Paper under double-blind review

ABSTRACT

Cheap model predictions can reduce the cost of estimating population means, but collecting more predictions does not remove a shared calibration bias. We study how to allocate a limited budget among trusted labels, predictions, and correlated paired observations when this bias lies in a known interval. Existing prediction-powered methods combine predictions with trusted labels to obtain valid inference, while recent minimax allocation theory optimizes acquisition under unrestricted prediction bias. These results leave unresolved how a finite bias bound changes which auxiliary observations are worth buying and how spending on them should scale with the budget. For Gaussian observations and continuous allocations, we reduce the exact fixed-design minimax risk over all measurable estimators to a bounded-normal-mean problem and derive two cost-normalized information criteria that distinguish regimes in which auxiliary data offer no improvement, optimal auxiliary spending remains bounded, or its share of the budget approaches one. We also derive sharp second-order risk expansions in the bounded-spending and critical regimes, characterize a $1/B$-wide transition window for total budget B, and prove that an exact optimum can be attained using at most three observation types.

1 INTRODUCTION

Model evaluation often requires estimating an average score from costly human annotations. Model-generated scores offer a cheaper alternative, but may systematically disagree with human ratings even when the two are strongly correlated (Fisch et al., 2024; Angelopoulos et al., 2025). A persistent calibration bias remains even as additional predictions reduce sampling noise. A limited budget must therefore be divided among trusted labels, cheap predictions, and paired human–model scores that help identify the discrepancy. The problem is to decide which observations to collect and how much to spend on each, especially when a shared bias makes their value depend on what else is collected.

Prediction-powered inference (PPI) combines trusted labels with predictions to obtain valid inference without requiring unbiased predictors (Angelopoulos et al., 2023a; Fisch et al., 2024). Active inference and cost-aware methods select examples to label and allocate resources between weak and strong raters (Zrnic & Candès, 2024a; Angelopoulos et al., 2025). MultiPPI optimizes allocation across sources with different costs and correlations, with minimax guarantees for known covariance (Cowen-Breen et al., 2026). It provides our closest unrestricted-bias baseline, but does not characterize how a finite bound on a shared discrepancy changes optimal spending.

A known bias bound allows an estimator to trade a controlled amount of residual bias for lower variance. Classical normal-mean theory and recent work on misspecification exploit this tradeoff through shrinkage (van Eeden & Zidek, 2001; 2004; Armstrong et al., 2025). Related studies choose experiments for subsequent shrinkage or for combining experimental and observational evidence (Rosenman & Miratrix, 2023; Epanomeritakis & Viviano, 2026). These results do not, however, determine how a fixed bound on a shared bias changes the optimal allocation as the total budget grows. In particular, can auxiliary data remain useful while their optimal expenditure stays bounded? Answering this question requires the full minimax risk, since a leading-order comparison can miss the benefit of such an allocation.

We formulate this acquisition problem for an unrestricted target mean θ and a shared discrepancy $\delta \in [-h, h]$, with known $0 < h < \infty$. We assume independent Gaussian observation streams

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.

with known costs, covariance matrices, and bias loadings, and choose continuous allocations before observing the data. Each purchased observation enters through its full covariance, preserving dependence within a pair. For a fixed allocation, a Gaussian decomposition reduces the exact minimax squared-error risk over all measurable estimators to a classical bounded-normal-mean problem. We then optimize this risk over allocations under a total budget B, without imposing a linear bias correction in advance.

The acquisition analysis uses two information rates per unit cost. The first, Γ_0, measures the best auxiliary target information if the discrepancy were known. The second, Γ_∞, measures the best target information from an auxiliary mixture when the discrepancy is treated as unrestricted. Comparing both with the trusted-label rate, normalized to one, separates the value of a small auxiliary purchase from the value of relying on auxiliary data at scale.

Our main contributions are:

- **Optimal acquisition regimes.** We establish the scalar phase diagram and its critical boundary. Trusted-only acquisition is optimal for every budget if and only if $\Gamma_0 \le 1$. When $\Gamma_\infty < 1 < \Gamma_0$, every exact optimum spends only $O(1)$ on auxiliary observations as $B \to \infty$; when $\Gamma_\infty > 1$, every optimal auxiliary budget share converges to one.
- **Second-order risk and critical windows.** Writing V_B for the optimal risk, we prove $V_B = B^{-1} - C_* B^{-2} + o(B^{-2})$ in the bounded-spending regime and characterize the positive constant C_*. We identify a common rank-one information structure at the nontrivial critical boundary, derive the sharp second-order risk there, and resolve a $1/B$-wide transition window around this boundary.
- **Sparse designs and source complementarity.** For every finite budget, an exact optimum can be attained using at most three observation types, including the trusted type when used, despite a generally nonconvex objective. Constructive examples show how opposite bias couplings can make individually unattractive sources useful together, and how calibration-only observations can change the acquisition regime.

2 Observation model and minimax risk

2.1 Observation types and information

Let $\theta \in \mathbb{R}$ be the target mean and let $\delta \in [-h, h]$ be a shared discrepancy, with known $0 < h < \infty$. An observation type specifies both the measurements returned by one purchase and their joint distribution. One observation of type j has the Gaussian model

$$Z_j \sim \mathcal{N}\left(F_j \begin{bmatrix} \theta \\ \delta \end{bmatrix}, \Sigma_j \right), \qquad F_j = [f_j \; g_j], \quad \Sigma_j \succ 0. \tag{1}$$

The loadings, covariances, and costs are known. Purchases are independent, with possible correlation within each observation. Trusted labels have no bias loading; predictions may have mean $\theta + \ell_j \delta$. A pair contains both measurements.

For $x_j > 0$ independent observations of type j, the sample mean $\bar{Z}_j$ has covariance Σ_j / x_j. We allow real $x_j \ge 0$ using the continuous approximate-design convention; $x_j = 0$ means that the type is absent. The total information is

$$M(x) = \sum_j x_j F_j^\top \Sigma_j^{-1} F_j = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0. \tag{2}$$

Each pair contributes once through its full covariance, preserving the dependence between its coordinates.

2.2 Exact risk for a fixed design

The remaining uncertainty about the bias is described by the classical bounded normal mean problem. For precision $r \ge 0$, define

$$\mathfrak{m}_h(r) = \inf_{\widehat{\delta}} \sup_{|\delta| \le h} \mathbb{E}_\delta \left[\left\{ \widehat{\delta}(\sqrt{r}\,\delta + G) - \delta \right\}^2 \right], \qquad G \sim \mathcal{N}(0,1). \tag{3}$$

At zero precision, the observation carries no information about δ, and $\mathfrak{m}_h(0) = h^2$. We use this minimax risk directly, without requiring a closed-form estimator (Casella & Strawderman, 1981; Fan & Zhang, 2007; Montiel Olea & Zubova, 2026).

Theorem 1 (Fixed-design minimax risk)**.** *For $a > 0$, let*

$$\kappa = \frac{b}{a}, \qquad \rho = c - \frac{b^2}{a} \geq 0. \tag{4}$$

The minimax squared-error risk for estimating θ, over all measurable estimators and all $(\theta, \delta) \in \mathbb{R} \times [-h, h]$, is

$$\mathcal{R}_{\mathrm{all}}(M) = \frac{1}{a} + \kappa^2 \mathfrak{m}_h(\rho). \tag{5}$$

If $a = 0$, the risk is infinite. The best affine estimator has risk

$$\mathcal{R}_{\mathrm{aff}}(M) = \frac{1}{J_h(M)}, \qquad J_h(M) = a - \frac{b^2}{c + h^{-2}}, \tag{6}$$

and satisfies

$$\mathcal{R}_{\mathrm{all}}(M) \leq \mathcal{R}_{\mathrm{aff}}(M) \leq \frac{5}{4} \mathcal{R}_{\mathrm{all}}(M). \tag{7}$$

The same inequalities hold after minimizing both risks over any common deterministic design class.

To interpret the identity, form the sufficient statistic $U = \sum_{j:x_j>0} x_j F_j^\top \Sigma_j^{-1} \bar{Z}_j$, which has distribution $\mathcal{N}(M(\theta, \delta)^\top, M)$. When $\rho > 0$, it yields two independent statistics,

$$T = \frac{U_\theta}{a} \sim \mathcal{N}(\theta + \kappa\delta, a^{-1}), \qquad D = \frac{U_\delta - (b/a)U_\theta}{\rho} \sim \mathcal{N}(\delta, \rho^{-1}). \tag{8}$$

The statistic T measures the target with residual bias, while D provides an independent estimate of the discrepancy. Here a^{-1} is the target sampling variance, ρ is the precision for learning the discrepancy, and κ measures how strongly discrepancy errors affect the target estimate. If $\widehat{\delta}_h(D)$ is a minimax bounded-mean rule, then $T - \kappa\widehat{\delta}_h(D)$ attains Equation (5). When $\rho = 0$, there is no independent information about δ, and the correction is zero. A diffuse proper prior on θ, combined with approximately least-favourable priors on δ, gives the matching lower bound; Section C provides the details, including singular information matrices.

Two special cases help interpret the identity. If $b = 0$, the risk is $1/a$ and no bias correction is needed. If $\rho = 0$ but $b \neq 0$, the data cannot separate the target from the discrepancy, leaving a contribution $\kappa^2 h^2$ to the risk. Increasing target precision is therefore only one part of the acquisition problem.

2.3 Affine estimation and extensions

The best affine estimate of δ is $h^2\rho D/(1 + h^2\rho)$, with worst-case risk $h^2/(1 + h^2\rho)$. Substituting it in $T - \kappa\widehat{\delta}(D)$ gives Equation (6). The factor $5/4$ follows from the classical comparison between linear and unrestricted Gaussian minimax estimation (Donoho et al., 1990; Donoho, 1994). Since J_h is concave, its reciprocal is convex wherever finite. For a compact polyhedral design set $\mathcal{X}$, the best affine allocation is therefore obtained from the semidefinite program

$$\begin{aligned} \min_{x \in \mathcal{X}, s, t} \quad & t \\ \text{s.t.} \quad & \begin{bmatrix} a(x) - s & b(x) \\ b(x) & c(x) + h^{-2} \end{bmatrix} \succeq 0, \qquad \begin{bmatrix} t & 1 \\ 1 & s \end{bmatrix} \succeq 0. \end{aligned} \tag{9}$$

This gives a tractable acquisition and estimation procedure whose worst-case risk is at most $5/4$ of the jointly optimized minimax value. The exact allocation analysis below uses Equation (5).

For an ellipsoidal discrepancy $\delta^\top H^{-1}\delta \leq 1$ in $\mathbb{R}^d$, with $H \succ 0$, write $M = \left[\begin{smallmatrix} a & b^\top \\ b & C \end{smallmatrix}\right]$, $k = b/a$, and $R = C - bb^\top/a$. The exact and affine risks are

$$a^{-1} + m_{H,k}(R) \quad \text{and} \quad \left\{a - b^\top (C + H^{-1})^{-1} b\right\}^{-1},$$

where $m_{H,k}(R)$ is the minimax risk for estimating $k^\top\delta$ from $W \sim \mathcal{N}(R\delta, R)$. The $5/4$ bound and semidefinite formulation extend, and a finite set of types without type-specific caps admits an exact continuous optimum using at most $(d + 1)(d + 2)/2$ types; see Theorem 11.

Corollary 2 (Independent tasks)**.** *Consider conditionally independent scalar observation blocks with separate interval discrepancies and a shared design. Under a product parameter set and additive squared loss with nonnegative weights, the exact minimax risk is the weighted sum of the blockwise values in Equation* (5)*. Minimizing the corresponding weighted sum of affine risks is convex and remains within* $5/4$ *of the jointly optimized minimax value.*

The corollary follows from products of approximately least-favourable priors. The acquisition theory below concerns multiple types sharing one scalar discrepancy.

3 Optimal budget allocation

3.1 Information per unit cost

Normalize the trusted information rate to one, so its information per unit expenditure is $A_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Dividing each auxiliary type's information in Equation (2) by its cost gives

$$A_j = \begin{bmatrix} p_j & q_j \\ q_j & r_j \end{bmatrix} \succeq 0, \qquad p_j \geq 0, \quad r_j > 0, \quad j = 1, \ldots, m. \tag{10}$$

The entries describe target information, target–bias coupling, and bias information, respectively. A source-only observation of $\theta + \ell_j \delta$, with $\ell_j \neq 0$, gives a rank-one matrix. A paired observation may give a positive-definite matrix. Calibration-only observations have $p_j = q_j = 0$ and $r_j > 0$. Any better pure-target type can be absorbed into the normalized trusted type.

Let $y_j \geq 0$ denote expenditure on auxiliary type j, chosen before observing the data. There are no type-specific caps, and $\sum_j y_j \leq B$; the remaining budget is spent on trusted observations. Define

$$E(y) = \sum_j (p_j - 1) y_j, \qquad Q(y) = \sum_j q_j y_j, \qquad R(y) = \sum_j r_j y_j. \tag{11}$$

Here $E(y)$ is the change in target information relative to spending the same amount on trusted data. The total information is $M_B(y) = \left[\begin{smallmatrix} B + E(y) & Q(y) \\ Q(y) & R(y) \end{smallmatrix}\right]$, so Theorem 1 gives

$$\mathcal{R}_B(y) = \frac{1}{B + E(y)} + \left\{ \frac{Q(y)}{B + E(y)} \right\}^2 \mathfrak{m}_h \left(R(y) - \frac{Q(y)^2}{B + E(y)} \right). \tag{12}$$

We set $\mathcal{R}_B(y) = +\infty$ when $B + E(y) = 0$. The optimal risk is

$$V_B = \min_{y \geq 0:\, \mathbf{1}^\top y \leq B} \mathcal{R}_B(y). \tag{13}$$

Two information rates determine the allocation regimes. For auxiliary proportions $\lambda \in \Delta_m$, where Δ_m is the probability simplex, write $A(\lambda) = \sum_j \lambda_j A_j = \left[\begin{smallmatrix} P(\lambda) & Q(\lambda) \\ Q(\lambda) & R(\lambda) \end{smallmatrix}\right]$ and define

$$\Gamma_0 = \max_j p_j, \qquad \Gamma_\infty = \max_{\lambda \in \Delta_m} s(\lambda), \qquad s(\lambda) = P(\lambda) - \frac{Q(\lambda)^2}{R(\lambda)}. \tag{14}$$

Γ_0 is the best auxiliary target information rate when the bias is known. Γ_∞ is the best rate when the bias must be treated as unrestricted. Although $\Gamma_\infty \leq \Gamma_0$, the mixture maximizing Γ_∞ may outperform every individual type by reducing the aggregate coupling $Q(\lambda)$.

Computing Γ_0 requires only a comparison of the p_j. The mixture rate is the value of a semidefinite program,

$$\max_{\lambda, t} \; t \quad \text{s.t.} \quad \lambda \geq 0, \quad \mathbf{1}^\top \lambda = 1, \quad \begin{bmatrix} P(\lambda) - t & Q(\lambda) \\ Q(\lambda) & R(\lambda) \end{bmatrix} \succeq 0. \tag{15}$$

Both rates are compared with the trusted rate of one. For every fixed finite $h > 0$, the regime boundaries depend only on these information rates; the bias bound determines the risk and expenditure within a regime.

3.2 ACQUISITION REGIMES

Theorem 3 (Optimal acquisition regimes)**.** *Under Equation* (10) *and* $0 < h < \infty$*, the following statements hold.*

1. ***Trusted-only acquisition.*** *Spending the entire budget on trusted observations is optimal for every* $B > 0$ *if and only if* $\Gamma_0 \leq 1$*. In this case,* $V_B = B^{-1}$*.*

2. ***Bounded auxiliary spending.*** *If* $\Gamma_\infty < 1 < \Gamma_0$*, every sequence of exact minimizers* y_B *has uniformly bounded auxiliary expenditure as* $B \to \infty$*. Define*

$$\mathcal{C}(y) = E(y) - Q(y)^2 \mathfrak{m}_h(R(y)), \qquad C_* = \max_{y \geq 0} \mathcal{C}(y). \tag{16}$$

   Then $C_* > 0$*, its maximizing set is nonempty and compact, and every accumulation point of* y_B *maximizes* $\mathcal{C}$*. Moreover,*

$$V_B = B^{-1} - C_* B^{-2} + o(B^{-2}). \tag{17}$$

   At least one limiting optimal allocation uses at most two auxiliary types.

3. ***Auxiliary dominance.*** *If* $\Gamma_\infty > 1$*, every sequence of exact minimizers satisfies*

$$\frac{\mathbf{1}^\top y_B}{B} \longrightarrow 1, \qquad BV_B \longrightarrow \Gamma_\infty^{-1}.$$

   The auxiliary proportions approach the maximizing set in Equation (14)*.*

The bounded-spending regime, also called saturation, explains why useful auxiliary data need not receive a fixed budget share. A small purchase can increase target information, while every auxiliary mixture has a lower asymptotic information rate than trusted data after accounting for the unknown discrepancy. The optimal compromise keeps auxiliary expenditure bounded. Its benefit appears in the B^{-2} term, which a comparison of leading B^{-1} risks would miss. The bound concerns absolute expenditure: a vanishing budget share would still allow spending of order $\sqrt{B}$, which is excluded for exact optima in this regime.

For fixed B, moving a small amount u from trusted observations to type j gives risk $B^{-1} - (p_j - 1)uB^{-2} + O(u^2)$. Thus any $p_j > 1$ makes a small purchase beneficial. At large budgets, a mixture λ receiving share t has normalized information $(1-t)A_0 + tA(\lambda)$, whose effective target information is $1 + t\{s(\lambda) - 1\}$. This determines whether a positive auxiliary share improves the leading risk. In the saturation regime, the uniform gap $1 - \Gamma_\infty$ also prevents optimal auxiliary expenditure from diverging. Expanding Equation (12) uniformly on the resulting bounded set yields Equations (16) and (17). Complete proofs appear in Section F.

3.3 SPARSE OPTIMAL DESIGNS

Proposition 4 (Finite-budget support)**.** *For every* $B > 0$*, an exact minimax allocation exists using at most three observation types in total, including the trusted type when used. If one type uniquely maximizes* p_j *among* $p_0 := 1, p_1, \ldots, p_m$*, then buying only that type is the unique optimum for all sufficiently small positive budgets.*

The support bound holds even though Equation (12) can be nonconvex. Increasing information in the positive-semidefinite order cannot increase minimax risk. An optimum can therefore be selected on an exposed face of the feasible information set, which lies in the three-dimensional space of symmetric 2×2 matrices. This gives the three-type bound. When all auxiliary types have the same nonzero ratio q_j/r_j, some finite-budget optimum uses at most two total types, and a limiting saturation optimum can use just one auxiliary type; see Corollary 29.

4 CRITICAL ALLOCATIONS AND TRANSITION WINDOWS

At $\Gamma_\infty = 1$, the best auxiliary mixtures match the trusted information rate when the bias is unrestricted. A finite bias bound can still improve risk, requiring a second-order analysis.

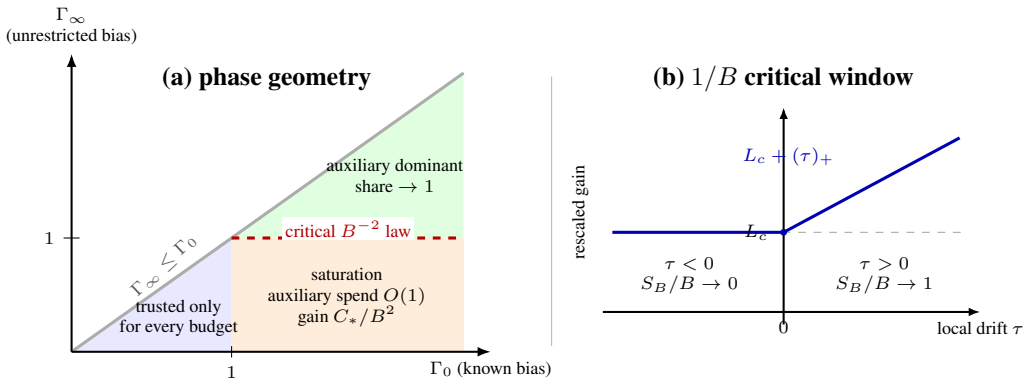


Figure 1: Acquisition regimes and the critical window. (a) Γ_0 determines whether auxiliary data can improve on trusted-only acquisition. Γ_∞ separates bounded auxiliary spending from a budget share approaching one. The dashed line is the critical boundary. (b) For observation types approaching this boundary with $B(\Gamma_{\infty,B} - 1) \to \tau$, the rescaled risk improvement converges to $L_c + (\tau)_+$, where $L_c = \pi^2/H_c^2$. Positive and negative drift lead to different limiting auxiliary budget shares.

4.1 The critical boundary

Let $\Lambda_c = \{\lambda \in \Delta_m : s(\lambda) = 1\}$ denote the critical mixtures.

Theorem 5 (Critical geometry and risk). *Suppose $\Gamma_\infty = 1$. If $\Gamma_0 = 1$, then $V_B = B^{-1}$ for every $B > 0$. If $\Gamma_0 > 1$, every $\lambda \in \Lambda_c$ has $Q(\lambda) \neq 0$, the quantity $H_c = h|R(\lambda)/Q(\lambda)|$ is independent of λ, and*

$$V_B = B^{-1} - \frac{\pi^2}{H_c^2 B^2} + o(B^{-2}). \tag{18}$$

In both cases, for every $B > 0$ an exact optimum exists that spends the entire budget on at most three auxiliary types. Every critical mixture satisfies

$$A(\lambda) - A_0 = \begin{bmatrix} Q(\lambda)/\sqrt{R(\lambda)} \\ \sqrt{R(\lambda)} \end{bmatrix} \begin{bmatrix} Q(\lambda)/\sqrt{R(\lambda)} \\ \sqrt{R(\lambda)} \end{bmatrix}^\top \succeq 0. \tag{19}$$

Equation (19) shows that replacing trusted expenditure by a critical auxiliary mixture cannot increase risk. The matrices $A(\lambda) - A_0$ also have the same one-dimensional range: averaging two different ranges would produce a mixture with effective target information greater than one. When $\Gamma_0 > 1$, this common direction has nonzero target–bias coupling and gives the radius H_c.

The second-order constant can be seen directly along a critical mixture λ_c. Write $Q_c = Q(\lambda_c)$ and $R_c = R(\lambda_c)$. When $\Gamma_0 > 1$, set $v = SQ_c^2/R_c$ and $w = Bv/(B + v)$ for expenditure $0 \le S \le B$. Then

$$\mathcal{R}_B(S\lambda_c) = \frac{1}{B} - \frac{\Psi_{H_c}(w)}{B^2}, \qquad \Psi_H(w) = w - w^2\mathfrak{m}_H(w). \tag{20}$$

The bounded-normal expansion gives $\Psi_H(w) \uparrow \pi^2/H^2$ as $w \to \infty$ (Bickel, 1981). Hence any diverging expenditure $S_B \le B$ along a fixed coupled critical mixture attains the same second-order coefficient. This includes both sublinear and linear expenditure. The theorem guarantees an exact optimum using only auxiliary observations. The second-order risk expansion also permits other expenditure scales. In the uncoupled example $A_1 = \mathrm{diag}(1, r)$, every split has risk $1/B$. The global comparison needed for Equation (18) is proved in Proposition 21.

4.2 A transition window of width $1/B$

To study observations close to the boundary, allow their information matrices to vary with the budget. For a fixed number of types, assume

$$A_{j,B} = \begin{bmatrix} p_{j,B} & q_{j,B} \\ q_{j,B} & r_{j,B} \end{bmatrix} \succeq 0, \qquad r_{j,B} \ge \underline{r} > 0, \quad p_{j,B} \le \overline{p} < \infty, \qquad \max_j \|A_{j,B} - A_j\| \to 0.$$

Write $A_B(\lambda) = \sum_j \lambda_j A_{j,B}$, with entries $P_B(\lambda)$, $Q_B(\lambda)$, and $R_B(\lambda)$, and let

$$s_B(\lambda) = P_B(\lambda) - \frac{Q_B(\lambda)^2}{R_B(\lambda)}, \qquad \Gamma_{\infty,B} = \max_{\lambda \in \Delta_m} s_B(\lambda).$$

Denote the corresponding exact optimal risk by $V_B^{\triangle}$.

Theorem 6 (Local critical window)**.** *Suppose the limiting observation types satisfy $\Gamma_\infty = 1 < \Gamma_0$, with common critical radius H_c. If*

$$B\{\Gamma_{\infty,B} - 1\} \longrightarrow \tau \in \mathbb{R}, \tag{21}$$

then

$$V_B^{\triangle} = \frac{1}{B} - \frac{\pi^2/H_c^2 + (\tau)_+}{B^2} + o(B^{-2}), \qquad (\tau)_+ = \max\{\tau, 0\}. \tag{22}$$

For exact minimizers y_B, let $S_B = \mathbf{1}^\top y_B$ and $\lambda_B = y_B/S_B$ when $S_B > 0$. If $\tau > 0$, then

$$\frac{S_B}{B} \to 1, \qquad B\{\Gamma_{\infty,B} - s_B(\lambda_B)\} \to 0. \tag{23}$$

If $\tau < 0$, then $S_B/B \to 0$. The theorem determines the limiting share; finer expenditure scales for $\tau \le 0$ depend on additional properties of the observation types.

A $1/B$ change in the information rate affects the B^{-2} risk term. Positive drift favors full auxiliary spending; negative drift allows diverging sublinear expenditure to retain the critical gain without a second-order penalty. The proof compares a diverging sublinear allocation with full auxiliary spending. A limiting share t adds at most $t\tau$ to π^2/H_c^2, giving $(\tau)_+$ after maximizing over $t \in [0, 1]$. Uniform bounds appear in Section F.6.

Corollary 7 (Local drift)**.** *Suppose, uniformly in j,*

$$A_{j,B} = A_j + B^{-1}\dot{A}_j + o(B^{-1}), \qquad \dot{A}_j = \begin{bmatrix} \dot{p}_j & \dot{q}_j \\ \dot{q}_j & \dot{r}_j \end{bmatrix}.$$

Writing $\dot{P}$, $\dot{Q}$, and $\dot{R}$ for the corresponding mixture entries, Equation (21) *holds with*

$$\tau = \max_{\lambda \in \Lambda_c} \left\{ \dot{P}(\lambda) - 2\frac{Q(\lambda)}{R(\lambda)}\dot{Q}(\lambda) + \frac{Q(\lambda)^2}{R(\lambda)^2}\dot{R}(\lambda) \right\}. \tag{24}$$

If $\tau > 0$, every accumulation point of the exact optimal auxiliary proportions maximizes this derivative over Λ_c.

For example, take $A_c = A_0 + zz^\top$ with both coordinates of z nonzero and $A_{1,B} = A_c + \tau A_0/B$. For sufficiently large B, this is a valid observation type with $\Gamma_{\infty,B} = 1 + \tau/B$, exhibiting both sides of the window.

5 Examples and allocation analysis

We illustrate the regimes with exact calculations, then examine sensitivity to information errors and integer counts.

5.1 A finite-budget example

Consider independent trusted observations $N(\theta, 1)$ costing 1 and source observations $N(\theta + \delta, 1)$ costing $1/4$, with $|\delta| \le 1/4$. The auxiliary information per unit cost is $A_1 = 4[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}]$, so $\Gamma_0 = 4$ and $\Gamma_\infty = 0$. The problem is therefore in the saturation regime.

For source expenditure $y \in [0, B]$, the affine information criterion is

$$J_{1/4}(y) = B - y + \frac{16y}{y + 4}, \qquad y^*_{\text{aff}} = \min\{B, 4\}.$$

Once $B \ge 4$, the best affine allocation buys 16 source observations and $B - 4$ trusted observations, giving worst-case MSE $1/(B + 4)$. Table 1 shows that the source count remains fixed as the budget grows. At $B = 16$, the mixed allocation reduces MSE by 20% relative to trusted-only acquisition, while source-only acquisition increases it by 25%. The latter has risk $1/(4B) + 1/16$: additional source observations reduce sampling variance but leave the shared-bias contribution unchanged. These are exact affine calculations; the unrestricted minimax optimum obeys $V_B \le \mathcal{R}^*_{\text{aff}} \le (5/4)V_B$ and need not coincide with the affine optimum.

Budget B	Counts (n_T, n_S)	Trusted only	Source only	Best affine
4	$(0, 16)$	0.250000	0.125000	0.125000
16	$(12, 16)$	0.062500	0.078125	0.050000
64	$(60, 16)$	0.015625	0.066406	0.014706

Table 1: Worst-case MSE at three budgets. The counts refer to the best affine allocation. Trusted-only and source-only risks are $1/B$ and $1/(4B) + 1/16$, respectively.

5.2 Complementary sources

The mixture criterion in Equation (14) accounts for two ways in which auxiliary types can help one another: cancelling opposite bias couplings and supplying calibration information.

Example 8 (Cancellation of bias coupling)**.** Let

$$A_+ = \begin{bmatrix} 5/3 & 4/3 \\ 4/3 & 4/3 \end{bmatrix}, \qquad A_- = \begin{bmatrix} 5/3 & -4/3 \\ -4/3 & 4/3 \end{bmatrix}.$$

Each type has effective target information $1/3 < 1$ when the discrepancy is unrestricted. Their equal mixture is $\mathrm{diag}(5/3, 4/3)$, giving $\Gamma_\infty = 5/3$. The opposite couplings cancel, and every optimal auxiliary budget share approaches one. Each type alone would instead produce saturation. Both matrices arise from positive-definite paired-observation covariances, as constructed in Section F.

Example 9 (Calibration changes the regime)**.** Let

$$A_1 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 100 \end{bmatrix}.$$

Type 1 alone has $\Gamma_0 = 2$ and $\Gamma_\infty = 0$, so its optimal expenditure is bounded. Type 2 measures only the discrepancy. Allocating auxiliary proportions $(3/4, 1/4)$ gives effective target information

$$\frac{3}{2} - \frac{(3/2)^2}{53/2} = \frac{75}{53} > 1.$$

Thus calibration changes the problem from saturation to auxiliary dominance, despite providing no direct target information.

These examples show why screening each type separately can discard useful combinations. The sign of q_j matters through the aggregate coupling, while a large r_j can make another type's target information usable.

5.3 Sensitivity and integer budgets

The stability of the regime classification depends on the distance from the thresholds. Suppose two collections of positive-semidefinite information matrices satisfy $r_j, \widetilde{r}_j \geq \underline{r} > 0$, $p_j, \widetilde{p}_j \leq \overline{p}$, and their corresponding entries differ by at most ε. Then

$$|\widetilde{\Gamma}_0 - \Gamma_0| \leq \varepsilon, \qquad |\widetilde{\Gamma}_\infty - \Gamma_\infty| \leq \left(1 + \sqrt{\overline{p}/\underline{r}}\right)^2 \varepsilon. \tag{25}$$

For a simultaneous entrywise error bound, write $K_\varepsilon = (1 + \sqrt{\overline{p}/\underline{r}})^2 \varepsilon$. The bounds imply trusted-only acquisition when $\widetilde{\Gamma}_0 + \varepsilon \leq 1$, saturation when $\widetilde{\Gamma}_\infty + K_\varepsilon < 1 < \widetilde{\Gamma}_0 - \varepsilon$, and auxiliary dominance when $\widetilde{\Gamma}_\infty - K_\varepsilon > 1$. If the error intervals cross a threshold, the bounds do not identify a strict regime. Identifying the local drift τ from these bounds requires errors of order $o(B^{-1})$. The perturbation argument is given in Proposition 30.

Integer counts preserve the leading risk regimes, but can change the second-order gain. If every observation costs at most $c_{\max}$, the three-type support bound gives $V_{\mathrm{cont}}(B) \leq V_{\mathbb{Z}}(B) \leq V_{\mathrm{cont}}(B - 3c_{\max})$ for $B > 3c_{\max}$; see Proposition 31. For the sharper saturation result, normalize the trusted replicate to cost and information one, and let an auxiliary replicate of type j cost an integer $c_j \geq 1$ and have information $c_j A_j$. Define

$$\mathcal{L} = \{(c_1 n_1, \ldots, c_m n_m) : n \in \mathbb{Z}_+^m\}, \qquad C_{\mathbb{Z}} = \max_{y \in \mathcal{L}} \mathcal{C}(y). \tag{26}$$

Proposition 10 (Integer second-order risk). *If $\Gamma_\infty < 1 < \Gamma_0$, the exact minimax value at integer budget N satisfies*

$$V_N^{\mathbb{Z}} = N^{-1} - C_{\mathbb{Z}} N^{-2} + o(N^{-2}), \qquad 0 \le C_{\mathbb{Z}} \le C_*. \tag{27}$$

The maximum defining $C_{\mathbb{Z}}$ is attained. For all sufficiently large N, every exact integer optimizer belongs to its maximizing set. The constant $C_{\mathbb{Z}}$ can be zero even when $C_* > 0$.

The distinction matters when the continuous optimum calls for a fractional observation. Restricting acquisition to whole observations can then remove a second-order benefit while leaving the leading $1/N$ risk unchanged.

The Gaussian model gives finite-sample guarantees for a specified bias interval, known covariances, and acquisition fixed before observing the data. These guarantees also cover bounded auxiliary counts in the saturation regime, where Gaussian approximations based on diverging counts do not determine the second-order constant. Extending the theory to estimated covariances, data-dependent bias bounds, and adaptive acquisition is a natural next step.

6 Related work

Prediction-powered acquisition. PPI uses trusted labels to correct prediction bias and obtain valid inference (Angelopoulos et al., 2023a). Active and cost-aware methods choose which labels to collect (Zrnic & Candès, 2024a; Angelopoulos et al., 2025), and MultiPPI optimizes allocation across correlated sources with known covariance (Cowen-Breen et al., 2026). Our analysis introduces a finite bound on their shared discrepancy and determines how it changes both source selection and the growth of optimal expenditure.

Estimation with bounded bias. Classical normal-mean theory quantifies the gain from a bounded parameter space (Casella & Strawderman, 1981). Combining estimators with bounded mean differences and adapting to misspecification use related shrinkage principles (van Eeden & Zidek, 2001; 2004; Armstrong et al., 2025). We optimize the resulting minimax risk over data acquisition. This yields the bounded-spending regime and its second-order gain, together with the critical transition that fixed-design estimation alone does not characterize.

Design with auxiliary information. Multifidelity methods exploit correlations through control variates and allocate samples according to cost and variance (Peherstorfer et al., 2016; Gorodetsky et al., 2020). Experimental design has also been adapted to shrinkage estimators and to combining experimental with potentially biased observational evidence (Rosenman & Miratrix, 2023; Epanomeritakis & Viviano, 2026). We jointly optimize acquisition and estimation under a fixed shared bias bound. The resulting analysis distinguishes useful finite auxiliary purchases from mixtures worth collecting at scale, and resolves the transition between them.

Optimal design and sparse support. Classical optimal design uses the geometry of information matrices to characterize efficient allocations (Elfving, 1952; Pukelsheim, 2006). Our objective includes the minimax error of estimating a bounded discrepancy, making it generally nonconvex. Using monotonicity in information, we prove that an exact optimum can be attained using at most three observation types. The bounds sharpen for saturation and aligned couplings. Further connections appear in Section G.

7 Conclusion

We developed a minimax acquisition theory for trusted and auxiliary observations with a shared bounded bias. An exact Gaussian reduction gives the risk for each design and a convex allocation problem for affine estimators with a $5/4$ guarantee. Two information rates determine the acquisition regimes, including bounded auxiliary spending. The second-order analysis resolves the saturation gain and critical window, and an exact optimum can be attained using at most three observation types. Future work may extend these results to learned covariance matrices and adaptive acquisition.

Reproducibility statement. The appendix provides proofs, boundary conventions, and the extensions to vector discrepancies and integer allocations. The values in Table 1 follow from the explicit risk formulas in Section 5; the remaining examples give their information matrices and calculations directly. The anonymous supplement implements the exact information ledger and reproduces the rational constructions used in the examples.

Ethics statement. This work is primarily theoretical and mathematical. It does not involve human subjects, personal data, crowdsourcing, or the release of datasets or models. Consequently, ethical concerns related to informed consent, privacy, and data licensing are not applicable.

AI use statement. A generative language-model was used for language polishing. All AI-assisted parts were reviewed and verified by the authors, who take full responsibility for the content of this paper.

REFERENCES

Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal design of experiments via regret minimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 126–135. PMLR, 2017. URL https://proceedings.mlr.press/v70/allen-zhu17e.html.

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

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

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

Timothy B. Armstrong, Patrick Kline, and Liyang Sun. Adapting to misspecification. *Econometrica*, 93(6):1981–2005, 2025. doi: 10.3982/ECTA21991.

Peter J. Bickel. Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *The Annals of Statistics*, 9(6):1301–1309, 1981. doi: 10.1214/aos/1176345646. URL https://doi.org/10.1214/aos/1176345646.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://web.stanford.edu/~boyd/cvxbook/.

Lawrence D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics*, 42(3):855–903, 1971. doi: 10.1214/aoms/1177693318.

George Casella and William E. Strawderman. Estimating a bounded normal mean. *The Annals of Statistics*, 9(4):870–878, 1981. doi: 10.1214/aos/1176345527.

Shuxiao Chen, Sai Li, Bo Zhang, and Ting Ye. Minimax rates and adaptivity in combining experimental and observational data. *Journal of Causal Inference*, 13(1):20240024, 2025. doi: 10.1515/jci-2024-0024. URL https://doi.org/10.1515/jci-2024-0024.

Xue Chen and Eric Price. Active regression via linear-sample sparsification. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 663–695. PMLR, 2019. URL https://proceedings.mlr.press/v99/chen19a.html.

Charlie Cowen-Breen, Alekh Agarwal, Stephen Bates, William W. Cohen, Jacob Eisenstein, Amir Globerson, and Adam Fisch. Multiple-prediction-powered inference. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=gJZ5rf2bS4.

Michał Dereziński, Kenneth L. Clarkson, Michael W. Mahoney, and Manfred K. Warmuth. Minimax experimental design: Bridging the gap between statistical and worst-case approaches to least squares regression. In *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1050–1069. PMLR, 2019. URL https://proceedings.mlr.press/v99/derezinski19b.html.

David L. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1): 238–270, 1994. doi: 10.1214/aos/1176325367.

David L. Donoho and Richard C. Liu. Geometrizing rates of convergence, III. *The Annals of Statistics*, 19(2):668–701, 1991. doi: 10.1214/aos/1176348115.

David L. Donoho, Richard C. Liu, and Brenda MacGibbon. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, 18(3):1416–1437, 1990. doi: 10.1214/aos/1176347758.

Yonina C. Eldar, Aharon Ben-Tal, and Arkadi Nemirovski. Robust mean-squared error estimation in the presence of model uncertainties. *IEEE Transactions on Signal Processing*, 53(1):168–181, 2005. doi: 10.1109/TSP.2004.838933.

Gustav Elfving. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, 23(2):255–262, 1952. doi: 10.1214/aoms/1177729442.

Aristotelis Epanomeritakis and Davide Viviano. Learning what to learn: Experimental design when combining experimental with observational evidence, 2026. URL https://arxiv.org/abs/2510.23434v6. arXiv preprint, version 6, 18 August 2026; first posted in 2025.

Jianqing Fan and Jin-Ting Zhang. A note on the bounded normal mean problem. In *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A. Doksum*, pp. 635–647. World Scientific, 2007. doi: 10.1142/9789812708298_0031. URL https://doi.org/10.1142/9789812708298_0031.

Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.

Adam Fisch, Joshua Maynez, R. Alex Hofer, Bhuwan Dhingra, Amir Globerson, and William W. Cohen. Stratified prediction-powered inference for effective hybrid evaluation of language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3541. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c9fcd02e6445c7dfbad6986abee53d0d-Abstract-Conference.html.

Xavier Fontaine, Pierre Perrault, Michal Valko, and Vianney Perchet. Online a-optimal design and active linear regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3374–3383. PMLR, 2021. URL https://proceedings.mlr.press/v139/fontaine21a.html.

Alex A. Gorodetsky, Gianluca Geraci, Michael S. Eldred, and John D. Jakeman. A generalized approximate control variate framework for multifidelity uncertainty quantification. *Journal of Computational Physics*, 408:109257, 2020. doi: 10.1016/j.jcp.2020.109257. URL https://doi.org/10.1016/j.jcp.2020.109257.

Eric Gourdin, Brigitte Jaumard, and Brenda MacGibbon. Global optimization decomposition methods for bounded parameter minimax risk evaluation. *SIAM Journal on Scientific Computing*, 15(1): 16–35, 1994. doi: 10.1137/0915002.

Dongning Guo, Yihong Wu, Shlomo Shamai, and Sergio Verdú. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4): 2371–2385, 2011. doi: 10.1109/TIT.2011.2111010.

Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4. URL https://www.cambridge.org/core/journals/canadian-journal-of-mathematics/article/equivalence-of-two-extremum-problems/B8B0626C11F52B0FD8C67C5D54BDDD43.

José Luis Montiel Olea and Ekaterina Zubova. Approximate minimax estimation of a bounded normal mean via stochastic mirror ascent, 2026. URL https://arxiv.org/abs/2607.05350. arXiv preprint.

Michael Oberst, Alexander D'Amour, Minmin Chen, Yuyan Wang, David Sontag, and Steve Yadlowsky. Understanding the risks and rewards of combining unbiased and possibly biased estimators, with applications to causal inference, 2022. URL https://arxiv.org/abs/2205.10467. arXiv preprint, revised May 2023.

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

Friedrich Pukelsheim. *Optimal Design of Experiments*, volume 50 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, 2006. doi: 10.1137/1.9780898719109.

Friedrich Pukelsheim and Sabine Rieder. Efficient rounding of approximate designs. *Biometrika*, 79 (4):763–770, 1992. doi: 10.1093/biomet/79.4.763. URL https://doi.org/10.1093/biomet/79.4.763.

Evan T. R. Rosenman and Luke Miratrix. Designing experiments toward shrinkage estimation. *Electronic Journal of Statistics*, 17(2):3406–3442, 2023. doi: 10.1214/23-EJS2179.

Evan T. R. Rosenman, Guillaume Basse, Art B. Owen, and Mike Baiocchi. Combining observational and experimental datasets using shrinkage estimators. *Biometrics*, 79(4):2961–2973, 2023. doi: 10.1111/biom.13827.

Jerome Sacks and Donald Ylvisaker. Linear estimation for approximately linear models. *The Annals of Statistics*, 6(5):1122–1137, 1978. doi: 10.1214/aos/1176344315.

Sabina J. Sloman, Ayush Bharti, Julien Martinelli, and Samuel Kaski. Bayesian active learning in the presence of nuisance parameters. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pp. 3245–3263, 2024. URL https://proceedings.mlr.press/v244/sloman24a.html.

A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959. doi: 10.1016/S0019-9958(59)90348-1.

Constance van Eeden and James V. Zidek. Estimating one of two normal means when their difference is bounded. *Statistics & Probability Letters*, 51(3):277–284, 2001. doi: 10.1016/S0167-7152(00)00159-0.

Constance van Eeden and James V. Zidek. Combining the data from two normal populations to estimate the mean of one when their means difference is bounded. *Journal of Multivariate Analysis*, 88(1):19–46, 2004. doi: 10.1016/S0047-259X(03)00049-6.

Harry L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, New York, 1968.

Douglas P. Wiens. Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, 31(3):353–371, 1992. doi: 10.1016/0378-3758(92)90142-F. URL https://doi.org/10.1016/0378-3758(92)90142-F.

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

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

A Appendix Overview

The appendix provides the technical details and proofs supporting the main results, together with extensions and further comparisons with related work. Section B summarizes the observation model, allocation conventions, and risk definitions used throughout the paper. Section C proves the scalar minimax reduction. Section D presents the extension to ellipsoidal vector discrepancies, with proofs in Section E. Section F establishes the optimal acquisition regimes and their second-order refinements. Section G provides further comparisons with related work.

B Model conventions and estimation

B.1 Observation types and continuous allocations

An observation type specifies the measurements returned by one purchase and their joint covariance. For independent replicates of $Z_j \sim N(F_j(\theta,\delta)^\top, \Sigma_j)$, a positive integer count x_j gives a sample mean with covariance Σ_j/x_j. Approximate design extends this information law to real $x_j \geq 0$. Information adds across independent purchases, while dependence within an observation is retained in Σ_j. An invertible transformation $(F_j, \Sigma_j) \mapsto (R_j F_j, R_j \Sigma_j R_j^\top)$ leaves $F_j^\top \Sigma_j^{-1} F_j$ unchanged. Thus the location and difference statistics derived in the proofs describe the same purchased observations and carry the same total information.

For example, a source-only observation $Z_j \sim N(\theta + \ell_j \delta, v_j)$ with $\ell_j \neq 0$ has the rank-one information matrix $v_j^{-1}(1,\ell_j)^\top(1,\ell_j)$. A paired observation can instead have a positive-definite information matrix. Both are covered by Theorem 3.

A type with $x_j = 0$ contributes no observation. All allocations are deterministic and chosen before the data used for estimation are collected. The main acquisition results assume a finite set of auxiliary types, one trusted type, known information matrices, and no type-specific caps. The trusted type receives the budget left after auxiliary purchases. Spending the full budget is optimal because additional independent information cannot increase minimax risk. Proposition 31 relates these continuous allocations to integer replication at fixed costs.

B.2 Risk and singular information

The target mean is unrestricted. The following conventions cover the boundary cases for $M = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right] \succeq 0$:

Boundary	Risk or information
$a = 0$	Positive semidefiniteness forces $b = 0$. The law is independent of θ, so the risk is $+\infty$.
$a > 0$, $\rho = 0$	The residual statistic contains no information about the discrepancy, and $\mathfrak{m}_h(0) = h^2$.
$h = 0$	The discrepancy is known, giving target information a.
$h = \infty$	The effective target information is the generalized Schur complement $a - b^2 c^\dagger$. If $c = 0$, positive semidefiniteness forces $b = 0$.
Zero loss weight	The corresponding task contributes zero risk.

All risks are total mean-squared errors. The unrestricted decision class includes all measurable estimators; allowing randomization leaves the minimax value unchanged under squared loss. The phase results assume $0 < h < \infty$. At $h = 0$, the discrepancy is known; at $h = \infty$, its estimation is governed by the generalized Schur complement. These two boundary models help interpret the information rates Γ_0 and Γ_∞.

B.3 Acquisition assumptions and extensions

The scalar theory allows rank-one source observations, correlated pairs, and calibration-only observations with no target loading. The condition $r_j > 0$ identifies the types carrying discrepancy

information. A type with $r_j = 0$ has $q_j = 0$ by positive semidefiniteness and supplies only target information; the best such type can be used to normalize the trusted rate.

The local critical-window theorem allows information matrices to vary with budget, with a fixed number of types and discrepancy-information entries uniformly bounded away from zero. For a nontrivial critical limit with $\Gamma_0 > 1$, the drift $B(\Gamma_{\infty,B} - 1)$ determines the additional second-order gain. Integer replication preserves the leading risk regimes, while its second-order saturation gain is characterized by the lattice constant in Proposition 10. The vector extension establishes the fixed-design risk, the affine guarantee, and a dimension-dependent support bound for ellipsoidal discrepancies.

These results connect the exact risk of estimating a bounded discrepancy to the allocation of a growing budget. The two information rates determine when auxiliary purchases are useful and whether their optimal expenditure stays bounded or grows with the budget. The critical geometry and local window then resolve the transition at second order.

B.4 Estimators for the selected design

For a chosen information matrix, form T and, when $\rho > 0$, D as in Equation (8). If $\widehat{\delta}_h(D)$ is a minimax rule for the bounded normal mean problem, then

$$\widehat{\theta}_{\mathrm{mm}} = T - \frac{b}{a}\widehat{\delta}_h(D) \tag{28}$$

attains the exact risk. Least-favourable-prior methods provide numerical approximations to this one-dimensional rule (Gourdin et al., 1994; Montiel Olea & Zubova, 2026). The affine alternative uses $\widehat{\delta}_{\mathrm{aff}}(D) = h^2\rho D/(1 + h^2\rho)$ and gives the $5/4$ risk guarantee. When $\rho = 0$, the discrepancy estimate is the constant zero.

The interval $[-h, h]$ bounds a persistent shared discrepancy, and the risk guarantee is uniform over this interval for the stated covariance model. Estimating the covariance or bias radius, and choosing purchases adaptively from the same data, introduce additional statistical uncertainty. These extensions motivate further work on learning and acquisition within the minimax framework.

C Proofs for the scalar information geometry

We prove Theorem 1 by separating target and discrepancy information, deriving the exact minimax risk, and then optimizing the affine rule. The argument covers singular information matrices. We write r for the residual precision denoted by ρ in the main text.

C.1 Sufficiency and orthogonalization

Fix a design and abbreviate its information matrix by

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0.$$

For the collected observations, expansion of the Gaussian log likelihood gives

$$\log p_{\theta,\delta}(z) = [\theta \quad \delta]\, U - \frac{1}{2}\,[\theta \quad \delta]\, M \begin{bmatrix} \theta \\ \delta \end{bmatrix} + C(z),$$

where $U \sim \mathcal{N}(M(\theta, \delta)^\top, M)$. This factorization establishes sufficiency of U, including when M is singular.

Suppose first that $a > 0$ and put $\kappa = b/a$ and $r = c - b^2/a$. Positive semidefiniteness gives $r \geq 0$. The linear statistics

$$T = U_\theta/a, \qquad V = U_\delta - (b/a)U_\theta$$

satisfy

$$\mathbb{E}T = \theta + \kappa\delta, \quad \mathrm{var}(T) = a^{-1}, \quad \mathbb{E}V = r\delta, \quad \mathrm{var}(V) = r, \quad \mathrm{cov}(T, V) = b - (b/a)a = 0.$$

They are jointly Gaussian and therefore independent. If $r > 0$, setting $D = V/r$ yields $D \sim \mathcal{N}(\delta, r^{-1})$. If $r = 0$, then $\mathbb{E}V = \mathrm{var}(V) = 0$, so $V = 0$ almost surely and the residual channel is absent. The convention $\mathfrak{m}_h(0) = h^2$ accounts for this absence of discrepancy information.

If $a = 0$, positive semidefiniteness forces $b = 0$. Equivalently, every collected observation has zero target loading because $a = \sum_j x_j f_j^\top \Sigma_j^{-1} f_j$. The data law is then independent of the unrestricted parameter θ, and the minimax risk is $+\infty$. All reciprocal-risk formulas below use this extended-value convention.

C.2 Exact all-estimator minimax risk

We first establish the upper bound. For any estimator $\widehat{\delta}(D)$, define $\widehat{\theta} = T - \kappa\widehat{\delta}(D)$. Independence gives

$$\begin{aligned}\mathbb{E}_{\theta,\delta}(\widehat{\theta} - \theta)^2 &= \mathbb{E}\{\varepsilon_T + \kappa(\delta - \widehat{\delta}(D))\}^2 \\ &= a^{-1} + \kappa^2 \mathbb{E}_\delta(\widehat{\delta}(D) - \delta)^2.\end{aligned}$$

Taking a minimax bounded-normal-mean rule proves

$$R_{\mathrm{all}}(x) \le a^{-1} + \kappa^2 \mathfrak{m}_h(r).$$

When $r = 0$, the same construction uses the constant residual estimate zero and the convention $\mathfrak{m}_h(0) = h^2$.

For the lower bound, a diffuse proper prior on θ removes the information that T provides about δ. This leaves precisely the bounded normal mean problem based on D. Let $\theta \sim \mathcal{N}(0, M_0^2)$ and independently let $\delta \sim \pi$, where π is any probability measure on $[-h, h]$. Write

$$\sigma_T^2 = a^{-1}, \qquad \rho_{M_0} = \frac{M_0^2}{M_0^2 + \sigma_T^2}, \qquad v_{M_0} = \frac{M_0^2 \sigma_T^2}{M_0^2 + \sigma_T^2}.$$

Conditional on (T, δ), elementary Gaussian conjugacy gives

$$\theta \mid T, \delta \sim \mathcal{N}(\rho_{M_0}(T - \kappa\delta), v_{M_0}).$$

Applying the conditional total-variance identity once more, now conditioning only on (T, D), shows that the Bayes risk for estimating θ is exactly

$$v_{M_0} + \rho_{M_0}^2 \kappa^2 \mathbb{E}[\mathrm{var}(\delta \mid T, D)]. \tag{29}$$

After θ is integrated out, $T \mid \delta \sim \mathcal{N}(\kappa\delta, M_0^2 + \sigma_T^2)$. This additional channel becomes uniformly uninformative as $M_0 \to \infty$. Since $T \perp D \mid \delta$, the chain rule gives

$$I(\delta; T \mid D) = I(\delta; T) + I(T; D \mid \delta) - I(T; D) \le I(\delta; T).$$

The Gaussian-channel capacity bound, followed by $\mathrm{var}(\delta) \le h^2$ and $\log(1 + x) \le x$, therefore yields

$$I(\delta; T \mid D) \le \frac{\kappa^2 h^2}{2(M_0^2 + \sigma_T^2)} \longrightarrow 0.$$

To quantify the effect on Bayes risk, let $\mu_{T,D} = \mathbb{E}(\delta \mid T, D)$ and $\mu_D = \mathbb{E}(\delta \mid D)$. Conditional total variance and the posterior-mean bound $|\mathbb{E}_P\delta - \mathbb{E}_Q\delta| \le 2h\ \mathrm{TV}(P, Q)$ give

$$\begin{aligned}0 &\le \mathbb{E}\,\mathrm{var}(\delta \mid D) - \mathbb{E}\,\mathrm{var}(\delta \mid T, D) \\ &= \mathbb{E}(\mu_{T,D} - \mu_D)^2 \\ &\le 4h^2 \mathbb{E}\,\mathrm{TV}^2(\mathcal{L}(\delta \mid T, D), \mathcal{L}(\delta \mid D)) \\ &\le 2h^2 I(\delta; T \mid D) \longrightarrow 0,\end{aligned}$$

where the final line applies Pinsker's inequality conditionally. Thus $\mathbb{E}\,\mathrm{var}(\delta \mid T, D) \to \mathbb{E}\,\mathrm{var}(\delta \mid D)$, the Bayes risk for estimating δ from D under π. Since $v_{M_0} \to a^{-1}$ and $\rho_{M_0} \to 1$, the limit of Equation (29) is

$$a^{-1} + \kappa^2 r_\pi(D).$$

Maximum risk dominates every proper-prior Bayes risk. Finally, compactness of $[-h, h]$ and the standard minimax–Bayes approximation for the bounded normal-mean problem provide a sequence π_n whose Bayes risks approach $\mathfrak{m}_h(r)$. Taking first $M_0 \to \infty$ and then $n \to \infty$ proves the lower bound. If $r = 0$, omit D; for the symmetric prior on $\{-h, h\}$, the Bayes risk in the diffuse-location limit tends to h^2, proving the same boundary value. This establishes Equation (5) over the full measurable decision class.

C.3 Optimization over all affine rules

Conditioning any affine rule based on the original observations on the jointly Gaussian sufficient score produces an affine score rule and cannot increase squared risk. This remains true for singular information matrices by using the Gaussian pseudoinverse conditional mean. It therefore suffices to optimize affine functions of the score. For $r > 0$, an arbitrary affine rule based on the sufficient statistic is $\widehat{\theta} = \alpha T + \beta D + c_0$. The supremum over $\theta \in \mathbb{R}$ is finite only if $\alpha = 1$. Its worst-case risk is then

$$a^{-1} + \frac{\beta^2}{r} + \left(h|\kappa + \beta| + |c_0|\right)^2. \tag{30}$$

Indeed, the first two terms are the noise variance, and maximizing the absolute deterministic bias $(\kappa + \beta)\delta + c_0$ over the symmetric interval gives the final term. Thus $c_0 = 0$ is optimal. Minimizing the remaining differentiable quadratic gives

$$\beta_* = -\frac{\kappa h^2 r}{1 + h^2 r}, \qquad \inf_{\beta} \left\{ \frac{\beta^2}{r} + h^2(\kappa + \beta)^2 \right\} = \frac{\kappa^2 h^2}{1 + h^2 r}.$$

Consequently the best affine rule is $T - \kappa\lambda_* D$, where $\lambda_* = h^2 r/(1 + h^2 r)$. When $r = 0$, D is absent; minimizing over the intercept in $T + c_0$ yields $a^{-1} + \kappa^2 h^2$, the continuous extension of the same formula. Finally,

$$\begin{aligned} R_{\mathrm{aff}} &= a^{-1} + \frac{(b/a)^2 h^2}{1 + h^2(c - b^2/a)} \\ &= \frac{c + h^{-2}}{a(c + h^{-2}) - b^2} = \frac{1}{a - b^2/(c + h^{-2})}, \end{aligned}$$

which proves Equation (6), including optimality among affine rules with an intercept.

The $5/4$ guarantee follows from the classical bounded-normal comparison of Donoho et al. (1990); Donoho (1994):

$$\mathfrak{m}_{h,\mathrm{aff}}(r) \le \frac{5}{4}\mathfrak{m}_h(r).$$

After standardizing the noise, the one-dimensional interval is compact, solid, orthosymmetric, and quadratically convex, so it satisfies the stated hypotheses of that comparison (and the zero-precision value follows by continuity). Applying it to the residual term and using $a^{-1} \ge 0$ yields

$$\begin{aligned} R_{\mathrm{aff}} &\le a^{-1} + \frac{5}{4}\kappa^2 \mathfrak{m}_h(r) \\ &\le \frac{5}{4}\{a^{-1} + \kappa^2 \mathfrak{m}_h(r)\} = \frac{5}{4} R_{\mathrm{all}}. \end{aligned}$$

The reverse inequality follows from inclusion of the affine class in the class of all measurable estimators. Taking infima over the same deterministic design set preserves both inequalities.

C.4 Convexity and semidefinite representation

Let $\gamma = h^{-2} > 0$. The map $(b, t) \mapsto b^2/t$ is a convex quadratic perspective on $t > 0$ (Boyd & Vandenberghe, 2004). Hence

$$J_\gamma(x) = a(x) - \frac{b(x)^2}{c(x) + \gamma}$$

is concave. It is positive whenever $a(x) > 0$: if $a > 0$, then $b^2 \le ac$ and $J_\gamma \ge a\gamma/(c + \gamma) > 0$. Since $s \mapsto 1/s$ is convex and decreasing on the positive half-line, $R_{\mathrm{aff}} = 1/J_\gamma$ is convex on its finite domain.

The first linear matrix inequality in Equation (9) is, by the Schur complement with $c + \gamma > 0$, equivalent to $s \le J_\gamma(x)$. The second is equivalent to $s > 0$ and $t \ge 1/s$; thus no open constraint needs to be added. At an optimum, both inequalities are tight. Thus the risk admits the stated semidefinite epigraph. When $\mathcal{X}$ is semidefinite representable, including the polyhedral budget and capacity sets considered here, the complete problem is an SDP. For any other compact convex $\mathcal{X}$, the same epigraph gives a convex optimization problem with the additional constraint $x \in \mathcal{X}$.

UNRESTRICTED-SHIFT BOUNDARY

When the discrepancy is unrestricted, the effective target information is the generalized Schur complement. Define

$$J_0 = a - b^2 c^\dagger, \qquad c^\dagger = \begin{cases} c^{-1}, & c > 0, \\ 0, & c = 0. \end{cases}$$

Positive semidefiniteness forces $b = 0$ when $c = 0$. With unrestricted δ, the target risk is finite exactly when $J_0 > 0$, in which case it is $1/J_0$. This includes a target-only experiment $(a > 0, b = c = 0)$, whose risk is $1/a$. A singular matrix with $c > 0$ therefore gives infinite target risk, whereas a target-only experiment remains estimable.

C.5 FISHER REPRESENTATION OF THE BOUNDED-NORMAL IMPROVEMENT

The second-order acquisition analysis uses the improvement over the unrestricted normal mean risk. This improvement has a useful Fisher-information representation. For $H > 0$, set

$$\Psi_H(w) := w - w^2 \mathfrak{m}_H(w), \qquad w \geq 0, \tag{31}$$

with $\Psi_H(0) = 0$. Let π be any probability measure supported on $[-H, H]$, let $X \sim \pi$, and for $w > 0$ let $Y = X + w^{-1/2} G$, where $G \sim N(0, 1)$ is independent. The density $f_{\pi,w}$ of Y is smooth and strictly positive. Differentiating its Gaussian convolution under the integral gives Tweedie's identity

$$\frac{f'_{\pi,w}(y)}{f_{\pi,w}(y)} = -w\{y - \mathbb{E}(X \mid Y = y)\}.$$

Because X is bounded, this score has at most linear growth and the Gaussian tails make every following integral finite. Squaring, integrating, and using conditional variance yields Brown's identity (Brown, 1971)

$$J(f_{\pi,w}) := \int_{\mathbb{R}} \frac{f'_{\pi,w}(y)^2}{f_{\pi,w}(y)}\, dy = w - w^2 \operatorname{mmse}_\pi(w).$$

The compact minimax–Bayes representation established above therefore gives

$$\Psi_H(w) = \inf_{\pi:\, \operatorname{supp}(\pi) \subset [-H,H]} J\big(\pi * N(0, w^{-1})\big), \qquad w > 0. \tag{32}$$

The representation holds for every compactly supported prior, including discrete priors.

We record three elementary consequences used later:

$$0 \leq \Psi_H(w) \leq w, \qquad \frac{\Psi_H(w)}{w} \longrightarrow 1 \quad (w \downarrow 0), \qquad \sup_{w \geq 0} \Psi_H(w) \leq \frac{\pi^2}{H^2}. \tag{33}$$

The first two follow from $0 \leq \mathfrak{m}_H(w) \leq \min\{H^2, w^{-1}\}$. For the last, use in Equation (32) the cosine density

$$g_H(x) = H^{-1} \cos^2\!\left(\frac{\pi x}{2H}\right) \mathbf{1}\{|x| \leq H\}, \qquad J(g_H) = \frac{\pi^2}{H^2}.$$

Adding independent Gaussian noise cannot increase location Fisher information: if $Y = X + Z$, the score identity $\rho_Y(Y) = \mathbb{E}\{\rho_X(X) \mid Y\}$ and conditional Jensen give $J(Y) \leq J(X)$. For g_H, both the density and its derivative vanish at the endpoints, so the integration by parts behind the score identity has no boundary term.

C.6 PRODUCT EXPERIMENTS

We now prove Corollary 2. Fix a design x and let $R_{j,\mathrm{all}}(x)$ denote the scalar minimax risk in block j. Applying the scalar minimax rules coordinatewise gives the upper bound

$$R^{\mathrm{prod}}_{\mathrm{all}}(x) \leq \sum_{j=1}^{d} g_j R_{j,\mathrm{all}}(x).$$

For the reverse bound, in each block choose a proper bias prior approaching its scalar least-favourable risk and an independent diffuse Gaussian prior on its unrestricted location. Take the product of these

priors. The raw experiment and posterior factorize across blocks, and additive squared loss makes the Bayes risk the sum of block Bayes risks. The same two-stage limits used above therefore give the reverse inequality. Hence equality holds. Applying the scalar $5/4$ comparison term by term and then taking the infimum over the common feasible design set proves Corollary 2. Each summand is understood in the extended-value sense

$$\frac{g_j}{J_j(x)} := \begin{cases} 0, & g_j = 0, \\ g_j/J_j(x), & g_j > 0,\ J_j(x) > 0, \\ +\infty, & g_j > 0,\ J_j(x) = 0. \end{cases}$$

It is convex by the preceding argument. Compactness then gives attainment whenever some feasible design has positive information in every block with strictly positive loss weight.

D ELLIPSOIDAL VECTOR DISCREPANCIES

The fixed-design reduction extends to a vector discrepancy constrained to an ellipsoid. It retains the same $5/4$ affine guarantee and yields a support bound depending only on the discrepancy dimension. Let now $\delta \in \mathbb{R}^d$ satisfy

$$\delta^\top H^{-1}\delta \le 1, \qquad H \succ 0,$$

and write the information of a fixed design as

$$M = \begin{bmatrix} a & b^\top \\ b & C \end{bmatrix} \succeq 0, \qquad R = C - \frac{bb^\top}{a}, \qquad k = \frac{b}{a}, \tag{34}$$

with $a > 0$. For the residual score $W \sim \mathcal{N}(R\delta, R)$, define

$$m_{H,k}(R) := \inf_{\widehat{L}} \sup_{\delta^\top H^{-1}\delta \le 1} \mathbb{E}_\delta\{\widehat{L}(W) - k^\top\delta\}^2.$$

Theorem 11 (Vector-discrepancy reduction and sparse design)**.** *For every $d \ge 1$,*

$$\mathcal{R}_{\mathrm{all}}(M) = \frac{1}{a} + m_{H,k}(R). \tag{35}$$

The best affine rule has risk

$$\mathcal{R}_{\mathrm{aff}}(M) = \frac{1}{a} + k^\top(R + H^{-1})^{-1}k = \frac{1}{J_H(M)}, \qquad J_H(M) := a - b^\top(C + H^{-1})^{-1}b, \tag{36}$$

and

$$\mathcal{R}_{\mathrm{all}}(M) \le \mathcal{R}_{\mathrm{aff}}(M) \le \frac{5}{4}\mathcal{R}_{\mathrm{all}}(M).$$

Consequently, affine acquisition over a semidefinite-representable design set is an SDP: replace the first 2×2 LMI in Equation (9) *by*

$$\begin{bmatrix} a - s & b^\top \\ b & C + H^{-1} \end{bmatrix} \succeq 0.$$

For a finite set of cost-normalized observation types, a budget $B > 0$, no type-specific caps, and at least one design with positive target information, an exact continuous all-estimator optimum exists using at most

$$D_d = \frac{(d+1)(d+2)}{2}$$

observation types. The same bound holds for an affine optimum.

The dimension-free constant follows from the classical Gaussian linear-functional comparison over a convex parameter set (Donoho, 1994). The support bound is determined by the dimension D_d of the symmetric information-matrix space and is independent of the number of available observation types. At $d = 1$, it recovers the scalar three-type bound.

E PROOFS FOR ELLIPSOIDAL VECTOR DISCREPANCIES

We prove Theorem 11 by extending the scalar orthogonalization to a vector discrepancy. After deriving the exact and affine risks, we establish convexity of the affine problem and the support bound for the exact optimum.

E.1 ORTHOGONAL SCORE EXPERIMENT

For a fixed design, the canonical Gaussian score satisfies

$$U = \begin{bmatrix} U_\theta \\ U_\delta \end{bmatrix} \sim \mathcal{N}\left(M \begin{bmatrix} \theta \\ \delta \end{bmatrix}, M \right), \qquad M = \begin{bmatrix} a & b^\top \\ b & C \end{bmatrix} \succeq 0.$$

Assume $a > 0$, put $k = b/a$, and define

$$T = \frac{U_\theta}{a}, \qquad W = U_\delta - \frac{b}{a} U_\theta, \qquad R = C - \frac{bb^\top}{a}.$$

Direct calculation gives

$$\begin{aligned} \mathbb{E}T &= \theta + k^\top \delta, & \operatorname{var}(T) &= a^{-1}, \\ \mathbb{E}W &= R\delta, & \operatorname{var}(W) &= R, \\ \operatorname{cov}(T, W) &= a^{-1} b^\top - a^{-1} b^\top = 0. \end{aligned}$$

The two statistics are jointly Gaussian, hence independent. This statement also covers singular R: then W is a Gaussian vector supported on $\operatorname{range}(R)$, and no inverse of R is introduced.

E.2 EXACT ALL-ESTIMATOR VALUE

Let $\widehat{L}(W)$ be any rule for $k^\top \delta$. The estimator $\widehat{\theta} = T - \widehat{L}(W)$ has, by independence,

$$\mathbb{E}_{\theta,\delta}(\widehat{\theta} - \theta)^2 = a^{-1} + \mathbb{E}_\delta\{\widehat{L}(W) - k^\top \delta\}^2.$$

Taking the infimum and supremum proves the upper bound in Equation (35).

For the reverse inequality, fix any proper prior π on the compact ellipsoid and independently let $\theta \sim \mathcal{N}(0, A^2)$. Write

$$\sigma^2 = a^{-1}, \qquad \rho_A = \frac{A^2}{A^2 + \sigma^2}, \qquad v_A = \frac{A^2 \sigma^2}{A^2 + \sigma^2}.$$

Conditional on (T, δ), Gaussian conjugacy gives

$$\theta \mid T, \delta \sim \mathcal{N}\{\rho_A (T - k^\top \delta), v_A\}.$$

Conditioning once more on (T, W), the posterior-variance identity yields

$$r_{\pi,A} = v_A + \rho_A^2 \, \mathbb{E}\{\operatorname{var}(k^\top \delta \mid T, W)\}, \tag{37}$$

where $r_{\pi,A}$ is the Bayes risk for estimating θ.

After integrating out θ, $T \mid \delta \sim \mathcal{N}(k^\top \delta, A^2 + \sigma^2)$. Moreover $T \perp W \mid \delta$, so the two chain rules give

$$I(\delta; T) = I(T; W) + I(\delta; T \mid W) \geq I(\delta; T \mid W).$$

The ellipsoid bound $|k^\top \delta| \leq L := \sqrt{k^\top H k}$ and the Gaussian-channel information bound give

$$I(\delta; T \mid W) \leq I(\delta; T) \leq \frac{L^2}{2(A^2 + \sigma^2)} \longrightarrow 0.$$

Let $\mu_{T,W} = \mathbb{E}(k^\top \delta \mid T, W)$ and $\mu_W = \mathbb{E}(k^\top \delta \mid W)$. Conditional total variance, the fact that $k^\top \delta \in [-L, L]$, and conditional Pinsker imply

$$\begin{aligned} 0 &\leq \mathbb{E} \operatorname{var}(k^\top \delta \mid W) - \mathbb{E} \operatorname{var}(k^\top \delta \mid T, W) \\ &= \mathbb{E}(\mu_{T,W} - \mu_W)^2 \leq 2L^2 I(\delta; T \mid W) \longrightarrow 0. \end{aligned}$$

Since $v_A \to a^{-1}$ and $\rho_A \to 1$, taking $A \to \infty$ in Equation (37) gives

$$a^{-1} + \mathbb{E}_\pi \operatorname{var}(k^\top \delta \mid W).$$

Maximum risk dominates this quantity for every proper π. The compact Gaussian minimax theorem supplies priors whose Bayes risks approach $m_{H,k}(R)$; taking that supremum proves the lower bound. If $a = 0$, positive semidefiniteness forces $b = 0$, the experiment is independent of the unrestricted target, and the risk is $+\infty$.

E.3 AFFINE MINIMAX RULE

Conditioning any affine estimator based on the original observations on the sufficient score produces an affine score estimator without increasing squared risk. An affine score rule with finite supremum over $\theta \in \mathbb{R}$ must have coefficient one on T. By symmetry of the ellipsoid its optimal intercept is zero, so write it as $T - z^\top W$. Its worst-case risk is

$$\frac{1}{a} + z^\top R z + (k - Rz)^\top H (k - Rz). \tag{38}$$

Indeed, the first two terms are the variances of the independent score noises, while

$$\sup_{\delta^\top H^{-1}\delta \le 1} \{(k - Rz)^\top \delta\}^2 = (k - Rz)^\top H (k - Rz)$$

by Cauchy–Schwarz after the change of variables $\delta = H^{1/2}u$.

Set $S = (H^{-1} + R)^{-1}$ and $z_\star = Sk$. The identity $(H^{-1} + R)S = I$ gives $k - Rz_\star = H^{-1}Sk$. Substituting in the two quadratic terms in Equation (38) gives

$$\begin{aligned} z_\star^\top R z_\star + (k - Rz_\star)^\top H (k - Rz_\star) &= k^\top S R S k + k^\top S H^{-1} S k \\ &= k^\top S (R + H^{-1}) S k = k^\top S k. \end{aligned}$$

For arbitrary z, completing the square around $z_\star$ adds the nonnegative term

$$(z - z_\star)^\top \{R + RHR\}(z - z_\star),$$

so $z_\star$ is globally optimal even when R is singular. This proves the first equality in Equation (36).

To obtain the regularized Schur-complement form, put $D = C + H^{-1}$ and $J = a - b^\top D^{-1} b$. Since

$$R + H^{-1} = D - \frac{bb^\top}{a},$$

the left side is positive definite. Congruence by $D^{-1/2}$ gives

$$I - \frac{D^{-1/2} b b^\top D^{-1/2}}{a} \succ 0,$$

so $b^\top D^{-1} b < a$ and $J > 0$. The rank-one inverse identity therefore gives:

$$\left(D - \frac{bb^\top}{a}\right)^{-1} = D^{-1} + \frac{D^{-1} b b^\top D^{-1}}{J}.$$

Writing $\xi = b^\top D^{-1} b = a - J$ and $k = b/a$, direct substitution yields

$$\frac{1}{a} + k^\top (R + H^{-1})^{-1} k = \frac{1}{a} + \frac{\xi}{a^2} + \frac{\xi^2}{a^2 J} = \frac{1}{J}.$$

The residual problem estimates the linear functional $k^\top \delta$ in a Gaussian experiment over a compact convex ellipsoid. The classical Gaussian linear-functional theorem of Donoho (1994) states that its minimax affine risk is at most $5/4$ times its unrestricted minimax risk. For singular $R = U_r \Lambda U_r^\top$, with $\Lambda \succ 0$, the nondegenerate statistic is explicitly

$$\Lambda^{-1/2} U_r^\top W \sim N(\Lambda^{1/2} U_r^\top \delta, I_r).$$

The components in the zero-eigenspace contain no observations, so the theorem applies to the nondegenerate Gaussian experiment above. If $R = 0$, both residual minimax risks equal $\sup_{\delta^\top H^{-1}\delta \le 1} (k^\top \delta)^2 = k^\top H k$, attained by the zero rule. Adding the nonnegative independent term a^{-1}, exactly as in the scalar proof, gives

$$\mathcal{R}_{\text{aff}}(M) \le a^{-1} + \frac{5}{4} m_{H,k}(R) \le \frac{5}{4} \mathcal{R}_{\text{all}}(M).$$

The opposite inequality follows because affine rules are a subclass of all measurable rules.

E.4 Convexity and sparse support

Convexity. The matrix fractional function

$$(b, D) \longmapsto b^\top D^{-1} b, \qquad D \succ 0,$$

is convex (Boyd & Vandenberghe, 2004). Hence $J_H(M) = a - b^\top (C + H^{-1})^{-1} b$ is concave in M, and $1/J_H(M)$ is convex on its positive domain. By the Schur complement,

$$\begin{bmatrix} a - s & b^\top \\ b & C + H^{-1} \end{bmatrix} \succeq 0 \quad \Longleftrightarrow \quad s \le J_H(M).$$

Together with $\left[\begin{smallmatrix} t & 1 \\ 1 & s \end{smallmatrix}\right] \succeq 0$, this proves the stated semidefinite epigraph.

Existence of an exact optimum. We first establish attainment of the minimum. Under the conditions of Theorem 11, information monotonicity allows a selected optimum to spend the full budget. After division by B, its information therefore lies in the convex hull of the cost-normalized information matrices. Let $\mathcal{K}$ be that convex hull of the finitely many observation information matrices in the vector space of symmetric $(d+1) \times (d+1)$ matrices, whose dimension is $D_d = (d+1)(d+2)/2$. The extended minimax risk is lower semicontinuous on the compact set $B\mathcal{K}$. To establish this property, for $a > 0$ represent the residual experiment equivalently as $Y_R = R^{1/2}\delta + Z$, where $Z \sim N(0, I_d)$. Multiplying by $R^{1/2}$ recovers W; conversely, $R^{\dagger/2} W + (I - P_R) Z_0$, with fresh standard Z_0, has the law of Y_R. Thus the added null-space noise is ancillary.

For each proper prior on the ellipsoid, its Bayes risk for $k^\top \delta$ is continuous in (k, R). Indeed, near any fixed k, all targets lie in one interval $[-L_0, L_0]$, so projecting every decision into that interval cannot increase loss. Continuity of the matrix square root and compactness of the ellipsoid make the joint laws of (δ, Y_R) converge in total variation, uniformly over the parameter. On the clipped decision class the squared losses are uniformly bounded and converge uniformly with k. More explicitly, if P_n and P are the joint laws under $(k_n, R_n) \to (k, R)$, every decision d taking values in $[-L_0, L_0]$ satisfies

$$\left| \mathbb{E}_{P_n}\{d(Y) - k_n^\top \delta\}^2 - \mathbb{E}_P\{d(Y) - k^\top \delta\}^2 \right| \le 8L_0^2 \|P_n - P\|_{\mathrm{TV}} + 4L_0 \sup_{\delta \in \mathcal{E}_H} |(k_n - k)^\top \delta| + o(1),$$

and the right side is independent of d and tends to zero. Taking the infimum over the common clipped decision class therefore preserves convergence. The compact Gaussian minimax theorem expresses $m_{H,k}(R)$ as the supremum of these continuous Bayes risks, so it is lower semicontinuous.

As $a \downarrow 0$, the oracle bound $\mathcal{R}_{\mathrm{all}}(M) \ge 1/a$ diverges; at $a = 0$, PSD forces $b = 0$ and the data law contains no θ, so the extended risk is $+\infty$. By assumption, a design with positive target information exists and has finite risk. The minimum is therefore attained on $B\mathcal{K}$.

Sparse support. Among the minimizing information matrices, choose $M_\star$ with maximum trace. No distinct $M \in B\mathcal{K}$ can satisfy $M - M_\star \succeq 0$: information monotonicity would make M another minimizer, while its trace would be larger. If $\dim \mathcal{K} < D_d$, ordinary Carathéodory already represents $M_\star/B$ using at most D_d observation types. If $\mathcal{K}$ is full dimensional, separate $\mathcal{K} - M_\star/B$ from the interior of the positive semidefinite cone. The separating normal is a nonzero positive semidefinite matrix, so $M_\star/B$ lies on a proper exposed face of $\mathcal{K}$. That face has dimension at most $D_d - 1$, and Carathéodory within its affine hull again uses at most D_d original information matrices. This proves the exact support claim. Applying the same argument to the Loewner-monotone affine risk proves its support claim as well.

F Proofs for optimal budget allocation

We establish information monotonicity and sparse support, then prove the allocation regimes and their second-order refinements. The final subsections treat complementary sources, aligned couplings, perturbations, and integer counts. Throughout, there are finitely many cost-normalized auxiliary information matrices

$$A_j = \begin{bmatrix} p_j & q_j \\ q_j & r_j \end{bmatrix} \succeq 0, \qquad p_j \ge 0, \quad r_j > 0, \qquad j = 1, \dots, m,$$

and the trusted information matrix $A_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. The same scalar discrepancy satisfies $\delta \in [-h, h]$, where $0 < h < \infty$. Observation streams are conditionally independent, expenditures are continuous and chosen before data collection, and there are no type-specific caps. For every $B > 0$, a design is

$$y \in \mathcal{Y}_B := \{y \in \mathbb{R}_+^m : \mathbf{1}^\top y \le B\};$$

the remaining $B - \mathbf{1}^\top y$ is spent on the trusted type. Because an additional observation may always be discarded, allowing unspent budget would give the same optimum.

For a nonnegative information level I, abbreviate

$$m_h(I) := \mathfrak{m}_h(I), \qquad m_h(0) := h^2. \tag{39}$$

Put $e_j := p_j - 1$. For $y \in \mathcal{Y}_B$, define

$$\begin{aligned} P(y) &:= p^\top y, & E(y) &:= e^\top y, \\ Q(y) &:= q^\top y, & R(y) &:= r^\top y, \\ a_B(y) &:= B + E(y), & I_B(y) &:= R(y) - \frac{Q(y)^2}{a_B(y)}. \end{aligned} \tag{40}$$

The total information is $M_B(y) = \left[\begin{smallmatrix} a_B(y) & Q(y) \\ Q(y) & R(y) \end{smallmatrix}\right]$. It is positive semidefinite. Whenever $a_B(y) > 0$, applying Theorem 1 gives

$$\mathcal{R}_B(y) = \frac{1}{a_B(y)} + \left\{\frac{Q(y)}{a_B(y)}\right\}^2 m_h(I_B(y)). \tag{41}$$

If $a_B(y) = 0$, positive semidefiniteness forces $Q(y) = 0$, the experiment contains no target information, and we set $\mathcal{R}_B(y) = +\infty$. We write

$$V_B := \min_{y \in \mathcal{Y}_B} \mathcal{R}_B(y); \tag{42}$$

existence of the minimum is established in Lemma 14 below.

F.1 Two bounded-normal facts

Continuity controls finite auxiliary allocations, while the small-noise limit determines the value of auxiliary mixtures whose expenditure grows.

Lemma 12 (Bounded-normal continuity and small-noise limit)**.** *The function $I \mapsto m_h(I)$ is nonincreasing, continuous, and h^4-Lipschitz on $[0, \infty)$. Moreover,*

$$I m_h(I) \longrightarrow 1 \qquad (I \to \infty). \tag{43}$$

Proof. The compact Gaussian minimax theorem represents $m_h(I)$ as the supremum, over priors supported on $[-h, h]$, of the Bayes posterior variance risk at Gaussian precision I (Ferguson, 1967). For each fixed prior, the derivative of this Bayes MMSE with respect to I is minus the expected squared posterior variance (Guo et al., 2011). A random variable in $[-h, h]$ has variance at most h^2, so every such Bayes-risk curve is nonincreasing and h^4-Lipschitz. The pointwise supremum has the same properties. At $I = 0$, the constant rule zero has maximum risk h^2, while the prior assigning equal mass to $\{-h, h\}$ has Bayes risk h^2; hence the convention in Equation (39) is the correct continuous value.

For Equation (43), the rescaled Gaussian observation $X = \delta + I^{-1/2} Z$ itself has risk I^{-1}, so $m_h(I) \le I^{-1}$. For a matching lower bound, fix $0 < a < h$ and use the proper density

$$\pi_a(u) = a^{-1} \cos^2\Big(\frac{\pi u}{2a}\Big) \mathbf{1}_{\{|u| \le a\}}.$$

Its zero extension is continuously differentiable, both the density and its first derivative vanish at the two endpoints, and direct integration gives $I_{\pi_a} = \int (\pi_a'/\pi_a)^2 \pi_a = \pi^2/a^2 < \infty$. These boundary properties permit integration by parts in the van Trees inequality (Van Trees, 1968). If $X \mid \delta$ has Gaussian precision I, then, for every estimator $T(X)$,

$$\mathbb{E}_{\pi_a}[(T(X) - \delta)\{\partial_\delta \log p(X \mid \delta) + \partial_\delta \log \pi_a(\delta)\}] = 1.$$

The score in braces has second moment $I + I_{\pi_a}$: the likelihood score has conditional mean zero, so its cross term with the prior score vanishes. Cauchy–Schwarz therefore yields the Bayes-risk bound

$$\mathbb{E}_{\pi_a}(T(X) - \delta)^2 \geq \frac{1}{I + I_{\pi_a}}.$$

Maximum risk dominates this proper-prior Bayes risk, and hence

$$\frac{I}{I + I_{\pi_a}} \leq Im_h(I) \leq 1.$$

Letting $I \to \infty$ proves the claim. □

F.2 Information monotonicity and sparse support

Lemma 13 (Information monotonicity)**.** *Let $M_1 \preceq M_2$ be two information matrices for the same Gaussian parameter (θ, δ) and the same parameter set. The minimax risk for estimating θ under M_2 is no larger than that under M_1.*

Proof. For any $M \succeq 0$, the canonical sufficient score experiment is

$$U_M \sim \mathcal{N}(M(\theta, \delta)^\top, M).$$

Put $D = M_2 - M_1 \succeq 0$ and consider independent scores U_{M_1} and U_D. Their joint Gaussian likelihood depends on the parameter through $U_{M_1} + U_D$ and $M_1 + D = M_2$; the sum is therefore sufficient and has the law of U_{M_2}. Thus the M_2 experiment is equivalent to observing the M_1 experiment plus an independent experiment of information D. One may discard the latter, so every rule available under M_1 is reproducible under M_2. Taking infimum maximum risks proves the assertion. This argument also covers singular matrices because it never inverts M_1 or D. □

Lemma 14 (At most three observation types)**.** *For every $B > 0$, there is an exact minimizer of Equation* (41) *using at most three observation types, including the trusted type when used.*

Proof. For a full-budget allocation, information per unit expenditure belongs to the compact polytope

$$\mathcal{K} := \operatorname{conv}\{A_0, A_1, \ldots, A_m\} \subset \mathbb{S}^2.$$

Conversely, every point of $\mathcal{K}$ is a feasible approximate design. On the subset with positive upper-left entry, the exact formula and Lemma 12 make the risk continuous. If a sequence of feasible matrices has upper-left entry $a \downarrow 0$, revealing the discrepancy gives the oracle lower bound $1/a \to +\infty$. With the extended value $+\infty$ at $a = 0$, the risk is therefore lower semicontinuous on the compact set $B\mathcal{K}$. It is finite at BA_0, so an exact minimizer exists.

Among all minimizing information matrices choose M_* with maximum trace. There is no $M \in B\mathcal{K}$ such that $M - M_* \succeq 0$ and $M \neq M_*$. Indeed, Lemma 13 would make M another minimizer, whereas $\operatorname{tr}(M) > \operatorname{tr}(M_*)$, contradicting the choice of M_*.

If $\dim \mathcal{K} \leq 2$, Carathéodory's theorem already represents M_*/B by at most three information matrices. Suppose instead that $\mathcal{K}$ has full dimension three in $\mathbb{S}^2$. The convex sets $\mathcal{K} - M_*/B$ and $\operatorname{int} \mathbb{S}^2_+$ are disjoint. Separating them gives a nonzero $H \in \mathbb{S}^2$ such that

$$\langle H, A - M_*/B \rangle \leq 0 \quad (A \in \mathcal{K}), \qquad \langle H, D \rangle \geq 0 \quad (D \succeq 0).$$

The second relation says $H \succeq 0$. Hence M_*/B lies in the exposed face

$$\mathcal{F} = \arg\max_{A \in \mathcal{K}} \langle H, A \rangle .$$

Because $H \neq 0$ and $\mathcal{K}$ is full-dimensional, $\mathcal{F}$ is a proper face of dimension at most two. It is the convex hull of the original information matrices lying in that face. Carathéodory's theorem inside the affine hull of $\mathcal{F}$ therefore represents M_*/B using at most three of those information matrices. Multiplying the weights by B gives the claimed design. □

F.3 Information rates and trusted-only acquisition

For $\lambda \in \Delta_m$, write

$$A(\lambda) = \sum_j \lambda_j A_j = \begin{bmatrix} P(\lambda) & Q(\lambda) \\ Q(\lambda) & R(\lambda) \end{bmatrix}$$

and

$$s(\lambda) := P(\lambda) - \frac{Q(\lambda)^2}{R(\lambda)}, \qquad \Gamma_0 = \max_j p_j, \qquad \Gamma_\infty = \max_{\lambda \in \Delta_m} s(\lambda). \tag{44}$$

Every $A(\lambda)$ is positive semidefinite and $R(\lambda) \geq \min_j r_j > 0$, so the displayed quantities are continuous and nonnegative. In particular, both maxima are attained. Since $s(\lambda) \leq P(\lambda) \leq \Gamma_0$,

$$\Gamma_\infty \leq \Gamma_0. \tag{45}$$

For $t \in (0, 1]$, direct substitution gives the useful identity

$$J_{\mathrm{Sch}}\{(1-t)A_0 + tA(\lambda)\} = 1 + t\{s(\lambda) - 1\}, \tag{46}$$

with the continuous value one at $t = 0$. Here $J_{\mathrm{Sch}}\left(\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]\right) = a - b^2/c$ when $c > 0$.

The SDP in Equation (15) also follows directly: because $R(\lambda) > 0$, its LMI is equivalent by the Schur complement to $t \leq s(\lambda)$. Maximizing t therefore returns exactly Γ_∞.

Lemma 15 (Exact trusted endpoint)**.** *Trusted-only acquisition is optimal for every $B > 0$ if and only if $\Gamma_0 \leq 1$. If $\Gamma_0 > 1$, it is not a local minimizer for any $B > 0$.*

Proof. First suppose $p_j \leq 1$ for every j. Reveal δ to the estimator. This oracle can only reduce minimax risk. Conditional on the revealed discrepancy, the Fisher information for the unrestricted location is the upper-left information entry

$$a_B(y) = B - \mathbf{1}^\top y + \sum_j p_j y_j \leq B.$$

The oracle Gaussian-location risk is $1/a_B(y) \geq 1/B$. Trusted-only acquisition attains $1/B$, so it is optimal over all estimators.

Conversely, let $p_j > 1$ for some j and put $e_j = p_j - 1 > 0$. Denote the jth coordinate vector by $\mathbf{e}^{(j)}$. Starting at the trusted-only design, spend $t > 0$ on type j and $B - t$ on the target. For $0 < t < B$, Equation (41) and $m_h(I) \leq h^2$ give

$$\mathcal{R}_B(t\mathbf{e}^{(j)}) \leq \frac{1}{B + e_j t} + \frac{q_j^2 h^2 t^2}{(B + e_j t)^2} = \frac{1}{B} - \frac{e_j}{B^2} t + O(t^2).$$

For all sufficiently small positive t, this is strictly below $1/B$. Thus trusted-only is not locally optimal at that budget. Since $B > 0$ was arbitrary, the necessity holds for every budget. □

Lemma 16 (Unique small-budget optimum)**.** *Put $p_0 = 1$, $q_0 = r_0 = 0$. If one index $k \in \{0, 1, \ldots, m\}$ uniquely maximizes p_i, then there is $B_0 > 0$ such that, for every $0 < B < B_0$, spending the full budget on type k is the unique exact minimax design.*

Proof. Represent a full-budget design by $\lambda \in \Delta_{m+1}$, including the trusted type. Write

$$P(\lambda) = \sum_{i=0}^m \lambda_i p_i, \quad Q(\lambda) = \sum_{i=0}^m \lambda_i q_i, \quad R(\lambda) = \sum_{i=0}^m \lambda_i r_i,$$

and

$$\kappa(\lambda) = \frac{Q(\lambda)}{P(\lambda)}, \qquad \rho(\lambda) = R(\lambda) - \frac{Q(\lambda)^2}{P(\lambda)}.$$

The uniquely maximizing type satisfies $p_k \geq p_0 = 1$, hence $p_k > 0$. The exact risk of the information matrix $B \sum_i \lambda_i A_i$, whenever $P(\lambda) > 0$, is

$$\mathcal{R}_B(\lambda) = \frac{1}{BP(\lambda)} + \kappa(\lambda)^2 m_h\{B\rho(\lambda)\}. \tag{47}$$

Let e_k be the kth simplex vertex and put

$$d = \min_{i \neq k}(p_k - p_i) > 0, \qquad t = 1 - \lambda_k.$$

Then $p_k - P(\lambda) \geq dt$ and $\|\lambda - e_k\|_1 = 2t$. Since $P(\lambda) \leq p_k$,

$$\frac{1}{BP(\lambda)} - \frac{1}{Bp_k} \geq \frac{dt}{Bp_k^2}. \tag{48}$$

On the neighborhood $t \leq 1/2$, $P(\lambda) \geq \lambda_k p_k \geq p_k/2$. There κ, κ^2, ρ are Lipschitz, $\rho \geq 0$ by positive semidefiniteness, and $0 \leq m_h \leq h^2$ together with Lemma 12 gives the second term in Equation (47) a Lipschitz constant $L < \infty$ uniformly over $0 < B \leq 1$. Hence

$$\mathcal{R}_B(\lambda) - \mathcal{R}_B(e_k) \geq \left\{ \frac{d}{Bp_k^2} - 2L \right\} t \qquad (t \leq 1/2).$$

If $t > 1/2$, the nonnegative residual term at λ, the bound $\kappa(e_k)^2 m_h\{B\rho(e_k)\} \leq h^2 \kappa(e_k)^2 =: K$, and Equation (48) instead give

$$\mathcal{R}_B(\lambda) - \mathcal{R}_B(e_k) \geq \frac{d}{2Bp_k^2} - K.$$

Both displays are strictly positive for all sufficiently small $B > 0$; designs with $P(\lambda) = 0$ have infinite risk. Hence e_k is the unique minimizer. □

F.4 Large-information limits and auxiliary dominance

Lemma 17 (Large-information local limit)**.** *Let* $\bar{M}_k = \left[\begin{smallmatrix} \bar{a}_k & \bar{b}_k \\ \bar{b}_k & \bar{c}_k \end{smallmatrix}\right] \succeq 0$ *be information matrices with* $\bar{a}_k > 0$ *and* $\bar{M}_k \to \bar{M}$*, where* $\bar{a} > 0$*,* $\bar{c} > 0$*, and* $\bar{\rho} := \bar{c} - \bar{b}^2/\bar{a} > 0$*. If* $B_k \to \infty$*, then*

$$B_k \mathcal{R}_{\mathrm{all}}(B_k \bar{M}_k) \longrightarrow \frac{1}{J_{\mathrm{Sch}}(\bar{M})}. \tag{49}$$

Proof. Put $\bar{\rho}_k = \bar{c}_k - \bar{b}_k^2/\bar{a}_k$. The exact scalar formula is

$$B_k \mathcal{R}_{\mathrm{all}}(B_k \bar{M}_k) = \frac{1}{\bar{a}_k} + \left(\frac{\bar{b}_k}{\bar{a}_k}\right)^2 B_k m_h(B_k \bar{\rho}_k).$$

For all large k, $\bar{\rho}_k$ is bounded away from zero. By Equation (43), $B_k m_h(B_k \bar{\rho}_k) \to 1/\bar{\rho}$. Elementary algebra then gives

$$\frac{1}{\bar{a}} + \frac{\bar{b}^2}{\bar{a}^2 \bar{\rho}} = \frac{\bar{c}}{\bar{a}\bar{c} - \bar{b}^2} = \frac{1}{\bar{a} - \bar{b}^2/\bar{c}} = \frac{1}{J_{\mathrm{Sch}}(\bar{M})}.$$

□

Lemma 18 (Singular local limit)**.** *Let* $\bar{M}_k = \left[\begin{smallmatrix} \bar{a}_k & \bar{b}_k \\ \bar{b}_k & \bar{c}_k \end{smallmatrix}\right] \succeq 0$ *satisfy* $\bar{a}_k > 0$ *and* $\bar{M}_k \to \bar{M}$*, where* $\bar{a} > 0$*,* $\bar{c} > 0$*, and* $J_{\mathrm{Sch}}(\bar{M}) = 0$*. If* $B_k \to \infty$*, then*

$$B_k \mathcal{R}_{\mathrm{all}}(B_k \bar{M}_k) \longrightarrow +\infty. \tag{50}$$

Proof. Put $\bar{\rho}_k = \bar{c}_k - \bar{b}_k^2/\bar{a}_k \geq 0$ and $\bar{\kappa}_k = \bar{b}_k/\bar{a}_k$. Singularity together with $\bar{a}, \bar{c} > 0$ gives $\bar{b}^2 = \bar{a}\bar{c} > 0$; hence $\bar{\rho}_k \to 0$ and $\bar{\kappa}_k^2 \to \bar{b}^2/\bar{a}^2 > 0$. Use the fixed interior prior from the proof of Lemma 12, whose Fisher information is $I_\pi < \infty$. Its van Trees bound gives $m_h(I) \geq (I + I_\pi)^{-1}$ for every $I \geq 0$. The exact risk identity therefore yields

$$B_k \mathcal{R}_{\mathrm{all}}(B_k \bar{M}_k) \geq \bar{\kappa}_k^2 B_k m_h(B_k \bar{\rho}_k) \geq \frac{\bar{\kappa}_k^2}{\bar{\rho}_k + I_\pi/B_k} \longrightarrow +\infty.$$

□

Proposition 19 (Supercritical auxiliary scaling)**.** *If $\Gamma_\infty > 1$, then every choice of exact minimizers y_B satisfies*

$$\frac{S_B}{B} \to 1, \qquad \operatorname{dist}\left(\frac{y_B}{S_B}, \Lambda_\infty\right) \to 0, \qquad BV_B \to \Gamma_\infty^{-1}, \tag{51}$$

where $S_B = \mathbf{1}^\top y_B$ and $\Lambda_\infty = \arg\max_{\lambda \in \Delta_m} s(\lambda)$. The normalized portfolio is well defined for all sufficiently large B.

Proof. Choose $\lambda_* \in \Lambda_\infty$ and spend the entire budget on the auxiliary mixture λ_*. Since $R(\lambda_*) > 0$ and $s(\lambda_*) = \Gamma_\infty > 1$, the matrix $A(\lambda_*)$ is positive definite; Lemma 17 gives

$$\limsup_{B\to\infty} BV_B \le \Gamma_\infty^{-1} < 1. \tag{52}$$

Now take an arbitrary sequence of exact minimizers and an arbitrary subsequence along which $t_B := S_B/B$ converges to $t \in [0,1]$. If $t = 0$, then $a_B(y_B)/B = 1 + t_B\{P(y_B/S_B) - 1\} \to 1$ whenever $S_B > 0$, while the second term in Equation (41) is nonnegative. Consequently $\liminf BV_B \ge 1$, contradicting Equation (52). The same contradiction is immediate if $S_B = 0$ infinitely often. Hence $t > 0$ and the auxiliary proportions have, after taking a further subsequence, a limit $\lambda \in \Delta_m$.

The information divided by B converges to $\bar{M} = (1-t)A_0 + tA(\lambda)$. If its upper-left entry is zero, the oracle bound gives $B\mathcal{R}_B(y_B) \ge B/a_B(y_B) \to +\infty$, contradicting Equation (52). Hence $\bar{a} > 0$. If $t = 1$ and $s(\lambda) = 0$, then the limit has positive diagonal entries and is singular, so Lemma 18 would give $BV_B \to \infty$, contradicting Equation (52). Otherwise $1 + t\{s(\lambda) - 1\} > 0$, so the limit is positive definite. Therefore Lemma 17 and Equation (46) yield

$$BV_B \longrightarrow \frac{1}{1 + t\{s(\lambda) - 1\}} \ge \frac{1}{1 + t(\Gamma_\infty - 1)} \ge \frac{1}{\Gamma_\infty}.$$

Together with Equation (52), equality must hold throughout. Since $\Gamma_\infty > 1$, this forces $t = 1$ and $s(\lambda) = \Gamma_\infty$. Every convergent subsequence has these properties, proving all three limits in Equation (51). □

F.5 The critical face

At the critical boundary, the leading risk matches trusted-only acquisition. The improvement is determined by the limiting bounded-normal gain. Recall $\Psi_H(w) = w - w^2 m_H(w)$.

Lemma 20 (Terminal bounded-normal improvement)**.** *For every $0 < H < \infty$, Ψ_H is strictly increasing and bounded on $[0,\infty)$, and*

$$L_H := \lim_{w\to\infty} \Psi_H(w) = \frac{\pi^2}{H^2}. \tag{53}$$

Proof. Rescaling $\eta = H\xi$ gives $m_H(w) = H^2 m_1(H^2 w)$ and $\Psi_H(w) = H^{-2}\Psi_1(H^2 w)$. The Fisher representation Equation (32) and Gaussian convolution show monotonicity. Indeed, if $0 < w_1 < w_2$, the output at precision w_1 is the output at precision w_2 plus independent Gaussian noise of variance $d = w_1^{-1} - w_2^{-1} > 0$. For every prior supported on $[-H, H]$, the output at precision w has variance at most $H^2 + w^{-1}$; the information–variance inequality and Equation (32) therefore give

$$\Psi_H(w) \ge \frac{1}{H^2 + w^{-1}} > 0, \qquad w > 0. \tag{54}$$

The score projection identity followed by Cauchy–Schwarz gives Stam's inequality $J_1^{-1} \ge J_2^{-1} + d$ (Stam, 1959). Applying it to priors approaching the infimum in Equation (32) gives

$$\Psi_H(w_1) \le \frac{\Psi_H(w_2)}{1 + d\Psi_H(w_2)} < \Psi_H(w_2),$$

where strictness follows from Equation (54). The cosine prior used in Equation (33) gives the uniform upper bound π^2/H^2, so the finite limit exists.

For its value, let $\rho(t)$ be the minimax risk from one $N(\eta, 1)$ observation over $|\eta| \le t$. Scaling the observation gives $m_H(w) = w^{-1}\rho(H\sqrt{w})$, and therefore $\Psi_H(w) = w\{1 - \rho(H\sqrt{w})\}$. Bickel's expansion gives $\rho(t) = 1 - \pi^2 t^{-2} + o(t^{-2})$ (Bickel, 1981). Substituting $t = H\sqrt{w}$ proves Equation (53). □

Put

$$s(\lambda) = P(\lambda) - \frac{Q(\lambda)^2}{R(\lambda)}, \qquad \Lambda_c = \{\lambda \in \Delta_m : s(\lambda) = 1\}.$$

For an auxiliary vector y, write $S = \mathbf{1}^\top y$ and define

$$D(y) := S - P(y) + \frac{Q(y)^2}{R(y)}, \qquad v(y) := \frac{Q(y)^2}{R(y)}, \tag{55}$$

with $D(0) = v(0) = 0$. When $\Gamma_\infty = 1$, homogeneity gives $D(y) \geq 0$.

Proposition 21 (Critical geometry and second-order value)**.** *Suppose* $\Gamma_\infty = 1$.

1. *If* $\Gamma_0 = 1$*, then* $V_B = B^{-1}$ *for every* $B > 0$.
2. *If* $\Gamma_0 > 1$*, every* $\lambda \in \Lambda_c$ *has* $Q(\lambda) \neq 0$*, and*

$$H_c := h\left|\frac{R(\lambda)}{Q(\lambda)}\right| \tag{56}$$

is independent of λ*. Moreover,*

$$V_B = B^{-1} - \frac{L_{H_c}}{B^2} + o(B^{-2}) = B^{-1} - \frac{\pi^2}{H_c^2 B^2} + o(B^{-2}). \tag{57}$$

For every B*, an exact optimum can be selected with* $S = B$ *and at most three auxiliary types. Every such selection satisfies*

$$\operatorname{dist}(y_B/B, \Lambda_c) \to 0, \qquad D(y_B) \to 0. \tag{58}$$

3. *In both cases* $BV_B \to 1$*, and for every* B *there exists an exact full-auxiliary optimum using at most three auxiliary types.*

Proof. Step 1: critical geometry. For every $\lambda_c \in \Lambda_c$, equality in the Schur complement gives

$$A(\lambda_c) - A_0 = \begin{bmatrix} Q_c^2/R_c & Q_c \\ Q_c & R_c \end{bmatrix} = z_c z_c^\top \succeq 0, \tag{59}$$

where $z_c = (Q_c/\sqrt{R_c}, \sqrt{R_c})^\top \neq 0$. If $z_c z_c^\top$ and $z_d z_d^\top$ are two critical increments, their average is also an admissible auxiliary mixture. Its Schur information is at least one by Equation (59) and at most one by $\Gamma_\infty = 1$; hence the sum of the two rank-one increments has rank one. Their ranges coincide. Consequently either all critical mixtures have $Q_c = 0, P_c = 1$, or all have $Q_c \neq 0$ and the ratio R_c/Q_c is common.

The first alternative is impossible when $\Gamma_0 > 1$. Choose a type with $p_j > 1$ and mix an ε fraction of it with a critical $\operatorname{diag}(1, R_c)$. Its Schur information is

$$1 + \varepsilon(p_j - 1) - \frac{\varepsilon^2 q_j^2}{(1-\varepsilon)R_c + \varepsilon r_j} > 1$$

for all sufficiently small $\varepsilon > 0$, a contradiction. This proves that Equation (56) is well defined and common.

Step 2: exact replacement and support. Start from any exact optimum with trusted expenditure $x_0 = B - S$. Replacing those units by a critical mixture changes the information by $x_0\{A(\lambda_c) - A_0\} \succeq 0$. Information monotonicity makes the replacement no worse, and optimality makes it another exact optimum, now with $S = B$. Restricting to the compact auxiliary hull, choose a trace-maximal optimum and repeat the exposed-face argument of Lemma 14; it uses at most three auxiliary information matrices. This construction establishes the claimed full-auxiliary optimum.

Step 3: the first-order value. Trusted-only acquisition gives $V_B \leq B^{-1}$. Given any sequence of designs, pass to a subsequence on which $S_B/B \to t \in [0, 1]$, and, if $t > 0$, their auxiliary proportions converge to λ. The normalized information tends to $(1-t)A_0 + tA(\lambda)$, whose generalized Schur information is

$$1 + t\{s(\lambda) - 1\} \leq 1.$$

If this number is positive, Lemma 17 gives a normalized risk at least one. If it is zero, Lemma 18 gives divergence. A zero target diagonal is handled by the oracle bound $\mathcal{R}_B \geq 1/a_B$, and $t = 0$ gives the same bound with $a_B/B \to 1$. Thus every subsequential lower limit is at least one, so $BV_B \to 1$.

If $\Gamma_0 = 1$, every $p_j \leq 1$. The oracle target precision of any design is therefore at most B, proving $V_B \geq 1/B$ and item 1.

Step 4: a global second-order envelope. Assume now $\Gamma_0 > 1$. If $Q(y) \neq 0$, put $H(y) = hR(y)/|Q(y)|$. The total information admits the exact decomposition

$$M_B(y) = \{B - D(y)\}A_0 + v(y) \begin{bmatrix} 1 \\ R(y)/Q(y) \end{bmatrix} \begin{bmatrix} 1 \\ R(y)/Q(y) \end{bmatrix}^\top.$$

The pair identity therefore gives, whenever $B > D(y)$,

$$\mathcal{R}_B(y) = \frac{1}{B - D(y)} - \frac{\Psi_{H(y)}\left(\frac{(B-D(y))v(y)}{B-D(y)+v(y)}\right)}{(B - D(y))^2}. \tag{60}$$

For fixed y, its limiting improvement is

$$\mathcal{C}(y) = P(y) - S - Q(y)^2 m_h(R(y)) = \Psi_{H(y)}(v(y)) - D(y), \tag{61}$$

with $\mathcal{C}(y) = -D(y)$ when $Q(y) = 0$.

We claim $\mathcal{C}(y) \leq L_{H_c}$ for every fixed y. Fix $\lambda_c \in \Lambda_c$, abbreviate its entries by P_c, Q_c, R_c, put $a_c := R_c/Q_c$, and compare

$$M_t = tA_0 + A(y), \qquad \widetilde{M}_t = tA(\lambda_c) + A(y) \succeq M_t.$$

Both designs have expenditure $t + S$. The finite-t Schur deficit of $\widetilde{M}_t$ relative to that exposure is

$$D_t(y; c) := (t + S) - tP_c - P(y) + \frac{\{tQ_c + Q(y)\}^2}{tR_c + R(y)} \geq 0, \tag{62}$$

$$D_t(y; c) = D_\infty(y; c) + O(t^{-1}), \qquad D_\infty(y; c) := S - P(y) + \frac{2Q(y)}{a_c} - \frac{R(y)}{a_c^2}. \tag{63}$$

The inequality in Equation (62) follows by applying $\Gamma_\infty = 1$ to the auxiliary mixture proportional to $t\lambda_c + y$. For the expansion, use $P_c - Q_c^2/R_c = 1$ and expand the displayed rational term in t^{-1}. In particular, $D_\infty(y; c) \geq 0$.

The coupled part of $\widetilde{M}_t$ has

$$v_t = \frac{\{tQ_c + Q(y)\}^2}{tR_c + R(y)} \longrightarrow \infty, \qquad H_t = h\frac{tR_c + R(y)}{|tQ_c + Q(y)|} \longrightarrow H_c.$$

Applying the exact decomposition Equation (60), with total budget $t + S$ and deficit $D_t(y; c)$, and then Lemma 20 gives

$$\lim_{t\to\infty} (t + S)^2\{(t + S)^{-1} - \mathcal{R}(\widetilde{M}_t)\} = L_{H_c} - D_\infty(y; c) \leq L_{H_c}.$$

Loewner monotonicity makes this improvement at least the corresponding one for M_t. Directly expanding Equation (41) for the fixed vector y shows that the latter improvement tends to $\mathcal{C}(y)$. Thus $\mathcal{C}(y) \leq L_{H_c}$. Conversely, along a critical ray $y = t\lambda_c$, $D = 0$, $H = H_c$, and $\mathcal{C}(y) = \Psi_{H_c}(tQ_c^2/R_c) \uparrow L_{H_c}$. Therefore

$$\sup_{y\geq 0} \mathcal{C}(y) = L_{H_c}. \tag{64}$$

Step 5: uniform second-order comparison. Select for each B a full-auxiliary exact optimizer y_B. Take any subsequence along which $y_B/B \to \bar{\lambda} \in \Delta_m$. The local, singular, and zero-target cases of Step 3, now with $t = 1$, show that $BV_B \to 1$ is possible only if $s(\bar{\lambda}) = 1$. Thus every subsequential limit lies in Λ_c, and

$$\operatorname{dist}(y_B/B, \Lambda_c) \to 0, \qquad \frac{D_B}{B} = 1 - s(y_B/B) \to 0, \tag{65}$$

where $D_B = D(y_B)$. Because Λ_c is compact and every one of its points has nonzero Q,

$$q_c := \min_{\lambda \in \Lambda_c} |Q(\lambda)| > 0.$$

Consequently $|Q(y_B)|/B \geq q_c/2$ for all sufficiently large B, while $D_B/B \to 0$ gives $B > D_B$. Thus the hypotheses used in the exact pair formula Equation (60) hold along the selected optimizers. Define $\alpha_B = B/(B - D_B)$ and $\mathcal{I}_B = B^2(B^{-1} - V_B)$. From Equation (60),

$$\mathcal{I}_B = \alpha_B^2 \Psi_{H(y_B)}\left(\frac{(B - D_B)v(y_B)}{B - D_B + v(y_B)}\right) - \alpha_B D_B. \tag{66}$$

PSD and the finite menu imply the uniform bound

$$\Psi_{H(y)}(w) \leq \frac{\pi^2}{H(y)^2} \leq \frac{\pi^2 p_{\max}}{h^2 r_{\min}} =: \bar{L}$$

for full-budget auxiliary aggregates; the case $Q = 0$ is zero. Since $\mathcal{I}_B \geq 0$ and eventually $\alpha_B \leq 2$, Equation (66) first yields $D_B \leq 4\bar{L}$, so $\alpha_B \to 1$. Monotonicity of Ψ_H and Equations (61) and (64) then give

$$\mathcal{I}_B \leq \mathcal{C}(y_B) + (\alpha_B^2 - 1)\bar{L} \leq L_{H_c} + o(1).$$

The critical design $B\lambda_c$ gives the matching lower limit through Equation (66). This proves Equation (57).

Step 6: selected optimizer refinement. Equality in the preceding comparison forces $\mathcal{C}(y_B) \to L_{H_c}$. By Equation (65), the common critical direction gives $H(y_B) \to H_c$ and $v(y_B) \asymp B$. Hence

$$\Psi_{H(y_B)}(v(y_B)) = H(y_B)^{-2}\Psi_1\{H(y_B)^2 v(y_B)\} \longrightarrow \frac{\pi^2}{H_c^2} = L_{H_c},$$

where the scaling identity makes the varying-H terminal passage explicit. Equation (61) then forces $D(y_B) \to 0$. □

Remark 22 (Risk and expenditure at criticality). For $A_1 = \mathrm{diag}(1, r)$, every split between trusted and auxiliary observations has risk $1/B$. Exact optima therefore include bounded expenditure, expenditure of order B^α for $0 < \alpha < 1$, and linear expenditure. On a coupled critical ray, every diverging expenditure up to B, including $\log B$ and a fixed positive fraction of the budget, attains the second-order coefficient in Equation (57). This coefficient alone does not distinguish the expenditure scales of such asymptotically optimal designs.

F.6 The local critical window

We now allow the finite auxiliary menu to depend on B. Use the notation from Theorem 6; in particular,

$$A_{j,B} \succeq 0, \qquad r_{j,B} \geq \underline{r} > 0, \qquad p_{j,B} \leq \overline{p} < \infty, \qquad \max_j \|A_{j,B} - A_j\| \to 0.$$

For $y \geq 0$, put $S = \mathbf{1}^\top y$ and define the budget-dependent information aggregates

$$P_B(y) = \sum_j p_{j,B} y_j, \qquad Q_B(y) = \sum_j q_{j,B} y_j, \qquad R_B(y) = \sum_j r_{j,B} y_j.$$

For $S > 0$, let

$$D_B(y) := S - P_B(y) + \frac{Q_B(y)^2}{R_B(y)} = S\{1 - s_B(y/S)\}, \qquad v_B(y) := \frac{Q_B(y)^2}{R_B(y)}, \tag{67}$$

and set $D_B(0) = v_B(0) = 0$. If $Q_B(y) \neq 0$, also let

$$H_B(y) := h\frac{R_B(y)}{|Q_B(y)|}.$$

When $Q_B(y) \neq 0$, the exact information decomposition is

$$M_B^{\triangle}(y) = \{B - D_B(y)\}A_0 + v_B(y)\begin{bmatrix} 1 \\ R_B(y)/Q_B(y) \end{bmatrix}\begin{bmatrix} 1 \\ R_B(y)/Q_B(y) \end{bmatrix}^\top. \tag{68}$$

Consequently, whenever $B > D_B(y)$, define

$$\alpha_B(y) := \frac{B}{B - D_B(y)}, \qquad w_B(y) := \frac{\{B - D_B(y)\}v_B(y)}{B - D_B(y) + v_B(y)}$$

and

$$\Phi_B(y) := \begin{cases} \Psi_{H_B(y)}\{w_B(y)\}, & Q_B(y) \neq 0, \\ 0, & Q_B(y) = 0. \end{cases}$$

The pair identity, with the second line also immediate when $Q_B(y) = 0$, then yields

$$\mathcal{I}_B(y) := B^2\{B^{-1} - \mathcal{R}_B^{\triangle}(y)\} = \alpha_B(y)^2\Phi_B(y) - \alpha_B(y)D_B(y). \tag{69}$$

Proof of Theorem 6. Write $L_c := \pi^2/H_c^2$ and $\mathcal{I}_B^* := B^2\{B^{-1} - V_B^{\triangle}\}$.

Step 1: uniform geometry and the varying-radius terminal limit. The lower bound on $r_{j,B}$ and the finite menu imply

$$\sup_{\lambda \in \Delta_m} |s_B(\lambda) - s(\lambda)| \longrightarrow 0. \tag{70}$$

Thus any sequence λ_B satisfying $s_B(\lambda_B) \to 1$ has all its accumulation points in Λ_c. Since $\Gamma_0 > 1$, the critical geometry in Proposition 21 gives $Q(\lambda) \neq 0$ and $hR(\lambda)/|Q(\lambda)| = H_c$ throughout Λ_c. Compactness therefore gives, uniformly as $\operatorname{dist}(\lambda_B, \Lambda_c) \to 0$,

$$\frac{hR_B(\lambda_B)}{|Q_B(\lambda_B)|} \to H_c, \qquad \frac{Q_B(\lambda_B)^2}{R_B(\lambda_B)} \text{ is bounded away from zero.} \tag{71}$$

In particular, if $H_B \to H_c$ and $z_B \to \infty$, then the scaling identity in Lemma 20 gives

$$\Psi_{H_B}(z_B) = H_B^{-2}\Psi_1(H_B^2 z_B) \longrightarrow L_c. \tag{72}$$

Positive semidefiniteness also gives $Q_B(\lambda)^2 \leq P_B(\lambda)R_B(\lambda)$. Hence, for all sufficiently large B,

$$0 \leq \Phi_B(y) \leq \frac{\pi^2\overline{p}}{h^2\underline{r}} =: \overline{L} \quad \text{whenever Equation (69) applies.} \tag{73}$$

Step 2: first-order localization of exact minimizers. Trusted-only acquisition is feasible, so $V_B^{\triangle} \leq B^{-1}$. Let y_B be exact minimizers and pass to a subsequence such that $S_B/B \to t \in [0,1]$. If $t > 0$, pass further so that $\lambda_B := y_B/S_B \to \lambda \in \Delta_m$. Uniform convergence of the information matrices shows that the information divided by B tends to

$$(1-t)A_0 + tA(\lambda),$$

whose generalized Schur information is $1 + t\{s(\lambda) - 1\} \leq 1$. The local and singular Gaussian limits used in Step 3 of the proof of Proposition 21 show that the normalized risk has lower limit at least one, with equality possible for $t > 0$ only if $s(\lambda) = 1$. If $t = 0$, the oracle bound $\mathcal{R}_B^{\triangle}(y_B) \geq \{B - S_B + P_B(y_B)\}^{-1}$ gives the same lower limit. Therefore

$$BV_B^{\triangle} \to 1, \qquad t > 0 \implies \lambda \in \Lambda_c. \tag{74}$$

It also follows from Equation (67) that, along every sequence of exact minimizers,

$$\frac{D_B(y_B)}{B} \to 0. \tag{75}$$

Indeed, on a subsequence with $S_B/B \to t > 0$, Equations (70) and (74) give $s_B(\lambda_B) \to 1$; on a subsequence with $t = 0$, the functions s_B are uniformly bounded and $D_B(y_B)/B = (S_B/B)\{1 - s_B(\lambda_B)\} \to 0$. Thus $B > D_B(y_B)$ eventually. When $t > 0$, Equation (71) also makes $Q_B(y_B) \neq 0$ eventually; when $t = 0$, the $Q_B(y_B) = 0$ convention in Equation (69) covers the remaining case.

Step 3: two matching constructions. Fix $\lambda_c \in \Lambda_c$, and put

$$\eta_B := B^{-1} + \max_j \|A_{j,B} - A_j\|, \qquad T_B := \eta_B^{-1/2}.$$

Then $T_B \to \infty$, $T_B \le \sqrt{B}$, and $T_B \max_j \|A_{j,B} - A_j\| \to 0$. For $y_B^\circ = T_B \lambda_c$,

$$D_B(y_B^\circ) = T_B\{1 - s_B(\lambda_c)\} = o(1), \qquad v_B(y_B^\circ) \asymp T_B \to \infty.$$

Moreover, $T_B/B \to 0$, so $w_B(y_B^\circ) \sim v_B(y_B^\circ)$. Equations (69), (71) and (72) give

$$\mathcal{I}_B(y_B^\circ) \to L_c. \tag{76}$$

If $\tau > 0$, choose $\lambda_B^* \in \arg\max_{\lambda \in \Delta_m} s_B(\lambda)$ and spend the entire budget on $y_B^* = B\lambda_B^*$. Uniform convergence implies that every accumulation point of λ_B^* belongs to Λ_c. Also

$$D_B(y_B^*) = B\{1 - \Gamma_{\infty,B}\} \to -\tau, \qquad v_B(y_B^*) \asymp B.$$

The same exact identity now yields

$$\mathcal{I}_B(y_B^*) \to L_c + \tau. \tag{77}$$

Combining the two constructions,

$$\liminf_{B\to\infty} \mathcal{I}_B^* \ge L_c + (\tau)_+. \tag{78}$$

Step 4: the uniform upper bound. Return to exact minimizers y_B. From $s_B(\lambda_B) \le \Gamma_{\infty,B}$, with the left side interpreted only when $S_B > 0$,

$$-D_B(y_B) = S_B\{s_B(\lambda_B) - 1\} \le \frac{S_B}{B} B\{\Gamma_{\infty,B} - 1\}. \tag{79}$$

This bounds $D_B(y_B)$ from below. In view of Equation (75), $\alpha_B(y_B) \to 1$; then Equations (69), (73) and (78) bounds $D_B(y_B)$ from above. Hence

$$|D_B(y_B)| = O(1). \tag{80}$$

Pass to a subsequence on which $S_B/B \to t$. If $t > 0$, then λ_B approaches Λ_c by Equation (74). Consequently $H_B(y_B) \to H_c$, $v_B(y_B) \asymp B$, and $w_B(y_B) \to \infty$. Thus $\Phi_B(y_B) \to L_c$, while Equation (79) gives

$$\limsup_B\{-D_B(y_B)\} \le t\tau.$$

Using Equation (69),

$$\limsup_B \mathcal{I}_B^* \le L_c + t\tau, \qquad t > 0. \tag{81}$$

Suppose instead that $t = 0$. If S_B is bounded, pass further so that $y_B \to y$. Uniform convergence of the information matrices and the same fixed-compact algebra as in Equation (94) give

$$\mathcal{I}_B^* \to \mathcal{C}(y) \le L_c,$$

where the inequality is Equation (64). If $S_B \to \infty$, pass further so that $\lambda_B \to \lambda$. Boundedness of $D_B(y_B)$ gives

$$1 - s_B(\lambda_B) = D_B(y_B)/S_B \to 0,$$

so $\lambda \in \Lambda_c$. Now $H_B(y_B) \to H_c$, $v_B(y_B) \asymp S_B \to \infty$, and $v_B(y_B) = o(B)$, whence $w_B(y_B) \to \infty$ and $\Phi_B(y_B) \to L_c$. Finally, Equation (79) gives $\limsup_B\{-D_B(y_B)\} \le 0$, because $S_B/B \to 0$. Therefore

$$\limsup_B \mathcal{I}_B^* \le L_c, \qquad t = 0. \tag{82}$$

Equations (81) and (82) imply

$$\limsup_B \mathcal{I}_B^* \le L_c + (\tau)_+.$$

Together with Equation (78), this proves Equation (22).

Step 5: optimizer selection. If $\tau > 0$, a subsequential limit $S_B/B \to t < 1$ would make Equations (81) and (82) strictly smaller than the established limit $L_c + \tau$. Hence $S_B/B \to 1$. Along such minimizers, $\Phi_B(y_B) \to L_c$, so Equation (69) gives $-D_B(y_B) \to \tau$. Therefore

$$B\{s_B(\lambda_B) - 1\} = \frac{B}{S_B}\{-D_B(y_B)\} \to \tau.$$

Subtracting this limit from $B\{\Gamma_{\infty,B} - 1\} \to \tau$ proves Equation (23).

If $\tau < 0$, any subsequential limit $t > 0$ would give $\limsup \mathcal{I}_B^* \le L_c + t\tau < L_c$, contradicting the established limit L_c. Thus $S_B/B \to 0$. When $\tau = 0$, the upper bounds agree for every t, so the second-order risk does not distinguish the limiting budget shares. □

Corollary 23 (Differentiable paths through the critical face)**.** *In addition to the assumptions of Theorem 6, suppose, uniformly over the finite menu,*

$$A_{j,B} = A_j + B^{-1}\dot{A}_j + o(B^{-1}), \qquad \dot{A}_j = \begin{bmatrix} \dot{p}_j & \dot{q}_j \\ \dot{q}_j & \dot{r}_j \end{bmatrix}. \tag{83}$$

For a mixture, use dots for the corresponding linear aggregates and define

$$\dot{s}(\lambda) := \dot{P}(\lambda) - 2\frac{Q(\lambda)}{R(\lambda)}\dot{Q}(\lambda) + \frac{Q(\lambda)^2}{R(\lambda)^2}\dot{R}(\lambda), \qquad \tau_* := \max_{\lambda\in\Lambda_c} \dot{s}(\lambda). \tag{84}$$

Then

$$B\{\Gamma_{\infty,B} - 1\} \to \tau_*,$$

so Equation (22) *holds with $\tau = \tau_*$. If $\tau_* > 0$, every accumulation point of the exact optimizer proportions λ_B belongs to* $\arg\max_{\lambda\in\Lambda_c} \dot{s}(\lambda)$.

Proof. Since all nuisance diagonals are uniformly bounded away from zero, Taylor expansion of the Schur complement is uniform on Δ_m:

$$s_B(\lambda) = s(\lambda) + B^{-1}\dot{s}(\lambda) + o(B^{-1}). \tag{85}$$

Evaluating at any maximizer of $\dot{s}$ on Λ_c gives

$$\liminf_B B\{\Gamma_{\infty,B} - 1\} \geq \tau_*.$$

Conversely, choose $\lambda_B^* \in \arg\max s_B$ and pass to a convergent subsequence. The preceding lower bound and Equation (85) force every limit into Λ_c; since $s(\lambda_B^*) \leq 1$,

$$\limsup_B B\{\Gamma_{\infty,B} - 1\} \leq \limsup_B \dot{s}(\lambda_B^*) \leq \tau_*.$$

This proves the drift formula and hence the value statement.

When $\tau_* > 0$, Equation (23) and Equation (85) imply

$$0 = B\{\Gamma_{\infty,B} - s_B(\lambda_B)\} + o(1) = B\{1 - s(\lambda_B)\} + \tau_* - \dot{s}(\lambda_B) + o(1).$$

Every accumulation point lies in Λ_c. Along any convergent subsequence, the two displayed terms have nonnegative lower limits; their sum can tend to zero only if $\dot{s}(\lambda_B) \to \tau_*$. This proves the selection claim. □

Example 24 (An exactly solvable local critical ray)**.** Let $z = (u, v)^\top$ with $uv \neq 0$, and set

$$A_c = A_0 + zz^\top = \begin{bmatrix} 1+u^2 & uv \\ uv & v^2 \end{bmatrix}, \qquad A_{1,B} = A_c + \frac{\tau}{B}A_0.$$

For every fixed $\tau \in \mathbb{R}$, $A_{1,B}$ is positive definite for all sufficiently large B. This single auxiliary type satisfies

$$\Gamma_{\infty,B} = 1 + u^2 + \frac{\tau}{B} - \frac{u^2v^2}{v^2} = 1 + \frac{\tau}{B}, \qquad \Gamma_0 = 1 + u^2 > 1, \qquad H_c = h\left|\frac{v}{u}\right|.$$

For the full-auxiliary design, $D_B = -\tau$, $v_B = Bu^2$, and its exact risk is

$$\frac{1}{B+\tau} - \frac{\Psi_{H_c}\left(\dfrac{(B+\tau)Bu^2}{B+\tau+Bu^2}\right)}{(B+\tau)^2}.$$

Its second-order improvement is $L_c + \tau$. On the other hand, spending $T_B = \sqrt{B}$ on the auxiliary type gives $D_B = -\tau/\sqrt{B} \to 0$ and coupled precision of order $\sqrt{B}$, hence second-order improvement L_c. These two allocations attain the competing second-order gains whose maximum gives $(\tau)_+$ in Equation (22).

F.7 BOUNDED AUXILIARY SPENDING

Assume throughout this subsection that

$$\Gamma_\infty < 1 < \Gamma_0, \qquad \Delta := 1 - \Gamma_\infty > 0. \tag{86}$$

The homogeneous form of the gap is, for every $y \geq 0$ with $S = \mathbf{1}^\top y > 0$,

$$S - \left\{ P(y) - \frac{Q(y)^2}{R(y)} \right\} = S\{1 - s(y/S)\} \geq \Delta S. \tag{87}$$

Lemma 25 (Uniform boundedness of every optimal auxiliary vector)**.** *For every selection $y_B \in \arg\min_{y \in \mathcal{Y}_B} \mathcal{R}_B(y)$, there are $B_0 < \infty$ and $K < \infty$, depending only on the observation types and h, such that $\mathbf{1}^\top y_B \leq K$ for every $B \geq B_0$.*

Proof. Every exact minimizer satisfies

$$\mathcal{R}_B(y_B) \leq \mathcal{R}_B(0) = B^{-1}. \tag{88}$$

Write $S_B = \mathbf{1}^\top y_B$. We first prove $S_B/B \to 0$. Otherwise there are $B_k \to \infty$ and a subsequence such that $t_k = S_{B_k}/B_k \to t > 0$. Passing to a further subsequence, the auxiliary proportions converge to some $\lambda \in \Delta_m$. As in the supercritical proof, the normalized information converges to $\bar{M} = (1-t)A_0 + tA(\lambda)$. If the target-information entry of $\bar{M}$ is zero, the oracle bound $B_k \mathcal{R}_{B_k}(y_{B_k}) \geq B_k / a_{B_k}(y_{B_k})$ diverges. If $\bar{M}$ has positive diagonal entries but is singular, Lemma 18 gives the same divergence. Both cases contradict the benchmark. In the remaining case, $J_{\mathrm{Sch}}(\bar{M}) = 1 + t\{s(\lambda) - 1\} > 0$, and Lemma 17 and Equation (46) give

$$B_k \mathcal{R}_{B_k}(y_{B_k}) \longrightarrow \frac{1}{1 + t\{s(\lambda) - 1\}} \geq \frac{1}{1 - t\Delta} > 1.$$

The last denominator is positive: this is immediate for $t < 1$; for $t = 1$ the present branch has $s(\lambda) > 0$, hence $\Gamma_\infty > 0$ and $\Delta < 1$. This contradicts Equation (88). Thus $S_B/B \to 0$.

Suppose now, for contradiction, that S_B is unbounded. Along a subsequence $B_k \to \infty$, we then have $S_k := S_{B_k} \to \infty$ and $S_k/B_k \to 0$. Pass again to a subsequence so that $y_{B_k}/S_k \to \lambda \in \Delta_m$, and abbreviate $e(\lambda) = P(\lambda) - 1$, $q(\lambda) = Q(\lambda)$, and $r(\lambda) = R(\lambda)$. Since all $r_j > 0$, $r(\lambda) > 0$. Directly from Equation (41),

$$\begin{aligned} &\frac{B_k^2}{S_k} \left\{ \mathcal{R}_{B_k}(y_{B_k}) - \frac{1}{B_k} \right\} \\ &\quad = -\frac{B_k e(y_{B_k}/S_k)}{B_k + S_k e(y_{B_k}/S_k)} + \left\{ \frac{B_k}{B_k + S_k e(y_{B_k}/S_k)} \right\}^2 S_k q(y_{B_k}/S_k)^2 m_h(I_k), \end{aligned} \tag{89}$$

where

$$\frac{I_k}{S_k} = r(y_{B_k}/S_k) - \frac{S_k q(y_{B_k}/S_k)^2}{B_k + S_k e(y_{B_k}/S_k)} \longrightarrow r(\lambda) > 0.$$

Thus $I_k \to \infty$ and, by Equation (43), $S_k m_h(I_k) \to 1/r(\lambda)$. The right-hand side of Equation (89) converges to

$$-e(\lambda) + \frac{q(\lambda)^2}{r(\lambda)} = 1 - s(\lambda) \geq \Delta > 0.$$

But Equation (88) makes the left-hand side nonpositive. This contradiction proves boundedness. If no common K existed, the preceding argument could be applied to a sequence with $S_B \to \infty$; hence the bound is uniform over all exact minimizers. □

Define the limiting improvement functional, with its continuous value at $y = 0$, by

$$\mathcal{C}(y) = E(y) - Q(y)^2 m_h(R(y)), \qquad \mathcal{C}(0) = 0. \tag{90}$$

Lemma 26 (Tail decay and positivity of the limiting improvement)**.** *Under Equation* (86)*, $\mathcal{C}$ is continuous and $\mathcal{C}(y) \to -\infty$ as $\mathbf{1}^\top y \to \infty$. Its maximum C_* is attained on a nonempty compact set, and $C_* > 0$.*

Proof. Continuity follows from Lemma 12 and the linearity of E, Q, R. For $y = t\lambda$ with $t > 0$ and $\lambda \in \Delta_m$,

$$\frac{\mathcal{C}(t\lambda)}{t} = P(\lambda) - 1 - tQ(\lambda)^2 m_h\{tR(\lambda)\}. \tag{91}$$

Because $R(\lambda) \geq \min_j r_j > 0$, the convergence in Equation (43) is uniform over $\lambda \in \Delta_m$ as $t \to \infty$. Hence the right-hand side of Equation (91) converges uniformly to

$$P(\lambda) - 1 - \frac{Q(\lambda)^2}{R(\lambda)} = s(\lambda) - 1 \leq -\Delta.$$

For all sufficiently large t, $\mathcal{C}(t\lambda) \leq -\Delta t/2$ uniformly in λ. This is the claimed tail decay, and continuity then gives a nonempty compact argmax.

Finally, $\Gamma_0 > 1$ supplies a type j with $e_j > 0$. With $\mathbf{e}^{(j)}$ denoting the jth coordinate vector, and since $m_h(I) \leq h^2$,

$$\mathcal{C}(t\mathbf{e}^{(j)}) \geq e_j t - q_j^2 h^2 t^2 > 0$$

for all sufficiently small $t > 0$ (with the conclusion immediate if $q_j = 0$). Thus $C_* > 0$. □

Proposition 27 (Limit portfolio and second-order value)**.** *Every accumulation point of exact minimizers maximizes $\mathcal{C}$, and*

$$V_B = B^{-1} - C_* B^{-2} + o(B^{-2}). \tag{92}$$

There exists a maximizer of $\mathcal{C}$ supported on at most two auxiliary types.

Proof. For every feasible y, define the exactly rescaled excess risk

$$\begin{aligned} G_B(y) &:= B^2\{\mathcal{R}_B(y) - B^{-1}\} \\ &= -\frac{BE(y)}{B + E(y)} + \left\{\frac{B}{B + E(y)}\right\}^2 Q(y)^2 m_h\left(R(y) - \frac{Q(y)^2}{B + E(y)}\right). \end{aligned} \tag{93}$$

On every fixed compact subset of $\mathbb{R}^m_+$, the denominators are positive for all sufficiently large B. The argument of m_h converges uniformly to $R(y)$, and Lemma 12 gives

$$G_B(y) \longrightarrow -\mathcal{C}(y) \tag{94}$$

uniformly on that compact set.

By Lemma 25, all exact minimizers for sufficiently large B lie in one fixed compact set. Let $y_{B_k} \to y_*$ be any convergent subsequence. Every fixed $y \geq 0$ is feasible once $B_k \geq \mathbf{1}^\top y$, and exact optimality gives $G_{B_k}(y_{B_k}) \leq G_{B_k}(y)$. Apply uniform convergence on a compact set containing y, y_*, and the minimizers to obtain $-\mathcal{C}(y_*) \leq -\mathcal{C}(y)$. Since y was arbitrary, y_* maximizes $\mathcal{C}$.

Choose one maximizer $\bar{y}$ of $\mathcal{C}$. It is feasible for all large B, so the preceding comparison gives

$$\limsup_B \min_{y \in \mathcal{Y}_B} G_B(y) \leq -C_*.$$

Conversely, boundedness and uniform convergence imply that every subsequence of minimizing values has lower limit at least $-C_*$. Therefore $\min G_B \to -C_*$, which is exactly Equation (92).

It remains to sharpen the limiting support. For every sufficiently large B, select the exact optimum supplied by Lemma 14. The boundedness lemma applies to this selection, so its trusted expenditure is $B - \mathbf{1}^\top y_B > 0$ for all large B. One of its at most three total observation types is therefore the trusted target, leaving at most two auxiliary types. There are only finitely many such pairs. Pass to a subsequence on which the pair is fixed and then to a convergent subsequence. Its limit maximizes $\mathcal{C}$ by the first part of the proof and uses at most those two auxiliary types. □

The trusted-only, auxiliary-dominant, and bounded-spending results together prove Theorem 3. The support and small-budget lemmas prove Proposition 4, and Proposition 21 proves Theorem 5.

F.8 Opposite correlations can cancel

For the two matrices in the main-text example,

$$\det(A_+) = \det(A_-) = \frac{4}{9} > 0,$$

so both are positive definite. Individually,

$$J_{\mathrm{Sch}}(A_+) = J_{\mathrm{Sch}}(A_-) = \frac{5}{3} - \frac{(4/3)^2}{4/3} = \frac{1}{3}.$$

At mixture weight λ on A_+ and $1-\lambda$ on A_-,

$$s(\lambda) = \frac{5}{3} - \frac{4}{3}(2\lambda - 1)^2.$$

It is maximized at $\lambda = 1/2$, where it equals $5/3$. Thus the two individual rates are below the trusted rate while $\Gamma_\infty = 5/3 > 1$. Equal spending cancels the off-diagonal entry and gives information $B\,\mathrm{diag}(5/3, 4/3)$, whose exact target risk is $3/(5B)$. Both matrices can be realized by paired Gaussian observations. Before normalizing by the trusted rate $3/2$, use the usual pair loading $F = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ and

$$\Sigma_+ = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}, \qquad \Sigma_- = \begin{bmatrix} 2 & 4 \\ 4 & 17/2 \end{bmatrix}.$$

Their determinants are one, and direct multiplication gives $F^\top \Sigma_\pm^{-1} F = \left[\begin{smallmatrix} 5/2 & \pm 2 \\ \pm 2 & 2 \end{smallmatrix}\right]$. Dividing by $3/2$ produces the main-text $A_\pm$. More generally, any positive-definite information matrix A under this invertible loading is realized by $\Sigma = FA^{-1}F^\top$.

F.9 Calibration-only observations can change the phase

For Example 9, positive semidefiniteness is immediate: $A_1 = 2(1,1)^\top(1,1)$ has rank one and $A_2 = \mathrm{diag}(0, 100)$. The first type has

$$J_{\mathrm{Sch}}(A_1) = 2 - \frac{2^2}{2} = 0.$$

Their stated mixture is

$$\frac{3}{4}A_1 + \frac{1}{4}A_2 = \begin{bmatrix} 3/2 & 3/2 \\ 3/2 & 53/2 \end{bmatrix},$$

and hence

$$J_{\mathrm{Sch}}\left(\frac{3}{4}A_1 + \frac{1}{4}A_2\right) = \frac{3}{2} - \frac{(3/2)^2}{53/2} = \frac{75}{53} > 1.$$

The calibration observation thus raises the mixture rate above the trusted rate, as stated in Example 9.

F.10 Aligned portfolios

Suppose now that $q_j/r_j = \ell \neq 0$ for every j. Put

$$a = \ell^{-1}, \qquad v = \begin{bmatrix} 1 \\ a \end{bmatrix}, \qquad \tau_j = \frac{q_j^2}{r_j} > 0, \qquad \sigma_j = p_j - \tau_j \geq 0.$$

Direct substitution verifies the decomposition

$$A_j = \sigma_j e_1 e_1^\top + \tau_j v v^\top. \tag{95}$$

Lemma 28 (Two total types for an aligned finite-budget optimum)**.** *For every $B > 0$, some exact aligned optimum uses at most two total observation types, counting the trusted target.*

Proof. The decomposition Equation (95) realizes the experiment as two independent channels: pure target precision u and common bounded-bias precision w. Per unit budget, feasible (u, w) rates form the polygon

$$\mathcal{P} = \mathrm{conv}\{(1,0), (\sigma_1, \tau_1), \ldots, (\sigma_m, \tau_m)\} \subset \mathbb{R}^2.$$

The exact minimax risk is coordinatewise nonincreasing in (u, w), since an added independent observation in either channel may be discarded. Among the optimal points of $B\mathcal{P}$, choose one maximizing $u + w$. It is not strictly dominated in the nonnegative orthant. If $\dim \mathcal{P} = 2$, the same separation argument as in Lemma 14 places it on an exposed face of dimension at most one; that face is generated by at most two original rate points. If $\dim \mathcal{P} \le 1$, ordinary Carathéodory in its affine hull also uses at most two points. Their spending weights give the asserted design. □

Corollary 29 (Aligned portfolio selection)**.** *If* $\max_j \sigma_j < 1 < \max_j(\sigma_j + \tau_j)$*, define*

$$H = \frac{h}{|\ell|}, \qquad \rho_j = \frac{1 - \sigma_j}{\tau_j}, \qquad \rho_* = \min_j \rho_j.$$

Then $\Gamma_\infty = \max_j \sigma_j$ *and*

$$C_* = \max_{w \ge 0} \{\Psi_H(w) - \rho_* w\}. \tag{96}$$

For every finite B, some exact optimum uses at most two total observation types. A limiting optimum can use one auxiliary type attaining ρ_*; if that type is unique, every accumulation point uses it exclusively.

Proof. For an auxiliary mixture,

$$A(\lambda) = \bar{\sigma} e_1 e_1^\top + \bar{\tau} v v^\top, \qquad \bar{\sigma} = \sum_j \lambda_j \sigma_j, \quad \bar{\tau} = \sum_j \lambda_j \tau_j > 0.$$

Its Schur information is exactly $\bar{\sigma}$, because its entries are $\bar{\sigma} + \bar{\tau}, a\bar{\tau}, a^2\bar{\tau}$. Consequently

$$\Gamma_\infty = \max_j \sigma_j, \qquad \Gamma_0 = \max_j (\sigma_j + \tau_j). \tag{97}$$

Assume the aligned strict-intermediate inequalities

$$\max_j \sigma_j < 1 < \max_j (\sigma_j + \tau_j). \tag{98}$$

They are exactly $\Gamma_\infty < 1 < \Gamma_0$. Let

$$d_j = 1 - \sigma_j > 0, \qquad \rho_j = d_j/\tau_j, \qquad \rho_* = \min_j \rho_j, \qquad H = |a|h = h/|\ell|.$$

For $W = \sum_j \tau_j y_j$, the aggregates in Equation (90) obey

$$E(y) = W - \sum_j d_j y_j, \qquad Q(y) = aW, \qquad R(y) = a^2 W.$$

The elementary rescaling of the bounded-normal problem gives

$$a^2 \mathfrak{m}_h(a^2 W) = \mathfrak{m}_H(W).$$

Indeed, multiplying both the parameter and its estimator by a changes the radius from h to $|a|h$ and multiplies squared error by a^2. Therefore

$$\begin{aligned} \mathcal{C}(y) &= W - W^2 \mathfrak{m}_H(W) - \sum_j d_j y_j \\ &= \Psi_H(W) - \sum_j d_j y_j. \end{aligned} \tag{99}$$

For every representation of W,

$$\sum_j d_j y_j = \sum_j \rho_j \tau_j y_j \ge \rho_* W.$$

Equality is achieved by assigning all of W to any one type attaining ρ_*, using expenditure $y_j = W/\tau_j$. Hence

$$\max_{y \geq 0} \mathcal{C}(y) = \max_{W \geq 0} \{\Psi_H(W) - \rho_* W\}. \tag{100}$$

The maximum is positive by Lemma 26, so every maximizing W is positive. Thus a limiting optimum can always be represented by one minimizing-ratio auxiliary type. If that type is unique, equality in the preceding weighted average is possible only when every other y_j is zero. Since every accumulation point maximizes $\mathcal{C}$, all accumulation points then use that unique type exclusively. Together with Lemma 28, this proves Corollary 29. □

F.11 Stability of the phase indices

Proposition 30 (Stability under information perturbations)**.** *Let A_j and $\widetilde{A}_j$ be two collections of positive-semidefinite information matrices for the same observation types. Suppose*

$$r_j, \widetilde{r}_j \geq \underline{r} > 0, \qquad p_j, \widetilde{p}_j \leq \overline{p},$$

and

$$\max_j \max\{|p_j - \widetilde{p}_j|, |q_j - \widetilde{q}_j|, |r_j - \widetilde{r}_j|\} \leq \varepsilon.$$

Then Equation (25) *holds.*

Proof. The first bound follows immediately by taking maxima of the top-left entries. For the second, fix $\lambda \in \Delta_m$, and let (p_t, q_t, r_t), $0 \leq t \leq 1$, be the line segment between the two corresponding mixture information matrices. It remains PSD, satisfies $p_t \leq \overline{p}$, and has $r_t \geq \underline{r}$. Therefore

$$\frac{|q_t|}{r_t} \leq \sqrt{\frac{p_t}{r_t}} \leq \sqrt{\frac{\overline{p}}{\underline{r}}} =: \zeta.$$

For $\phi(p, q, r) = p - q^2/r$,

$$|\partial_p \phi| + |\partial_q \phi| + |\partial_r \phi| = 1 + \frac{2|q|}{r} + \frac{q^2}{r^2} \leq (1 + \zeta)^2.$$

Each coordinate of the two mixture information matrices differs by at most ε. Integrating the derivative along the segment gives

$$|s(\lambda) - \widetilde{s}(\lambda)| \leq (1 + \zeta)^2 \varepsilon.$$

Taking suprema over the same simplex proves the claim. □

F.12 Integer replication

Proof of Proposition 10. At integer budget N, an auxiliary replication vector $n \in \mathbb{Z}_+^m$ produces expenditure $y = (c_1 n_1, \ldots, c_m n_m) \in \mathcal{L}$. Because the costs c_j and N are integers and a trusted replicate costs one, every remaining unit $N - \mathbf{1}^\top y$ can be spent on the trusted type. Thus

$$V_N^{\mathbb{Z}} = \min_{\substack{y \in \mathcal{L} \\ \mathbf{1}^\top y \leq N}} \mathcal{R}_N(y). \tag{101}$$

By Lemma 26, $\mathcal{C}(y) \to -\infty$ as $\mathbf{1}^\top y \to \infty$. The lattice $\mathcal{L}$ is discrete, so $\mathcal{C}$ attains its maximum on $\mathcal{L}$. Since $0 \in \mathcal{L}$, and since $\mathcal{L} \subset \mathbb{R}_+^m$,

$$0 = \mathcal{C}(0) \leq C_{\mathbb{Z}} \leq C_*.$$

Let y_N be any exact integer minimizers. Trusted-only acquisition is integer feasible, so $\mathcal{R}_N(y_N) \leq N^{-1}$. The proof of Lemma 25 uses only this benchmark, the strict mixture gap Equation (87), and the exact risk formula; it does not use divisibility or continuity of the feasible set. Applying the same boundedness argument to the integer minimizers gives constants $N_0, K < \infty$ such that

$$\mathbf{1}^\top y_N \leq K \qquad (N \geq N_0) \tag{102}$$

uniformly over all exact integer minimizers.

Choose any $y^\star \in \arg\max_{y\in\mathcal{L}} \mathcal{C}(y)$, and enlarge K, if necessary, so that $\mathbf{1}^\top y^\star \le K$. The set $\mathcal{L}_K := \{y \in \mathcal{L} : \mathbf{1}^\top y \le K\}$ is finite. On it, the exact rescaling in Equation (93) gives uniformly

$$N^2\{\mathcal{R}_N(y) - N^{-1}\} \longrightarrow -\mathcal{C}(y). \tag{103}$$

The design $y^\star$ is feasible for all sufficiently large N, so exact optimality and Equation (103) give

$$\limsup_N N^2\{V_N^{\mathbb{Z}} - N^{-1}\} \le -C_{\mathbb{Z}}.$$

Conversely, Equation (102) places every large-N optimizer in the finite set $\mathcal{L}_K$. Passing to a constant subsequence and using Equation (103) gives the reverse lower bound. Hence

$$N^2\{V_N^{\mathbb{Z}} - N^{-1}\} \to -C_{\mathbb{Z}},$$

which is Equation (27).

Finally, within the finite set $\mathcal{L}_K$, the values of $\mathcal{C}$ outside its lattice argmax are separated from $C_{\mathbb{Z}}$ by a positive gap. Uniform convergence in Equation (103), together with comparison to $y^\star$, excludes all such points for sufficiently large N. Thus every exact integer optimizer eventually belongs to $\arg\max_{y\in\mathcal{L}} \mathcal{C}(y)$. □

Proposition 31 (Risk bounds for integer allocations)**.** *Suppose every observation type $j = 0, 1, \ldots, m$, including the trusted type $j = 0$, has per-replicate cost $0 < c_j \le c_{\max}$, and every continuous optimum can be selected with at most D total observation types, again counting the trusted type. Let $V_{\mathrm{cont}}(B)$ and $V_{\mathbb{Z}}(B)$ denote the minimax values with, respectively, nonnegative real and nonnegative integer replication counts. For $B > Dc_{\max}$,*

$$V_{\mathrm{cont}}(B) \le V_{\mathbb{Z}}(B) \le V_{\mathrm{cont}}(B - Dc_{\max}). \tag{104}$$

Here $D = 3$ for the scalar portfolio (trusted type included) and $D = (d+1)(d+2)/2$ for the vector extension.

Proof. The left inequality follows because integer designs form a subset of continuous designs. For the right inequality, take a D-supported continuous optimum at budget $B - Dc_{\max}$, and round every positive replication count upward to the next integer. At most D counts are rounded, so the resulting cost is no more than

$$B - Dc_{\max} + Dc_{\max} = B.$$

Rounding upward adds a positive semidefinite information matrix. Information monotonicity therefore makes the rounded design's risk no larger than $V_{\mathrm{cont}}(B - Dc_{\max})$, proving the upper bound. □

In particular, whenever $V_{\mathrm{cont}}(B) \sim c/B$, the integer and continuous values have ratio tending to one. Thus the leading trusted-only, saturation, critical, and auxiliary-dominant value regimes survive fixed integer costs. The lattice constant in Proposition 10 characterizes the corresponding second-order saturation gain.

G Further related work

Prediction-powered inference and source acquisition. Prediction-powered inference estimates a population quantity by combining a prediction-based estimate with a correction from trusted labels. PPI++ improves the efficiency of this correction, while cross-prediction-powered inference allows the predictor to be trained using the available labeled sample (Angelopoulos et al., 2023b; Zrnic & Candès, 2024b). Stratified PPI and active inference direct labeling effort toward informative observations (Fisch et al., 2024; Zrnic & Candès, 2024a). Cost-aware evaluation further accounts for the relative prices of weak and strong raters, and MultiPPI optimizes acquisition across several correlated prediction sources (Angelopoulos et al., 2025; Cowen-Breen et al., 2026). These developments establish how predictions and selective labeling can improve inference. Our question concerns the additional information supplied by a known finite bound on a shared discrepancy. This bound changes the optimal correction and produces an acquisition regime in which auxiliary observations remain useful even though their optimal expenditure stays bounded.

Bounded bias and combining evidence. The geometry of a parameter space governs the minimax risk of Gaussian estimation and the performance of linear procedures (Donoho & Liu, 1991; Donoho, 1994). Bounded differences between normal means provide a direct model for combining a trusted estimate with an auxiliary estimate (van Eeden & Zidek, 2001; 2004). In the integration of experimental and observational data, shrinkage and adaptive methods study the tradeoff between a low-variance, potentially biased estimate and an unbiased estimate (Oberst et al., 2022; Rosenman et al., 2023; Chen et al., 2025). Adaptation to misspecification makes the cost of uncertainty about the bias itself explicit (Armstrong et al., 2025). These estimation questions supply the statistical foundation for our acquisition problem. We optimize the exact risk over observations with different costs and shared bias loadings, obtaining expenditure laws and second-order gains that describe how the preferred combination changes with the total budget.

Robust design and nuisance information. Robust experimental design controls estimation error under departures from a working regression model (Sacks & Ylvisaker, 1978; Wiens, 1992). Ellipsoidal uncertainty also leads to minimax linear estimation problems with semidefinite formulations (Eldar et al., 2005). More recent design methods account for the shrinkage estimator used after an experiment and for biased external evidence available alongside experimental data (Rosenman & Miratrix, 2023; Epanomeritakis & Viviano, 2026). A related Bayesian active-learning question is how to balance information about a target against information about nuisance parameters (Sloman et al., 2024). In our model, calibration-only observations make this interaction concrete: their value comes from improving the use of other sources. The shared discrepancy structure yields explicit information rates for this complementarity and identifies when calibration changes the optimal acquisition regime.

Approximate design, sparse support, and integer allocation. Classical design theory relates optimality to the geometry of information matrices and their supporting faces (Elfving, 1952; Kiefer & Wolfowitz, 1960; Pukelsheim, 2006). Computational design and active regression develop efficient sample-selection procedures, including regret-based design, linear-sample sparsification, and minimax regression sampling (Allen-Zhu et al., 2017; Chen & Price, 2019; Dereziński et al., 2019). Online design additionally addresses unknown, observation-dependent noise levels (Fontaine et al., 2021). Our support result uses information monotonicity to obtain an exact optimum with at most three observation types despite the generally nonconvex risk. The affine formulation complements this existence result with a convex allocation problem and a uniform $5/4$ guarantee. Rounding theory connects approximate designs to finite replication counts (Pukelsheim & Rieder, 1992); here, sparse support gives a direct risk comparison, while the lattice constant $C_{\mathbb{Z}}$ identifies the exact second-order effect of integer costs in the saturation regime.