When Shared Inputs Prevent Heavy Tails from Slowing Model Selection: A Minimax Separation

Anonymous authors
Paper under double-blind review

Abstract

Selecting among models with similar performance can require many evaluations, especially when rare large losses affect their ranking. We study how evaluating models on common inputs changes the optimal rate of selection by Expected Shortfall (ES), the average loss over a fixed upper-tail fraction, and what remains achievable from independent evaluations.

For K candidates and one reference, we characterize model selection under synchronized evaluation on common inputs and independent evaluation, using the same marginal loss distributions and n observations per model. The reference loss has a bounded pth moment, with $1 < p < 2$, while differences between candidate and reference losses on the same input have uniformly bounded L^r norms, with $2 \le r \le \infty$. Over this class with fixed moment bounds, we establish matching minimax rates for expected simple regret, defined as the expected ES gap between the selected candidate and the best candidate. The rates are $\sqrt{\log K/n}$ under synchronized evaluation and $(\log K/n)^{\gamma(p,r)}$ under independent evaluation, where $\gamma(p,r) = p(1-1/r)/(p+2-2p/r) < 1/2$. The slower minimax rate under independent sampling persists with bounded loss differences, even when the reference marginal and every model's population quantile at the ES level are known.

In controlled experiments, we compare how many evaluations are needed to select the best model with the same accuracy. As models become closer in tail risk, this number grows faster with independent samples than with shared inputs. Comparisons on fixed distributions isolate the effect of sampling. On held-out medical claims, synchronization raises the probability of correct selection from 0.144 to 0.831 at equal evaluation cost; language model benchmarks also show gains from synchronized sampling. These results guide evaluation design: coordinating inputs across models can reduce the budget needed to achieve a given expected ES gap from the best candidate.

1 Introduction

Model evaluations can use separately sampled questions from the same benchmark, or independently drawn simulation inputs. Once these loss records have been collected, reevaluating all models on a common sample may be impractical. When planning a new evaluation, however, we can choose between common inputs and a separate sample for each model. We ask whether this choice changes the optimal rate of tail-risk model selection at a fixed evaluation budget, and show that shared inputs can yield a faster rate than independent evaluation even when the two evaluation schemes use the same marginal loss distributions.

Common random numbers and paired comparisons exploit shared inputs (Glasserman & Yao, 1992; Chick & Inoue, 2001; Saad et al., 2023). We characterize the statistical cost of losing that pairing. For absolute prediction error, a large outcome can make every model's loss large. If all predictions lie in a fixed bounded interval, however, their absolute errors on the same input differ by at most the width of that interval. To emphasize large losses in model selection, we measure tail risk by Expected

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.

Shortfall (ES), the average loss over the worst $1-\alpha$ fraction of inputs for a fixed $\alpha \in (0,1)$. Among K candidates with losses $L_1, \ldots, L_K$, we seek the one with the smallest ES.

To describe how candidate losses differ on the same input, we use a reference model with loss L_0. Subtracting its ES leaves the candidate ordering unchanged. We assume that L_0 has a bounded pth moment, with $1 < p < 2$, while the contrasts $L_j - L_0$ have uniformly bounded L^r norms, with $2 \le r \le \infty$. These conditions allow individual losses to have infinite variance, while differences on the same input have uniformly bounded second moments.

Synchronized evaluation observes all models on the same n inputs, whereas independent evaluation draws n inputs separately for each model. We compare the two designs with the same marginal loss distributions and $(K+1)n$ loss evaluations, including those of the reference. Their performance is measured by simple regret, the ES gap between the selected candidate and the best candidate. Let $\mathfrak{R}_{\mathrm{sync}}(n,K)$ and $\mathfrak{R}_{\mathrm{ind}}(n,K)$ denote the corresponding minimax expected regrets over the same class. For any fixed positive moment bounds, $K \ge 8$, and sufficiently large $n/\log K$, matching upper and lower bounds give

$$\begin{aligned} \mathfrak{R}_{\mathrm{sync}}(n,K) &\asymp \sqrt{\frac{\log K}{n}}, \\ \mathfrak{R}_{\mathrm{ind}}(n,K) &\asymp \left(\frac{\log K}{n}\right)^{\gamma(p,r)}. \end{aligned} \tag{1}$$

Here $\gamma(p,r) = p(1-1/r)/(p+2-2p/r) < 1/2$, and $\asymp$ denotes comparison up to positive constants depending only on the fixed class parameters. Thus, even without pairs, $\gamma(p,r)$ exceeds $(p-1)/p$, the exponent obtained from marginal moments alone; pairing raises it to $1/2$.

The Rockafellar–Uryasev (RU) representation expresses relative ES as the expected difference of scores evaluated at population quantiles (Rockafellar & Uryasev, 2000). The contrast condition controls second moments of paired score differences under synchronized evaluation. Under independent evaluation, it controls the difference between the candidate and reference truncation biases, rather than requiring separate bounds on the two biases. Balancing this relative bias against marginal sampling fluctuations yields the rate exponent $\gamma(p,r)$. We also control the population RU objective error from estimating the quantiles, allowing atoms and nonunique minimizers without a density condition at the ES threshold. Together, these bounds support sequential evaluation procedures that adaptively choose which candidates to evaluate and when to stop, while guaranteeing a prescribed probability of selecting the best candidate.

Our experiments show that shared inputs increasingly reduce the evaluations needed for reliable selection as models become closer in tail risk. Comparisons on fixed distributions show gains from sharing inputs while holding the empirical ES selector fixed. On held-out medical claims, at equal evaluation cost, synchronization raises the probability of correct selection from 0.144 to 0.831 and reduces mean simple regret from 4.227 to 0.126 in nominal dollars. Comparisons without replacement on medical and language model benchmarks show that sharing inputs can also improve selection with the same number of candidate evaluations under both designs. When candidate losses have been collected separately, subtracting the empirical reference ES computed on each candidate's inputs can improve selection without new candidate evaluations. This adjustment lowers mean regret on the medical evaluation and one language model benchmark, but raises it on another.

2 Two Evaluation Experiments

Fix a tail level $\alpha \in (0,1)$ and write $\tau = 1-\alpha$. For $x \in \mathbb{R}$, write $x_+ := \max\{x, 0\}$. For an integrable loss Z, Expected Shortfall has the Rockafellar–Uryasev (RU) representation (Rockafellar & Uryasev, 2000; 2002)

$$\mathrm{ES}_\alpha(Z) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{\tau}\mathbb{E}[(Z-t)_+] \right\}. \tag{2}$$

We evaluate K candidate models, indexed by $[K] := \{1, \ldots, K\}$, and one reference model, all fixed before evaluation. Let L_j denote candidate j's loss and L_0 the reference loss, with joint law P for $(L_0, \ldots, L_K)$. For $j = 0, \ldots, K$, let $\rho_j(P)$ denote model j's ES. For $j \in [K]$, let $\theta_j(P)$ denote its ES minus the reference ES. We write $\rho_\star(P)$ for the smallest candidate ES and $\mathcal{J}^\star(P)$ for the set of

candidate indices attaining it:

$$\begin{aligned} \rho_j(P) &= \mathrm{ES}_\alpha(L_j), & \theta_j(P) &= \rho_j(P) - \rho_0(P), \\ \rho_\star(P) &= \min_{j\in[K]} \rho_j(P), & \mathcal{J}^\star(P) &= \arg\min_{j\in[K]} \rho_j(P). \end{aligned} \tag{3}$$

When the best candidate is unique, denote its index by $j^\star(P)$. When P is fixed, we suppress it from the notation. Since ES is nonlinear, in general $\theta_j \neq \mathrm{ES}_\alpha(L_j - L_0)$. Subtracting the reference ES leaves the candidate ordering unchanged. The reference need not minimize ES; its role is to define the loss contrasts below.

Moment assumptions. We impose the same moment bounds on the loss distributions in both experiments. Fix $1 < p < 2$ and $2 \le r \le \infty$, with $1/\infty = 0$. For $M_0, B > 0$, let $\mathcal{P}_{K,p,r}(M_0, B)$ be the class of joint laws of $(L_0, \dots, L_K)$ satisfying

$$\mathbb{E}|L_0|^p \le M_0, \tag{4}$$

$$\max_{j\in[K]} \|L_j - L_0\|_{L^r} \le B, \tag{5}$$

where (5) uses the essential-supremum norm when $r = \infty$ and $\|X\|_{L^r} := (\mathbb{E}|X|^r)^{1/r}$ otherwise. Each contrast bound connects a candidate to the same reference. The contrast bound concerns losses on the same input; similar ES values alone do not imply it. For absolute-error loss, predictions in a common interval of width B yield (5) with $r = \infty$ (Section 5.2).

Because $r > p$, all model losses satisfy

$$\max_{0\le j\le K} \|L_j\|_{L^p} \le M_0^{1/p} + B, \qquad \overline{M} := (M_0^{1/p} + B)^p. \tag{6}$$

Thus all ES values are finite and $\mathbb{E}|L_j - L_0|^2 \le B^2$, although marginal variances need not be uniformly bounded. Class parameters are fixed across K and sample size. The cross-fitted estimators and adaptive procedures use known moment orders and bounds; conservative values may define a larger class.

The class permits discrete losses, atoms at the ES threshold, and nonunique minimizers in the RU representation. For $j = 0, \dots, K$, let $F_{L_j}(t) := \mathbb{P}(L_j \le t)$ and $q_j := \inf\{t : F_{L_j}(t) \ge \alpha\}$ be the lower generalized α-quantile. This quantile minimizes the RU objective in (2) for $Z = L_j$.

Synchronized evaluation. Synchronized evaluation draws n independent loss vectors with law P, each containing the reference and all candidate losses on one common input:

$$\mathcal{E}_{\mathrm{sync},n}(P) = P^{\otimes n}. \tag{7}$$

Independent evaluation. Let P_j denote the marginal law of L_j for $j = 0, \dots, K$. Independent evaluation draws n observations per model, independently across models:

$$\mathcal{E}_{\mathrm{ind},n}(P) = P_0^{\otimes n} \otimes \bigotimes_{j=1}^{K} P_j^{\otimes n}. \tag{8}$$

Selectors observe only the sampled losses, without the underlying inputs or labels. Proposition 2 gives an equivalent description: the reference obeys the p-moment bound, and each candidate marginal admits a coupling with it whose difference has L^r norm at most B. Thus the contrast constraint restricts the marginals even without observed pairs. Both experiments use $N = (K+1)n$ loss evaluations with the same marginals and target (Figure 1).

Fixed-budget selection risk. For a distribution class $\mathcal{P}$ and $\mathsf{e} \in \{\mathrm{sync}, \mathrm{ind}\}$, at fixed n, let $\widehat{j}$ range over measurable selectors taking values in $[K]$ and using the observations from experiment e, with an optional independent random seed. The minimax expected simple regret is

$$\mathfrak{R}_{\mathsf{e}}(n, K; \mathcal{P}) := \inf_{\widehat{j}} \sup_{P\in\mathcal{P}} \mathbb{E}_{\mathcal{E}_{\mathsf{e},n}(P)}\big[\rho_{\widehat{j}}(P) - \rho_\star(P)\big]. \tag{9}$$

For sampled data, $\mathbb{P}_P$ and $\mathbb{E}_P$ include the specified sampling protocol and algorithmic randomization on instance P.

In the fixed-confidence analysis, procedures adapt sampling and stop once they can guarantee a prescribed probability of correct selection. Each model evaluation costs one unit (Appendix B).

We write $\mathbf{1}_A$ for the indicator of an event A. Throughout, $a_n \lesssim b_n$ denotes an inequality up to a constant depending only on the fixed class parameters. In the fixed-confidence bounds, $\widetilde{O}$ and $\widetilde{\Theta}$ additionally suppress logarithmic and iterated-logarithmic factors.

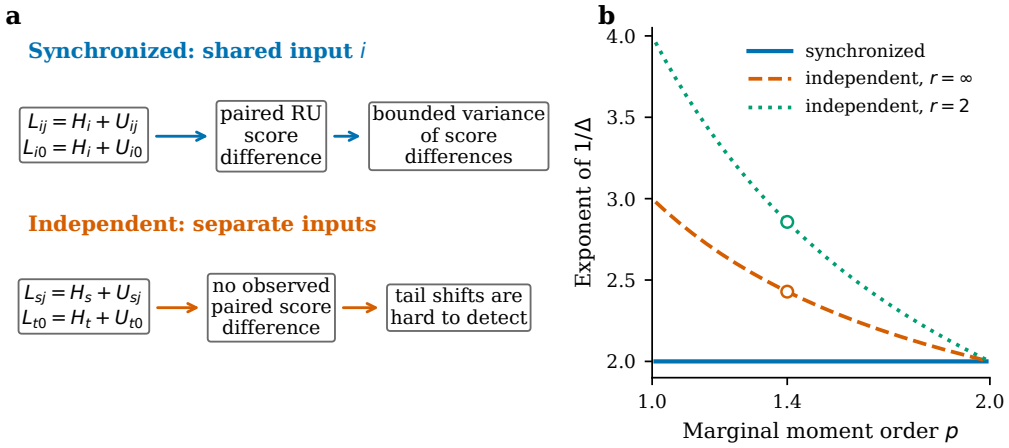

Figure 1: Why shared inputs change sample complexity. **(a)** For input i, $L_{i,j} = H_i + U_{i,j}$ combines shared variation H_i with model-specific variation $U_{i,j}$. Under the contrast bound, paired RU score differences have bounded variance. Independent marginal samples make the lower-bound tail shifts hard to detect. **(b)** For a common positive ES gap $\Delta > 0$, fixed K, and a fixed target probability of correct selection, evaluation counts scale as Δ^{-2} (synchronized) and $\Delta^{-1/\gamma(p,r)}$ (independent), up to logarithmic factors. The vertical axis shows these exponents; circles mark $p = 1.4$. All exponents converge to 2 as $p \uparrow 2$.

3 Main Minimax Results

The comparison is governed by

$$\gamma(p, r) := \frac{p(1 - 1/r)}{p + 2 - 2p/r} < \frac{1}{2}. \tag{10}$$

The strict inequality holds throughout $1 < p < 2$ and $2 \le r \le \infty$. It is the exponent obtained by balancing the tail bias and stochastic error under independent sampling.

For fixed M_0 and B, we abbreviate $\mathcal{P}_{K,p,r}(M_0, B)$ as $\mathcal{P}_{K,p,r}$.

Theorem 1 (Minimax regret rates)**.** *Fix $\alpha \in (0, 1)$, $1 < p < 2$, $2 \le r \le \infty$, and $M_0, B > 0$. There are constants $c_{\mathrm{up}}, c_{\mathrm{low}}, c, C > 0$, depending only on (α, p, r, M_0, B), such that the following bounds hold. For $K \ge 1$ and $n \ge c_{\mathrm{up}} \log(2K)$,*

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{P}_{K,p,r}) \le C\sqrt{\frac{\log(2K)}{n}}, \tag{11}$$

$$\mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{P}_{K,p,r}) \le C\left(\frac{\log(2K)}{n}\right)^{\gamma(p,r)}. \tag{12}$$

For $K \ge 8$ and $n \ge c_{\mathrm{low}} \log K$,

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{P}_{K,p,r}) \ge c\sqrt{\frac{\log K}{n}}, \tag{13}$$

$$\mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{P}_{K,p,r}) \ge c\left(\frac{\log K}{n}\right)^{\gamma(p,r)}. \tag{14}$$

The upper and lower bounds match for $K \ge 8$ whenever both sample-size conditions hold. The independent lower bound also holds when selectors are additionally given P_0 and the lower generalized quantiles $(q_j)_{j=0}^{K}$.

Interpretation and scope. The theorem compares worst-case selection under the same moment constraints. The independent lower bound remains valid when P_0 and $(q_j)_{j=0}^K$ are known (Appendix G.2), so the exponent gap is not due to estimating the reference distribution or the ES thresholds. Under the conditions for matching bounds in Theorem 1, writing $N = (K+1)n$ gives

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{P}_{K,p,r}) \asymp \sqrt{\frac{K \log K}{N}}, \qquad \mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{P}_{K,p,r}) \asymp \left(\frac{K \log K}{N}\right)^{\gamma(p,r)}. \tag{15}$$

Since $\gamma(p, r) < 1/2$, the budget required to achieve a smaller worst-case expected regret grows faster under independent evaluation.

Pairing each candidate with the reference on its own independently sampled inputs suffices for the square-root rate (Corollary 26). With n observations per candidate, this uses $2Kn$ loss evaluations. Full synchronization uses $(K+1)n$ by evaluating the reference only once per shared input. With bounded contrasts, ranking candidates by empirical ES on the common sample attains the synchronized expected-regret rate without sample splitting or moment-dependent tuning (Corollary 24).

4 Why the Observation Protocol Changes the Rate

The decomposition $L_j = H + U_j$ and $L_0 = H + U_0$ provides intuition: individual losses may inherit the heavy tails of H, while the contrast assumption bounds the second moments of $U_j - U_0$. For means, pairing cancels H. For relative ES, we use the difference between Rockafellar–Uryasev scores at the two models' quantiles.

Paired scores yield the square-root rate. The RU score is

$$\ell(z, t) = t + \tau^{-1}(z - t)_+. \tag{16}$$

At population quantiles, $\mathbb{E}\ell(L_j, q_j) = \mathrm{ES}_\alpha(L_j)$. The paired score difference is

$$D_j = \ell(L_j, q_j) - \ell(L_0, q_0). \tag{17}$$

The positive-part map is one-Lipschitz, so

$$|D_j| \le \tau^{-1}|L_j - L_0| + (1 + \tau^{-1})|q_j - q_0|. \tag{18}$$

The contrast bound and the uniform quantile bound implied by (6) give uniformly bounded second moments for paired score differences, even when individual scores have no such bound.

Both estimators use two-fold cross-fitting: quantiles are fitted on one fold and scores evaluated on the other; the symmetric versions average both directions (Appendix K). Let $m = \lfloor n/2 \rfloor$ and $Q := 1 + (\overline{M}/\min\{\alpha, \tau\})^{1/p}$. For $j = 0, \dots, K$, let $\widehat{q}_j$ be an empirical RU minimizer over $[-Q, Q]$ on the fitting fold, and write $L_{i,j}$ for evaluation-fold losses. For one fold direction, synchronized data use shared rows and

$$\widehat{\theta}_j^{\mathrm{sync}} := \frac{1}{m} \sum_{i=1}^{m} \mathrm{clip}_{H_K}(\ell(L_{i,j}, \widehat{q}_j) - \ell(L_{i,0}, \widehat{q}_0)).$$

Here clip_{H_K} truncates to $[-H_K, H_K]$, with $H_K \asymp \sqrt{n/\log(2K)}$ up to fixed class constants. The second-moment bound controls clipping bias; Bernstein's inequality controls the maximum fluctuation of the clipped averages across candidates. Together they give the $\sqrt{\log(2K)/n}$ rate.

Independent sampling creates a bias–variance tradeoff. Without observed pairs, we estimate the candidate and reference score means separately:

$$\widehat{\theta}_j^{\mathrm{ind},T} := \frac{1}{m} \sum_{i=1}^{m} \ell(L_{i,j} \wedge T, \widehat{q}_j) - \frac{1}{m} \sum_{i=1}^{m} \ell(L_{i,0} \wedge T, \widehat{q}_0),$$

where $T > 0$ is the winsorization threshold and $z \wedge T := \min\{z, T\}$.

For $T \ge 2Q + 1$, conditional on the fitting data, the expected maximum absolute deviation of the evaluation-fold score averages from their conditional means is

$$O\left(T^{1-p/2}\sqrt{\frac{\log(2K)}{n}} + T\frac{\log(2K)}{n}\right). \tag{19}$$

For T above the population quantiles, truncation reduces $\mathrm{ES}_\alpha(L_j)$ by $\tau^{-1}\mathbb{E}(L_j - T)_+$. The coupling bound in Section 2, Hölder's inequality, and Markov's inequality give

$$|\mathbb{E}(L_j - T)_+ - \mathbb{E}(L_0 - T)_+| \le \|L_j - L_0\|_{L^r}\mathbb{P}(\max(L_j, L_0) > T)^{1-1/r} \lesssim T^{-a}, \tag{20}$$

where $a = p(1 - 1/r)$. For fixed p, a larger contrast moment order r improves this bias bound through the exponent a. Balancing the relative truncation bias T^{-a} with the sampling term $T^{1-p/2}\sqrt{\log(2K)/n}$ gives

$$T \asymp \left(\frac{n}{\log(2K)}\right)^{1/d}, \qquad d = p + 2 - \frac{2p}{r}. \tag{21}$$

Both terms then have order $(\log(2K)/n)^{\gamma(p,r)}$; the remaining Bernstein term $T\log(2K)/n$ is smaller. Marginal moments alone would give the weaker bound $(\log(2K)/n)^{(p-1)/p}$. The contrast bound improves the exponent to $\gamma(p, r) > (p - 1)/p$, even without observed pairs. To complete both upper bounds, write $G_j(t) := \mathbb{E}\ell(L_j, t)$ and $e_j := G_j(\widehat{q}_j) - \rho_j \ge 0$. Lemma 3 gives $\mathbb{E}\max_{0\le j\le K} e_j \lesssim \sqrt{\log(2K)/n}$; subtracting the reference leaves $e_j - e_0$, preserving both rates without a density condition.

For either procedure, let $\widehat{\theta}_j$ denote the final estimate after averaging the two fold directions. Select $\widehat{j} \in \arg\min_{j\in[K]} \widehat{\theta}_j$, with a fixed rule for ties. Then

$$\rho_{\widehat{j}} - \rho_\star \le 2\max_{j\in[K]} |\widehat{\theta}_j - \theta_j|.$$

Thus the uniform estimation bounds give (11) and (12).

Lower bounds for the two sampling designs. The two lower bounds use different hard families within the same class. Let $\Delta > 0$ denote the common positive ES gap. Under synchronized sampling, for a fixed residual-noise scale σ, a local paired perturbation of size u changes relative ES by order σu and has per-observation Kullback–Leibler (KL) divergence of order u^2, so Fano's inequality yields (13). Under independent sampling, moving a rare tail component of mass π and scale s_{tail} by b gives an ES difference of order πb and per-observation KL divergence of order at most $\pi b^2/s_{\mathrm{tail}}^2$. Choosing $\pi s_{\mathrm{tail}}^p \asymp 1$ and $\pi b^r \asymp 1$ for $r < \infty$ (with $b \asymp 1$ for $r = \infty$) yields

$$\Delta \asymp s_{\mathrm{tail}}^{-a}, \qquad \text{per-observation KL} \lesssim s_{\mathrm{tail}}^{-d} \asymp \Delta^{1/\gamma(p,r)}. \tag{22}$$

Fano's inequality gives a regret lower bound of order Δ when $n\Delta^{1/\gamma(p,r)}$ is a sufficiently small multiple of $\log K$. This yields (14). For the known-quantile strengthening, a separate family has common P_0, all $q_j = 0$, and the same ES-gap and KL orders (Appendix G.2).

Adaptive fixed-confidence selection. The same rate separation reappears under fixed confidence, where the estimators above provide confidence bounds for eliminating candidates. On a unique-best instance with $K \ge 2$, let $\Delta_j := \rho_j - \rho_{j^\star}$, and write N_{sync} and N_{ind} for the random total evaluation costs at stopping. For $\delta \in (0, 1/4)$, the restarted elimination procedures in Appendix I select the best candidate with probability at least $1 - \delta$ and achieve, up to logarithmic and iterated-logarithmic factors in K, $1/\delta$, and the inverse positive gaps,

$$\mathbb{E}_P N_{\mathrm{sync}} = \widetilde{O}\left(\sum_{j\ne j^\star} \Delta_j^{-2}\right), \qquad \mathbb{E}_P N_{\mathrm{ind}} = \widetilde{O}\left(\sum_{j\ne j^\star} \Delta_j^{-1/\gamma(p,r)}\right). \tag{23}$$

For every sufficiently small gap vector, we construct a synchronized instance with a matching lower bound. The independent lower bound matches equal or comparable gaps. Extending the latter to heterogeneous gaps and optimizing the logarithmic factors remain open.

5 Empirical Evidence

We test the theory in three ways: whether evaluation budgets follow the predicted scaling with the ES gap, whether synchronization improves finite-budget selection, and whether evaluating the reference helps after independent evaluation. Scaling experiments use the lower-bound families in Appendix F, while the other comparisons in the main text use the same empirical ES selector across protocols. For synchronized i.i.d. observations with bounded contrasts, Corollary 24 gives its square-root rate. Appendix M compares estimators and shows winsorization gains with independent samples.

5.1 Controlled Scaling and Fixed-Distribution Comparisons

To test how the required evaluation budget varies with the ES gap, we fix $K = 8$, $p = 1.4$, $\alpha = 0.95$, and $r = \infty$. The protocols use their respective hard families under the same fixed moment bounds. Let PCS denote the probability of correct selection. On these equal-gap families, $\rho_{\widehat{j}} - \rho_\star = \Delta \mathbf{1}_{\{\widehat{j} \neq j^\star\}}$, so $\mathbb{E}[\rho_{\widehat{j}} - \rho_\star] = \Delta(1 - \mathrm{PCS})$.

We estimate the total budget $N_{0.90}(\Delta)$ needed for 90% correct selection. Over six prespecified gaps, we fit $\log N_{0.90}(\Delta)$ against $\log(1/\Delta)$ by least squares, with slopes b_{sync} and b_{ind} for the two protocols. Separate preliminary simulations select five budgets per protocol and gap, with 8,000 replications then run at each budget. Simultaneous PCS confidence bands yield 95% intervals for the slopes and their difference (Appendix L.1). Appendix L.2 reports the $r = 2$ experiment.

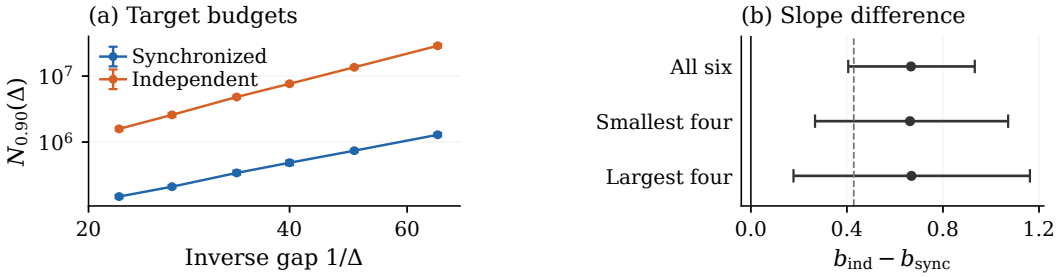

Figure 2: Controlled scaling for $r = \infty$. **(a)** Budgets for PCS 0.90 on each protocol's hard family. **(b)** Slope differences $b_{\mathrm{ind}} - b_{\mathrm{sync}}$ using all six gaps, the smallest four, or the largest four. Error bars derive from simultaneous 95% PCS bands; the dashed line marks the predicted difference $17/7 - 2 = 3/7$.

The synchronized and independent slopes are 1.969 and 2.636, with difference 0.667 and 95% interval $[0.405, 0.933]$. The difference intervals also lie above zero for the smallest four and largest four gaps (Figure 2(b)). The independent slope interval $[2.509, 2.769]$ lies above the predicted asymptotic exponent $17/7$. Because these experiments use different hard families, we next hold the distribution fixed to isolate the effect of synchronization at finite budgets.

Sampling comparisons on fixed distributions. Both comparisons use $K = 8$, $\Delta = 0.03$, $n = 1{,}024$, and the same empirical ES selector across protocols. For a mixture from the independent lower bound, synchronization raises PCS from 0.134 to 0.786; all failed synchronized samples lack a rare tail observation. To rule out perfect identification from a single tail observation, a second mixture uses bounded random loss differences with overlapping distributions. Synchronization lowers mean regret from 0.02595 to 0.02310 (paired reduction 0.00285, 95% CI $[0.00180, 0.00387]$). Appendix L.3 gives both constructions and complete results.

5.2 Medical Insurance Claims Case Study

To test whether the advantage of shared inputs extends to a real prediction task, we use the public Society of Actuaries Group Medical Insurance Database (Society of Actuaries, 2004), training on 1997 data, choosing the budget on 1998 data, and evaluating on held-out 1999 data. We compare seven fixed configurations of histogram gradient boosting using the 0.95 ES of absolute prediction error, with the 1997 empirical median as reference.

Predictions are clipped to the range from zero to the largest 1997 outcome, while outcomes remain unclipped. Writing Y for annual paid charges, X for the predictor vector, and $\widehat{m}_j(X)$ for forecast j, the reverse triangle inequality $\big||Y - \widehat{m}_j(X)| - |Y - \widehat{m}_0(X)|\big| \leq |\widehat{m}_j(X) - \widehat{m}_0(X)|$ shows that this setting satisfies the theorem's contrast condition with $r = \infty$. On the 1999 empirical population, paired RU score differences have variances only 1.27×10^{-5}–1.76×10^{-5} of the corresponding summed marginal variances. Synchronization improves PCS and mean regret with and without replacement (Table 1). Without replacement, the probability of selecting within one nominal dollar of the best ES rises from 0.296 to 0.969 (Appendices N and P).

Table 1: Medical claims: $K = 7$, $n = 8{,}192$, and $\alpha = 0.95$. Regret is in nominal dollars. Counts match within each design: $(K + 1)n$ loss observations with replacement, including the reference, and $Kn = 57{,}344$ distinct candidate–input evaluations without replacement.

Sampling	Protocol	PCS	Mean simple regret
With replacement	Synchronized	0.831	0.126
	Independent	0.144	4.227
Without replacement	Synchronized	0.811	0.159
	Independent	0.161	4.298

5.3 Language Model Evaluation

We use Open LLM Leaderboard records for 28 instruction-tuned models (7–9 billion parameters) on MMLU-Pro and ARC-Challenge (Wang et al., 2024; Clark et al., 2018; Hugging Face, 2024). We consider two reference configurations, using Mistral and OpenOrca in turn; in each, the reference is excluded from the corresponding pool of 27 candidate models. We select by the 0.95 ES of the correct answer's negative log probability after normalizing over answer options, emphasizing questions on which that probability is low. Here questions are the evaluation inputs: candidates share sampled questions under synchronization and sample questions independently otherwise (Figure 3).

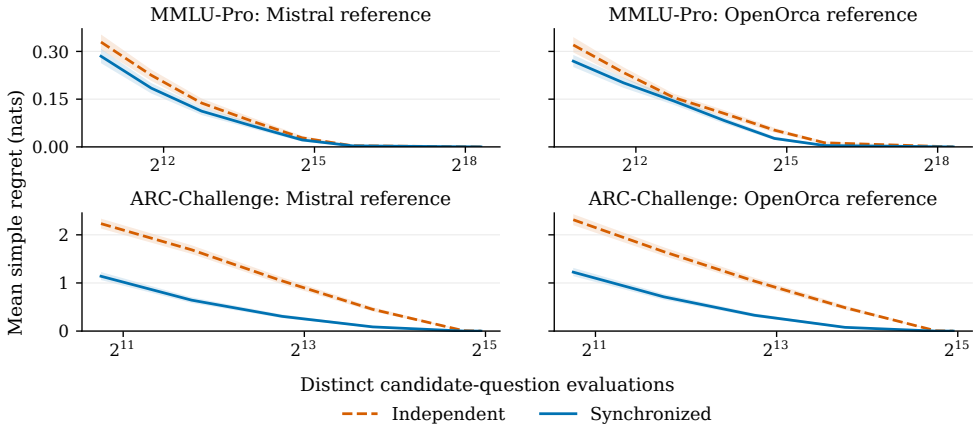

Figure 3: Language model selection without replacement. Both protocols rank the $K = 27$ non-reference candidates by empirical ES. The axis counts Kn distinct candidate–question evaluations. Shading shows pointwise 95% percentile-bootstrap intervals from 1,000 replications.

At $n = 512$ on ARC-Challenge, synchronization reduces mean regret from 0.451 to 0.087 (Mistral) and from 0.487 to 0.078 (OpenOrca). At $n = 2{,}048$ on MMLU-Pro, the OpenOrca values fall from 0.01291 to 0.00432; both Mistral values are below 0.004, with a difference interval containing zero. At population quantiles, variances of RU score differences between candidates and the reference are 0.14–0.62 times the corresponding sums of marginal score variances on ARC-Challenge, versus 0.49–0.91 on MMLU-Pro, consistent with the stronger gains on ARC-Challenge. Appendices O and P give these diagnostics, sampling details, and uncertainty intervals.

5.4 Reference Evaluation on Existing Inputs

When candidate losses have already been collected independently, we ask whether evaluating the reference on those same inputs can recover some of the benefit of pairing. For i.i.d. pairs, Corollary 26 gives the square-root rate using clipped, cross-fitted RU scores. Here we subtract a separate reference ES estimate, computed on each candidate's inputs, from that candidate's empirical ES.

On separate medical samples ($n_0 = 8{,}192$ per candidate), evaluating the reference on every initial input lowers mean regret from 4.430 to 1.991 without new candidate evaluations. Assuming

reference evaluations cost 5% of candidate evaluations, completion costs the equivalent of 2,823 candidate evaluations on average (4.92% of the initial candidate cost). The rule lowers mean regret in both ARC-Challenge settings but raises it in both MMLU-Pro settings. With all records retained, adjustment raises the mean squared error of the estimated ES gap between the best candidate and its closest competitor by 29% (Mistral) and 87% (OpenOrca). The correction leaves expected gap bias unchanged, but in both MMLU-Pro settings its variance exceeds twice its covariance with the unadjusted gap error, even with population quantiles fixed (Appendix P).

6 Related Work

Paired selection and model similarity. Best-arm identification studies fixed-budget and fixed-confidence selection (Audibert et al., 2010; Garivier & Kaufmann, 2016). The closest results with joint observations exploit dependence across models: Saad et al. (2023) use the variance of jointly observed coordinate differences, while Synchronized Successive Rejects (SySRs) uses paired evaluations for best-model identification with bounded scores (Lyu et al., 2026). Kim & Ramdas (2025) construct confidence sets for coordinates with the smallest means, with local minimax guarantees and a robust finite-variance variant. These works exploit joint observations for selection based on means or bounded scores. We instead establish matching minimax rates for synchronized and independent evaluation over the same heavy-tailed ES class, characterizing the cost of losing pairing.

Budgets and observation design. Zhou et al. (2025) combine adaptive evaluation and low-rank factorization to select language model configurations by average score. Anchor Points selects representative examples (Vivek et al., 2024); tinyBenchmarks uses small evaluation sets and item response theory (Maia Polo et al., 2024). Multiple-Prediction-Powered Inference (MultiPPI) chooses which variables to observe jointly and how to allocate samples to estimate a linear function of their means (Cowen-Breen et al., 2026), with minimax guarantees under finite variances and known covariance. These works reduce evaluation cost through adaptive allocation or additional structure. Our comparison instead fixes the marginals, target, and observations per model, and asks how synchronized versus independent sampling changes the minimax ES selection rate.

Heavy-tailed means and tail risk. Robust estimation and selection under heavy-tailed moment assumptions are established for means and conditional value at risk (CVaR) (Bubeck et al., 2013; Prashanth et al., 2020; Kagrecha et al., 2022). For independent arms, Agrawal et al. (2021) establish asymptotically optimal fixed-confidence identification as the error probability tends to zero. Holland & Haress (2021, Remark 5) explain how validation through the Rockafellar–Uryasev objective avoids local quantile regularity conditions, with bounds depending on the second moment of its score. Bartl & Eckstein (2026) study robust nonparametric ES estimation, and Dufour & He (2026) develop inference for differences of distribution functionals from dependent and overlapping samples. Our synchronized analysis controls the second moment of the paired score difference, even when marginal scores have no uniform second-moment bound.

7 Scope and Conclusion

Expected Shortfall (ES) depends only on marginal losses, but its minimax selection difficulty also depends on the observation protocol. Over the same moment-constrained class and at equal evaluation budgets, synchronized and independent sampling have different minimax exponents. This separation is informational: even with bounded contrasts, a known reference marginal, and known quantiles, independent observations cannot attain the synchronized rate uniformly over the class. The result suggests a design principle: preserve input-level pairing when paired score differences have stronger moment bounds than marginal scores. For ES, pairing each candidate with a reference suffices for the square-root rate, while full synchronization reuses reference evaluations across candidates.

Limitations and future work. Synchronization requires coordination and may reduce distinct inputs at a fixed budget. We assume fixed candidates and i.i.d. sampling from $\mathcal{P}_{K,p,r}(M_0, B)$; our cross-fitted and adaptive procedures use known moment orders and bounds, and constants deteriorate as $\alpha \uparrow 1$. Future work includes unknown-moment adaptation, finite-sample contrast verification, adaptive data reuse, and balancing shared inputs against coordination cost.

AI Use Statement

During the preparation of this manuscript, generative AI tools were used to assist with drafting and polishing portions of the text, including grammar correction, sentence rephrasing, and language editing, and to support retrieval and discovery of related literature. All AI-assisted text, retrieved sources, and citations were reviewed and verified by the author. Generative AI tools were not used for research ideation or execution, generating synthetic data, or proving mathematical claims. The author takes full responsibility for the final content of this work.

Ethics Statement

The medical insurance analysis uses a public historical dataset released by the Society of Actuaries. We use no claimant identifiers and report only aggregate model-evaluation results. The findings are specific to this historical dataset and should not be interpreted as current medical-cost estimates or clinical evidence.

Reproducibility Statement

The appendices provide complete proofs and experimental protocols. The anonymous supplement's `README.md` maps the reported results to code, source tables, and reconstruction instructions. It also identifies archived sensitivity analyses omitted from the paper.

References

Shubhada Agrawal, Wouter M. Koolen, and Sandeep Juneja. Optimal best-arm identification methods for tail-risk measures. In *Advances in Neural Information Processing Systems*, volume 34, pp. 25578–25590, 2021.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Proceedings of the 23rd Conference on Learning Theory*, pp. 41–53. Omnipress, 2010. URL https://www.learningtheory.org/colt2010/papers/59Audibert.pdf.

Daniel Bartl and Stephan Eckstein. Optimal nonparametric estimation of the expected shortfall risk. *SIAM Journal on Financial Mathematics*, 17(3):691–727, 2026. doi: 10.1137/24M1704518. URL https://epubs.siam.org/doi/10.1137/24M1704518.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013. doi: 10.1109/TIT.2013.2277869. URL https://arxiv.org/abs/1209.1727.

Stephen E. Chick and Koichiro Inoue. New procedures to select the best simulated system using common random numbers. *Management Science*, 47(8):1133–1149, August 2001. doi: 10.1287/mnsc.47.8.1133.10229. URL https://pubsonline.informs.org/doi/10.1287/mnsc.47.8.1133.10229.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try ARC, the AI2 reasoning challenge, 2018. URL https://arxiv.org/abs/1803.05457.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404. URL https://doi.org/10.1093/biomet/26.4.404.

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.

Jean-Marie Dufour and Tianyu He. Nonparametric methods for comparing distribution functionals for dependent samples with applications to welfare indices. Working Paper 2026s-14, CIRANO, 2026. URL https://doi.org/10.54932/HLSP1201.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 998–1027. PMLR, 2016. URL https://proceedings.mlr.press/v49/garivier16a.html.

Paul Glasserman and David D. Yao. Some guidelines and guarantees for common random numbers. *Management Science*, 38(6):884–908, 1992. doi: 10.1287/mnsc.38.6.884. URL https://pubsonline.informs.org/doi/10.1287/mnsc.38.6.884.

Matthew J. Holland and El Mehdi Haress. Learning with risk-averse feedback under potentially heavy tails. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 892–900. PMLR, 2021. URL https://proceedings.mlr.press/v130/holland21b.html.

Hugging Face. Open LLM leaderboard: Archived item-level evaluation results. Hugging Face Datasets, 2024. URL https://huggingface.co/open-llm-leaderboard.

Anmol Kagrecha, Jayakrishnan Nair, and Krishna P. Jagannathan. Statistically robust, risk-averse best arm identification in multi-armed bandits. *IEEE Transactions on Information Theory*, 68 (8):5248–5267, August 2022. doi: 10.1109/TIT.2022.3163524. URL https://ieeexplore.ieee.org/document/9745056/.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016. URL https://jmlr.org/papers/v17/kaufman16a.html.

Ilmun Kim and Aaditya Ramdas. Locally minimax optimal confidence sets for the best model, 2025. URL https://arxiv.org/abs/2503.21639.

Zifan Lyu, Chahine Nejma, Tobias Wegel, Fanny Yang, and Florian E. Dorner. Cutting LLM evaluation costs with SySRs: A bandit algorithm that provably exploits model similarity, 2026. URL https://arxiv.org/abs/2606.07726. Published at ICML 2026.

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

Pascal Massart. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. doi: 10.1214/aop/1176990746. URL https://doi.org/10.1214/aop/1176990746.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake VanderPlas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL https://jmlr.org/papers/v12/pedregosa11a.html.

L. A. Prashanth, Krishna Jagannathan, and Ravi Kumar Kolla. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5577–5586. PMLR, 2020. URL https://proceedings.mlr.press/v119/l-a-20a.html.

R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000. doi: 10.21314/JOR.2000.038. URL https://doi.org/10.21314/JOR.2000.038.

R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002. doi: 10.1016/S0378-4266(02)00271-6.

El Mehdi Saad, Gilles Blanchard, and Nicolas Verzelen. Covariance-adaptive best arm identification. In *Advances in Neural Information Processing Systems*, volume 36, pp. 73287–73298, 2023. doi: 10.52202/075280-3204.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Birkhäuser, 2015. doi: 10.1007/978-3-319-20828-2. URL https://doi.org/10.1007/978-3-319-20828-2.

Jonathan Scarlett and Volkan Cevher. An introductory guide to Fano's inequality with applications in statistical estimation. In Miguel R. D. Rodrigues and Yonina C. Eldar (eds.), *Information-Theoretic Methods in Data Science*, chapter 16, pp. 487–528. Cambridge University Press, 2021. doi: 10.1017/9781108616799.017. URL https://doi.org/10.1017/9781108616799.017.

Society of Actuaries. Medical large claims experience study. Society of Actuaries, 2004. URL https://www.soa.org/resources/experience-studies/2000-2004/research-medical-large-claims-experience-study/.

Rajan Vivek, Kawin Ethayarajh, Diyi Yang, and Douwe Kiela. Anchor Points: Benchmarking models with much fewer examples. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1576–1601. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.eacl-long.95. URL https://aclanthology.org/2024.eacl-long.95/.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, and Wenhu Chen. MMLU-Pro: A more robust and challenging multi-task language understanding benchmark. In *Advances in Neural Information Processing Systems: Datasets and Benchmarks Track*, volume 37, pp. 95266–95290, 2024. doi: 10.52202/079017-3018. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/ad236edc564f3e3156e1b2feafb99a24-Abstract-Datasets_and_Benchmarks_Track.html.

Jin Peng Zhou, Christian K. Belardi, Ruihan Wu, Travis Zhang, Carla P. Gomes, Wen Sun, and Kilian Q. Weinberger. On speeding up language model evaluation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2407.06172.

A Reader's Guide and Proof Map

Appendix B fixes the decision criteria, and Appendix C collects lemmas shared by the two fixed-budget upper bounds. The proof paths for the main results are:

Claim	Upper bound or procedure	Lower-bound proof
Synchronized fixed-budget rate	Appendix D	Appendix F: noisy-contrast family
Independent fixed-budget rate	Appendix E	Appendix F: remote-tail family
Each candidate paired with the reference	Appendix H	Appendix H
Adaptive fixed-confidence upper bounds	Appendix I	—
Fixed-confidence gap lower bounds	—	Appendix J

Appendix G gives the empirical ES guarantee for bounded contrasts, the independent lower bound with known quantiles, and the fixed-budget extension to $p = 1$. Appendix K gives the estimator implementations. Appendices L–P give experimental protocols and results. The supplement's `README.md` provides the corresponding file map.

B Formal Setup and Decision Criteria

This appendix fixes notation used in the proofs and makes the adaptive accounting convention precise. The target, observation laws, and fixed-budget risk are those in Section 2. Write $\tau := 1 - \alpha$ and $\kappa := \min\{\alpha, \tau\}$. For $1 < p < 2$ and $2 \le r \le \infty$, with $1/\infty := 0$, define

$$a := p\left(1 - \frac{1}{r}\right), \qquad d := p + 2 - \frac{2p}{r}, \qquad \gamma := \gamma(p, r) = \frac{a}{d}. \tag{24}$$

Then $0 < \gamma < 1/2$.

We use the class and notation of Section 2; the definitions below also allow $K \ge 1$. The fixed moment bounds imply $\mathbb{E}\,|L_j|^p \le \overline{M}$ by (6).

For probability laws P, Q on $\mathbb{R}$, let $\Gamma(P, Q)$ denote the set of couplings with marginals P and Q. Define the extended transport costs

$$\mathsf{W}_r^{\mathrm{ext}}(P, Q) := \inf_{\lambda \in \Gamma(P,Q)} \left(\int |x - y|^r \, \lambda(dx, dy) \right)^{1/r}, \quad 2 \le r < \infty,$$
$$\mathsf{W}_\infty^{\mathrm{ext}}(P, Q) := \inf_{\lambda \in \Gamma(P,Q)} \operatorname*{ess\,sup}_{(x,y)\sim\lambda} |x - y|.$$

The value may be infinite, allowing marginals without finite rth moments.

For a class $\mathcal{Q}$ of joint laws, write $\Pi_{\mathrm{ind}}(\mathcal{Q}) := \{\bigotimes_{j=0}^K P_j : P \in \mathcal{Q}\}$.

Proposition 2 (Marginal form of the product class)**.** *For every K, p, r, M_0, B,*

$$\Pi_{\mathrm{ind}}(\mathcal{P}_{K,p,r}(M_0, B)) = \left\{ \bigotimes_{j=0}^K P_j : \int |x|^p P_0(dx) \le M_0, \quad \max_{j\in[K]} \mathsf{W}_r^{\mathrm{ext}}(P_j, P_0) \le B \right\}.$$

Proof. Every admissible joint law supplies a coupling of (P_0, P_j) with cost at most B, proving one inclusion. Conversely, let $F_{L_j}^{-1}$ be the lower generalized quantile function associated with P_j, draw $U \sim \mathrm{Unif}(0, 1)$, and set $L_j = F_{L_j}^{-1}(U)$ for $0 \le j \le K$.

For $2 \le r < \infty$ and $M > 0$, define $\mathrm{clip}_M(x) := \max\{-M, \min\{x, M\}\}$. The standard one-dimensional monotone rearrangement inequality (Santambrogio, 2015, Theorem 2.9) gives, for every

$\lambda \in \Gamma(P_0, P_j)$,

$$\int_0^1 \left|\mathrm{clip}_M\left(F_{L_j}^{-1}(u)\right) - \mathrm{clip}_M\left(F_{L_0}^{-1}(u)\right)\right|^r du \\ \leq \int |\mathrm{clip}_M(x) - \mathrm{clip}_M(y)|^r \, \lambda(dx, dy) \leq \int |x-y|^r \, \lambda(dx, dy).$$

Fatou's lemma as $M \to \infty$ shows that the common quantile coupling has cost no larger than any λ. For $r = \infty$, fix $\lambda \in \Gamma(P_0, P_j)$ with finite essential-supremum cost and let $(X, Y) \sim \lambda$. For every $2 \leq r' < \infty$, the finite-order result gives

$$\left\|F_{L_j}^{-1}(U) - F_{L_0}^{-1}(U)\right\|_{L^{r'}} \leq \|X - Y\|_{L^{r'}} \leq \|X - Y\|_\infty .$$

The left-hand side increases to $\left\|F_{L_j}^{-1}(U) - F_{L_0}^{-1}(U)\right\|_\infty$ as $r' \to \infty$. Taking this limit and then the infimum over such λ proves optimality for the essential-supremum cost. Thus the common quantile coupling attains each extended transport cost and satisfies all candidate–reference contrast bounds, proving the reverse inclusion. □

Write $\mathcal{P} := \mathcal{P}_{K,p,r}(M_0, B)$ for the fixed class.

B.1 Counting Evaluations under Adaptive Sampling

Throughout the fixed-confidence analysis, assume $K \geq 2$. At adaptive stage t, an algorithm chooses a nonempty candidate subset $A_t \subseteq [K]$. In the synchronized experiment, it evaluates the reference and all candidates in A_t on one common fresh random input. In the independent experiment, it obtains fresh mutually independent observations from the requested candidate arms and, when requested, the reference. We charge one unit for every arm–input evaluation. A synchronized round that evaluates the reference therefore costs $|A_t| + 1$. Let N_j be the number of evaluations of candidate j and let N be the total number of candidate and reference evaluations. For an algorithm $\mathcal{A}$, write $N_{\mathcal{A}}$ for this total cost under $\mathcal{A}$.

Following standard fixed-confidence terminology (Kaufmann et al., 2016), a stopping algorithm is δ-PAC on the unique-best instances in $\mathcal{P}$ if it stops almost surely on each such instance and its output $\widehat{j}$ satisfies

$$\inf_{P \in \mathcal{P}:\, |\mathcal{J}^\star(P)|=1} \mathbb{P}_P(\widehat{j} = j^\star(P)) \geq 1 - \delta.$$

For an observation model $\mathsf{e} \in \{\mathrm{sync}, \mathrm{ind}\}$ and a subclass $\mathcal{Q} \subseteq \mathcal{P}$, let $\mathfrak{A}_{\mathsf{e}}(\delta; \mathcal{Q})$ be the algorithms that are δ-PAC on $\mathcal{Q}$. Their minimax expected stopping cost is

$$\mathfrak{C}_{\mathsf{e}}(\delta; \mathcal{Q}) := \inf_{\mathcal{A} \in \mathfrak{A}_{\mathsf{e}}(\delta; \mathcal{Q})} \sup_{P \in \mathcal{Q}} \mathbb{E}_P N_{\mathcal{A}}. \tag{25}$$

For fixed $0 < c_- \leq c_+ < \infty$ and $\Delta > 0$, define

$$\mathcal{Q}_K^{\mathrm{eq}}(\Delta; c_-, c_+) := \left\{ P \in \mathcal{P}_{K,p,r}(M_0, B) : \begin{array}{c} |\mathcal{J}^\star(P)| = 1, \\ c_-\Delta \leq \rho_j(P) - \rho_{j^\star(P)}(P) \leq c_+\Delta \\ \forall j \in [K] \setminus \{j^\star(P)\} \end{array} \right\}. \tag{26}$$

We choose fixed constants c_-, c_+ so that the displayed hard families lie in this band.

C Shared Technical Lemmas

C.1 Quantile Bounds and RU Objective Excess

For a loss Z with cdf F_Z, define $q_\alpha(Z) := \inf\{t : F_Z(t) \geq \alpha\}$. It satisfies $F_Z(q_\alpha(Z)-) \leq \alpha \leq F_Z(q_\alpha(Z))$. Since every arm has p-th moment at most $\overline{M}$, its lower generalized quantile lies in $[-Q_0, Q_0]$, where

$$Q_0 := \left(\frac{\overline{M}}{\kappa}\right)^{1/p}, \qquad Q := Q_0 + 1.$$

Indeed, if $q_\alpha(Z) \geq 0$, then $\mathbb{P}(Z \geq q_\alpha(Z)) \geq \tau$; if $q_\alpha(Z) < 0$, then $\mathbb{P}(Z \leq q_\alpha(Z)) \geq \alpha$.

Let $\ell(z,t) := t + (z-t)_+/\tau$ and $G_Z(t) := \mathbb{E}\ell(Z,t)$. Given m observations from arm j, let $\widehat{q}_j$ be any measurable minimizer of the empirical RU objective over $[-Q,Q]$; the smallest minimizer is a canonical tie-break.

Lemma 3 (Uniform RU objective excess)**.** *For $J \geq 1$ arms, whose samples may be dependent across arms but are i.i.d. within each arm, set $e_j := G_{L_j}(\widehat{q}_j) - \mathrm{ES}_\alpha(L_j) \geq 0$. Then, for every $\delta \in (0,1)$,*

$$\mathbb{P}\left(\max_{1 \leq j \leq J} e_j > \frac{2Q}{\tau} \sqrt{\frac{\log(2J/\delta)}{2m}} \right) \leq \delta. \tag{27}$$

Moreover, for a constant $C_{\mathrm{RU}} < \infty$ depending only on α,

$$\mathbb{E} \max_{1 \leq j \leq J} e_j \leq C_{\mathrm{RU}} Q \min\left\{ 1, \sqrt{\frac{\log(2J)}{m}} \right\}. \tag{28}$$

Proof. Fix one arm and write $Z \sim P$, with cdf F_Z. Let P_m, $\widehat{F}_m$, and $\widehat{q}$ denote its empirical law, empirical cdf, and constrained empirical RU minimizer, respectively. For an integrable function f, write $Pf := \int f\, dP$ and $P_m f := \int f\, dP_m$. Put $h_t(z) := \ell(z,t) - \ell(z,0)$. Subtracting the term independent of t leaves the minimizers unchanged and makes $h_t(z)$ uniformly bounded over $|t| \leq Q$ and $z \in \mathbb{R}$. Then

$$h_t(z) = t - \frac{1}{\tau} \int_0^t \mathbf{1}_{\{z > u\}}\, du, \qquad \sup_{|t| \leq Q} |(P_m - P)h_t| \leq \frac{Q}{\tau} \|\widehat{F}_m - F_Z\|_\infty.$$

The lower generalized α-quantile minimizes the population RU objective and lies in $[-Q,Q]$. Empirical optimality therefore gives

$$0 \leq G_Z(\widehat{q}) - \mathrm{ES}_\alpha(Z) \leq 2 \sup_{|t| \leq Q} |(P_m - P)h_t| \leq \frac{2Q}{\tau} \|\widehat{F}_m - F_Z\|_\infty.$$

The Dvoretzky–Kiefer–Wolfowitz inequality with Massart's sharp constant (Massart, 1990) and a union bound prove (27). Integrating the same bound, using $\|\widehat{F}_m - F_Z\|_\infty \leq 1$, proves (28). □

Lemma 4 (Dual representation of Expected Shortfall)**.** *For every integrable loss Z,*

$$\mathrm{ES}_\alpha(Z) = \sup\{\mathbb{E}[WZ] : 0 \leq W \leq \tau^{-1} \text{ almost surely}, \ \mathbb{E}W = 1\}.$$

If Z has a continuous law and q is its unique α-quantile, then $W^\star = \tau^{-1}\mathbf{1}_{\{Z > q\}}$ is an optimizer.

Proof. For every feasible W and $t \in \mathbb{R}$,

$$\mathbb{E}[WZ] = t + \mathbb{E}[W(Z-t)] \leq t + \tau^{-1}\mathbb{E}[(Z-t)_+].$$

Taking the infimum over t gives the upper bound. For the reverse inequality, let q be an α-quantile, so $\mathbb{P}(Z < q) \leq \alpha \leq \mathbb{P}(Z \leq q)$, and put

$$p_> := \mathbb{P}(Z > q), \qquad p_= := \mathbb{P}(Z = q), \qquad \lambda_q := \begin{cases} (\tau - p_>)/p_=, & p_= > 0, \\ 0, & p_= = 0. \end{cases}$$

The quantile inequalities give $p_> \leq \tau \leq p_> + p_=$; when $p_= = 0$, they give $p_> = \tau$. Hence $0 \leq \lambda_q \leq 1$ and $\lambda_q p_= = \tau - p_>$. Therefore

$$W^\star := \tau^{-1}\left(\mathbf{1}_{\{Z > q\}} + \lambda_q \mathbf{1}_{\{Z = q\}}\right)$$

is feasible, and

$$\mathbb{E}[W^\star Z] = \frac{1}{\tau}\mathbb{E}[Z\mathbf{1}_{\{Z>q\}}] + \frac{q}{\tau}(\tau - p_>) = q + \frac{1}{\tau}\mathbb{E}[(Z-q)_+] = \mathrm{ES}_\alpha(Z),$$

where the last equality uses that every α-quantile minimizes the RU objective. Thus $W^\star$ attains the supremum. For a continuous law, $p_= = 0$ and this reduces to $W^\star = \tau^{-1}\mathbf{1}_{\{Z>q\}}$. □

Lemma 5 (ES continuity and bounds on simple regret)**.** *For integrable X, Y on a common probability space,*

$$|\mathrm{ES}_\alpha(X) - \mathrm{ES}_\alpha(Y)| \le \tau^{-1}\mathbb{E}\,|X - Y|\,.$$

If $X - Y \in L^r$ for $2 \le r \le \infty$, then

$$|\mathrm{ES}_\alpha(X) - \mathrm{ES}_\alpha(Y)| \le \tau^{-1/r}\|X - Y\|_{L^r}.$$

Consequently, on $\mathcal{P}_{K,p,r}(M_0, B)$, every $j \in [K]$ satisfies

$$|\theta_j(P)| \le B\tau^{-1/r}, \qquad 0 \le \rho_{\widehat{j}} - \rho_\star \le 2B\tau^{-1/r}.$$

Proof. By Lemma 4, feasible weights satisfy $0 \le W \le \tau^{-1}$ and $\mathbb{E}W = 1$. The first bound follows from $|\mathbb{E}[W(X - Y)]| \le \tau^{-1}\mathbb{E}\,|X - Y|$. For $2 \le r < \infty$, let $r' = r/(r-1)$. Then

$$\mathbb{E}W^{r'} \le \tau^{-(r'-1)}, \qquad \|W\|_{L^{r'}} \le \tau^{-1/r}.$$

Hölder's inequality proves the second bound; for $r = \infty$, use $\mathbb{E}W = 1$. Taking suprema in both directions and applying the contrast bound gives the stated bounds on relative ES and regret. □

C.2 A Maximal Bernstein Inequality

Lemma 6 (Maximal Bernstein inequality)**.** *Let $\mathcal{F}$ be a sigma-field. For $1 \le j \le J$ and $1 \le i \le m$, suppose that, conditionally on $\mathcal{F}$, the variables $X_{i,j}$ are independent over i, centered, and satisfy*

$$\mathbb{E}[X_{i,j}^2 \mid \mathcal{F}] \le v, \qquad |X_{i,j}| \le h \quad \textit{almost surely}.$$

Dependence across j is allowed. Put $\overline{X}_j = m^{-1}\sum_i X_{i,j}$. Then there is a universal constant $C < \infty$ such that, for every $\delta \in (0,1)$, conditionally on $\mathcal{F}$,

$$\mathbb{P}\left(\max_{j \le J} |\overline{X}_j| > \sqrt{\frac{2vs}{m}} + \frac{2hs}{3m} \,\middle|\, \mathcal{F}\right) \le \delta, \qquad s := \log\frac{2J}{\delta}, \tag{29}$$

$$\mathbb{E}\left[\max_{j \le J} |\overline{X}_j| \mid \mathcal{F}\right] \le C\left\{\sqrt{\frac{v\log(2J)}{m}} + \frac{h\log(2J)}{m}\right\}. \tag{30}$$

Proof. The conditional scalar Bernstein inequality and a union bound give (29). For the expectation, the Bernstein bound for the moment generating function gives, for $0 < \lambda < 3/h$,

$$\mathbb{E}[e^{\lambda m \overline{X}_j} \mid \mathcal{F}] \le \exp\left\{\frac{\lambda^2 m v}{2(1 - \lambda h/3)}\right\},$$

and the same holds for $-\overline{X}_j$. Applying log-sum-exp to the $2J$ signed averages and choosing λ of order $\min\{\sqrt{\log(2J)/(mv)}, 1/h\}$ gives (30). □

D Synchronized Fixed-Budget Upper Bound

This section proves the synchronized upper bound in three steps: cross-fit constrained empirical RU minimizers, establish a uniform conditional second-moment bound for the observed candidate–reference score differences, and apply maximal Bernstein concentration to their clipped averages.

Split the n synchronized full-vector observations into a fitting fold and an evaluation fold, each of size $m := \lfloor n/2 \rfloor$. On the fitting fold, let $\widehat{q}_0, \ldots, \widehat{q}_K$ be measurably selected constrained empirical RU minimizers. Write $\widehat{q} = (\widehat{q}_0, \ldots, \widehat{q}_K)$ and let $\mathcal{F}_{\mathrm{q}}$ be the sigma-field generated by that fold. On the evaluation fold, define

$$D_{i,j} := \ell(L_{i,j}, \widehat{q}_j) - \ell(L_{i,0}, \widehat{q}_0), \qquad j \in [K]. \tag{31}$$

Unlike D_j in (17), $D_{i,j}$ uses fitted quantiles; its conditional mean need not equal θ_j.

Lemma 7 (Conditional second-moment bound)**.** *There is a finite constant V_{sync}, depending only on (α, p, M_0, B), such that, conditionally on the fitting fold,*

$$\max_{j \le K} \mathbb{E}[D_{i,j}^2 \mid \mathcal{F}_{\mathrm{q}}] \le V_{\mathrm{sync}}.$$

Proof. For arbitrary x, y, t, s, $|(x-t)_+ - (y-s)_+| \le |x-y| + |t-s|$. Since every constrained empirical RU minimizer lies in $[-Q, Q]$, we use $|\ell(x,t) - \ell(y,s)| \le |x-y|/\tau + 2Q(1+1/\tau)$ for $t, s \in [-Q, Q]$. The contrast bound (5) and $r \ge 2$ imply $\mathbb{E}\,|L_j - L_0|^2 \le B^2$. Squaring the preceding display proves the claim. □

For $H > 0$, write $\mathrm{clip}_H(x) := \max\{-H, \min\{x, H\}\}$. Both selectors break ties in favor of the smallest index. For expected risk, choose $s_{\mathrm{sync},K} := \log(2K)$ and $H_K := \sqrt{V_{\mathrm{sync}} m / s_{\mathrm{sync},K}}$, and define

$$\widehat{\theta}_j^{\mathrm{sync}} := \frac{1}{m} \sum_{i=1}^{m} \mathrm{clip}_{H_K}(D_{i,j}), \qquad \widehat{j}_{\mathrm{sync}} := \min \arg\min_{j \in [K]} \widehat{\theta}_j^{\mathrm{sync}}. \tag{32}$$

Theorem 8 (Synchronized uniform upper bound)**.** *There are constants $c, C < \infty$, depending only on the fixed class constants, such that if $n \ge c \log(2K)$, then*

$$\sup_{P \in \mathcal{P}_{K,p,r}} \mathbb{E}_P \max_{j \le K} \left| \widehat{\theta}_j^{\mathrm{sync}} - \theta_j(P) \right| \le C \sqrt{\frac{\log(2K)}{n}}. \tag{33}$$

Consequently,

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{P}_{K,p,r}) \le C \sqrt{\frac{\log(2K)}{n}}. \tag{34}$$

For a prescribed error probability $\delta \in (0, 1/2)$, replace $s_{\mathrm{sync},K}$ by $s = \log(2K/\delta)$ and H_K by $\sqrt{V_{\mathrm{sync}} m / s}$. If $n \ge cs$, then, for every $P \in \mathcal{P}_{K,p,r}$,

$$\mathbb{P}_P \left(\max_{j \le K} \left| \widehat{\theta}_j^{\mathrm{sync}} - \theta_j(P) \right| > C \sqrt{\frac{s}{n}} \right) \le \delta. \tag{35}$$

Proof. Condition on the fitting fold and put $\mu_j(\widehat{q}) := \mathbb{E}[D_{i,j} \mid \mathcal{F}_{\mathrm{q}}] = G_{L_j}(\widehat{q}_j) - G_{L_0}(\widehat{q}_0)$. By Lemma 7, clipping contributes bias at most $\mathbb{E}\big[|D_{i,j}| \mathbf{1}_{\{|D_{i,j}| > H\}} \,\big|\, \mathcal{F}_{\mathrm{q}}\big] \le V_{\mathrm{sync}}/H$. The centered clipped scores have conditional variance at most V_{sync} and absolute value at most $2H$. Combining the clipping-bias bound with Lemma 6, with $H = H_K$, gives

$$\mathbb{E}\left[\max_{j \le K} \left| \widehat{\theta}_j^{\mathrm{sync}} - \mu_j(\widehat{q}) \right| \,\middle|\, \mathcal{F}_{\mathrm{q}} \right] \le C \sqrt{\frac{V_{\mathrm{sync}} \log(2K)}{m}}.$$

For $0 \le j \le K$, put $e_j := G_{L_j}(\widehat{q}_j) - \rho_j(P)$. Since $e_j \ge 0$,

$$\mu_j(\widehat{q}) - \theta_j(P) = e_j - e_0, \qquad \max_{j \le K} |\mu_j(\widehat{q}) - \theta_j(P)| \le \max_{0 \le k \le K} e_k.$$

Lemma 3 makes this term $O(\{\log(2K)/m\}^{1/2})$ in expectation. Combining the bounds proves (33). For the selector, choose any $j^\circ \in \mathcal{J}^\star(P)$. Deterministically, $\rho_{\widehat{j}_{\mathrm{sync}}} - \rho_\star = \theta_{\widehat{j}_{\mathrm{sync}}} - \theta_{j^\circ} \le 2 \max_{j \le K} \left| \widehat{\theta}_j^{\mathrm{sync}} - \theta_j \right|$, which proves (34). For (35), use Lemma 3 with failure probability $\delta/2$, the clipping-bias bound V_{sync}/H, and (29) with failure probability $\delta/2$. Together these terms are bounded by $C(s/m)^{1/2}$. □

Remark 9 (Why the proof uses clipped scores). A median-of-means estimator gives the same prescribed-confidence bound (Bubeck et al., 2013). For expected simple regret without an additional $\log n$ factor, clipped scores are convenient because the clipping bias and the maximal Bernstein fluctuation are both controlled directly at $s_{\mathrm{sync},K} = \log(2K)$.

E Independent Fixed-Budget Upper Bound

This section proves the independent upper bound by decomposing the error into relative winsorization bias, stochastic fluctuation, and RU objective excess. The marginal truncation step follows standard heavy-tail arguments (Bubeck et al., 2013); the contrast moment bound controls the relative bias, whose balance with stochastic error yields the exponent $\gamma(p, r)$.

For $T > 0$, define upper winsorization by $w_T(z) := z \wedge T = \min\{z, T\}$. Split every arm's independent sample into a fitting fold and an evaluation fold of size $m = \lfloor n/2 \rfloor$. Let $\widehat{q}_j$ be a measurably selected constrained empirical RU minimizer on the first fold, and let $\mathcal{F}_{\mathrm{q}}$ be the sigma-field generated by all fitting folds. On the evaluation fold, set

$$U_{i,j}^{(T)} := \ell(w_T(L_{i,j}), \widehat{q}_j), \qquad \widehat{G}_j^{(T)} := \frac{1}{m} \sum_{i=1}^{m} U_{i,j}^{(T)},$$

and define

$$\widehat{\theta}_j^{\mathrm{ind},T} := \widehat{G}_j^{(T)} - \widehat{G}_0^{(T)}, \qquad \widehat{j}_{\mathrm{ind}} := \min \arg\min_{j \in [K]} \widehat{\theta}_j^{\mathrm{ind},T}. \tag{36}$$

Lemma 10 (Winsorized score variance and range)**.** *There are constants $C_V, C_R < \infty$, depending only on the fixed class constants, such that for every $T \geq 2Q + 1$, conditionally on the quantile fold,*

$$\max_{0 \leq j \leq K} \mathrm{Var}(U_{i,j}^{(T)} \mid \mathcal{F}_{\mathrm{q}}) \leq C_V T^{2-p}, \tag{37}$$

$$\max_{0 \leq j \leq K} \left| U_{i,j}^{(T)} - \mathbb{E}[U_{i,j}^{(T)} \mid \mathcal{F}_{\mathrm{q}}] \right| \leq C_R T. \tag{38}$$

Proof. Because $\widehat{q}_j \in [-Q, Q]$ and $w_T(L_j) \leq T$, $\left| U_{i,j}^{(T)} \right| \leq C_R T$ and $(w_T(L_j) - \widehat{q}_j)_+ \leq \min\{(L_j)_+, T\} + Q$. For $T \geq 1$,

$$\mathbb{E} \min\{(L_j)_+, T\}^2 \leq \mathbb{E}[L_j^2 \mathbf{1}_{\{0 < L_j \leq T\}}] + T^2 \mathbb{P}(L_j > T) \leq 2\overline{M} T^{2-p}.$$

Absorbing the fixed Q term into T^{2-p} proves the variance bound. □

Lemma 11 (Uniform relative winsorization bias)**.** *There is a constant $C_b < \infty$ such that, uniformly over $P \in \mathcal{P}_{K,p,r}$, for every $T \geq 2Q + 1$ and $j \in [K]$,*

$$|\theta_j(P) - [\mathrm{ES}_\alpha(w_T(L_j)) - \mathrm{ES}_\alpha(w_T(L_0))]| \leq C_b T^{-a}. \tag{39}$$

Proof. Since T exceeds both relevant quantiles, $\mathrm{ES}_\alpha(Z) - \mathrm{ES}_\alpha(w_T(Z)) = \tau^{-1}\mathbb{E}[(Z - T)_+]$. Under the coupling specified by P, let $A_T = \{\max(L_j, L_0) > T\}$. Then $|(L_j - T)_+ - (L_0 - T)_+| \leq |L_j - L_0| \, \mathbf{1}_{A_T}$. For finite r, Hölder's inequality and Markov's inequality give

$$\mathbb{E}[|L_j - L_0| \mathbf{1}_{A_T}] \leq B \mathbb{P}(A_T)^{1-1/r} \leq B(2\overline{M} T^{-p})^{1-1/r}.$$

For $r = \infty$, the same argument gives $\mathbb{E}[|L_j - L_0| \mathbf{1}_{A_T}] \leq B\mathbb{P}(A_T)$. Division by τ proves the claim. □

For expected risk, let $s_{\mathrm{ind},K} := \log(2(K+1))$ and $T_K := 2Q + 1 + (m/s_{\mathrm{ind},K})^{1/d}$.

Theorem 12 (Independent uniform upper bound)**.** *There are constants $c, C < \infty$, depending only on the fixed class constants, such that if $n \geq c \log(2K)$, then the estimator (36) with $T = T_K$ satisfies*

$$\sup_{P \in \mathcal{P}_{K,p,r}} \mathbb{E}_P \max_{j \leq K} \left| \widehat{\theta}_j^{\mathrm{ind},T_K} - \theta_j(P) \right| \leq C \left(\frac{\log(2K)}{n} \right)^{\gamma}. \tag{40}$$

Consequently,

$$\mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{P}_{K,p,r}) \leq C \left(\frac{\log(2K)}{n} \right)^{\gamma}. \tag{41}$$

For prescribed $\delta \in (0, 1/2)$, put $s = \log(2(K+1)/\delta)$ and $T_{K,\delta} := 2Q + 1 + (m/s)^{1/d}$. If $n \geq cs$, then, for every $P \in \mathcal{P}_{K,p,r}$,

$$\mathbb{P}_P \left(\max_{j \leq K} \left| \widehat{\theta}_j^{\mathrm{ind},T_{K,\delta}} - \theta_j(P) \right| > C \left(\frac{s}{n} \right)^{\gamma} \right) \leq \delta. \tag{42}$$

Proof. Condition on the fitting folds and let $\mu_j^{(T)}(\widehat{q}_j) := \mathbb{E}[U_{i,j}^{(T)} \mid \mathcal{F}_{\mathrm{q}}]$. Lemmas 10 and 6, applied to the $K+1$ arm averages, give

$$\mathbb{E}\left[\max_{0\le j\le K}\left|\widehat{G}_j^{(T)} - \mu_j^{(T)}(\widehat{q}_j)\right| \,\middle|\, \mathcal{F}_{\mathrm{q}}\right] \le C\left\{T^{1-p/2}\sqrt{\frac{\log(2K)}{m}} + T\frac{\log(2K)}{m}\right\}.$$

The same bound holds after taking expectations over the fitting data. For every $t \le T$,

$$G_L(t) - G_{w_T(L)}(t) = \frac{1}{\tau}\mathbb{E}[(L-T)_+]. \tag{43}$$

Because $T > Q$, the raw and winsorized objectives differ by this constant throughout the optimization interval, and their lower generalized quantiles coincide in $[-Q_0, Q_0]$. Hence

$$G_{w_T(L_j)}(\widehat{q}_j) - \mathrm{ES}_\alpha(w_T(L_j)) = G_{L_j}(\widehat{q}_j) - \mathrm{ES}_\alpha(L_j) =: e_j.$$

Lemma 3 bounds the maximal objective excess by $C\{\log(2K)/m\}^{1/2}$ in expectation. Since $e_j, e_0 \ge 0$, $|e_j - e_0| \le \max\{e_j, e_0\}$. Lemma 11 adds CT^{-a}. Thus

$$\mathbb{E}\max_{j\le K}\left|\widehat{\theta}_j^{\mathrm{ind},T} - \theta_j\right| \le C\left\{T^{-a} + T^{1-p/2}\sqrt{\frac{s_{\mathrm{ind},K}}{m}} + T\frac{s_{\mathrm{ind},K}}{m} + \sqrt{\frac{s_{\mathrm{ind},K}}{m}}\right\}. \tag{44}$$

Put $x = s_{\mathrm{ind},K}/m$ and $T \asymp x^{-1/d}$. The first two terms are of order x^γ:

$$T^{-a} \asymp x^{a/d} = x^\gamma, \qquad T^{1-p/2}x^{1/2} \asymp x^{1/2-(1-p/2)/d} = x^\gamma.$$

Moreover, $1 - 1/d - \gamma = (1-p/r)/d > 0$, so $Tx \asymp x^{1-1/d} = o(x^\gamma)$ as $x \downarrow 0$, and $x^{1/2} \le x^\gamma$ for $0 < x \le 1$ because $\gamma < 1/2$. This proves (40). The simple-regret result follows from the same deterministic selection inequality as in Theorem 8.

For (42), combine Lemma 3, the union Bernstein bound (29), and Lemma 11 with $T = T_{K,\delta}$. The same exponent calculation completes the proof. □

F Matching Fixed-Budget Lower Bounds on a Common Class

We construct two families, each with K hypotheses, in a single class with moment bounds independent of K. The independent family concentrates the decision-relevant change in a remote component, while the synchronized family uses bounded contrasts with a nonvanishing residual-noise scale.

F.1 Information Lemmas

Lemma 13 (Fano bound)**.** *Let $K \ge 2$, let V be uniform on $[K]$, and let the observation Z have law Q_v conditional on $V = v$. If there is a law $\bar{Q}$ such that*

$$\frac{1}{K}\sum_{v=1}^{K}\mathrm{KL}(Q_v\|\bar{Q}) \le \frac{1}{8}\log K,$$

then every estimator $\widehat{V}$ satisfies

$$\sup_{v\in[K]} Q_v(\widehat{V} \ne v) \ge 1 - \frac{1}{8} - \frac{\log 2}{\log K}.$$

The right-hand side is at least $1/2$ for $K \ge 8$.

Proof. The information-radius identity bounds the mutual information $I(V;Z)$ by $K^{-1}\sum_v \mathrm{KL}(Q_v\|\bar{Q})$. Fano's inequality (Scarlett & Cevher, 2021) gives the stated average-error bound, and maximum error is no smaller than average error. □

F.2 Independent Evaluation: A Remote-Tail Family

Let Φ denote the standard normal cdf. Let $s_{\mathrm{tail}} \geq 1$ be the scale introduced in Section 4. Put

$$A := 4s_{\mathrm{tail}}, \qquad \pi := s_{\mathrm{tail}}^{-p}, \qquad b := \begin{cases} s_{\mathrm{tail}}^{p/r}, & r < \infty, \\ 1, & r = \infty. \end{cases}$$

Let $G_{\mathrm{cen}} := \mathsf{N}(0,1)$ and $G_{\mu,s_{\mathrm{tail}}} := \mathsf{N}(\mu, s_{\mathrm{tail}}^2)$ for $\mu \in \mathbb{R}$, and define

$$F_0 := (1-\pi)G_{\mathrm{cen}} + \pi G_{A,s_{\mathrm{tail}}}, \qquad F_1 := (1-\pi)G_{\mathrm{cen}} + \pi G_{A-b,s_{\mathrm{tail}}}. \tag{45}$$

We call the components centered at A and $A-b$ remote: they move away from the central component as s_{tail} grows, while their probability π decreases. For a distribution F, write $\mathrm{ES}_\alpha(F) := \mathrm{ES}_\alpha(Z)$ for $Z \sim F$.

Lemma 14 (Uniform class membership)**.** *There exist constants $M_{0,\mathrm{rem}}, Q_{\mathrm{rem}}$ and $s_{\mathrm{tail},0} < \infty$, depending only on (α, p, r), such that for every $s_{\mathrm{tail}} \geq s_{\mathrm{tail},0}$:*

(i) both F_0 and F_1 have pth moment at most $M_{0,\mathrm{rem}}$;

(ii) their lower generalized α-quantiles lie in $[-Q_{\mathrm{rem}}, Q_{\mathrm{rem}}]$;

(iii) under the coupling that shares the mixture label and Gaussian draw, let $Z_i \sim F_i$ denote the coupled variables. Their displacement is b on the remote event and zero otherwise, so $\mathbb{E}|Z_1 - Z_0|^r = \pi b^r = 1$ for finite r, while $|Z_1 - Z_0| \leq 1$ for $r = \infty$.

All constants are independent of K.

Proof. Since $b \leq s_{\mathrm{tail}}$ for $s_{\mathrm{tail}} \geq 1$, there is a constant $C_p < \infty$, depending only on p, such that

$$\mathbb{E}_{G_{A-t,s_{\mathrm{tail}}}} |Z|^p = s_{\mathrm{tail}}^p \mathbb{E}\,|4 - t/s_{\mathrm{tail}} + W|^p \leq C_p s_{\mathrm{tail}}^p, \qquad 0 \leq t \leq b,$$

where $W \sim \mathsf{N}(0,1)$. Multiplication by $\pi = s_{\mathrm{tail}}^{-p}$ gives a uniform marginal p-moment bound.

The moment bound places both lower generalized α-quantiles in a fixed interval $[-Q_{\mathrm{rem}}, Q_{\mathrm{rem}}]$.

For the coupling, draw $I_{\mathrm{rem}} \sim \mathrm{Bernoulli}(\pi)$. If $I_{\mathrm{rem}} = 0$, use the same G_{cen} draw. If $I_{\mathrm{rem}} = 1$, draw one $W \sim \mathsf{N}(0,1)$ and set $Z_0 = A + s_{\mathrm{tail}} W$ and $Z_1 = A - b + s_{\mathrm{tail}} W$. The contrast identities follow immediately. □

Lemma 15 (Finite KL at the remote-tail scale)**.** *For every $s_{\mathrm{tail}} \geq 1$,*

$$\mathrm{KL}(F_1 \| F_0) \leq \frac{\pi b^2}{2s_{\mathrm{tail}}^2} = \frac{1}{2} s_{\mathrm{tail}}^{-d}.$$

Proof. Revealing the mixture label gives KL divergence $\pi b^2/(2s_{\mathrm{tail}}^2)$ by the equal-variance Gaussian formula. Marginalization cannot increase KL, and substituting π and b gives $s_{\mathrm{tail}}^{-d}/2$. □

Lemma 16 (Expected-Shortfall separation)**.** *There are constants $0 < c_{\mathrm{ES}} \leq C_{\mathrm{ES}} < \infty$ and $s_{\mathrm{tail},1} \geq s_{\mathrm{tail},0}$, depending only on (α, p, r), such that for all $s_{\mathrm{tail}} \geq s_{\mathrm{tail},1}$,*

$$c_{\mathrm{ES}} \pi b \leq \mathrm{ES}_\alpha(F_0) - \mathrm{ES}_\alpha(F_1) \leq C_{\mathrm{ES}} \pi b.$$

Thus the separation has order s_{tail}^{-a}.

Proof. Use the coupling from Lemma 14 and write $Z_0 = Z_1 + bI_{\mathrm{rem}}$. Let q_1 be the α-quantile of Z_1. By Lemma 4, $W_1 = \tau^{-1}\mathbf{1}_{\{Z_1 > q_1\}}$ is a dual optimizer. Hence

$$\mathrm{ES}_\alpha(Z_0) - \mathrm{ES}_\alpha(Z_1) \geq \mathbb{E}[W_1(Z_0 - Z_1)] = \frac{\pi b}{\tau} \mathbb{P}(A - b + s_{\mathrm{tail}} W > q_1).$$

For all sufficiently large s_{tail}, $q_1 \leq Q_{\mathrm{rem}} \leq s_{\mathrm{tail}}$ and $b \leq s_{\mathrm{tail}}$, so $(q_1 - A + b)/s_{\mathrm{tail}} \leq -2$ and the conditional exceedance probability is at least $\Phi(2)$. For the upper bound, let W_0 be the dual optimizer for Z_0. Then

$$\mathrm{ES}_\alpha(Z_0) - \mathrm{ES}_\alpha(Z_1) \leq \mathbb{E}[W_0(Z_0 - Z_1)] \leq \frac{\pi b}{\tau}.$$

Finally, $\pi b = s_{\mathrm{tail}}^{-p(1-1/r)} = s_{\mathrm{tail}}^{-a}$, including the convention $r = \infty$. □

For each $v \in [K]$, define a joint law P_v^{rem} as follows. Draw $I_{\mathrm{rem}} \sim \mathrm{Bernoulli}(\pi)$. If $I_{\mathrm{rem}} = 0$, draw $C \sim G_{\mathrm{cen}}$ and set all $K+1$ losses equal to C. If $I_{\mathrm{rem}} = 1$, draw $W \sim \mathsf{N}(0,1)$ and set

$$L_0 = A + s_{\mathrm{tail}} W, \qquad L_v = A - b + s_{\mathrm{tail}} W, \qquad L_j = A + s_{\mathrm{tail}} W \quad (j \neq v).$$

Thus arm v has marginal F_1 and every other candidate has marginal F_0.

Theorem 17 (Independent K-arm Fano lower bound)**.** *There are constants $c, c_0 > 0$, depending only on (α, p, r), such that for every $K \geq 8$ and $n \geq c_0 \log K$,*

$$\mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{P}_{K,p,r}(M_{0,\mathrm{rem}}, 1)) \geq c \left(\frac{\log K}{n} \right)^{\gamma}.$$

Proof. Lemma 14 and the construction above place the entire hard family in the displayed class with constants independent of K, n, and v: non-special contrasts vanish, while the special contrast has rth moment one for finite r and essential supremum one for $r = \infty$. Let Q_v^{rem} be the independent observation law under P_v^{rem}, and let $\bar{Q}^{\mathrm{rem}} := \bigotimes_{j=0}^{K} F_0^{\otimes n}$. Only arm v differs, so Lemma 15 gives

$$\mathrm{KL}(Q_v^{\mathrm{rem}} \| \bar{Q}^{\mathrm{rem}}) = n \, \mathrm{KL}(F_1 \| F_0) \leq \frac{n}{2} s_{\mathrm{tail}}^{-d}.$$

Choose $s_{\mathrm{tail}} = (4n/\log K)^{1/d}$. For $n \geq c_0 \log K$, this is at least $s_{\mathrm{tail},1}$ and the average KL is at most $(1/8)\log K$. Lemma 13 implies that every selector errs with probability at least $1/2$ for some v. Under P_v^{rem}, arm v is uniquely best and every incorrect arm has gap

$$\Delta_{\mathrm{tail}} := \mathrm{ES}_\alpha(F_0) - \mathrm{ES}_\alpha(F_1) \geq c_{\mathrm{ES}} s_{\mathrm{tail}}^{-a} = c_{\mathrm{ES}} 4^{-\gamma} \left(\frac{\log K}{n} \right)^{\gamma}.$$

Taking the supremum over the hard family and then the infimum over selectors gives the result. For each fixed (n, K), the reference marginal F_0 is the same under every hypothesis and the comparison law. Revealing it leaves the KL comparison, and hence the Fano lower bound, unchanged. □

Corollary 18 (The remote family under both experiments)**.** *Let $s_{\mathrm{tail}} = (4n/\log K)^{1/d}$ and $\mathcal{H}_{n,K}^{\mathrm{rem}} := \{P_v^{\mathrm{rem}} : v \in [K]\}$ denote the remote-tail family at this scale. Write $\pi = s_{\mathrm{tail}}^{-p}$ and $\Delta_{\mathrm{tail}} := \mathrm{ES}_\alpha(F_0) - \mathrm{ES}_\alpha(F_1)$. Under the conditions of Theorem 17,*

$$\mathfrak{R}_{\mathrm{ind}}(n, K; \mathcal{H}_{n,K}^{\mathrm{rem}}) \geq \frac{1}{2} \Delta_{\mathrm{tail}} \geq \frac{c_{\mathrm{ES}}}{2} 4^{-\gamma} \left(\frac{\log K}{n} \right)^{\gamma},$$

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{H}_{n,K}^{\mathrm{rem}}) = \left(1 - \frac{1}{K} \right) \Delta_{\mathrm{tail}} (1 - \pi)^n.$$

This compares the two observation experiments on the same family of joint distributions.

Proof. The independent lower bound follows from the Fano argument in Theorem 17, restricted to its hard family. Let E_{nr} be the event that no synchronized observation comes from the remote component. On E_{nr}^c, any observation from the remote component identifies the special arm v exactly. Conditional on E_{nr}, the observation law is identical for every $v \in [K]$.

For any possibly randomized selector, let $a_j := \mathbb{P}(\widehat{j} = j \mid E_{\mathrm{nr}})$. Since $\sum_{j=1}^{K} a_j = 1$,

$$\sup_{v \in [K]} \mathbb{P}_{P_v^{\mathrm{rem}}}(\widehat{j} \neq v \mid E_{\mathrm{nr}}) = 1 - \min_{v \in [K]} a_v \geq 1 - \frac{1}{K}.$$

Selecting uniformly on E_{nr} and selecting the revealed arm on E_{nr}^c attains equality. Since every incorrect selection has regret Δ_{tail} and $\mathbb{P}(E_{\mathrm{nr}}) = (1-\pi)^n$, the synchronized identity follows. □

Remark 19 (A family without a uniform variance bound). Each fixed Gaussian mixture has a finite second moment, but the sequence has no uniform second-moment bound: the remote component contributes $\pi s_{\mathrm{tail}}^2 = s_{\mathrm{tail}}^{2-p} \to \infty$. Thus the sequence satisfies uniform reference pth-moment and contrast-moment bounds without a uniform variance bound.

F.3 Synchronized Evaluation: A Noisy-Contrast Family

A deterministic-shift family $L_j = C + \mu_j$ cannot yield a synchronized lower bound because one full-vector observation reveals every difference $\mu_j - \mu_k$. We instead use bounded contrasts with a nonvanishing residual-noise scale. Fix $0 < \sigma \le \sigma_{\max}$ and let $C \sim \mathsf{N}(0,1)$. For $\zeta \in (0,1)$, let $Z_\zeta \in \{-1,1\}$ satisfy $\mathbb{P}(Z_\zeta = 1) = \zeta$, independently of C, and define $\Psi_\sigma(\zeta) := \mathrm{ES}_\alpha(C + \sigma Z_\zeta)$.

Lemma 20 (ES differences under Bernoulli perturbations)**.** *Fix $\sigma_{\max} < \infty$. There are constants $0 < c_\alpha \le C_\alpha < \infty$, depending only on $(\alpha, \sigma_{\max})$, such that for every $0 < \sigma \le \sigma_{\max}$ and $\zeta, \zeta' \in [1/4, 1/2]$,*

$$c_\alpha \sigma \left|\zeta - \zeta'\right| \le \left|\Psi_\sigma(\zeta) - \Psi_\sigma(\zeta')\right| \le C_\alpha \sigma \left|\zeta - \zeta'\right|.$$

Moreover, Ψ_σ is strictly increasing.

Proof. For $t \in \mathbb{R}$, define

$$G(\zeta, t) := t + \frac{1}{\tau}\left\{(1-\zeta)\mathbb{E}(C - \sigma - t)_+ + \zeta \mathbb{E}(C + \sigma - t)_+\right\}.$$

The loss distribution has an everywhere positive density, so $G(\zeta, \cdot)$ has a unique minimizer t_ζ. Define

$$h_\sigma(t) := \frac{1}{\tau}\left\{\mathbb{E}(C + \sigma - t)_+ - \mathbb{E}(C - \sigma - t)_+\right\}.$$

For $\zeta > \zeta'$, optimality of t_ζ and $t_{\zeta'}$ gives the two-sided inequality

$$(\zeta - \zeta') h_\sigma(t_\zeta) \le \Psi_\sigma(\zeta) - \Psi_\sigma(\zeta') \le (\zeta - \zeta') h_\sigma(t_{\zeta'}).$$

The almost-sure bounds $C - \sigma_{\max} \le C + \sigma Z_\zeta \le C + \sigma_{\max}$ imply that all t_ζ, $\zeta \in [1/4, 1/2]$, lie in $[-Q_{\mathrm{sync}}, Q_{\mathrm{sync}}]$ for a finite Q_{sync} depending only on $(\alpha, \sigma_{\max})$. By Fubini,

$$\mathbb{E}(C + \sigma - t)_+ - \mathbb{E}(C - \sigma - t)_+ = \int_{-\sigma}^{\sigma} \mathbb{P}(C > t - u)\, du.$$

For $t \in [-Q_{\mathrm{sync}}, Q_{\mathrm{sync}}]$ and $|u| \le \sigma_{\max}$, the integrand is at least $\mathbb{P}(C > Q_{\mathrm{sync}} + \sigma_{\max}) > 0$ and at most one. Hence, uniformly over these minimizers,

$$\frac{2\sigma}{\tau}\mathbb{P}(C > Q_{\mathrm{sync}} + \sigma_{\max}) \le h_\sigma(t_\zeta) \le \frac{2\sigma}{\tau}.$$

Substitution into this inequality proves the two-sided bound and strict monotonicity; exchanging ζ and ζ' gives the absolute-value form. □

Fix $\zeta_0 = 1/2$ and $\delta_0 \in (0, 1/4]$. For each hypothesis $v \in [K]$, define the joint law P_v^{sync} by drawing $C \sim \mathsf{N}(0,1)$ and, independently of C, mutually independent signs $Z_1, \ldots, Z_K \in \{-1, 1\}$ with

$$\zeta_{v,j} := \zeta_0 - \delta_0 \mathbf{1}_{\{j=v\}}, \qquad \mathbb{P}(Z_j = 1) = \zeta_{v,j}.$$

Set

$$L_0 = C, \qquad L_j = C + \sigma Z_j, \quad j \in [K]. \tag{46}$$

Proposition 21 (Uniform class membership)**.** *For fixed $\sigma_{\max}$ there is a constant $M_{0,\mathrm{sync}} < \infty$, independent of K, v, and δ_0, such that the family (46), with $0 < \sigma \le \sigma_{\max}$ and $0 < \delta_0 \le 1/4$, belongs to $\mathcal{P}_{K,p,r}(M_{0,\mathrm{sync}}, \sigma_{\max})$ for every $2 \le r \le \infty$.*

Proof. The reference is standard normal, so its pth moment is a fixed constant. The contrasts equal σZ_j, hence are bounded by σ. Since $\zeta_{v,j} \in [1/4, 1/2]$, their variances are at least $3\sigma^2/4$, so the residual noise does not degenerate. □

For $u, v \in [0,1]$, let $\mathrm{kl}(u,v) := \mathrm{KL}(\mathrm{Bernoulli}(u) \| \mathrm{Bernoulli}(v))$ denote binary relative entropy, allowing the value $+\infty$. For $u, v \in (0,1)$, this equals $u \log(u/v) + (1-u)\log\{(1-u)/(1-v)\}$.

Lemma 22 (Synchronized KL bound)**.** *Let $P_\circ^{\mathrm{sync}}$ be the law obtained from the preceding construction by setting $\delta_0 = 0$, with the law of C and all independence assumptions unchanged. Then*

$$\mathrm{KL}(P_v^{\mathrm{sync}} \| P_\circ^{\mathrm{sync}}) \le \mathrm{kl}(\zeta_0 - \delta_0, \zeta_0) \le 4\delta_0^2.$$

Proof. In the joint law of $(C, Z_1, \dots, Z_K)$, only the distribution of Z_v changes. The loss vector is a measurable function of $(C, Z_1, \dots, Z_K)$, so data processing gives the first inequality. The second follows from the chi-squared divergence bound $\mathrm{KL}(P\|Q) \le \chi^2(P\|Q)$ and

$$\chi^2(\mathrm{Bernoulli}(1/2 - \delta_0)\|\,\mathrm{Bernoulli}(1/2)) = 4\delta_0^2.$$

□

Theorem 23 (Synchronized K-arm lower bound)**.** *There are constants $c, c_0 > 0$, depending only on $(\alpha, p, r, \sigma_{\max})$, such that for every $K \ge 8$, $n \ge c_0 \log K$, and $0 < \sigma \le \sigma_{\max}$,*

$$\mathfrak{R}_{\mathrm{sync}}(n, K; \mathcal{P}_{K,p,r}(M_{0,\mathrm{sync}}, \sigma_{\max})) \ge c\sigma\sqrt{\frac{\log K}{n}}.$$

Proof. Let $Q_v^{\mathrm{sync}} = (P_v^{\mathrm{sync}})^{\otimes n}$ and $\bar{Q}^{\mathrm{sync}} = (P_\circ^{\mathrm{sync}})^{\otimes n}$. By Lemma 22, $\mathrm{KL}(Q_v^{\mathrm{sync}}\|\bar{Q}^{\mathrm{sync}}) \le 4n\delta_0^2$. Choose $\delta_0 := \sqrt{\log K/(32n)}$. For $n \ge c_0 \log K$, this is at most $1/4$, and the average KL is at most $(1/8)\log K$. Fano's inequality gives an error probability of at least $1/2$ for some hypothesis. By Lemma 20, arm v is uniquely best and every other arm has gap

$$\Psi_\sigma(\zeta_0) - \Psi_\sigma(\zeta_0 - \delta_0) \ge c_\alpha \sigma \delta_0.$$

Proposition 21 places the family in the displayed fixed class, completing the minimax lower bound. □

Completion of Theorem 1. Set $\sigma_{\max} = 1$ and $\sigma_0 = 1/2$, and define $M_{0,\star} := \max\{M_{0,\mathrm{rem}}, M_{0,\mathrm{sync}}\}$ and $B_\star := \max\{1, \sigma_{\max}\}$. Lemma 14 and Proposition 21 place both lower-bound families in $\mathcal{P}_{K,p,r}(M_{0,\star}, B_\star)$, with constants independent of K and n. For the prescribed $M_0, B > 0$, define

$$s := \min\left\{1, \left(\frac{M_0}{M_{0,\star}}\right)^{1/p}, \frac{B}{B_\star}\right\} > 0.$$

Scaling every loss by s maps the common class into $\mathcal{P}_{K,p,r}(M_0, B)$: the reference moment is at most $s^p M_{0,\star} \le M_0$, and each contrast norm is at most $sB_\star \le B$, including when $r = \infty$.

In either experiment, this invertible transformation gives a bijection between measurable selectors before and after scaling, with any auxiliary random seed unchanged. Positive homogeneity of ES multiplies every selection regret by s. Hence, for $\mathsf{e} \in \{\mathrm{sync}, \mathrm{ind}\}$,

$$\mathfrak{R}_{\mathsf{e}}(n, K; \mathcal{P}_{K,p,r}(M_0, B)) \ge s\,\mathfrak{R}_{\mathsf{e}}(n, K; \mathcal{P}_{K,p,r}(M_{0,\star}, B_\star)).$$

Theorems 23 and 17, with $\sigma = \sigma_0$ in the synchronized construction, give the lower bounds on the common class. Since s is independent of K and n, these imply the required lower bounds on the target class. Theorems 8 and 12 give its upper bounds. Substituting $n = N/(K+1)$ proves (15).

The independent lower bound with the additional information P_0 and $(q_j)_{j=0}^K$ is proved in Appendix G.2.

G Additional Fixed-Budget Results

G.1 Synchronized Empirical ES

Corollary 24 (Empirical ES under bounded contrasts)**.** *Fix $\alpha \in (0,1)$, $1 < p < 2$, and $M_0, B > 0$. Under synchronized evaluation, define, for $j = 0, \dots, K$,*

$$\widehat{\rho}_j^{\mathrm{emp}} := \inf_{t\in\mathbb{R}} \frac{1}{n}\sum_{i=1}^n \ell(L_{i,j}, t), \qquad \widehat{j}_{\mathrm{emp}} := \min \arg\min_{j\in[K]} \widehat{\rho}_j^{\mathrm{emp}}.$$

There are constants $c, C > 0$, depending only on (α, p, M_0, B), such that, for $K \ge 2$ and $n \ge c\log(2K)$,

$$\sup_{P\in\mathcal{P}_{K,p,\infty}(M_0,B)} \mathbb{E}_P\left[\rho_{\widehat{j}_{\mathrm{emp}}} - \rho_\star\right] \le C\sqrt{\frac{\log(2K)}{n}}.$$

The selector uses all observations and requires no moment-dependent tuning.

Proof. Put $\kappa = \min\{\alpha, \tau\}$ and

$$Q_{\mathrm{emp}} := \left(\frac{4\overline{M}}{\kappa}\right)^{1/p}.$$

This radius is used only in the proof. For each arm, write

$$P_{n,j} f := \frac{1}{n}\sum_{i=1}^{n} f(L_{i,j}), \qquad \widehat{F}_{n,j}(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{L_{i,j}\le t\}}.$$

Define

$$h_t(z) := \ell(z,t) - \ell(z,0), \qquad c_j := \inf_{|t|\le Q_{\mathrm{emp}}} P_j h_t, \qquad \widehat{c}_j := \inf_{|t|\le Q_{\mathrm{emp}}} P_{n,j} h_t.$$

The moment bound places a population RU minimizer in this interval, so

$$\rho_j = \tau^{-1}\mathbb{E}(L_j)_+ + c_j.$$

The empirical-process calculation in the proof of Lemma 3 gives

$$|\widehat{c}_j - c_j| \le \sup_{|t|\le Q_{\mathrm{emp}}} |(P_{n,j} - P_j)h_t| \le \frac{Q_{\mathrm{emp}}}{\tau}\|\widehat{F}_{n,j} - F_{L_j}\|_\infty.$$

Set

$$D_{i,j}^{(0)} := \ell(L_{i,j},0) - \ell(L_{i,0},0) = \frac{(L_{i,j})_+ - (L_{i,0})_+}{\tau},$$

and define the constrained comparison

$$\widetilde{\theta}_j := \frac{1}{n}\sum_{i=1}^{n} D_{i,j}^{(0)} + \widehat{c}_j - \widehat{c}_0.$$

The exact decomposition gives

$$\begin{aligned}\widetilde{\theta}_j - \theta_j = \frac{1}{n}\sum_{i=1}^{n} D_{i,j}^{(0)} - \mathbb{E}D_{1,j}^{(0)} \\ + (\widehat{c}_j - c_j) - (\widehat{c}_0 - c_0).\end{aligned}$$

Since $|D_{i,j}^{(0)}| \le B/\tau$, Hoeffding's inequality and the Dvoretzky–Kiefer–Wolfowitz inequality, followed by union bounds and integration, yield

$$\mathbb{E}_P \max_{j\in[K]} |\widetilde{\theta}_j - \theta_j| \le C\sqrt{\frac{\log(2K)}{n}}.$$

Let $\mathcal{G}$ be the event

$$\max_{0\le j\le K} \|\widehat{F}_{n,j} - F_{L_j}\|_\infty \le \kappa/4.$$

Markov's inequality gives $F_{L_j}(-Q_{\mathrm{emp}}) \le \kappa/4$ and $F_{L_j}(Q_{\mathrm{emp}}) \ge 1 - \kappa/4$. On $\mathcal{G}$, this implies $\widehat{F}_{n,j}(-Q_{\mathrm{emp}}) < \alpha < \widehat{F}_{n,j}(Q_{\mathrm{emp}})$ for every arm. Thus all empirical RU minimizers lie in $[-Q_{\mathrm{emp}}, Q_{\mathrm{emp}}]$, and hence

$$\widehat{\rho}_j^{\mathrm{emp}} - \widehat{\rho}_0^{\mathrm{emp}} = \widetilde{\theta}_j.$$

Moreover,

$$\mathbb{P}_P(\mathcal{G}^c) \le 2(K+1)\exp(-n\kappa^2/8).$$

On $\mathcal{G}$, the deterministic selection inequality bounds regret by twice the maximum relative estimation error above. On $\mathcal{G}^c$, bounded contrasts and the monotonicity and translation equivariance of ES give regret at most $2B$. Therefore,

$$\mathbb{E}_P[\rho_{\widehat{j}_{\mathrm{emp}}} - \rho_\star] \le C\sqrt{\frac{\log(2K)}{n}} + 4B(K+1)\exp(-n\kappa^2/8).$$

For sufficiently large c, the second term is absorbed into the first whenever $n \ge c\log(2K)$. □

G.2 Independent Lower Bound with Known Quantiles

For the strengthened lower bound in Theorem 1, allow selectors to use P_0 and the lower generalized quantiles $(q_j)_{j=0}^K$ in addition to observations from $\mathcal{E}_{\mathrm{ind},n}(P)$. We construct a family with the same reference marginal and the same α-quantiles under every hypothesis, while changing only the special candidate's tail law.

For sufficiently large $s_{\mathrm{tail}} \geq 1$, set

$$\pi := s_{\mathrm{tail}}^{-p} < \tau, \qquad \eta := \frac{1}{4} s_{\mathrm{tail}}^{p/r-1}.$$

On $[0,1]$, define

$$f_\eta(u) := 1 + \eta(1-2u), \qquad F_\eta(u) := u + \eta u(1-u).$$

Since $\eta \leq 1/4$, this density is strictly positive. Let $I \sim \mathrm{Bernoulli}(\pi)$ and $U \sim \mathrm{Unif}[0,1]$ be independent. Under hypothesis $v \in [K]$, define

$$L_0 = I s_{\mathrm{tail}}(1+U), \qquad L_j = L_0 \quad (j \neq v), \qquad L_v = I s_{\mathrm{tail}}\big(1 + F_\eta^{-1}(U)\big).$$

The comparison law uses $L_j = L_0$ for every candidate.

Put $V = F_\eta^{-1}(U)$. Then $U - V = \eta V(1-V) \in [0, \eta/4]$, and hence

$$\mathbb{E}|L_0|^p = C_p := \int_0^1 (1+u)^p \, du, \qquad \|L_v - L_0\|_{L^r} \leq \frac{1}{16},$$

with the same bound for $r = \infty$. Multiplying all losses by

$$s := \min\left\{1, \left(\frac{M_0}{C_p}\right)^{1/p}, 16B\right\} > 0$$

therefore places the family in $\mathcal{P}_{K,p,r}(M_0, B)$. This factor is independent of n and K.

Before scaling, every marginal has mass $1 - \pi > \alpha$ at zero. Its lower generalized α-quantile is therefore zero, and the common positive ES gap is exactly

$$\Delta = \frac{\pi s_{\mathrm{tail}} \eta}{6\tau} = \frac{s_{\mathrm{tail}}^{-a}}{24\tau}.$$

Only the special candidate marginal differs from the comparison law. Its KL divergence is bounded by

$$\begin{aligned} \pi \int_0^1 f_\eta(u) \log f_\eta(u) \, du &\leq \pi \int_0^1 (f_\eta(u) - 1)^2 \, du \\ &= \frac{\pi \eta^2}{3} = \frac{s_{\mathrm{tail}}^{-d}}{48}. \end{aligned}$$

Scaling leaves KL unchanged. Under independent evaluation, the observation KL is thus at most $n s_{\mathrm{tail}}^{-d}/48$.

Choose $s_{\mathrm{tail}} = (n/\log K)^{1/d}$. For $n \geq c_0 \log K$ with sufficiently large c_0, all preceding conditions hold, and the observation KL is at most $(1/48)\log K$. Lemma 13 gives error probability at least $1/2$ for some hypothesis when $K \geq 8$. After scaling, every incorrect selection has regret $s\Delta$, so the expected regret is at least

$$\frac{s}{48\tau}\left(\frac{\log K}{n}\right)^{\gamma(p,r)}.$$

For each fixed (n, K), the reference marginal is identical under all hypotheses and the comparison law, and $q_j = 0$ for every $j = 0, \ldots, K$. Revealing P_0 and $(q_j)_{j=0}^K$ therefore leaves the testing problem and the lower bound unchanged.

G.3 The First-Moment Endpoint

For this subsection, define $\mathcal{P}_{K,1,r}(M_0, B)$ by (4)–(5) with $p = 1$. The corresponding constants are

$$\overline{M} = M_0 + B, \qquad Q = 1 + \frac{\overline{M}}{\kappa}, \qquad a = 1 - \frac{1}{r}, \qquad d = 3 - \frac{2}{r}.$$

Proposition 25 (Fixed-budget rates at $p = 1$)**.** *Fix $\alpha \in (0,1)$, $2 \le r \le \infty$, and $M_0, B > 0$. There are constants $c, C, c_0 > 0$, depending only on these parameters, such that, for $K \ge 8$ and $n \ge c_0 \log K$,*

$$c\sqrt{\frac{\log K}{n}} \le \mathfrak{R}_{\mathrm{sync}}\big(n, K; \mathcal{P}_{K,1,r}(M_0, B)\big) \le C\sqrt{\frac{\log K}{n}},$$

and

$$c\left(\frac{\log K}{n}\right)^{\gamma(1,r)} \le \mathfrak{R}_{\mathrm{ind}}\big(n, K; \mathcal{P}_{K,1,r}(M_0, B)\big) \le C\left(\frac{\log K}{n}\right)^{\gamma(1,r)},$$

where $\gamma(1,r) = (1-1/r)/(3-2/r)$. For the class $\mathcal{P}_{K,1,\infty}(M_0, B)$, the conclusion of Corollary 24 also holds with $p = 1$.

Proof. Every marginal has first absolute moment at most $\overline{M} = M_0 + B$, so the quantile bounds and the proof of Lemma 3 remain valid. The synchronized upper-bound proof uses only this quantile bound and the second moment of the paired score difference.

For independent evaluation, the variance and relative-bias arguments give, with $x = \log(2K)/n$,

$$C\left\{T^{-(1-1/r)} + T^{1/2}\sqrt{x} + Tx + \sqrt{x}\right\}.$$

Choosing $T = 2Q + 1 + x^{-1/(3-2/r)}$ gives the stated exponent. Thus the proofs of Theorems 8 and 12 extend to this endpoint. For bounded contrasts, the proof of Corollary 24 also applies with the first-moment bound.

The construction in Appendix G.2 remains valid at $p = 1$ and gives the independent lower bound. For the synchronized lower bound, Jensen's inequality gives

$$\mathcal{P}_{K,3/2,r}(M_0^{3/2}, B) \subseteq \mathcal{P}_{K,1,r}(M_0, B).$$

The synchronized lower bound in Theorem 1, applied to the smaller class, is also a lower bound on the larger class. □

H Pairing Each Candidate with the Reference

Let P_{0j} denote the joint law of (L_0, L_j) under P. For each candidate j, observe n independent pairs with law P_{0j}, independently across candidates. The observation law is

$$\mathcal{E}_{\mathrm{pair},n}(P) := \bigotimes_{j=1}^{K} P_{0j}^{\otimes n}. \tag{47}$$

Define $\mathfrak{R}_{\mathrm{pair}}(n, K; \mathcal{P})$ by (9), with selectors based on these pairs.

Corollary 26 (Pairing each candidate with the reference)**.** *Fix $\alpha \in (0,1)$, $1 < p < 2$, and $M_0, B > 0$. For every $2 \le r \le \infty$, $K \ge 8$, and $n \ge 1$,*

$$\mathfrak{R}_{\mathrm{pair}}\big(n, K; \mathcal{P}_{K,p,r}(M_0, B)\big) \asymp \min\left\{1, \sqrt{\frac{\log K}{n}}\right\}, \tag{48}$$

where the comparison constants depend only on (α, p, M_0, B). The lower bound also holds when the reference marginal P_0 is known exactly. Under unit evaluation costs, $N = 2Kn$, so equivalently the risk is of order $\min\{1, \sqrt{K \log K / N}\}$.

Proof of Corollary 26. **Upper bound.** For $n = 1$, select any fixed candidate; its regret is at most $2B/\tau$ by Lemma 5. For $n \geq 2$, proceed as follows. Write $(L_{i,0}^{(j)}, L_{i,j})$ for pair i associated with candidate j. Split each sample of n pairs into two deterministic folds of size $m = \lfloor n/2 \rfloor$, discarding one pair when n is odd. On the first fold, fit $\widehat{q}_j$ and $\widehat{q}_0^{(j)}$ by minimizing their respective empirical RU objectives over $[-Q, Q]$, with measurable tie-breaking. Let $\mathcal{F}_{\mathrm{q}}$ be the sigma-field generated by all fitting folds. On the second fold, indexed by $i = 1, \ldots, m$, set

$$D_{i,j}^{\mathrm{pair}} := \ell(L_{i,j}, \widehat{q}_j) - \ell(L_{i,0}^{(j)}, \widehat{q}_0^{(j)}).$$

The pointwise argument in Lemma 7 bounds these paired scores' conditional second moments by V_{sync}. With $H_K = \sqrt{V_{\mathrm{sync}} m / \log(2K)}$, define

$$\widehat{\theta}_j^{1\to 2} := \frac{1}{m} \sum_{i=1}^{m} \mathrm{clip}_{H_K}\left(D_{i,j}^{\mathrm{pair}}\right), \qquad \mu_j := G_{L_j}(\widehat{q}_j) - G_{L_0}(\widehat{q}_0^{(j)}).$$

Conditional on $\mathcal{F}_{\mathrm{q}}$, the score differences are independent over i with mean μ_j. The same clipping-and-Bernstein argument as in Theorem 8 therefore gives

$$\mathbb{E}_P\left[\max_{j\in[K]} |\widehat{\theta}_j^{1\to 2} - \mu_j| \,\middle|\, \mathcal{F}_{\mathrm{q}}\right] \leq C\sqrt{\frac{\log(2K)}{m}}.$$

Put $e_j := G_{L_j}(\widehat{q}_j) - \rho_j$ and $e_0^{(j)} := G_{L_0}(\widehat{q}_0^{(j)}) - \rho_0$. These quantities are nonnegative, and

$$\mu_j - \theta_j = e_j - e_0^{(j)}, \qquad \max_{j\in[K]} |\mu_j - \theta_j| \leq \max_{j\in[K]} \max\{e_j, e_0^{(j)}\}.$$

Lemma 3, applied to the $2K$ fitted objectives, bounds the expectation on the right by $C\sqrt{\log(4K)/m}$. Define $\widehat{\theta}_j^{2\to 1}$ by interchanging the folds and set $\widehat{\theta}_j^{\mathrm{pair}} := (\widehat{\theta}_j^{1\to 2} + \widehat{\theta}_j^{2\to 1})/2$. The same bound holds in each direction; the triangle inequality gives

$$\sup_{P\in\mathcal{P}_{K,p,r}} \mathbb{E}_P \max_{j\in[K]} |\widehat{\theta}_j^{\mathrm{pair}} - \theta_j| \leq C\sqrt{\frac{\log K}{n}}.$$

Select the smallest index minimizing $\widehat{\theta}_j^{\mathrm{pair}}$. The deterministic inequality $\rho_{\widehat{j}} - \rho_\star \leq 2\max_{j\in[K]} |\widehat{\theta}_j^{\mathrm{pair}} - \theta_j|$ and the deterministic bound $2B/\tau$ prove the upper bound for $n \geq 2$. The case $n = 1$ follows from $\log K > 1$. All constants used here depend only on (α, p, M_0, B), not on r.

Lower bound. Use the family in (46) with $\sigma_{\max} = 1$, $\sigma = 1/2$, and $\zeta_0 = 1/2$. For the prescribed moment bounds, set

$$s := \min\left\{1, \left(\frac{M_0}{M_{0,\mathrm{sync}}}\right)^{1/p}, \frac{B}{\sigma}\right\} > 0,$$

and let $\widetilde{P}_v$ and $\widetilde{P}_\circ$ be the laws obtained by multiplying every loss under P_v^{sync} and $P_\circ^{\mathrm{sync}}$ by s. Their reference pth moments are at most M_0, and all contrasts are bounded by $s\sigma \leq B$; hence these laws belong to $\mathcal{P}_{K,p,r}$ for every permitted r. Only the pair marginal indexed by v differs from the comparison law. Tensorization, data processing, invariance of KL under positive scaling, and Lemma 22 give

$$\begin{aligned}\mathrm{KL}\Big(\mathcal{E}_{\mathrm{pair},n}(\widetilde{P}_v) \,\Big\|\, \mathcal{E}_{\mathrm{pair},n}(\widetilde{P}_\circ)\Big) \\ = n\sum_{j=1}^{K} \mathrm{KL}\Big((\widetilde{P}_v)_{0j} \,\Big\|\, (\widetilde{P}_\circ)_{0j}\Big) \leq 4n\delta_0^2.\end{aligned}$$

Choose $\delta_0 = \min\{1/4, \sqrt{\log K/(32n)}\}$. Then $0 < \delta_0 \leq 1/4$ and the last bound is at most $(1/8)\log K$ for every $n \geq 1$. Lemma 13 yields error probability at least $1/2$ for some v, including for randomized selectors. By Lemma 20 and positive homogeneity of ES, candidate v is uniquely best under $\widetilde{P}_v$, and every other candidate has gap at least $sc_\alpha\sigma\delta_0$. Consequently,

$$\mathfrak{R}_{\mathrm{pair}}(n, K; \mathcal{P}_{K,p,r}) \geq \frac{sc_\alpha\sigma}{2}\delta_0 \geq \frac{sc_\alpha\sigma}{2\sqrt{32}} \min\left\{1, \sqrt{\frac{\log K}{n}}\right\}.$$

The constants s and c_α can be chosen independently of r, K, and n, proving the stated lower bound. The reference law is $\mathsf{N}(0, s^2)$ under every hypothesis and the comparison law. Revealing it therefore leaves the Fano argument unchanged. Substituting $n = N/(2K)$ gives the cost formulation. □

I Adaptive Fixed-Confidence Upper Bounds

When all confidence bounds are valid, elimination gives the desired gap-dependent cost. To also control expected cost, we combine finite capped elimination runs with restarts whose error probabilities are summable. For a fixed instance with unique best arm, define

$$\Delta_j := \rho_j - \rho_{j^\star}, \qquad j \in [K], \qquad \Delta_{\min} := \min_{j \neq j^\star} \Delta_j.$$

Thus $\Delta_{j^\star} = 0$ and $\Delta_j > 0$ for every $j \neq j^\star$.

Within this section, R indexes a capped run and h a stage within that run; m_h, A_h, $\varepsilon_{R,h}$, and $\beta_{R,h}$ denote the fold size, active set, stage error budget, and confidence radius. The exponent ν describes how evaluation counts grow as gaps shrink; it is 2 for synchronized evaluation and $1/\gamma$ for independent evaluation.

I.1 A Capped Elimination Run

Fix a nonnegative integer cap R, a run-level error budget $\varepsilon_R \in (0, 1/2)$, and a sufficiently large fixed initial fold size m_0. Start with $A_0 = [K]$. At internal stage $h = 0, \dots, R$, let $m_h = 2^h m_0$ and allocate

$$\varepsilon_{R,h} := \frac{\varepsilon_R 2^{-(h+1)}}{\sum_{u=0}^{R} 2^{-(u+1)}}.$$

Use two fresh folds of size m_h for each active candidate and the reference. At each stage, use the prescribed-confidence estimator from the corresponding upper-bound theorem with $n = 2m_h$ and failure probability $\varepsilon_{R,h}$, including its confidence-dependent threshold. In the synchronized run, set

$$\beta^{\mathrm{sync}}_{R,h} := C_{\mathrm{s}} \sqrt{\frac{s_{R,h}}{m_h}}, \qquad s_{R,h} := \log \frac{2K}{\varepsilon_{R,h}}. \tag{49}$$

In the independent run, set

$$\beta^{\mathrm{ind}}_{R,h} := C_{\mathrm{i}} \left(\frac{s_{R,h}}{m_h} \right)^{\gamma}. \tag{50}$$

C_{s} and C_{i} are sufficiently large fixed constants inherited from the synchronized and independent deviation bounds. At stage h, $\widehat{\theta}_{j,h}$ denotes the corresponding estimate for arm j, and $\beta_{R,h}$ denotes $\beta^{\mathrm{sync}}_{R,h}$ or $\beta^{\mathrm{ind}}_{R,h}$ according to the experiment. Let c_{dev} be a fixed constant large enough for both deviation bounds. If $m_h < c_{\mathrm{dev}} s_{R,h}$, make no elimination. Otherwise, update

$$A_{h+1} := \left\{ j \in A_h : \widehat{\theta}_{j,h} - \beta_{R,h} \leq \min_{k \in A_h} (\widehat{\theta}_{k,h} + \beta_{R,h}) \right\}. \tag{51}$$

The run returns immediately when one candidate remains. If more than one remains after stage R, it returns no decision.

Lemma 27 (Correct elimination on a good run)**.** *Suppose that $|\widehat{\theta}_{j,h} - \theta_j| \leq \beta_{R,h}$ for every active arm at every executed stage satisfying $m_h \geq c_{\mathrm{dev}} s_{R,h}$. Then the best arm is never removed. At each such stage, every active arm j with $\Delta_j > 4\beta_{R,h}$ is removed. A run whose cap allows these conditions to hold for every remaining suboptimal arm returns the best arm.*

Proof. At a stage with $m_h < c_{\mathrm{dev}} s_{R,h}$, no arm is removed. At any other stage, if $j^\star \in A_h$, every $k \in A_h$ satisfies

$$\widehat{\theta}_{j^\star,h} - \beta_{R,h} \leq \theta_{j^\star} \leq \theta_k \leq \widehat{\theta}_{k,h} + \beta_{R,h}.$$

Hence $j^\star \in A_{h+1}$, and induction shows that the best arm is never removed. If $\Delta_j > 4\beta_{R,h}$, then

$$\widehat{\theta}_{j,h} - \beta_{R,h} \geq \theta_j - 2\beta_{R,h} > \theta_{j^\star} + 2\beta_{R,h} \geq \widehat{\theta}_{j^\star,h} + \beta_{R,h},$$

so j is removed. □

I.2 Restarts with Summable Error Probabilities

Let

$$\varepsilon_R := \frac{7\delta}{8}\, 2^{-3(R+1)}, \qquad R = 0, 1, 2, \dots.$$

Run the capped procedures independently in increasing order of R, always restarting from all candidates with fresh data, and stop at the first returned candidate. Since $\sum_R \varepsilon_R < \delta$, a wrong output has probability at most δ by Lemma 27.

Put

$$A_\delta := \log \frac{16K}{\delta}, \qquad \Lambda(P,\delta) := A_\delta + \log\left(e + \frac{1}{\Delta_{\min}}\right), \tag{52}$$

$$\mathcal{L}(P,\delta) := 1 + \log\left(e + \frac{1}{\Delta_{\min}}\right) + \log(e + A_\delta).$$

Lemma 28 (Deterministic restart threshold)**.** *There are constants $c_1, c_2 > 0$, depending only on fixed class and radius constants, and, for every $\nu > 0$, a constant $C_\nu < \infty$, depending only on ν and those constants, with the following property. Define*

$$R_\nu := \min\left\{R \geq 0 : 2^R m_0 \geq c_1 \Delta_{\min}^{-\nu}(A_\delta + R + 1)\right\}. \tag{53}$$

Then $R_\nu < \infty$, its defining inequality holds for every $R \geq R_\nu$, and

$$R_\nu \leq C_\nu \left\{1 + \log\left(e + \frac{1}{\Delta_{\min}}\right) + \log(e + A_\delta)\right\} = C_\nu \mathcal{L}(P,\delta). \tag{54}$$

Moreover, uniformly for $h \leq R$,

$$s_{R,h} \leq A_\delta + c_2(R+1). \tag{55}$$

Proof. Because $\sum_{u=0}^{R} 2^{-(u+1)} \leq 1$ and $\varepsilon_R = (7\delta/8)2^{-3(R+1)}$, $\varepsilon_{R,h} \geq \varepsilon_R 2^{-(h+1)}$, which proves (55) for a numerical c_2. The sequence $2^R/(A_\delta + R + 1)$ is nondecreasing, so the defining inequality persists at every larger cap.

Set $\omega := \max\{1, c_1 m_0^{-1} \Delta_{\min}^{-\nu}\}$, $A := A_\delta + 1 \geq 1$, and $\overline{R} := \lceil 2\log_2(4\omega A)\rceil$. Then $2^{\overline{R}} \geq 16\omega^2 A^2$ and $A + \overline{R} \leq 10\omega A$. Therefore

$$2^{\overline{R}} \geq \omega(A_\delta + \overline{R} + 1) \geq c_1 m_0^{-1} \Delta_{\min}^{-\nu}(A_\delta + \overline{R} + 1).$$

Multiplying by m_0 verifies the defining inequality at $\overline{R}$, so $R_\nu \leq \overline{R}$. The definitions of ω and A then give (54). □

Lemma 5 bounds every gap by a class-dependent constant. Hence the arm-specific sample thresholds below also imply the preliminary requirement $m_h \geq c_{\mathrm{dev}} s_{R,h}$ after constants are enlarged.

Lemma 29 (Generic restarted elimination bound)**.** *Let $\mathsf{H}_{R,h}$ denote the complete history before stage h of run R, including earlier runs and the algorithm's randomization. Suppose that, for fixed $C, \nu > 0$ and every unique-best instance $P \in \mathcal{P}$, each executed stage with $m_h \geq c_{\mathrm{dev}} s_{R,h}$ satisfies, almost surely,*

$$\mathbb{P}_P\left(\max_{j \in A_h} |\widehat{\theta}_{j,h} - \theta_j| > \beta_{R,h} \,\middle|\, \mathsf{H}_{R,h}\right) \leq \varepsilon_{R,h}, \qquad \beta_{R,h} \leq C(s_{R,h}/m_h)^{1/\nu}.$$

Then the restart schedule is δ-PAC on the unique-best instances in $\mathcal{P}$. For every such P, it stops almost surely and satisfies

$$\mathbb{E}_P N \leq C_\nu\, \mathcal{L}(P,\delta)\Lambda(P,\delta) \sum_{j \neq j^\star} \Delta_j^{-\nu} + C_\nu K. \tag{56}$$

Proof. We separate the argument into correctness, the cost of good capped runs, and the tail cost of later restarts. Let $\mathcal{R}_R$ be the event that run R is reached. Let $\mathcal{G}_R$ be the event that, at every executed stage h of run R satisfying $m_h \geq c_{\mathrm{dev}} s_{R,h}$, $|\widehat{\theta}_{j,h} - \theta_j| \leq \beta_{R,h}$ for all $j \in A_h$. Define $\mathcal{B}_R := \mathcal{R}_R \cap \mathcal{G}_R^c$.

The conditional coverage bound and the stagewise error allocation give $\mathbb{P}(\mathcal{B}_R) \le \varepsilon_R \mathbb{P}(\mathcal{R}_R) \le \varepsilon_R$. A wrong output can occur only on some $\mathcal{B}_R$, so its probability is at most $\sum_R \varepsilon_R < \delta$.

Let R_ν be from Lemma 28 and put $H_\nu := A_\delta + c_2(R_\nu + 1)$. For each $j \neq j^\star$, set $m^\star_{j,\nu} := C_0 \Delta_j^{-\nu} H_\nu$, with C_0 sufficiently large. On a good run,

$$m_h \ge C_0 \Delta_j^{-\nu} s_{R,h} \quad \Longrightarrow \quad \beta_{R,h} \le C C_0^{-1/\nu} \Delta_j \le \frac{\Delta_j}{8},$$

where the preliminary condition $m_h \ge c_{\mathrm{dev}} s_{R,h}$ follows from the preceding bound on the gaps. Lemma 27 therefore removes arm j.

For $R \le R_\nu$, (55) and geometric growth of m_h bound the cumulative evaluations of arm j on a good run by $C \min\{2^R m_0, m^\star_{j,\nu}\}$. Before stopping on a good run, the best arm remains active, $S_h := A_h \setminus \{j^\star\}$ is nonempty, and $|A_h| + 1 = |S_h| + 2 \le 3|S_h|$. Thus, after absorbing the fixed two-fold factor, the best-arm and reference costs can be charged to active suboptimal arms. All good runs with $R \le R_\nu$ therefore cost at most

$$C(1 + R_\nu) \sum_{j \neq j^\star} m^\star_{j,\nu}.$$

With the constant c_1 in Lemma 28 chosen large enough, every good run with cap $R \ge R_\nu$ returns the best arm. Hence, for $R > R_\nu$,

$$\mathcal{R}_R \subseteq \mathcal{B}_{R-1}, \qquad \mathbb{P}(\mathcal{R}_R) \le \varepsilon_{R-1}.$$

The events $\mathcal{R}_R$ are decreasing, so $\mathbb{P}(\text{the procedure never stops}) = \lim_R \mathbb{P}(\mathcal{R}_R) = 0$. A capped run with cap R costs at most $CK2^R m_0$. Combining the costs of good and bad runs with $R \le R_\nu$ and all runs with $R > R_\nu$ gives

$$\mathbb{E}_P N \le C(1 + R_\nu) \sum_{j \neq j^\star} m^\star_{j,\nu} + CKm_0 \sum_{R \le R_\nu} 2^R \varepsilon_R + CKm_0 \sum_{R > R_\nu} 2^R \varepsilon_{R-1}.$$

Both series are bounded because $\varepsilon_R = (7\delta/64)2^{-3R}$. Finally, Lemma 28 gives $R_\nu \le C_\nu \mathcal{L}(P, \delta)$ and $H_\nu \le C_\nu \Lambda(P, \delta)$. Substituting the definition of $m^\star_{j,\nu}$ proves (56). □

Corollary 30 (Restarted elimination upper bounds)**.** *Both restarted procedures are δ-PAC on the unique-best instances in $\mathcal{P}$. For every such P, they stop almost surely and satisfy*

$$\mathbb{E}_P N_{\mathrm{sync}} \le C\, \mathcal{L}(P, \delta) \Lambda(P, \delta) \sum_{j \neq j^\star} \Delta_j^{-2} + CK, \tag{57}$$

$$\mathbb{E}_P N_{\mathrm{ind}} \le C\, \mathcal{L}(P, \delta) \Lambda(P, \delta) \sum_{j \neq j^\star} \Delta_j^{-1/\gamma} + CK. \tag{58}$$

Proof. Each stage uses fresh data independent of the preceding history. Conditional on that history, Theorems 8 and 12 therefore provide the coverage required by Lemma 29. When $m_h \ge c_{\mathrm{dev}} s_{R,h}$, the synchronized radius (49) is at most a constant multiple of $(s_{R,h}/m_h)^{1/2}$. Apply Lemma 29 first with $\nu = 2$, then with $\nu = 1/\gamma$ and the independent radius (50). □

J Fixed-Confidence Lower Bounds

The lower-bound argument first converts the δ-PAC requirement into armwise information constraints. It then applies the noisy-contrast family to synchronized heterogeneous gaps and the remote-tail family to independent evaluation. The independent lower bound matches on equal or comparable-gap classes; matching for arbitrary heterogeneous gaps remains open.

J.1 Adaptive Information Inequality

The next lemma gives a conditional-kernel version of the standard bandit change-of-measure argument (Kaufmann et al., 2016, Lemma 1). We include a proof for the present observation model.

Lemma 31 (Adaptive change of measure)**.** *Consider data distributions P, Q under the same sampling and stopping rules. Let T be the stopping time, assumed almost surely finite under both distributions. Let H_t contain the algorithm's random seed and the actions, observations, and stopping indicators through round t. The seed has the same law under P and Q and is independent of the underlying loss samples. Let A_t be the action or arm subset determined by H_{t-1}, and let $P_{A_t}(\cdot \mid \mathsf{H}_{t-1})$ and $Q_{A_t}(\cdot \mid \mathsf{H}_{t-1})$ be the conditional observation kernels. For a terminal decision $\widehat{j}$ and an event E measurable at stopping, the data-processing inequality gives*

$$\mathrm{kl}(\mathbb{P}_P(E), \mathbb{P}_Q(E)) \le \mathbb{E}_P \sum_{t \le T} \mathrm{KL}(P_{A_t}(\cdot \mid \mathsf{H}_{t-1}) \| Q_{A_t}(\cdot \mid \mathsf{H}_{t-1})). \tag{59}$$

In particular, suppose that, almost surely under P, the two conditional observation kernels coincide whenever $j \notin A_t$, and their KL divergence is at most κ_j whenever $j \in A_t$. If N_j counts the observations containing j up to stopping, then

$$\kappa_j \mathbb{E}_P N_j \ge \mathrm{kl}(\mathbb{P}_P(E), \mathbb{P}_Q(E)). \tag{60}$$

Proof. Let $T_m = T \wedge m$. The finite-horizon chain rule, conditional on the predictable actions, gives

$$\mathrm{KL}(\mathrm{Law}_P(\mathsf{H}_{T_m}) \| \mathrm{Law}_Q(\mathsf{H}_{T_m})) = \mathbb{E}_P \sum_{t=1}^{T_m} \mathrm{KL}(P_{A_t} \| Q_{A_t}).$$

Set $E_m = E \cap \{T \le m\}$. Its indicator is measurable with respect to H_{T_m}, so data processing gives the analogue of (59) for E_m. Almost-sure stopping implies $\mathbb{P}_P(E_m) \to \mathbb{P}_P(E)$ and $\mathbb{P}_Q(E_m) \to \mathbb{P}_Q(E)$. Lower semicontinuity of binary relative entropy and monotone convergence on the right give (59). If only observations containing j differ, the right-hand side is at most $\kappa_j \mathbb{E}_P N_j$, proving (60). □

J.2 Synchronized Heterogeneous-Gap Lower Bound

Fix $\sigma > 0$ and use the function Ψ_σ from Lemma 20. On $[1/4, 1/2]$, its increments are comparable to $\sigma\,|\zeta - \zeta'|$.

Theorem 32 (Synchronized lower bound for prescribed gaps)**.** *Fix $\delta \in (0, 1/4)$. There are fixed class constants and a constant $c > 0$, independent of δ, such that, for every sufficiently small gap vector $(\Delta_j)_{j=2}^K$, one can construct an instance P with best arm 1 and suboptimality gaps exactly $(\Delta_j)_{j=2}^K$ for which every δ-PAC synchronized algorithm satisfies*

$$\mathbb{E}_P N \ge c\sigma^2 \sum_{j=2}^{K} \Delta_j^{-2} \log \frac{1}{\delta}. \tag{61}$$

Proof. Choose $\zeta_\star = 3/8$. For each $j \ge 2$, choose $u_j \in (0, 1/8)$ so that

$$\Psi_\sigma(\zeta_\star + u_j) - \Psi_\sigma(\zeta_\star) = \Delta_j.$$

For sufficiently small Δ_j, existence follows from continuity and strict monotonicity; Lemma 20 also gives $u_j \asymp \Delta_j/\sigma$. Draw $C \sim \mathsf{N}(0,1)$ and, independently of C, mutually independent signs $Z_1, \dots, Z_K \in \{-1, 1\}$ with

$$\mathbb{P}(Z_1 = 1) = \zeta_\star, \qquad \mathbb{P}(Z_j = 1) = \zeta_\star + u_j, \quad j = 2, \dots, K.$$

Set

$$L_0 = C, \qquad L_j = C + \sigma Z_j, \quad j \in [K].$$

This instance belongs to a fixed class with constants independent of K.

For each $j \ge 2$, define $Q^{(j)}$ by changing $\mathbb{P}(Z_j = 1)$ to $\zeta_\star - u_j$ and leaving the other distributions unchanged. Arm j is then uniquely best. An observation containing arm j has one-step KL divergence at most, for a fixed numerical constant C_{kl},

$$\mathrm{kl}(\zeta_\star + u_j, \zeta_\star - u_j) \le C_{\mathrm{kl}} u_j^2 \lesssim \Delta_j^2/\sigma^2,$$

whereas observations excluding arm j have identical laws. For $E = \{\widehat{j} = 1\}$, the PAC property gives $\mathbb{P}_P(E) \ge 1 - \delta$ and $\mathbb{P}_{Q^{(j)}}(E) \le \delta$. Applying Lemma 31 yields

$$\mathbb{E}_P N_j \ge c\sigma^2 \Delta_j^{-2}\, \mathrm{kl}(1-\delta, \delta) \ge c'\sigma^2 \Delta_j^{-2} \log \frac{1}{\delta}.$$

Summing the armwise bounds and using $N \ge \sum_{j=2}^K N_j$ proves the theorem. □

J.3 Independent Comparable-Gap Lower Bound

For $s_{\mathrm{tail}} \geq 1$, retain A, π, and b from Appendix F and extend (45) to $t \in \{0, 1, 2\}$ by

$$F_t := (1 - \pi)\mathsf{N}(0, 1) + \pi\mathsf{N}(A - tb, s_{\mathrm{tail}}^2).$$

Introducing the mixture label and applying the data-processing inequality gives

$$\mathrm{KL}(F_0 \| F_2) \leq 2\pi b^2 / s_{\mathrm{tail}}^2 = 2s_{\mathrm{tail}}^{-d}. \tag{62}$$

Lemma 33 (Moment bounds and ES gaps for F_0, F_1, F_2)**.** *There are fixed constants $0 < c \leq C < \infty$, $M_{0,3}, B_3$, and $s_{\mathrm{tail},3} < \infty$, depending only on (α, p, r), such that for every $s_{\mathrm{tail}} \geq s_{\mathrm{tail},3}$, any collection consisting of reference marginal F_0 and K candidate marginals chosen from $\{F_0, F_1, F_2\}$ admits a common coupling satisfying the contrast bounds in $\mathcal{P}_{K,p,r}(M_{0,3}, B_3)$, uniformly in K. Moreover,*

$$c\pi b \leq \mathrm{ES}_\alpha(F_0) - \mathrm{ES}_\alpha(F_1) \leq C\pi b, \qquad c\pi b \leq \mathrm{ES}_\alpha(F_1) - \mathrm{ES}_\alpha(F_2) \leq C\pi b.$$

Proof. Draw $I_{\mathrm{rem}} \sim \mathrm{Bernoulli}(\pi)$ and $W \sim \mathsf{N}(0, 1)$ independently. On $\{I_{\mathrm{rem}} = 0\}$, set all arms equal to W; on $\{I_{\mathrm{rem}} = 1\}$, set the reference to $A + s_{\mathrm{tail}}W$ and a level-t arm to $A - tb + s_{\mathrm{tail}}W$. Since $b = o(s_{\mathrm{tail}})$ for finite r and $b = 1$ for $r = \infty$, $A - 2b \geq 2s_{\mathrm{tail}}$ for all sufficiently large s_{tail}. The moment and quantile-bound arguments of Lemma 14 are therefore uniform over $t \in \{0, 1, 2\}$. The largest contrast is $2b$ on an event of probability π, giving $\pi(2b)^r = 2^r$ for finite r and essential supremum two for $r = \infty$.

For the ES gaps, couple $Z_t \sim F_t$ and $Z_{t+1} \sim F_{t+1}$ for $t = 0, 1$ so that $Z_t = Z_{t+1} + bI_{\mathrm{rem}}$. The upper dual bound is $\pi b/\tau$. For the lower bound, use the dual optimizer of Z_{t+1} as in Lemma 16. The remote center is at least $A - 2b \geq 2s_{\mathrm{tail}}$, while all quantiles remain in a fixed compact interval, so the conditional exceedance probability is bounded below by a positive constant. □

Theorem 34 (Independent equal-gap adaptive lower bound)**.** *Fix $\delta \in (0, 1/4)$. There are fixed class constants and a constant $c > 0$, independent of δ, such that, for every sufficiently small $\Delta > 0$ and $K \geq 2$, there is an independent-evaluation instance P with $K - 1$ suboptimal arms, all with gaps comparable to Δ, on which every δ-PAC algorithm satisfies*

$$\mathbb{E}_P N \geq c(K - 1)\Delta^{-1/\gamma} \log \frac{1}{\delta}. \tag{63}$$

Proof. Choose s_{tail} so that $\Delta \asymp \pi b = s_{\mathrm{tail}}^{-a}$, equivalently $s_{\mathrm{tail}}^d \asymp \Delta^{-1/\gamma}$. Let the reference have marginal F_0, let arm one have marginal F_1, and let each arm $j \geq 2$ have marginal F_0. Lemma 33 places this base instance and the alternatives below in one fixed class, and all suboptimal gaps are comparable to Δ.

For each $j \geq 2$, define $Q^{(j)}$ by changing only arm j from F_0 to F_2. Arm j is uniquely best, and (62) bounds each observation's KL divergence by $2s_{\mathrm{tail}}^{-d}$. With $E = \{\widehat{j} = 1\}$, Lemma 31 gives

$$2s_{\mathrm{tail}}^{-d}\mathbb{E}_P N_j \geq \mathrm{kl}(1 - \delta, \delta) \geq c \log \frac{1}{\delta}.$$

Thus $\mathbb{E}_P N_j \geq cs_{\mathrm{tail}}^d \log(1/\delta)$ for each $j \geq 2$. Summing these bounds and using $N \geq \sum_{j=2}^K N_j$ and $s_{\mathrm{tail}}^d \asymp \Delta^{-1/\gamma}$ proves (63). □

Corollary 35 (Evaluation counts for comparable ES gaps)**.** *There exist fixed $c_-, c_+ > 0$ and a sequence of classes with the same numerical constants such that, for all sufficiently small Δ and $\delta \in (0, 1/4)$,*

$$\mathfrak{C}_{\mathrm{sync}}(\delta; \mathcal{Q}_K^{\mathrm{eq}}(\Delta; c_-, c_+)) = \widetilde{\Theta}\left(K\Delta^{-2} \log \frac{1}{\delta}\right),$$

$$\mathfrak{C}_{\mathrm{ind}}(\delta; \mathcal{Q}_K^{\mathrm{eq}}(\Delta; c_-, c_+)) = \widetilde{\Theta}\left(K\Delta^{-1/\gamma} \log \frac{1}{\delta}\right).$$

Here $\widetilde{\Theta}$ suppresses additional logarithmic or iterated-logarithmic factors in K, $1/\delta$, and $1/\Delta$, beyond the displayed $\log(1/\delta)$. The powers of Δ match; optimizing the additional logarithmic factors remains open.

Proof. For the synchronized construction, take equal base gaps of order Δ. Lemma 20 then keeps every positive gap under a one-arm alternative within fixed multiples of Δ. For the independent construction, Lemma 33 makes both consecutive gaps among F_0, F_1, F_2 of order $\pi b \asymp \Delta$, and hence the same holds for every positive gap under a one-arm alternative. Thus fixed c_-, c_+ can be chosen so that all base and alternative instances in both constructions belong to $\mathcal{Q}_K^{\mathrm{eq}}(\Delta; c_-, c_+)$. Take the maxima of the reference-moment and contrast bounds for the synchronized and independent hard families. Both lower constructions then lie in the same fixed class, while the uniform elimination upper bounds apply to that class. Combining the corresponding upper and lower bounds proves the claim. □

Remark 36 (Boundary of the independent adaptive lower bound). The upper bound (58) holds for arbitrary gap vectors, whereas Corollary 35 gives a matching lower bound only on equal or comparable-gap bands. In the remote construction, a tail component of mass π and scale s_{tail} shifted by b_j gives $\Delta_j \asymp \pi b_j$ and, writing κ_j for the corresponding per-arm KL divergence, $\kappa_j \lesssim \pi b_j^2 / s_{\mathrm{tail}}^2$. Choosing these parameters so that both moment bounds are attained up to constants gives $\pi \asymp s_{\mathrm{tail}}^{-p}$ and $b_j \asymp s_{\mathrm{tail}}^{p/r}$ for $r < \infty$, with $b_j \asymp 1$ for $r = \infty$. Hence $\kappa_j \lesssim \Delta_j^{1/\gamma}$ at essentially one gap scale. Using a separate tail component for each candidate, with mass π_j and scale $s_{\mathrm{tail},j}$ satisfying $\pi_j s_{\mathrm{tail},j}^p \asymp 1$, would contribute $\sum_j \pi_j s_{\mathrm{tail},j}^p \asymp K$ to the common reference moment, violating a fixed class constant. A matching lower bound for arbitrary gap vectors would require a reference construction with tail components at several scales or a different information argument.

K Two-Fold Estimators and Fixed Threshold Multipliers

The following corollaries cover symmetric two-fold estimation and winsorization with a fixed threshold multiplier. These implementations preserve the minimax exponents. For $n \geq 2$, let $m = \lfloor n/2 \rfloor$ and form two deterministic folds $\mathcal{I}_1$ and $\mathcal{I}_2$, each of size m, discarding one observation per model when n is odd.

Corollary 37 (Symmetric two-fold synchronized evaluation)**.** *Let $\widehat{\theta}_j^{1\to 2}$ be the synchronized estimator from Appendix D, with RU minimizers fitted on $\mathcal{I}_1$, scores evaluated on $\mathcal{I}_2$, and clipping level $H_K = \sqrt{V_{\mathrm{sync}} m / \log(2K)}$. Define $\widehat{\theta}_j^{2\to 1}$ by interchanging the two folds, and set $\widehat{\theta}_j^{\mathrm{sym}} = (\widehat{\theta}_j^{1\to 2} + \widehat{\theta}_j^{2\to 1})/2$. There are constants $0 < c, C < \infty$, depending only on the fixed class constants, such that, for $n \geq c \log(2K)$,*

$$\sup_{P \in \mathcal{P}_{K,p,r}} \mathbb{E}_P \max_{j \leq K} |\widehat{\theta}_j^{\mathrm{sym}} - \theta_j(P)| \leq C \sqrt{\frac{\log(2K)}{n}}. \tag{64}$$

Its selector has expected simple regret at most $C\sqrt{\log(2K)/n}$. For $\delta \in (0, 1/2)$ and $n \geq c \log(4K/\delta)$, replacing $\log(2K)$ in the clipping threshold by $\log(4K/\delta)$ gives the deviation bound (35) for $(\widehat{\theta}_j^{\mathrm{sym}})_{j \in [K]}$, with logarithmic factor $\log(4K/\delta)$.

Proof. Each direction obeys Theorem 8. The triangle inequality controls their average in expectation, and allocating failure probability $\delta/2$ to each direction gives the deviation claim. Each fitting fold remains independent of its evaluation fold. □

Corollary 38 (Fixed deterministic winsorization multiplier)**.** *Fix $\lambda > 0$ independently of n, K, the data, and the selected arm. Set $T_{\lambda,K} := 2Q + 1 + \lambda\{m/\log(2(K+1))\}^{1/d}$. Let $\widehat{\theta}_j^{\lambda,1\to 2}$ denote the estimator in (36), with RU minimizers fitted on $\mathcal{I}_1$, scores evaluated on $\mathcal{I}_2$, and threshold $T_{\lambda,K}$. Define $\widehat{\theta}_j^{\lambda,2\to 1}$ analogously and set*

$$\widehat{\theta}_j^{\lambda,\mathrm{sym}} := \frac{\widehat{\theta}_j^{\lambda,1\to 2} + \widehat{\theta}_j^{\lambda,2\to 1}}{2}.$$

There is a constant $C_\lambda < \infty$, depending only on λ and the fixed class constants, such that the following bound holds for either $\widehat{\theta}_j^{\lambda,1\to 2}$ or $\widehat{\theta}_j^{\lambda,\mathrm{sym}}$; write $\widehat{\theta}_j^{\lambda}$ for the selected version:

$$\sup_{P \in \mathcal{P}_{K,p,r}} \mathbb{E}_P \max_{j \leq K} |\widehat{\theta}_j^{\lambda} - \theta_j(P)| \leq C_\lambda \left\{ \frac{\log(2K)}{n} \right\}^{\gamma(p,r)}. \tag{65}$$

For $\delta \in (0, 1/2)$, replace $\log(2(K+1))$ in $T_{\lambda,K}$ by $\log(4(K+1)/\delta)$. Then, for either estimator version and every $P \in \mathcal{P}_{K,p,r}$,

$$\mathbb{P}_P\left(\max_{j\le K}|\widehat{\theta}_j^{\lambda} - \theta_j(P)| > C_\lambda\left\{\frac{\log(4(K+1)/\delta)}{n}\right\}^{\gamma(p,r)}\right) \le \delta.$$

Proof. Put $x = \log(2(K+1))/m$. The bound on estimation error (44) is $C\{T^{-a} + T^{1-p/2}x^{1/2} + Tx + x^{1/2}\}$. For $x \le 1$, the stated threshold is within fixed λ-dependent multiples of $x^{-1/d}$; the first two terms are $O_\lambda(x^\gamma)$, while $1-1/d-\gamma = (1-p/r)/d > 0$ and $\gamma < 1/2$ make the remaining terms no larger. If $x > 1$, the threshold, winsorized scores, and targets are bounded by fixed class- and λ-dependent constants, which are absorbed by $x^\gamma \ge 1$. The expected bound for the symmetric version follows by the triangle inequality. For the high-probability claim, put $x_\delta = \log(4(K+1)/\delta)/m$, and let c_{ind} denote the constant c in Theorem 12. When $x_\delta \le 2/c_{\mathrm{ind}}$, repeat its proof with the modified threshold. The Bernstein, RU-excess, and relative truncation-bias bounds give the same rate, with constants that may also depend on λ. Apply these bounds with error probability $\delta/2$ in each fold direction and take a union bound. Otherwise, the same deterministic bound is absorbed by x_δ^γ, after enlarging C_λ. □

L Controlled Experiments

L.1 Scaling with Bounded Contrasts

Families and estimators. We use $K = 8$, $p = 1.4$, $r = \infty$, and $\alpha = 0.95$. Synchronized evaluation uses the noisy-contrast construction in Appendix F. Independent evaluation uses its remote-tail construction, except that

$$\pi = c_\pi s_{\mathrm{tail}}^{-p}, \qquad c_\pi = (1-\alpha)^{(p+2)/2} \approx 0.00614114.$$

Here c_π is fixed across gaps and budgets; $A = 4s_{\mathrm{tail}}$ and $b = 1$, with s_{tail} chosen to match each population ES gap. The independent marginal pth-moment bound is approximately 0.904782, below the common bound 2.4299107478 used for both protocols; both also satisfy $B = 1$. We use the symmetric two-fold estimators in Corollaries 37 and 38, with the independent multiplier fixed at $\lambda = 2$ as described in Appendix M. For each fold of size $m = n/2$, fitted quantiles are restricted to $[-Q, Q]$, with $Q \approx 17.52245$. The synchronized score-clipping and independent upper-winsorization thresholds are

$$H = \sqrt{\frac{V_{\mathrm{sync}} m}{\log(2K)}}, \qquad T = 2Q + 1 + 2\left\{\frac{m}{\log(2(K+1))}\right\}^{1/(p+2)},$$

where $V_{\mathrm{sync}} \approx 1.08402 \times 10^6$ is the fixed score second-moment bound used by the implementation. Each estimator averages the two directions of fitting and evaluation. Budgets count all $N = (K+1)n$ model evaluations, including the reference.

Experimental design. The six gaps are $\{0.015, 0.020, 0.025, 0.030, 0.0375, 0.045\}$. An independent pilot locates five budgets per protocol and gap by inverting its fitted PCS curve at levels 0.78, 0.86, 0.90, 0.94, and 0.98. The pilot selects 8,000 replications per budget as the smallest count in $\{1000, 2000, 4000, 6000, 8000\}$ with predicted slope-difference interval halfwidths at most 0.30 over all six gaps and 0.60 over each four-gap subset. We fix the budgets, random seeds, and analysis before running all 60 combinations, giving 480,000 replications in total. The pilot observations are excluded from the analysis. The primary target is PCS 0.90; PCS 0.95 is secondary. The two prespecified subsets contain the smallest four and largest four gaps. The supplementary configuration file records the full-precision constants, family parameters, budgets, and random seeds.

Target budgets and slopes. For target q_{PCS}, we use weighted isotonic regression to fit PCS as a nondecreasing function of budget, with replication counts as weights. Linear interpolation of the fitted PCS values in $\log N$ gives the target budget $N_{q_{\mathrm{PCS}}}(\Delta)$. We regress $\log N_{q_{\mathrm{PCS}}}(\Delta)$ on $\log(1/\Delta)$ over each specified set of gaps.

Confidence intervals. Conditional on the pilot, all budgets and isotonic weights are fixed. We construct one-sided Clopper–Pearson bounds (Clopper & Pearson, 1934), allocating $0.05/(2 \cdot 60)$ to each tail across all 60 combinations. We apply the same isotonic fit and interpolation to the lower and upper PCS bounds. Both operations preserve pointwise order, so the resulting curves enclose the interpolant of the isotonic fit to the true grid PCS values. The crossing of the upper curve gives the lower target-budget bound; the crossing of the lower curve gives the upper bound. All target crossings lie within the tested budget ranges. To obtain slope bounds, we minimize and maximize the least-squares slope over the intervals for the log target budgets. The slope is linear in these ordinates, so each extremum chooses an endpoint according to the sign of $x_i - \bar{x}$, where $\Delta^{(i)}$ is the ith selected gap, $x_i = \log(1/\Delta^{(i)})$, and $\bar{x}$ is the mean of these values. For slope intervals $[b_{\mathsf{e}}^-, b_{\mathsf{e}}^+]$, $\mathsf{e} \in \{\mathrm{sync}, \mathrm{ind}\}$, the difference interval is

$$[b_{\mathrm{ind}}^- - b_{\mathrm{sync}}^+,\ b_{\mathrm{ind}}^+ - b_{\mathrm{sync}}^-].$$

These intervals cover the slopes of the isotonic-interpolated true PCS curves on the fixed budget grids. Both PCS targets and all three sets of gaps share the same simultaneous event, giving at least 95% joint coverage conditional on the pilot, and hence unconditionally.

Simulation. To reduce simulation cost, we sample mixture-component counts and the tail observations needed to compute the RU scores. Low observations with zero positive-part contributions to the RU scores are represented by their counts; the full samples are generated when the quantile conditions needed for this reduction fail. Under synchronization, the reduction is used only after checking that the clipping threshold exceeds a deterministic bound on every paired score difference. This preserves the distribution of the selection rule, including quantile estimation and tie-breaking.

Results. Table 2 reports both slopes and their difference for each prespecified set of gaps and PCS target. The difference intervals are positive throughout.

Table 2: Fitted slopes and independent-minus-synchronized differences for $r = \infty$. Brackets give simultaneous 95% intervals derived from the PCS confidence bands.

PCS	Gaps used	Synchronized	Independent	Difference
0.90	All six	1.969 [1.836, 2.103]	2.636 [2.509, 2.769]	0.667 [0.405, 0.933]
0.90	Smallest four	1.916 [1.690, 2.131]	2.578 [2.398, 2.762]	0.663 [0.267, 1.072]
0.90	Largest four	2.025 [1.756, 2.297]	2.694 [2.474, 2.918]	0.669 [0.177, 1.162]
0.95	All six	2.003 [1.827, 2.154]	2.639 [2.523, 2.769]	0.636 [0.369, 0.942]
0.95	Smallest four	1.851 [1.614, 2.077]	2.574 [2.394, 2.783]	0.724 [0.317, 1.169]
0.95	Largest four	2.191 [1.922, 2.413]	2.736 [2.518, 2.951]	0.546 [0.106, 1.029]

Using all six gaps, the independent slope interval $[2.509, 2.769]$ lies above the asymptotic exponent $17/7$. The ratio of $N_{0.90}(\Delta)\Delta^{17/7}$ at the smallest gap to its value at the largest gap is $1.260\ [1.123, 1.415]$. For synchronization, the corresponding ratio for $N_{0.90}(\Delta)\Delta^2$ is $0.958\ [0.857, 1.081]$. Figure 4 shows how the normalized budgets vary across the tested gaps.

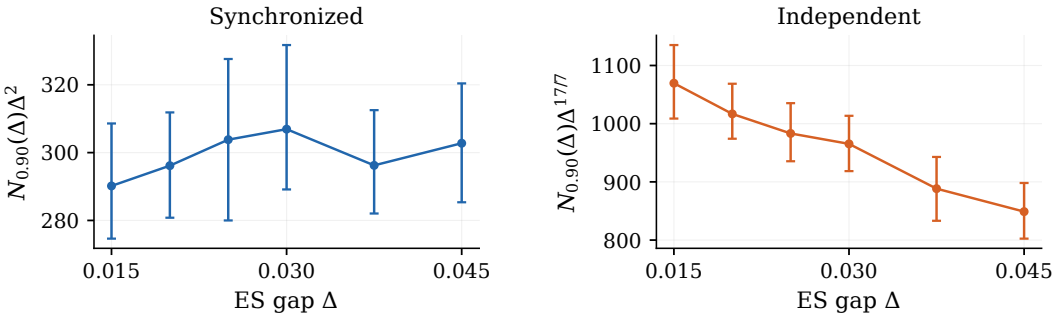

Figure 4: Target budgets for $r = \infty$, normalized by the theoretical exponents, 2 for synchronized sampling and $17/7$ for independent sampling. Error bars are 95% intervals obtained from the same simultaneous PCS bands as Table 2.

L.2 Scaling with a Finite Contrast Moment

Design and intervals. We also examine $r = 2$, keeping $K = 8$, $p = 1.4$, and $\alpha = 0.95$. The synchronized family is unchanged. The independent construction is as above, but with $c_\pi = (1-\alpha)^{2/(2-p/2)} \approx 0.00996358$ and $b = s_{\text{tail}}^{p/2}$. Its marginal pth-moment bound is approximately 0.944153, and its contrast L^2 norm is $\sqrt{c_\pi} < 1$. Both protocols therefore satisfy the common marginal moment bound 2.4299107478 and contrast bound $B = 1$. We use the same symmetric two-fold estimation rules, with thresholds evaluated at $r = 2$. The six gaps are $\{0.035, 0.0425, 0.050, 0.060, 0.0725, 0.0875\}$. Separate pilot simulations determine six budgets per protocol and gap, fixed before 1,000 replications are run at each of the 72 combinations. Target-budget point estimates use the isotonic fit and log-budget interpolation described above. For confidence intervals, we allocate $0.05/(2 \cdot 72)$ to each tail of the Clopper–Pearson bounds across all combinations. After isotonic fitting, we bracket each target budget by the last tested budget where the upper PCS bound is below the target and the first where the lower bound reaches it. Thus these intervals use tested budget endpoints rather than interpolated crossings of the confidence bounds. We obtain slope ranges by the same endpoint optimization as above. All 24 target-budget intervals lie within the tested range, and the slope ranges have at least 95% joint coverage.

Table 3: Scaling experiment for $r = 2$: fitted slopes of $\log N_{q_{\text{PCS}}}(\Delta)$ against $\log(1/\Delta)$. Brackets give ranges obtained from simultaneous 95% PCS bands using tested budget endpoints.

PCS target	Synchronized [range]	Independent [range]
0.90	1.856 [0.732, 2.787]	3.061 [2.299, 3.832]
0.95	1.911 [0.820, 2.875]	3.088 [2.071, 4.195]

The ranges in Table 3 contain the predicted budget exponents 2 and $20/7$. The independent ranges exclude 2, but the two protocols' ranges overlap. Figure 5 shows the target budgets.

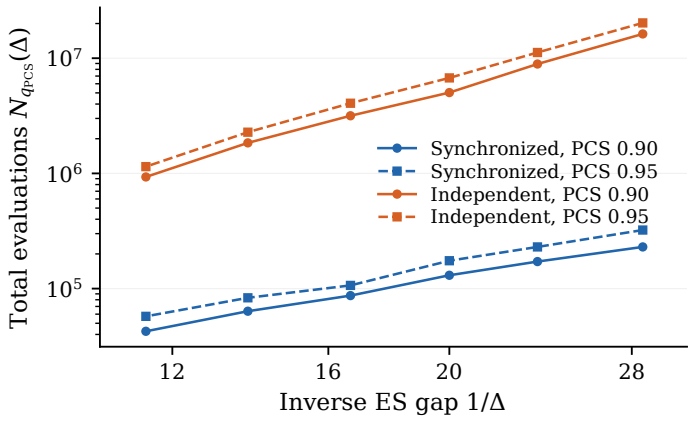

Figure 5: Target budgets for $r = 2$ at PCS 0.90 and 0.95 on each protocol's hard family. Lines connect estimates across the six gaps for visual guidance.

L.3 Sampling Comparisons on Fixed Distributions

Purpose and construction. We compare synchronized and independent sampling on the same loss distribution. We use the remote-tail joint law P_v^{rem} from Appendix F, with $p = 1.4$, $r = \infty$, $K = 8$, ES level $\alpha = 0.95$, and common suboptimality gap $\Delta = 0.03$. The exact gap equation gives remote scale $s_{\text{tail}} = 104.0080$, mixture probability $\pi = 0.00150005$, remote location $A = 4s_{\text{tail}} = 416.0322$, and shift $b = 1$. We take $v = 4$ and break empirical-ES ties by the smallest candidate index.

In each replication, synchronized evaluation observes n complete loss vectors from P_v^{rem}. Independent evaluation draws the reference and all candidate marginals independently from the corresponding

product projection. The reference sample is shared across the two protocol outcomes, but remains independent of every candidate sample within independent evaluation; this coupling supports paired inference without changing the product law. Both protocols use the same empirical ES selector, sample size per model, total cost $(K+1)n$, population ES differences, and best candidate.

Budget selection. A pilot with independent random seeds used 200 replications at each candidate budget. We retained budgets with estimated PCS in $(0.05, 0.95)$ under both protocols and in $[0.65, 0.90]$ under synchronization. Among these budgets, we selected the one whose synchronized PCS was closest to 0.80, breaking ties toward the smaller budget. This fixed $n = 1024$ before the evaluation outcomes were generated. The archived protocol records the full pilot grid.

Evaluation results. The evaluation uses 1,000 new replications at $n = 1024$. Table 4 reports PCS and maximum estimation error. The intervals jointly resample protocol outcomes within each replication using 10,000 paired bootstrap draws. Since all positive gaps equal 0.03, mean simple regret is $0.03(1 - \mathrm{PCS})$.

Table 4: Selection on a fixed distribution. Mean maximum error averages $\max_{j\in[K]} |\widehat{\theta}_j - \theta_j|$ over replications, where $\widehat{\theta}_j = \widehat{\rho}_j^{\mathrm{emp}} - \widehat{\rho}_0^{\mathrm{emp}}$ and $\theta_j = \rho_j - \rho_0$. Differences are synchronized minus independent for PCS and independent minus synchronized for error. Brackets are 95% paired percentile-bootstrap intervals.

Outcome	Synchronized	Independent	Difference [95% CI]
PCS	0.786	0.134	0.652 [0.619, 0.685]
Mean max. error	0.0193	23.0791	23.0598 [22.5242, 23.6047]

Sampling the rare tail component. Synchronized selection is correct in all 786 replications containing a rare-tail observation. In the remaining 214, all empirical ES estimates tie and the fixed tie rule misses the best candidate. This is the identification mechanism in Corollary 18.

Random loss differences. We next examine a distribution in which candidate–reference loss differences remain bounded, but one tail observation does not determine the best candidate. Let $I \sim \mathrm{Bernoulli}(\pi)$ and $Z \sim \mathsf{N}(0,1)$ be independent, with $\pi = s_{\mathrm{tail}}^{-p}$, and set

$$H = \begin{cases} Z, & I = 0, \\ s_{\mathrm{tail}}(4+Z), & I = 1, \end{cases} \qquad L_0 = H, \qquad L_j = H + U_j.$$

Conditional on I, the residuals are independent across candidates and independent of Z. For $I = 0$, all U_j are uniform on $[-1, 1]$. For $I = 1$, ordinary candidates remain uniform, while the best candidate v has density

$$f_v(u) = \tfrac{1}{2}(1 - \kappa_{\mathrm{res}} u)\mathbf{1}_{[-1,1]}(u), \qquad \kappa_{\mathrm{res}} = \tfrac{1}{2}.$$

Thus $|L_j - L_0| \le 1$, and $\mathbb{E}|H|^p \le \mathbb{E}|Z|^p + \mathbb{E}|4+Z|^p$ gives a reference moment bound independent of $s_{\mathrm{tail}} \ge 1$. The tilted residual is stochastically smaller, but all hypotheses have the same support and strictly positive residual densities. Even if the mixture labels and residuals are revealed, a sample with exactly one tail observation assigns each candidate posterior probability at most $(1+\kappa_{\mathrm{res}})/(1 + \kappa_{\mathrm{res}} + (K-1)(1-\kappa_{\mathrm{res}})) = 0.3$ under a uniform prior over the best candidate.

We retain $K = 8$, $p = 1.4$, $\alpha = 0.95$, and $\Delta = 0.03$. Deterministic integration of the marginal RU objectives, checked by a separate adaptive quadrature calculation, gives $s_{\mathrm{tail}} = 28.9229237850$ and $\pi = 0.00900038173319$. Both protocols use this new law's marginals, the empirical ES selector, and $(K+1)n$ evaluations including the reference. For paired comparisons, they share base residual uniforms and a reference stream; candidate marginals remain mutually independent and independent of the reference within the independent protocol. Before generating outcomes, we fixed 1,000 replications and $n \in \{64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768\}$, using nested prefixes within each replication. The main comparison uses $n = 1024$, matching the preceding experiment. Intervals use 2,000 paired bootstrap resamples of whole replications.

Figure 6 shows the complete budget curve. At $n = 1024$, PCS is 0.230 under synchronization and 0.135 under independent sampling, giving the mean-regret reduction 0.00285 $[0.00180, 0.00387]$ reported in the main text. At $n = 64$ and $n = 256$, the regret-difference intervals include zero. At the specified diagnostic budget $n = 128$, only 54 of the 358 replications containing exactly one tail observation select the best candidate: PCS 0.1508 with Wilson 95% interval $[0.1175, 0.1916]$. Mixture labels are used only for this diagnostic.

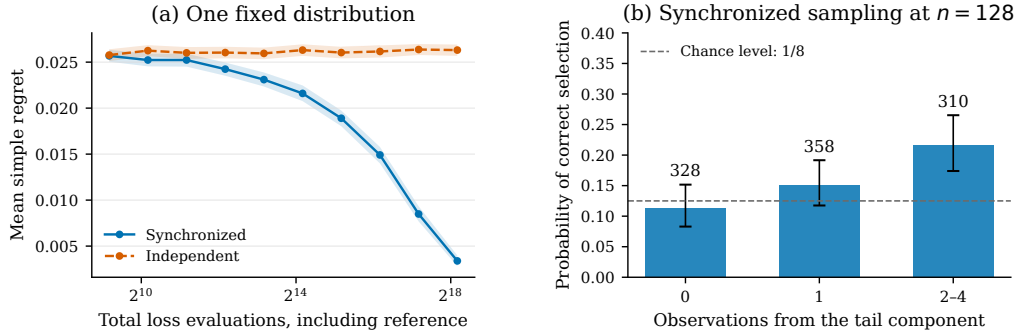

Figure 6: Selection with bounded random loss differences on one fixed distribution. **(a)** Mean simple regret against the total number of loss evaluations, including the reference; shaded bands are pointwise 95% percentile-bootstrap intervals. **(b)** Synchronized PCS at $n = 128$, conditional on the number of observations from the tail component. Numbers above bars are replication counts, and error bars are Wilson 95% intervals. The displayed groups contain 996 replications; the remaining four contain at least five tail observations and are retained in the supplementary table.

M Estimator Comparisons and Robustness Checks

Design. We compare alternative estimators separately under synchronized and independent sampling. For each protocol, we use its primary $r = \infty$ family at gaps 0.015, 0.030, and 0.045. At each gap, the three evaluation budgets are taken from the lower endpoint, point estimate, and upper endpoint of the preliminary budget estimate for 90% correct selection. Each of the resulting 18 settings receives 300 fresh replications.

Within each protocol, all estimators receive the same total model-evaluation budget.

Estimators. The synchronized rule uses the symmetric two-fold estimator in Corollary 37. The independent rule uses the symmetric two-fold marginal estimator with a common upper winsorization threshold from Corollary 38. On a separate exploratory grid, eligible multipliers retained at least 90% of the population gap in every prespecified setting. For each eligible multiplier, we computed the mean maximum estimation error in each setting and minimized the largest of these values, breaking ties toward the smaller multiplier. This selected $\lambda = 2$, held fixed across gaps, protocols, and benchmarks. Symmetric SySRs–RU applies the SySRs successive-rejection phase schedule (Lyu et al., 2026) to the same symmetric RU score matrix, while symmetric covariance–RU applies covariance-adaptive elimination (Saad et al., 2023) to that matrix. The truncated empirical CVaR baseline uses all n observations per candidate, clips losses to $[-n^{0.4}, n^{0.4}]$, and ranks candidates by empirical ES (Kagrecha et al., 2022). The remaining baseline adapts the median CVaR estimator of Kagrecha et al. (2022): it upper-winsorizes each marginal sample, splits it into 48 blocks of nearly equal size, and takes the median of the empirical ES estimates computed within the blocks. For the independent Gaussian-mixture family, we use the marginal moment bound

$$M_{\mathrm{b}} := \mathbb{E}|Z|^p + 2^{p-1}\pi\big(A^p + s_{\mathrm{tail}}^p \mathbb{E}|Z|^p\big), \qquad Z \sim \mathsf{N}(0,1).$$

With n observations per model and $\delta = 0.05$, the winsorization threshold is

$$T = \frac{3}{2} + \left(\frac{M_{\mathrm{b}}}{\min\{\alpha, \tau\}}\right)^{1/p} + \left(\frac{M_{\mathrm{b}} n}{\log(2(K+1)/\delta)}\right)^{1/p}.$$

The block count is $\lceil 8 \log(2(K+1)/\delta) \rceil = 48$ for $K = 8$.

Each comparison pairs the proposed estimator with a baseline on the same replications. For each setting, we report 95% percentile-bootstrap intervals for differences in correct-selection probability, mean maximum estimation error, and mean simple regret, using 2,000 paired resamples. For averages across settings, we give equal weight to the nine settings within each protocol and use 3,000 bootstrap resamples, resampling paired replications separately within each setting. These baseline intervals are not adjusted for multiple comparisons.

Synchronized comparisons. Relative to symmetric SySRs–RU, the synchronized estimator's average PCS difference is 0.0004, with a 95% interval of $[-0.0015, 0.0022]$. The comparison did not detect a difference. Relative to symmetric covariance–RU, the average PCS difference is 0.0207, with a 95% interval of $[0.0130, 0.0293]$.

Independent comparisons. On this grid, the independent estimator and the truncated empirical CVaR baseline made identical choices in all 2,700 replications.

Relative to the median of blockwise ES estimates, the independent estimator's average PCS difference is 0.1063, with a 95% interval of $[0.0926, 0.1204]$.

M.1 Clipping and Winsorization

The scaling and baseline experiments above recorded no clipping or winsorization. To examine active winsorization under the bounded contrast assumption, we use

$$L_j = C + \mu_j + U_j, \qquad j = 0, \dots, K.$$

Here C has a Pareto distribution with tail index 1.6 and scale 1. The variables U_j are independent and uniformly distributed on $[-1, 1]$, independently of C. For each replication, we draw the best candidate $j^\star$ uniformly from $[K]$, set $\mu_{j^\star} = -\Delta$, and set all other offsets to zero, including μ_0. The population ES gap is therefore exactly Δ. Synchronized sampling shares C across models, whereas independent sampling draws each model's marginal separately.

We fix $K = 8$, $p = 1.4$, $r = \infty$, and $\alpha = 0.95$. For both gaps, valid common bounds are

$$M_0 = \left\{ 8^{1/p} + (p+1)^{-1/p} \right\}^p, \qquad B = 2.6,$$

with $\overline{M}$ given by (6). We use

$$Q = \left(\frac{\overline{M}}{\min\{\alpha, \tau\}} \right)^{1/p} + 1.5 \approx 65.670723, \qquad V_{\mathrm{sync}} = 2(B/\tau)^2 + 8Q^2(1 + \tau^{-1})^2.$$

The symmetric two-fold estimators use these constants, with the independent threshold multiplier fixed at $\lambda = 2$. Within each protocol, the two versions differ only in truncation, sharing observations, fitted quantiles, folds, and tie-breaking.

We use gaps $\Delta \in \{0.3, 0.6\}$ and sample sizes $n \in \{25{,}000, 50{,}000, 100{,}000\}$ per model, giving six settings per protocol. With gaps and sample sizes fixed, a separate pilot selects 1,500 replications per setting for precision; the main experiment uses fresh seeds and excludes pilot data.

For each independent setting, we count replications in which winsorization selects correctly but the untruncated version does not, and replications with the reverse outcome. We construct Clopper–Pearson bounds for these two probabilities, allocating error $0.05/(4 \cdot 6)$ to each tail. Subtracting the appropriate endpoints gives intervals for the six PCS differences with at least 95% simultaneous coverage.

Winsorization occurs in every independent replication, changing 0.0276%–0.0310% of losses across settings. PCS gains range from 0.0453 to 0.3213, and all six simultaneous intervals lie above zero. The supplement reports the results for every setting. At $\Delta = 0.6$ and $n = 100{,}000$, PCS rises from 0.4000 to 0.7213 at the same evaluation cost. Mean simple regret, measured against the original population ES, decreases by 0.1928, with a simultaneous 95% interval of $[0.1628, 0.2213]$. No synchronized scores are clipped, and both synchronized versions select correctly in all recorded replications.

N SOA Medical Insurance Claims Case Study

Data and task. The Society of Actuaries Group Medical Insurance Database contains annual claimant records from seven insurers (Society of Actuaries, 2004). We use 1,241,438 records from 1997 for training, 1,460,854 from 1998 for development, and 1,591,738 from 1999 for the held-out comparison. For the original with-replacement study, the model pool and evaluation protocol were specified before the 1999 file was accessed. The outcome is annual total paid medical charges, and the predictors are age, sex, relationship, preferred provider organization status, and exposure. Claimant identifiers, diagnoses, charge subtotals, and all other outcome components are excluded. The statistical unit is a claimant-year, not a complete insured population. Pooling across insurers may introduce dependence between records.

Loss construction. Seven histogram gradient boosting models, implemented with scikit-learn's `HistGradientBoostingRegressor` version 1.5.2 (Pedregosa et al., 2011), predict the median; the reference is the 1997 empirical median. Table 6(a) lists the seven configurations, all trained on 1997 data. Predictions are clipped to the finite interval from zero to the largest 1997 outcome, while outcomes are never clipped. The conditional median is the canonical point forecast under absolute-error loss, and the prediction range defined by the training data supplies the deterministic bound on loss differences below. Write Y for annual paid charges, X for the predictor vector, and $\widehat{m}_j(X)$ for forecast j. For the absolute-error loss $L_j = |Y - \widehat{m}_j(X)|$, the reverse-triangle inequality gives

$$|L_j - L_0| \le |\widehat{m}_j(X) - \widehat{m}_0(X)|, \tag{66}$$

so the fixed model pool has deterministically bounded contrasts. Selection minimizes 0.95 ES across the seven challengers.

Development and held-out evaluation. The 1998 pilot used 200 replications at each candidate budget. Among budgets with synchronized PCS in $[0.20, 0.95]$ and independent PCS in $[0.05, 0.95]$, the selection rule chose the budget whose synchronized PCS was closest to 0.75, breaking ties toward smaller n. This fixed $n = 8{,}192$ before the 1999 file was accessed. The held-out comparison then used 1,000 replications on the 1999 finite population. Synchronized sampling draws complete loss rows with replacement; independent sampling draws each marginal column independently with replacement. The population, marginal laws, finite-population ES differences, empirical ES selector, number of observations per model, and total number of sampled loss observations are fixed across protocols.

Table 5 reports the held-out results. Confidence intervals jointly resample the two protocol outcomes by replication ID using 10,000 percentile-bootstrap draws.

Table 5: Held-out SOA 1999 results. Mean maximum error averages $\max_{j\in[K]} |\widehat{\theta}_j - \theta_j|$ over replications, where $\widehat{\theta}_j = \widehat{\rho}_j^{\mathrm{emp}} - \widehat{\rho}_0^{\mathrm{emp}}$ and $\theta_j = \rho_j - \rho_0$. Differences are synchronized minus independent for PCS and independent minus synchronized for error and regret. Brackets are 95% paired percentile-bootstrap intervals. Errors and regret are in nominal dollars.

Outcome	Synchronized	Independent	Difference [95% CI]
PCS	0.831	0.144	$0.687\ [0.655, 0.718]$
Mean max. error	9.328	3787.222	$3777.894\ [3685.715, 3870.681]$
Mean simple regret	0.126	4.227	$4.101\ [3.792, 4.416]$

Risk gaps and score variability. We compute candidate ES values, gaps, and RU score variances under the uniform empirical distribution of the 1999 loss rows. Let q_j be its lower 0.95 quantile for arm j and set $D_j = \ell(L_j, q_j) - \ell(L_0, q_0)$. The reported variance ratio is

$$\frac{\mathrm{Var}(D_j)}{\mathrm{Var}(\ell(L_j, q_j)) + \mathrm{Var}(\ell(L_0, q_0))}.$$

These full-population quantiles are used only for this diagnostic; the selection procedure uses quantiles fitted from its sampled losses. Table 6(b) reports the results.

Table 6: SOA model configurations and diagnostics on the 1999 empirical population. HGB denotes histogram gradient boosting. Panel (a) gives the seven fixed parameter combinations; panel (b) retains their names in the archived protocol and reports the diagnostics. ES and gaps are in nominal dollars; the reference ES is 18,361.314.

(a) Model parameters

Configuration	Maximum iterations	Learning rate	Maximum leaf nodes	Minimum samples per leaf	ℓ_2 regularization
HGB 1	80	0.08	7	2,000	50
HGB 2	120	0.06	15	1,000	20
HGB 3	160	0.05	31	500	10
HGB 4	200	0.04	63	250	3
HGB 5	140	0.05	31	2,000	30
HGB 6	180	0.04	31	250	0.3
HGB 7	220	0.03	63	500	10

(b) Risk gaps and score variability

Configuration	Archived name	ES	Gap above best	RU variance ratio
HGB 1	`hgb_coarse`	18,209.245	16.142	1.266×10^{-5}
HGB 2	`hgb_shallow`	18,200.102	6.999	1.523×10^{-5}
HGB 3	`hgb_balanced`	18,195.012	1.909	1.660×10^{-5}
HGB 4	`hgb_fine`	18,193.103	0.000	1.760×10^{-5}
HGB 5	`hgb_sparse`	18,195.842	2.739	1.648×10^{-5}
HGB 6	`hgb_low_reg`	18,195.849	2.746	1.628×10^{-5}
HGB 7	`hgb_deep_reg`	18,193.633	0.530	1.734×10^{-5}

O Language Model Benchmarks

O.1 Benchmark Construction and Eligibility

The benchmark population was fixed from metadata before option likelihoods or aggregate scores were inspected. Eligible entries had public weights, a value of `True` in the archived leaderboard's `Official Providers` metadata field, an instruction or chat interface, 7–9 billion parameters, bfloat16 evaluation records, and complete item-level results for both tasks. We retained all 28 eligible models, including the reference. MMLU-Pro is the primary benchmark because its 12,032 aligned items support the full budget grid. ARC-Challenge provides an additional comparison on a smaller benchmark of 1,172 items (Wang et al., 2024; Clark et al., 2018). All evaluations use five-shot prompting, the recorded model-specific chat template, and the same evaluation-harness revision. The archived manifest records the complete model population and evaluated revisions.

The references were chosen without performance information: the primary reference is the earliest released eligible model, with canonical name as the tie-breaker, and the sensitivity reference is the next model under that order. This rule selects `mistralai/Mistral-7B-Instruct-v0.1` and `Open-Orca/Mistral-7B-OpenOrca`. Each reference configuration selects among the remaining 27 models and uses a separate Monte Carlo sampling stream. For fixed candidate samples, subtracting a common reference ES leaves their empirical-ES ranking unchanged.

O.2 Loss, Target, and Sampling Experiments

Let s_{ijc} be the archived log-likelihood assigned by model j to option c on item i, and let y_i denote the correct option. We use

$$L_{ij} = -s_{ijy_i} + \log \sum_c \exp(s_{ijc}), \tag{67}$$

computed by log-sum-exp without a probability floor. For each reference 0, if I is drawn uniformly from the task's finite item population, write $L_j := L_{Ij}$. The target remains $\theta_j = \mathrm{ES}_{0.95}(L_j) - \mathrm{ES}_{0.95}(L_0)$, distinct from $\mathrm{ES}_{0.95}(L_j - L_0)$. For each task and reference configuration, the full aligned matrix defines the finite-population targets and best candidate.

Across these four finite populations, the maximum absolute candidate–reference loss difference ranges from 29.39 to 66.35 nats. At population quantiles, the ratio of the variance of each paired RU score difference to the sum of its two marginal score variances ranges from 0.137 to 0.623 on ARC-Challenge and from 0.486 to 0.912 on MMLU-Pro.

For sampling with replacement, for each per-model sample size $n \in \{64, 128, 256, 512, 1024, 2048\}$, we draw items uniformly with replacement from the complete empirical item distribution. Synchronized evaluation uses the same draws for all 28 models; independent evaluation draws items separately for each model. Both protocols supply $28n$ loss observations to the selector, counting repeated draws separately. This count measures resampling from the archived records. If model–question losses are cached, repeated draws reuse an existing loss and require no additional model call. Every combination of task, reference, protocol, method, and sample size has 1,000 replications.

O.3 Selection Results with Replacement

Figure 7 compares synchronized and independent sampling using the same empirical ES selector.

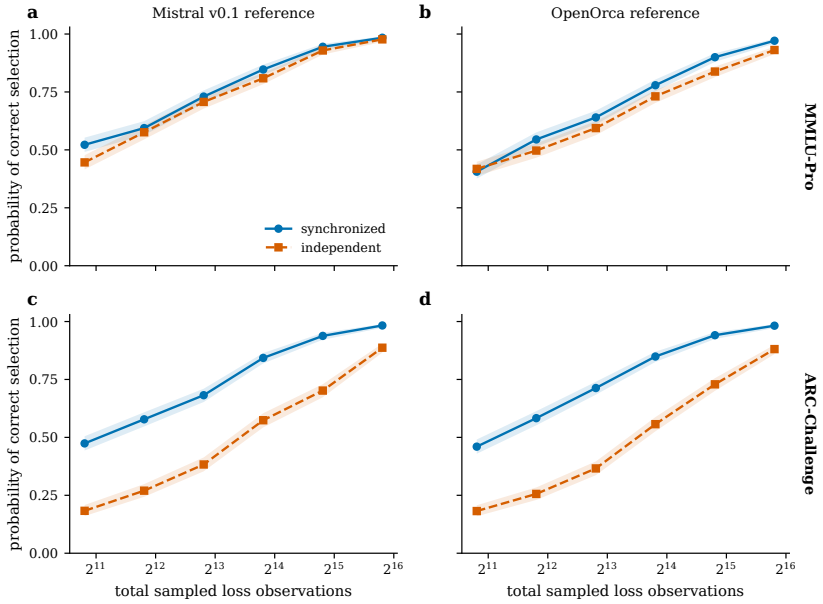

Figure 7: Synchronized versus independent language model selection with replacement and the same empirical-ES rule. Rows are MMLU-Pro and ARC-Challenge; columns correspond to the two reference configurations, each selecting among the remaining 27 models. The horizontal axis counts sampled loss observations across all 28 models, including repeated draws. Shaded regions are pointwise 95% Wilson score intervals from 1,000 replications.

At $n = 2{,}048$ draws per model, replacing empirical ES with the symmetric two-fold RU estimator changes the reported PCS by at most 0.005 in each of the four settings under either sampling protocol. This sensitivity analysis sets its quantile bound and contrast scale from the complete loss matrix before resampling. Its budgets count sampled losses and exclude the cost of obtaining that matrix. The supplementary materials give the full settings, error summaries, and diagnostics of dependence between losses. Appendix P reports sampling without replacement and reference evaluation on existing candidate inputs.

P Sampling without Replacement and Additional Evaluations

We use the same medical and language model loss matrices, candidates, and references as in Appendices N and O, with ES level $\alpha = 0.95$. We fixed the sampling and spending rules before generating these new Monte Carlo replications, after the original analysis of the data. ES values

computed from the entire matrices determine the best candidate and the regret of each selection; the selectors do not receive these values.

Each setting has 1,000 replications. Mean regrets and paired differences use 95% percentile-bootstrap intervals based on 2,000 resamples of replications. Success probabilities use Wilson intervals. All intervals are pointwise.

P.1 Sampling without Replacement

We draw n inputs uniformly without replacement for each candidate. Synchronized sampling uses the same inputs for every candidate; independent sampling draws a separate sample for each candidate. Samples from different candidates may overlap. Both protocols select the candidate with the smallest empirical ES, including the appropriate fraction of the observation at the tail boundary and breaking ties by the smallest original candidate index.

The count is Kn distinct candidate–input evaluations. No reference evaluation is required for this ranking. This count differs from the $(K+1)n$ sampled loss observations used in the original with-replacement comparisons.

The medical sample sizes are $n \in \{512, 1024, 2048, 4096, 8192, 16384, 32768\}$. Both language model tasks use $n \in \{64, 128, 256, 512, 1024\}$ and the full question population; MMLU-Pro additionally uses $n = 2048$. The population sizes are 1,172 for ARC-Challenge and 12,032 for MMLU-Pro.

Table 7 reports paired differences at the sample sizes discussed in the main text. The MMLU-Pro/Mistral interval includes zero. At the full question population, both language model sampling protocols have zero regret.

Table 7: Mean simple regret reductions without replacement. A reduction is independent minus synchronized regret. Medical values are in nominal dollars; language model values are in nats. Mistral and OpenOrca identify the reference excluded from the candidate set. Intervals are pointwise paired 95% bootstrap intervals.

Setting	n	Regret reduction	95% CI
Medical claims	8,192	4.139	$[3.812, 4.481]$
ARC-Challenge / Mistral	512	0.36421	$[0.31189, 0.41451]$
ARC-Challenge / OpenOrca	512	0.40968	$[0.35788, 0.46055]$
MMLU-Pro / Mistral	2,048	0.00037	$[-0.00292, 0.00365]$
MMLU-Pro / OpenOrca	2,048	0.00859	$[0.00396, 0.01346]$

We also report $\mathbb{P}(\rho_{\hat{j}} - \rho_\star \le \varepsilon)$. The medical tolerances are $\varepsilon \in \{0, 0.25, 0.5, 1, 2, 5, 10\}$ nominal dollars, and the language model tolerances are $\varepsilon \in \{0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2\}$ nats. All positive language model gaps exceed 0.2, so success at every tested tolerance is equivalent to correct selection.

P.2 Additional Evaluations on Existing Records

Within each replication, the procedures start from the same n_0 records per candidate, sampled independently across candidates and without replacement within each candidate. These initial samples are generated separately from the preceding comparison without replacement. We report $n_0 = 8192$ for medical claims, $n_0 = 2048$ for MMLU-Pro, and $n_0 = 512$ for ARC-Challenge, and repeat the comparisons with one quarter of each initial sample size.

The initial candidate cost Kn_0 is already incurred. Obtaining a candidate loss on a new input costs one unit, and obtaining a reference loss costs $c_{\mathrm{ref}} \in \{0, 0.05, 0.25, 1\}$ units. We use additional budget caps

$$\mathcal{B} = Kn_0u, \qquad u \in \{0, 1/64, 1/32, 1/16, 1/8, 1/4, 1/2, 1, 2\}.$$

We apply these assumed costs to observations drawn from the stored loss matrices, with inputs and labels already available. Each procedure reuses its own available losses and pays only for model–input pairs it has not already obtained.

Table 8: Additional evaluations at unit candidate and reference costs and cap $\mathcal{B} = Kn_0$. Cost is measured in candidate-evaluation units. Regret is in nominal dollars for medical claims and nats for language models. Costs and regrets are averages over 1,000 replications; regret intervals are pointwise 95% bootstrap intervals.

Additional evaluations	Mean cost	Mean regret [95% CI]
Medical claims: $\mathcal{B} = 57{,}344$		
None	0.0	4.430 [4.109, 4.738]
Independent candidate observations	57,344.0	4.261 [3.947, 4.575]
Reference on existing inputs	56,467.0	1.991 [1.847, 2.137]
Candidates on common inputs	57,341.1	0.154 [0.128, 0.181]
MMLU-Pro / Mistral: $\mathcal{B} = 55{,}296$		
None	0.0	0.00475 [0.00219, 0.00767]
Independent candidate observations	55,296.0	0.00037 [0.00000, 0.00110]
Reference on existing inputs	11,953.7	0.01548 [0.01110, 0.02040]
Candidates on common inputs	55,285.7	0.00073 [0.00000, 0.00183]
MMLU-Pro / OpenOrca: $\mathcal{B} = 55{,}296$		
None	0.0	0.00968 [0.00646, 0.01309]
Independent candidate observations	55,296.0	0.00091 [0.00000, 0.00213]
Reference on existing inputs	11,954.1	0.03830 [0.03203, 0.04522]
Candidates on common inputs	55,285.0	0.00152 [0.00030, 0.00305]
ARC-Challenge / Mistral: $\mathcal{B} = 13{,}824$		
None	0.0	0.47045 [0.42191, 0.52119]
Independent candidate observations	13,824.0	0.00839 [0.00315, 0.01468]
Reference on existing inputs	1,172.0	0.26594 [0.23100, 0.30273]
Candidates on common inputs	13,816.7	0.00315 [0.00000, 0.00734]
ARC-Challenge / OpenOrca: $\mathcal{B} = 13{,}824$		
None	0.0	0.43814 [0.39185, 0.48715]
Independent candidate observations	13,824.0	0.00785 [0.00210, 0.01463]
Reference on existing inputs	1,172.0	0.17904 [0.15252, 0.20768]
Candidates on common inputs	13,816.4	0.00105 [0.00000, 0.00315]

Table 8 compares the procedures under a common spending limit; actual costs may differ.

We compare using the initial records alone with three ways to collect more losses: adding independent candidate observations, evaluating the reference on existing candidate inputs, and evaluating candidates on common inputs. For additional independent sampling, we allocate the budget equally across candidates. Each candidate receives $\lfloor \mathcal{B}/K \rfloor$ new observations, or all remaining inputs if fewer are available. We sample without replacement and select using empirical ES computed from all initial and additional observations.

For reference evaluation, let $\mathcal{I}_j^{(m)}$ contain the first $m \le n_0$ inputs in candidate j's initial random sample. Let $\widehat{\mathrm{ES}}_\alpha$ denote empirical ES. For $m \ge 1$, the selector minimizes

$$\widehat{\theta}_j^{(m)} := \widehat{\mathrm{ES}}_\alpha\big((L_{i,j})_{i \in \mathcal{I}_j^{(m)}}\big) - \widehat{\mathrm{ES}}_\alpha\big((L_{i,0})_{i \in \mathcal{I}_j^{(m)}}\big).$$

Thus it estimates the same target $\theta_j = \rho_j - \rho_0$. The reference is evaluated once per distinct input and reused across candidates, so the additional cost is

$$c_{\mathrm{ref}} \left| \bigcup_{j=1}^{K} \mathcal{I}_j^{(m)} \right|.$$

We choose the largest m whose cost does not exceed $\mathcal{B}$, using input identities and costs, without inspecting loss values. Selection uses only these m paired observations per candidate. The procedure stops after obtaining references for all n_0 initial observations and leaves any remaining budget unused.

To evaluate candidates on common inputs, we draw a new random input sequence without replacement and use the longest part affordable for all candidates. We reuse initial candidate losses when available

and pay for the missing model–input pairs. Selection uses empirical ES on these common inputs; other initial records do not enter the selector. If the budget cannot cover one paired observation per candidate or one common input for all candidates, the corresponding procedure retains the initial selection.

Before simulation, we selected two main comparisons of evaluating the reference against adding independent candidate observations, both at $c_{\mathrm{ref}} = 1$ and $\mathcal{B} = Kn_0$. For medical claims at $n_0 = 8192$, the mean regret reduction is 2.270 $[1.920, 2.617]$. For MMLU-Pro with the Mistral reference at $n_0 = 2048$, it is -0.01512 $[-0.01967, -0.01074]$. Positive reductions favor evaluating the reference.

At small budgets, selection from only the paired observations or common inputs can omit useful initial records and perform worse than the initial selection. Even after the reference is evaluated on all initial inputs, however, regret increases in both reported MMLU-Pro settings.

How reference evaluation changes rankings. We diagnose complete reference evaluation using all initial candidate observations and the original 1,000 replications in each setting. This is a post-hoc analysis of the existing results. Let $\widehat{\rho}_j$ denote candidate j's empirical ES and $\widehat{\rho}_0^{(j)}$ the reference empirical ES on its inputs. For $j \neq j^\star$, define

$$X_j = (\widehat{\rho}_j - \widehat{\rho}_{j^\star}) - \Delta_j, \qquad Y_j = \widehat{\rho}_0^{(j)} - \widehat{\rho}_0^{(j^\star)}.$$

The unadjusted and adjusted gap errors are X_j and $X_j - Y_j$. Equal sample sizes from the same reference population give $\mathbb{E}Y_j = 0$, so adjustment leaves the expected gap bias unchanged. Its effect on mean squared error (MSE) satisfies the exact identity

$$\mathbb{E}[(X_j - Y_j)^2] - \mathbb{E}[X_j^2] = \mathbb{E}[Y_j^2] - 2\mathbb{E}[X_j Y_j]. \tag{68}$$

We compute these quantities for every competitor and summarize the one with the smallest positive population ES gap in Table 9.

We also replace empirical ES by the sample mean of $S_j := \ell(L_j, q_j)$ for $j = 0, \ldots, K$, where q_j is an α-quantile of the full finite population. These quantiles are used only for diagnosis. Because candidate batches are independent and have equal size, the ratio of adjusted to unadjusted gap variances is

$$\frac{\mathrm{Var}(S_j - S_0) + \mathrm{Var}(S_{j^\star} - S_0)}{\mathrm{Var}(S_j) + \mathrm{Var}(S_{j^\star})}.$$

The common finite-population sampling factor cancels. We calculate this ratio from the complete loss matrices; the empirical-ES MSE ratios instead use the 1,000 replications.

Table 9: Ranking changes after evaluating the reference on all initial inputs. Ratios compare adjusted with unadjusted gap error for the closest competitor to the population best. Values below one indicate a reduction. Corrected and disrupted count selections changing from incorrect to correct and from correct to incorrect, respectively, out of 1,000 replications.

Setting	n_0	Gap MSE ratio	Fixed-quantile variance ratio	Corrected	Disrupted
Medical claims	8,192	3.28×10^{-5}	3.50×10^{-5}	192	118
ARC / Mistral	512	0.664	0.726	234	124
ARC / OpenOrca	512	0.535	0.530	232	85
MMLU-Pro / Mistral	2,048	1.287	1.325	9	37
MMLU-Pro / OpenOrca	2,048	1.871	1.934	24	107

In MMLU-Pro, the reference contribution $\mathbb{E}[Y_j^2]$ exceeds the cancellation term $2\mathbb{E}[X_j Y_j]$, and misranking the closest competitor becomes more frequent. The same deterioration occurs with population quantiles fixed, where the variance ratios exceed one. For medical claims, the closest gap is only 0.530 dollars: despite a large MSE reduction, its misranking probability changes from 0.491 to 0.475. Changes between two incorrect selections account for a net 1.795 of the 2.439-dollar mean-regret reduction. The same regret directions persist at the smaller initial sample sizes. The supplementary results contain all settings, competitor comparisons, and selection transitions.