QUANTUM ALGORITHMS FOR RISK-SENSITIVE BANDITS: MATCHING UPPER AND LOWER BOUNDS FOR QUANTUM CVAR ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-armed bandit algorithms decide which option to try next when each choice both yields an outcome and reveals information. Standard methods rank options by average reward, but finance, medicine, and safety-critical control care about lower-tail performance, average performance in the worst α fraction of cases, formalized by the conditional value-at-risk. Quantum algorithms accelerate average-reward bandits, and risk-sensitive bandit theory is well developed classically, but the intersection has been empty. We give the first quantum algorithms for risk-sensitive bandits, with matching lower bounds. Our core primitive estimates conditional value-at-risk using quadratically fewer oracle queries than the analyzed classical sampling baseline, and an amplitude-aware refinement adds a further square-root advantage in the tail level by exploiting the fact that the estimated quantity is intrinsically small deep in the tail. Building on it we obtain best-arm identification that improves quadratically in the arm gap, and regret with no polynomial gap dependence. Adversary-method lower bounds show the estimator is optimal in both parameters.

1 INTRODUCTION

Multi-armed bandits are the canonical model of sequential decision-making under uncertainty, and the overwhelming majority of the literature ranks arms by expected reward. In many of the settings that motivate the model, however, the average is the wrong objective. A portfolio, a dosing policy, or a controller in a safety-critical loop is judged by what happens on its worst days, not on a typical one. Risk-sensitive bandits address this by replacing the mean with a risk measure, most commonly the conditional value-at-risk CVaR_α, the average reward over the worst α-fraction of cases, defined by the Rockafellar-Uryasev formula (1); when the α-quantile carries an atom this is not literally a conditional expectation, since the atom must be allocated fractionally (Rockafellar & Uryasev, 2000; Artzner et al., 1999). The classical theory here is mature (Cassel et al., 2018; Tamkin et al., 2019).

Independently, quantum access to reward distributions is known to accelerate bandit learning. Given a qsample oracle preparing a superposition whose amplitudes encode the reward distribution, amplitude estimation (Brassard et al., 2002) estimates a mean to accuracy ε with $\widetilde{O}(1/\varepsilon)$ queries rather than the classical $\Theta(1/\varepsilon^2)$ samples, and this quadratic advantage propagates into bandit algorithms (Wang et al., 2021; Wan et al., 2023; Wu et al., 2023). Every one of these results optimizes the mean.

The intersection has been empty. To our knowledge no prior work gives quantum algorithms for bandits under a risk objective, and the closest work on quantum CVaR estimation (Skarlatos & Konofaos, 2025) states that a matching lower bound for d-dimensional CVaR *subgradient* estimation remains open. This paper addresses the scalar CVaR-value problem in both directions: we give algorithms, and we prove lower bounds establishing that the core estimation primitive is optimal.

Two features of CVaR make this more than a substitution exercise. First, CVaR_α is not linear in the distribution, so the correctness arguments of classical bandit algorithms must be checked for hidden uses of linearity, and the estimator must expose an interface that survives composition. Second, the quantity to be estimated is intrinsically small: near the α-quantile the expected shortfall has

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.

Table 1: Query and sample complexity for bandit learning, with $H = \sum_{i \neq i^\star} \Delta_i^{-2}$, Δ_i the gap in the relevant objective, and all sums over $i \neq i^\star$ since $\Delta_{i^\star} = 0$. The classical CVaR rows instantiate standard elimination and optimism analyses with Proposition 9; they are stated asymptotically because its universal constants are not pinned down, and they are upper bounds only, with no matching classical lower bound claimed or cited (Appendix B).

Objective	Access	Best-arm identification	Regret
Mean	classical	$\Theta(\sum_{i \neq i^\star} \Delta_i^{-2})$	$\Theta(\sum_{i \neq i^\star} \Delta_i^{-1} \log T)$
Mean	quantum	$\widetilde{O}(\sqrt{H})$ (Wang et al., 2021)	$O(K \log T)$ (Wan et al., 2023)
CVaR_α	classical	$O(\sum_{i \neq i^\star} \alpha^{-1} \Delta_i^{-2})$	$O(\sum_{i \neq i^\star} \alpha^{-1} \Delta_i^{-1} \log T)$
CVaR_α	quantum	$\widetilde{O}(\sum_{i \neq i^\star} \alpha^{-1/2} \Delta_i^{-1})$ **(ours)**	$\widetilde{O}(K \alpha^{-1/2})$ **(ours)**
CVaR_α estimation, classical		$O(\alpha^{-1} \varepsilon^{-2})$ (Proposition 9)	
CVaR_α estimation, quantum		$\widetilde{\Theta}(\alpha^{-1/2} \varepsilon^{-1})$ **(ours, matching)**	

magnitude $O(\alpha)$, and amplitude estimation is cheaper for small amplitudes, which converts one factor of α into $\sqrt{\alpha}$. The same fact drives our lower bound, where a CVaR gap of ε is purchased with a mass shift of only $\alpha\varepsilon$, leaving the oracles nearly indistinguishable.

Contributions.

- A quantum CVaR estimator (Lemma 1) achieving $\widetilde{O}(\alpha^{-1}\varepsilon^{-1})$ queries via noisy ternary search on the concave Rockafellar-Uryasev objective, with explicit constants and no density assumption, and an amplitude-aware refinement (Lemma 2) achieving $\widetilde{O}(\alpha^{-1/2}\varepsilon^{-1})$.
- Bandit algorithms built on this primitive: best-arm identification in $\widetilde{O}(\sum_{i \neq i^\star} \alpha^{-1/2} \Delta_i^{-1})$ queries (Theorem 4), a quadratic-in-gap improvement, and a regret algorithm (Theorem 5) whose bound carries no polynomial gap dependence at all, because the per-arm query cost and the per-query regret cancel exactly.
- Lower bounds via the quantum adversary method: $\Omega(\varepsilon^{-1}\alpha^{-1/2})$ for CVaR estimation (Theorem 6), matching Lemma 2 up to logarithmic factors and, to our knowledge, the first lower bound for scalar CVaR-value estimation; it does not settle the d-dimensional subgradient bound left open by Skarlatos & Konofaos (2025). And $\Omega(\alpha^{-1/2}\sqrt{H})$ for CVaR best-arm identification (Theorem 7).
- An analysis of the remaining $\sqrt{K}$ structural gap (Section 6): closing it requires variable-time amplitude amplification over a *coherent* CVaR estimator, and we show that three natural constructions do not currently supply one.

Table 1 situates the results. We are explicit throughout about a limitation shared with the quantum-query bandit literature: the advantage is asymptotic, and Section 7 quantifies the crossover rather than leaving it implicit.

2 Preliminaries

Risk-sensitive bandits. There are K arms; arm i has an unknown reward distribution ν_i supported on $[0, 1]$. Fix a tail level $\alpha \in (0, 1)$ and write $C_i := \mathrm{CVaR}_\alpha(\nu_i)$. We adopt the lower-tail convention, so CVaR_α is the average of the worst α-fraction of rewards and larger is better, taken throughout in the Rockafellar-Uryasev sense (1), which coincides with a conditional expectation only when F is continuous at the α-quantile. We assume a unique optimum $i^\star := \arg\max_i C_i$ and write $\Delta_i := C_{i^\star} - C_i > 0$ and $H := \sum_{i \neq i^\star} \Delta_i^{-2}$.

The Rockafellar-Uryasev form. All of our algorithms are built on the variational identity of Rockafellar & Uryasev (2000). For $X \sim \nu$ supported on $[0, 1]$,

$$\mathrm{CVaR}_\alpha(\nu) \;=\; \max_{t \in [0,1]} \phi(t), \qquad \phi(t) \;:=\; t - \tfrac{1}{\alpha}\,\mathbb{E}\big[(t - X)_+\big], \tag{1}$$

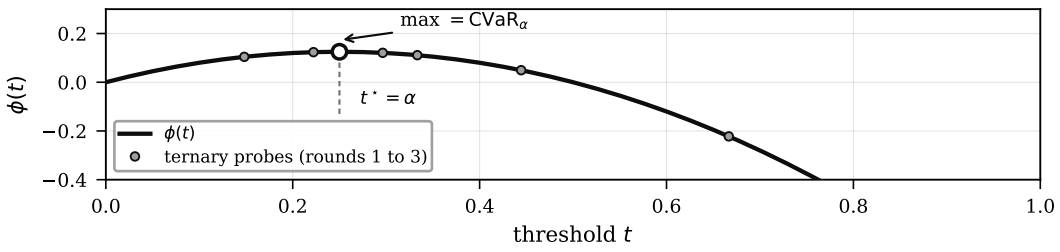

Figure 1: The structural fact the paper turns on. The Rockafellar-Uryasev objective is concave in a single variable, so CVaR estimation reduces to a search whose comparisons are trustworthy; each probe is one amplitude estimation.

with the maximum attained at the α-quantile. Three properties of ϕ are used repeatedly, each proved in Appendix A. (i) ϕ is concave, since $t \mapsto (t - X)_+$ is convex and expectation preserves convexity. (ii) ϕ is $1/\alpha$-Lipschitz, because $\phi'(t) = 1 - F(t)/\alpha$ for F the CDF of ν; this needs no density, only the CDF. (iii) Writing $a_t := \mathbb{E}[(t - X)_+]$ for the quantity actually estimated, $a_t \le \mathbb{P}(X < t) = F(t^-)$, and since $F(t^{\star -}) \le \alpha$ by definition of the quantile, $a_{t^\star} \le \alpha$ exactly, with no continuity assumption. Property (iii) is the source of the $\sqrt{\alpha}$ advantage; Figure 1 shows the search it enables.

Quantum access model. We use the standard qsample model (Wang et al., 2021; Wan et al., 2023; Wu et al., 2023): for each arm i a unitary O_i with $O_i|0\rangle = \sum_x \sqrt{\nu_i(x)}\,|x\rangle|\psi_{i,x}\rangle$ for arbitrary garbage states $|\psi_{i,x}\rangle$. Access is through the single *indexed* oracle

$$\mathcal{O} \;:\; |i\rangle|0\rangle \;\longmapsto\; |i\rangle\, O_i|0\rangle \;=\; |i\rangle \sum_x \sqrt{\nu_i(x)}\,|x\rangle|\psi_{i,x}\rangle, \tag{2}$$

together with $\mathcal{O}^\dagger$ and controlled versions, one query being one use of any of these. This is the model of Wang et al. (2021, Eq. 2), in which the index register may be held in superposition. Our upper bounds do not need that freedom, since Algorithms 1-4 always present the index register in a computational-basis state and so run verbatim given K separate oracles; the lower bounds and Section 6 do use the indexed form. A classical pull is recovered by preparing and measuring, so quantum access generalizes classical access at the same per-query accounting, the convention under which regret is defined in Section 4. Inverse access is necessary rather than merely convenient: Tang & Wright (2025) show the quadratic speedup of amplitude amplification and estimation is unattainable from forward-only access, and every result below invokes amplitude estimation.

Amplitude estimation. We use canonical amplitude estimation (Brassard et al., 2002, Theorem 12) as a black box. With M applications of the Grover iterate, it outputs an estimate $\tilde{a}$ of an amplitude a satisfying $|\tilde{a} - a| \le 2\pi\sqrt{a(1-a)}/M + (\pi/M)^2$ with probability at least $8/\pi^2$. Two features matter. The error decays as $1/M$ rather than $1/\sqrt{M}$, giving the quadratic advantage in ε; and the leading term carries $\sqrt{a(1-a)}$, so estimating a *small* amplitude to a given additive accuracy is cheaper. One subtlety is easy to lose in the folklore statement: the output is one of two grid values straddling a, both satisfying the bound. The good set is an interval, so the median of a majority of good runs remains in it and standard median amplification applies; see Appendix C.

3 Quantum CVaR estimation

3.1 A first estimator via noisy ternary search

Equation (1) turns CVaR estimation into maximization of a concave univariate function, and concavity means a two-point comparison reveals which side of the interval contains the maximizer. Our estimator performs a ternary search on ϕ, evaluating each probe by estimating the single amplitude $a_t = \mathbb{E}[(t - X)_+]$, which one oracle call, a comparator and a payload rotation prepare.

Algorithm 1: CVaREst$(O_\nu, \alpha, \varepsilon, \delta)$

$R \leftarrow \lceil \ln(1/(\alpha\varepsilon))/\ln(3/2) \rceil$; $\eta \leftarrow \varepsilon/(2(2R+1))$; $M \leftarrow \lceil 4\pi(2R+1)/(\alpha\varepsilon) \rceil$;
$k \leftarrow \lceil 5.184 \ln((2R+1)/\delta) \rceil$
$l \leftarrow 0, r \leftarrow 1, \mathcal{E} \leftarrow \emptyset$
Function `Eval(`t`)`**:**
  Prepare A_t encoding $a_t = \mathbb{E}[(t - X)_+]$ as an amplitude (one oracle call, comparator, payload rotation)
  Run amplitude estimation with M Grover iterates, k times; $\hat{a}_t \leftarrow$ median
  $\hat{\phi}(t) \leftarrow t - \hat{a}_t/\alpha$; add $(t, \hat{\phi}(t))$ to $\mathcal{E}$; **return** $\hat{\phi}(t)$
for $j = 1, \ldots, R$ **do**
  $m_1 \leftarrow l + (r-l)/3$; $m_2 \leftarrow r - (r-l)/3$
  if `Eval(`m_1`)` $\geq$ `Eval(`m_2`)` **then** $r \leftarrow m_2$
  else $l \leftarrow m_1$
end
`Eval(`$(l+r)/2$`)`
return $\hat{C} \leftarrow \max\{\hat{\phi}(t) : (t, \hat{\phi}(t)) \in \mathcal{E}\}$

Lemma 1 (CVaR estimation)**.** *Let $\varepsilon, \delta \in (0,1)$ and $\alpha \in (0,1)$ with $\alpha\varepsilon \leq 0.1$ and $\delta \leq 1/2$. Algorithm 1 outputs $\hat{C}$ with $\mathbb{P}(|\hat{C} - \mathrm{CVaR}_\alpha(\nu)| \leq \varepsilon) \geq 1 - \delta$ using at most*

$$Q \;\leq\; 140\,(2R+1)^2\,\frac{\ln\big((2R+1)/\delta\big)}{\alpha\varepsilon} \;=\; O\!\left(\frac{\log^2\!\big(\frac{1}{\alpha\varepsilon}\big)\log\frac{\log(1/\alpha\varepsilon)}{\delta}}{\alpha\varepsilon}\right) \tag{3}$$

queries, where $R = \lceil \ln(1/(\alpha\varepsilon))/\ln(3/2) \rceil$ is the number of search rounds. The constant 140 is proved in Appendix C under the regime conditions stated above; the complementary regimes $\alpha\varepsilon > 0.1$ and $\delta > 1/2$ are not covered by it and require a separate argument.

The proof is in Appendix C; one step is specific to the risk setting. Each probe is accurate to $\eta := \varepsilon/(2(2R+1))$ after median amplification, but a noisy comparison can discard the wrong third of the interval. Concavity bounds the damage: the trisection points are a $\frac{1}{2}$-weighted average toward any interior maximizer, so an errant cut loses at most 2η of objective value and losses accumulate additively rather than compounding. After R rounds the interval has width at most $\alpha\varepsilon$, and the Lipschitz property contributes a further $\varepsilon/2$.

One comparison is worth making. Classically, a standard estimator attains accuracy ε using $O(\alpha^{-1}\varepsilon^{-2})$ samples, for every distribution on $[0,1]$ and with no continuity assumption (Proposition 9, whose harder side we take from Mhammedi et al. (2020, Theorem 11)); it is an upper bound for a specified estimator, and no matching classical lower bound is claimed or cited. Lemma 1 therefore improves the ε dependence quadratically but *matches* the α^{-1} dependence. It would be easy to present that α^{-1} as an advantage; it is not.

3.2 An amplitude-aware estimator, and a $\sqrt{\alpha}$ advantage

The α^{-1} in Lemma 1 arises because we estimate a_t to additive accuracy $\alpha\eta$ and pay $\Theta(1/(\alpha\eta))$ for it. But by property (iii) of Section 2, the amplitude near the optimum is itself of order α, and the error term of amplitude estimation carries $\sqrt{a}$. Estimating an amplitude $a \approx \alpha$ to additive accuracy $\alpha\eta$ costs

$$M \;\approx\; \frac{2\pi\sqrt{a}}{\alpha\eta} \;\approx\; \frac{2\pi\sqrt{\alpha}}{\alpha\eta} \;=\; \frac{2\pi}{\sqrt{\alpha}\,\eta}, \tag{4}$$

so one factor of α cancels. Realizing this requires confining the search to a region where all probed amplitudes are provably $O(\alpha)$, which is what the two-phase algorithm below does.

Lemma 2 (Amplitude-aware CVaR estimation)**.** *Let $\varepsilon \in (0,1)$, $\delta \in (0, 1/2]$, $\alpha \in (0, 1/4]$ with $\alpha\varepsilon \leq 10^{-2}$. Algorithm 2 (Appendix D) outputs $\hat{C}$ with $\mathbb{P}(|\hat{C} - \mathrm{CVaR}_\alpha(\nu)| \leq \varepsilon) \geq 1 - \delta$ using at most*

$$Q' \;\leq\; 900\,(2R_2+1)^2\,\frac{\ln((B+2R_2+3)/\delta)}{\sqrt{\alpha}\,\varepsilon} \;=\; \widetilde{O}\!\left(\frac{1}{\sqrt{\alpha}\,\varepsilon}\right) \tag{5}$$

queries, where $R_2 = \lceil \ln(1/(\alpha\varepsilon))/\ln(3/2) \rceil$ and $B = \lceil \log_2(4/(\alpha\varepsilon)) \rceil$ is the number of Phase-1 bisection rounds, so that the logarithm here is identical to the one used in Appendix D.

Phase 1 localizes the α-quantile by a three-way bisection whose tests are amplitude estimations of $F(m)$ at accuracy $\alpha\varepsilon/8$, returning LOW, HIGH, or BAND. The tests are designed so that LOW certifies $F(m) < \alpha$ and HIGH certifies $F(m) > \alpha$, maintaining the invariant $F(l) < \alpha < F(r)$ and hence $t^\star \in (l, r]$; the bisection is entered only on LOW at the origin, which establishes the invariant at initialization, while HIGH and BAND there return $\mathrm{CVaR}_\alpha \approx 0$ immediately. Phase 2 runs the ternary search of Lemma 1 restricted to $[0, r_0]$, where monotonicity of F guarantees every probed amplitude is at most 2α, so amplitude-aware costs apply. The fine-accuracy tests of Phase 1 are not as expensive as they look: they need fine accuracy only where $F \le 3\alpha$, and there the amplitude-aware cost is again $O(1/(\sqrt{\alpha}\varepsilon))$. Two exit conditions require care and are handled in Appendix D: a BAND exit, where the missed optimum costs at most $\varepsilon/4$, and an atom exit, where F jumps across the band.

Against the classical $O(\alpha^{-1}\varepsilon^{-2})$, Lemma 2 improves both parameters, quadratically in ε and by $\sqrt{\alpha}$ in the tail level; Theorem 6 shows this is optimal up to logarithmic factors.

4 Bandit algorithms

We now compose the estimator with classical bandit schemes. The composition is legitimate because of the following property, which we verify by inspection of the proofs in Appendix E.

Proposition 3 (Interface-only composition)**.** *The correctness and complexity arguments for Algorithms 3 and 4 (Appendix E) use only: (i) that each C_i is a fixed real number, (ii) the (ε, δ) guarantee of the estimator, and (iii) differences $C_{i^\star} - C_i$. They never use linearity or additivity of the objective.*

This matters twice: it is what makes the estimator swap sound at all, since CVaR_α is not linear in the distribution, and it means the upgrade from Lemma 1 to Lemma 2 propagates to every bound below by substitution. Throughout we use accuracy schedule $\varepsilon_r = 2^{-r}$ and failure allocation $\delta_{i,r} = \delta/(Kr(r+1))$, so that $\sum_{i,r} \delta_{i,r} \le \delta$ by telescoping. Both bandit algorithms start at round

$$r_{\min} \;:=\; \max\{1,\, \lceil \log_2(100\alpha) \rceil\} \;\le\; 5 \qquad (\alpha \le 1/4), \tag{6}$$

which is what keeps every estimator call inside Lemma 2's domain $\alpha\varepsilon \le 10^{-2}$. Write E for the event that every estimator call in the run meets its guarantee; $\mathbb{P}(E) \ge 1 - \delta$.

Theorem 4 (Best-arm identification)**.** *Let $\alpha \le 1/4$ and let Algorithm 3 start at $r_{\min}$ as in* (6)*. On event E it outputs $i^\star$, eliminates each suboptimal arm i by round $\bar{r}_i = \max\{r_{\min}, \lfloor \log_2(4/\Delta_i) \rfloor + 1\}$, and uses at most*

$$\begin{aligned} Q_A \;\le\; & \sum_{i \ne i^\star} 57600\,(2J_i+1)^2\, \frac{\ln\big((B_i + 2J_i + 3)K\bar{r}_i(\bar{r}_i+1)/\delta\big)}{\sqrt{\alpha}\,\Delta_i} \\ & +\; 57600\,(2J_\star+1)^2\, \frac{\ln\big((B_\star + 2J_\star + 3)K\bar{r}_{\max}(\bar{r}_{\max}+1)/\delta\big)}{\sqrt{\alpha}\,\Delta_{\min}} \;=\; \widetilde{O}\Big(\sum_{i \ne i^\star} \tfrac{1}{\sqrt{\alpha}\Delta_i}\Big) \end{aligned} \tag{7}$$

queries, where $J_i = \lceil \ln(1/(\alpha\varepsilon_{\bar{r}_i}))/\ln(3/2) \rceil$ and $B_i = \lceil \log_2(4/(\alpha\varepsilon_{\bar{r}_i})) \rceil$ are the search-round and bisection counts Lemma 2 supplies at arm i's terminal accuracy $\varepsilon_{\bar{r}_i} = 2^{-\bar{r}_i}$. The second display is the survivor's contribution: $\bar{r}_{\max} = \max_i \bar{r}_i$ is the last round $i^\star$ is estimated, and $J_\star, B_\star$ are J_i, B_i evaluated at $\varepsilon_{\bar{r}_{\max}}$. The classical counterpart is $O(\sum_{i \ne i^\star} \alpha^{-1}\Delta_i^{-2})$.

Two features of the bound deserve comment. The prefactor is Lemma 2's, so J_i and B_i are that lemma's round and bisection counts rather than Lemma 1's R and $2R+1$. And $57600 = 900 \cdot 64$, where 900 is Lemma 2's constant and 64 comes from the geometric sum starting at $r_{\min}$, which carries $2^{r_{\min}} \le 32$; that start is also why J_i, B_i are indexed by the arm, since terminal accuracies differ across arms.

The algorithm is successive elimination (Even-Dar et al., 2002) driven by CVaR estimates: at round r every surviving arm is estimated to accuracy ε_r, and any arm falling more than $2\varepsilon_r$ below the best surviving estimate is discarded. The optimal arm survives because $\hat{C}_{i^\star} \ge C_{i^\star} - \varepsilon_r \ge C_j - \varepsilon_r \ge \hat{C}_j - 2\varepsilon_r$. A suboptimal arm is eliminated once $\varepsilon_r < \Delta_i/4$ *strictly*, which is why $\bar{r}_i$ uses a floor plus one rather than a ceiling.

Theorem 5 (Regret). *Let every query present the index register of $\mathcal{O}$ in a computational-basis state, so that each of the T queries names an arm I_t, and let regret be the horizon-T quantity* $\mathrm{Regret}(T) = \sum_{t=1}^{T} \Delta_{I_t}$. *Let $\alpha \le 1/4$, let Algorithm 4 start at $r_{\min}$ as in* (6)*, and let $T \ge Q_{\mathrm{init}}$, the deterministic cost of the K initialization calls. On event E, Algorithm 4 issues exactly T queries and satisfies*

$$\mathrm{Regret}(T) \;\le\; \sum_{i \neq i^\star} 57600\,(2J_i+1)^2\,\frac{\ln\big((B_i+2J_i+3)KL_i(L_i+1)/\delta\big)}{\sqrt{\alpha}} + \frac{3600\,(2J_s+1)^2\ln\big((B_s+2J_s+3)K/\delta\big)}{\sqrt{\alpha}} \;=\; O\Big(\tfrac{K}{\sqrt{\alpha}}\log^2\tfrac{T}{\alpha}\log\tfrac{K\log T}{\delta}\Big), \tag{8}$$

with no polynomial dependence on any gap Δ_i, *where L_i is the last round arm i is played and J_i, B_i are as in Theorem 4 at ε_{L_i}. The second display is the terminal term: if the next estimator call on the selected arm s does not fit in the remaining budget, those rounds are played on s and the selection inequality bounds their regret. The classical counterpart is $O(\sum_{i \neq i^\star} \alpha^{-1}\Delta_i^{-1}\log T)$.*

Both theorems are stated for $\alpha \le 1/4$, the domain on which Lemma 2 is proved. For $\alpha > 1/4$ the same algorithms driven by Lemma 1 give bounds of the same form up to constants, since $\alpha^{-1} \le 2\alpha^{-1/2}$ there; we do not state those constants.

The cancellation is the sharpest qualitative consequence of the quadratic speedup. Under optimism a suboptimal arm i is selected only while its confidence radius exceeds $\Delta_i/2$, so it receives $O(1/(\sqrt{\alpha}\Delta_i))$ queries, each incurring regret Δ_i; the product is $O(1/\sqrt{\alpha})$, independent of Δ_i. Classically the analogous product is $O(\alpha^{-1}\Delta_i^{-1})$, which diverges as gaps shrink. This mirrors the $\sqrt{T} \to \log T$ improvement of Wan et al. (2023) for the mean, now for a risk objective.

On the regret convention: one query is charged one unit of regret, following Wan et al. (2023). A reader who regards coherent queries as strictly more powerful than classical pulls may not accept this accounting; Theorem 4 is convention-free, which is one reason we present both results.

5 Lower bounds

We now show that Lemma 2 is optimal, using the quantum adversary method (Ambainis, 2002). The construction rests on one observation about the geometry of CVaR. For $m \in (0, \alpha]$ let ν_m place mass m at 0 and $1-m$ at 1; the worst α-fraction is then all of the mass at 0 together with $\alpha - m$ of the mass at 1, so

$$\mathrm{CVaR}_\alpha(\nu_m) \;=\; \frac{m\cdot 0 + (\alpha - m)\cdot 1}{\alpha} \;=\; 1 - \frac{m}{\alpha}. \tag{9}$$

The sensitivity $\mathrm{d}C/\mathrm{d}m = -1/\alpha$ is the crux: a CVaR gap of ε is purchased with a mass shift of only $\alpha\varepsilon$, and since the oracle encodes $\sqrt{m}$ as an amplitude that shift produces an even smaller angular separation. This is the structural fact that made Lemma 2 fast, viewed from the other side.

Theorem 6 (CVaR estimation lower bound). *Let $\varepsilon \le 1/16$ and $\alpha \le 1/4$. Any quantum algorithm that, given $\{O_\nu, O_\nu^\dagger,$ controlled-$O_\nu\}$, outputs $\hat{C}$ with $|\hat{C} - \mathrm{CVaR}_\alpha(\nu)| \le \varepsilon$ with probability at least $2/3$ for every ν on $[0,1]$ must make at least*

$$T \;\ge\; \frac{1}{16\,\varepsilon\sqrt{\alpha}} \;=\; \Omega\Big(\frac{1}{\varepsilon\sqrt{\alpha}}\Big) \tag{10}$$

queries. Combined with Lemma 2, quantum CVaR estimation is $\widetilde{\Theta}(1/(\varepsilon\sqrt{\alpha}))$.

Proof sketch. Take $m_0 = \alpha/2$ and $m_1 = m_0 + 3\alpha\varepsilon$, so by (9) the CVaRs differ by 3ε and any ε-accurate estimator distinguishes them with probability $2/3$. The oracle is specified only on $|0\rangle$, so the lower bound may choose the completion; taking O_{m_1} to differ from O_{m_0} by a rotation θ inside the two-dimensional span of the coin states gives $\|O_{m_1} - O_{m_0}\| \le \theta$. A Hellinger identity bounds $1 - \cos\theta \le \frac{21}{4}\alpha\varepsilon^2$, and Jordan's inequality converts this to $\theta \le 5.09\sqrt{\alpha}\,\varepsilon$: the mass shift $\alpha\varepsilon$ enters the angle only as its square root, which is the whole source of the bound. A hybrid argument over the three query types then gives $\||\psi_T^{m_0}\rangle - |\psi_T^{m_1}\rangle\| \le T\theta$, while distinguishing with probability $2/3$ forces Euclidean distance at least 0.338. Details in Appendix F. □

Theorem 7 (Best-arm identification lower bound)**.** *Let $\alpha \le 1/4$, let $0 < \delta < 1/2$, and let the gaps $\Delta_i \le 1/2$ be realizable by the two-point family. Any quantum algorithm identifying the CVaR-optimal arm with probability at least $1-\delta$ requires*

$$T \;\ge\; c_\delta \sqrt{\frac{H}{\alpha}}, \qquad c_\delta = \frac{1 - 2\sqrt{\delta(1-\delta)}}{9}, \qquad H = \sum_{i \neq i^\star} \Delta_i^{-2} \tag{11}$$

queries, that is $T = \Omega(\alpha^{-1/2}\sqrt{H})$.

The proof adapts the weighted adversary argument of Wang et al. (2021), whose bound consumes the oracle family only through the two-by-two matrix $A(a)^\dagger A(b) - I$ at the single differing arm, with no coupling across arms; it therefore transfers once that matrix is recomputed for mass-parameterized coins. Its diagonal entry is quadratic in the gap while the off-diagonal entry is linear and carries $\sqrt{\alpha}$, so the off-diagonal governs and produces $\sqrt{\alpha} \cdot \sqrt{\sum_{i \neq i^\star} \Delta_i'^{-2}}$. That the off-diagonal is genuinely linear is essential: were both entries quadratic the bound would collapse. Appendix G gives the constants, and the Cauchy-Schwarz step where the diagonal's $K-1$ distinct states must be handled with care.

6 The remaining $\sqrt{K}$ gap

Theorems 4 and 7 agree in the tail parameter but not in structure: the upper bound is $\sum_{i \neq i^\star} \Delta_i^{-1}$ while the lower bound is $\sqrt{\sum_{i \neq i^\star} \Delta_i^{-2}}$, which for K equal gaps differ by $\sqrt{K}$. We do not close this gap; we characterize what closing it would require.

The gap is not an artifact of loose analysis. It is the difference between estimating arms one at a time, which any elimination scheme does, and amplifying toward the best arm in superposition. Wang et al. (2021) achieve $\sqrt{H}$ for the mean by the latter route, using variable-time amplitude amplification (VTAA) (Ambainis, 2010; Childs et al., 2017; Chakraborty et al., 2019) over a subroutine that thresholds a *native* amplitude. Our thresholded quantity is not an amplitude; evaluating it requires running the estimator of Section 3. Closing the gap therefore requires VTAA over a coherent CVaR estimator, and Appendix H shows that three natural routes do not supply one: coherent execution of the adaptive search has no known implementation with the interface VTAA requires; a non-adaptive replacement built on the integral representation of Thomas & Learned-Miller (2019) must touch $\Theta(1/\varepsilon)$ bins and loses the quadratic advantage; and composition with Blanchet et al. (2025) is valid but is costed by a worst-case inner second moment rather than the local variance near the quantile, so it does not recover the $\sqrt{\alpha}$. We claim no impossibility, and regard this as the central open problem raised by this work.

7 Numerical evaluation

Because the results are query-complexity bounds rather than claims about a physical device, we evaluate them by simulation, following Wu et al. (2023), drawing amplitude-estimation outcomes from the true measurement distribution of Brassard et al. (2002) rather than a Gaussian surrogate. One distinction organizes the experiments: Algorithm 1's query count is *deterministic*, namely $(2R+1)k(2M+1)$ fixed in advance by $(\alpha, \varepsilon, \delta)$, so it is computed rather than measured, while its accuracy is stochastic and is what we measure. Details are in Appendix I.

Figure 3(c) locates the crossover: with the constants of Lemma 2 and the classical radius of Appendix B, $\varepsilon^* \approx 3.6 \times 10^{-6}$ at $\alpha = 0.25$ and $\varepsilon^* \approx 1.9 \times 10^{-5}$ at $\alpha = 0.005$. Its location is a constant-factor artifact of the uniform accuracy schedule and of canonical rather than iterative amplitude estimation, neither intrinsic to the result, and the asymptotic separation is unaffected since the exponents differ. The same caveat applies to the existing quantum bandit literature and is in our view under-reported there.

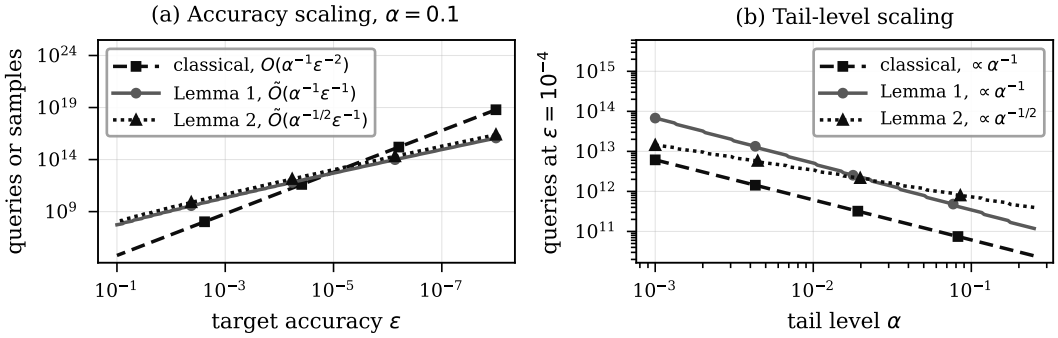

Figure 2: Query complexity with the constants of Lemmas 1 and 2 and the classical radius of Proposition 9 at the explicit constant fixed in Appendix B; no fitting is involved, and series are distinguished by line style and marker so the figure survives greyscale printing. (a) At $\alpha = 0.1$ both quantum curves have slope -1 against the classical -2. (b) At $\varepsilon = 10^{-4}$, Lemma 2 improves the exponent to $\alpha^{-1/2}$ while Lemma 1 matches the classical α^{-1}. At these constants the quantum curves cross near $\alpha \approx 0.022$: the $\sqrt{\alpha}$ gain overtakes the larger prefactor only below that, which is why panel (a) shows Lemma 2 above Lemma 1 throughout.

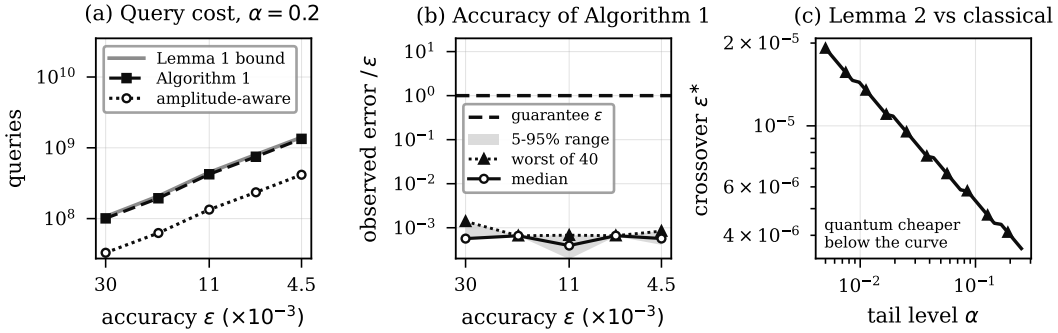

Figure 3: Cost, accuracy, and practical reach, at $\alpha = 0.2$ on $X \sim \mathrm{Unif}[0,1]$, where $\mathrm{CVaR}_\alpha = \alpha/2$ and the maximizer is interior. (a) Algorithm 1's query count is deterministic and sits within 7% of the Lemma 1 bound. The third curve is an oracle-tuned variant that sizes each amplitude estimation to the exact a_t; it is not a run of Algorithm 1, which cannot know a_t, but its stable $3.1\times$ advantage measures what amplitude-awareness is worth, the premise Lemma 2 formalizes. (b) Algorithm 1 run verbatim at its prescribed fixed M: all 40 repetitions at every accuracy met the ε guarantee, worst error $1.4 \times 10^{-3}\varepsilon$; Appendix I explains why the margin is structural. (c) The accuracy below which Lemma 2's bound falls below the classical bound of Appendix B; a comparison of upper bounds, not of measured costs.

8 Implications and Applications

Where the model is realistic. The qsample oracle is the binding assumption. Wang et al. (2021) observe that it is available whenever the reward is produced by a program rather than by nature: given a reversible circuit computing a reward from random bits, one query to that circuit yields one oracle query. This covers Monte Carlo evaluation of engineered systems, precisely where tail risk is the operative concern, such as derivative valuation, power-network reliability under simulated contingencies, and controller safety evaluation. It does not cover rewards from physical sampling that cannot be run in superposition, such as clinical trials.

Beyond CVaR. Proposition 3 is the reusable part: any risk measure admitting an estimator with the same (ε, δ) interface inherits Theorems 4 and 5 verbatim, with only the query-cost function changing. This needs solitary-arm optimality and the estimator interface established separately, plus a variational or integral representation to drive a search; it does not cover distortion risk measures in

general, for which the optimal policy can be a mixture of arms. The estimation layer, not the bandit layer, is where a new risk measure requires work.

Implications for the quantum bandit literature. The regret cancellation of Theorem 5 is not specific to CVaR: whenever an estimator achieves inverse-accuracy rather than inverse-square-accuracy scaling, the per-arm query cost and the per-query regret cancel and polynomial gap dependence disappears. And carrying explicit constants through to a crossover (Figure 3(c)) is, to our knowledge, not standard practice here; a reader cannot otherwise assess what an asymptotic separation would mean in practice.

8.1 Limitations

Collected in one place: (i) the results assume the qsample oracle of Section 8, a restriction inherited from all prior quantum bandit work; (ii) the speedups are asymptotic, with the crossover quantified in Section 7, and we claim no practical advantage at moderate accuracy; (iii) the multi-arm upper and lower bounds differ by $\sqrt{K}$ (Section 6); (iv) the regret result depends on the query-charging convention of Section 4, though best-arm identification does not; (v) constants are explicit but unoptimized; and (vi) the experiments cover Algorithm 1 only, so the $\alpha^{-1/2}$ claim of Algorithm 2 rests on the proof alone.

9 Related work

Risk-sensitive bandits. Classical algorithms for CVaR and related criteria are well developed (Cassel et al., 2018; Tamkin et al., 2019; Chow et al., 2015), with estimation primitives from Brown (2007), Thomas & Learned-Miller (2019) and Mhammedi et al. (2020); Proposition 9, which our classical baselines instantiate, takes its harder side from the last of these. All of it is classical.

Quantum bandits. Wang et al. (2021) give best-arm identification for the mean with a matching adversary lower bound, whose machinery we adapt in Theorem 7; Wan et al. (2023) obtain logarithmic regret; Wu et al. (2023) extend to heavy tails via an estimator swap and doubling trick that Section 4 follows in structure. Liu et al. (2026) prove a minimax $\Omega(K \log(T/K))$ regret lower bound, with which Theorem 5 is consistent, carrying $\log^2$ factors in T and linear dependence on K; we claim no optimality in K or T. All optimize the mean, the axis on which we differ.

Quantum estimation of risk measures. Woerner & Egger (2019) give circuits for financial risk measures including CVaR, focusing on construction rather than query complexity. Skarlatos & Konofaos (2025) study quantum subgradient estimation for CVaR and state that a matching lower bound for the d-dimensional subgradient problem is open; Theorem 6 proves one for scalar CVaR-value estimation and does not settle that question. Montanaro (2015) and Kothari & O'Donnell (2023) give mean estimators whose amplitude-aware structure we exploit, and Blanchet et al. (2025) treat nested expectations (Section 6).

10 Conclusion and future work

We have given the first quantum algorithms for risk-sensitive bandits, with lower bounds proving the core estimation primitive optimal in both accuracy and tail parameter. The technical core is one observation: the quantity CVaR requires estimating is intrinsically small in the tail, which yields the $\sqrt{\alpha}$ advantage and, read in reverse, supplies the hard instance.

Four directions follow. **Closing the $\sqrt{K}$ gap** is the most substantial, and by Section 6 needs a coherent, amplitude-aware, non-adaptive estimator for variable-time amplification to wrap; a barrier showing adaptivity to be necessary would be equally informative. **Tightening constants** is more tractable and moves the crossover toward accuracies that arise in real Monte Carlo work; Appendix I identifies one source of slack. **Broader risk measures** follow Section 8: which distortion risk measures admit a Lemma 2-style estimator is open. Finally, **richer decision problems** such as contextual and linear risk-sensitive bandits are natural targets, though the estimator swap needs revisiting where the objective couples across arms.

Reproducibility statement

Complete proofs appear in Appendices A-G with explicit constants. The numerical evaluation is specified in Appendix I: the seed, the exact amplitude-estimation sampling procedure, and the script that checks the numerical constants.

AI disclosure section

In compliance with the ICLR 2027 conference policy, the authors disclose the use of Large Language Models (LLMs) in the preparation of this manuscript. Specifically, AI tools were utilized to aid and polish the writing, as well as for retrieval and discovery purposes, such as identifying relevant related work. AI tools were not used for research ideation or execution, drafting core sections, generating synthetic datasets, or proving mathematical claims. The authors assume full responsibility for the originality, contents, and accuracy of this paper.

References

Andris Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64(4):750–767, 2002.

Andris Ambainis. Variable time amplitude amplification and a faster quantum algorithm for solving systems of linear equations. *arXiv preprint arXiv:1010.4458*, 2010.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

Jose Blanchet, Yassine Hamoudi, Mario Szegedy, and Guanyang Wang. Non-linear quantum Monte Carlo. *arXiv preprint arXiv:2502.05094*, 2025.

Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.

David B Brown. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, 35(6):722–730, 2007.

Asaf Cassel, Shie Mannor, and Assaf Zeevi. A general approach to multi-armed bandits under risk criteria. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *PMLR*, pp. 1295–1306, 2018.

Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The power of block-encoded matrix powers: improved regression techniques via faster Hamiltonian simulation. In *International Colloquium on Automata, Languages, and Programming*, 2019.

Andrew M Childs, Robin Kothari, and Rolando D Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950, 2017.

Yinlam Chow, Aviv Tamar, Shie Mannor, and Marco Pavone. Risk-sensitive and robust decision-making: a CVaR optimization approach. In *Advances in Neural Information Processing Systems*, 2015.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *Computational Learning Theory*, pp. 255–270, 2002.

Robin Kothari and Ryan O'Donnell. Mean estimation when you have the source code; or, quantum Monte Carlo methods. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1186–1215, 2023.

Maoli Liu, Zhuohua Li, and John C. S. Lui. Quantum multi-armed bandits and linear bandits: Lower bounds and algorithms. *arXiv preprint arXiv:2608.14319*, 2026.

Zakaria Mhammedi, Benjamin Guedj, and Robert C. Williamson. PAC-Bayesian bound for the conditional value at risk. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Ashley Montanaro. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A*, 471(2181):20150301, 2015.

R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

Vasilis Skarlatos and Nikos Konofaos. Quantum subgradient estimation for conditional value-at-risk optimization. *arXiv preprint arXiv:2510.04736*, 2025.

Alex Tamkin, Ramtin Keramati, Christoph Dann, and Emma Brunskill. Distributionally-aware exploration for CVaR bandits. In *NeurIPS 2019 Workshop on Safety and Robustness in Decision Making*, 2019.

Ewin Tang and John Wright. Amplitude amplification and estimation require inverses. *arXiv preprint arXiv:2507.23787*, 2025.

Philip Thomas and Erik Learned-Miller. Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*, PMLR, pp. 6225–6233, 2019.

Zongqi Wan, Zhijie Zhang, Tongyang Li, Jialin Zhang, and Xiaoming Sun. Quantum multi-armed bandits and stochastic linear bandits enjoy logarithmic regrets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023. arXiv:2205.14988.

Daochen Wang, Xuchen You, Tongyang Li, and Andrew M Childs. Quantum exploration algorithms for multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 10102–10110, 2021.

Stefan Woerner and Daniel J Egger. Quantum risk analysis. *npj Quantum Information*, 5(1):15, 2019.

Yulian Wu, Chaowen Guan, Vaneet Aggarwal, and Di Wang. Quantum heavy-tailed bandits. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2023. arXiv:2301.09680.

A PRELIMINARIES: PROOFS

We collect the properties of the Rockafellar-Uryasev objective $\phi(t) = t - \frac{1}{\alpha}\mathbb{E}[(t-X)_+]$ used throughout, for X supported on $[0,1]$ and $\alpha \in (0,1)$. Let F denote the CDF of X.

Proposition 8. *(i) The supremum in* (1) *is attained in* $[0,1]$*, at the* α*-quantile* $t^\star = \inf\{t : F(t) \geq \alpha\}$*. (ii)* ϕ *is concave. (iii)* $|\phi'(t)| \leq 1/\alpha$ *wherever the derivative exists, and* ϕ *is* $1/\alpha$*-Lipschitz on* $[0,1]$*. (iv) The payload satisfies* $a_t := \mathbb{E}[(t-X)_+] \leq \mathbb{P}(X < t) = F(t^-)$ *for* $t \in [0,1]$*, and consequently* $a_{t^\star} \leq F(t^{\star -}) \leq \alpha$ *for every distribution on* $[0,1]$*, with or without an atom at* $t^\star$*. Moreover* $\mathrm{Var}\big((t^\star - X)_+\big) \leq a_{t^\star} \leq \alpha$*.*

Proof. (i) For $t < 0$ we have $\phi(t) = t < 0 = \phi(0)$. For $t > 1$, $(t-X)_+ = t - X$, so $\phi'(t) = 1 - 1/\alpha \leq 0$. Hence the supremum is attained in $[0,1]$. Standard convex analysis identifies the maximizer as the α-quantile.

(ii) For each fixed x, $t \mapsto (t-x)_+ = \max(t-x, 0)$ is a maximum of two affine functions and hence convex. Expectation preserves convexity, so $t \mapsto \mathbb{E}[(t-X)_+]$ is convex and ϕ is the difference of an affine and a convex function, hence concave.

(iii) The function $g(t) := \mathbb{E}[(t-X)_+]$ is convex, so its one-sided derivatives exist everywhere, and where differentiable $g'(t) = \mathbb{P}(X \leq t) = F(t)$. Hence $\phi'(t) = 1 - F(t)/\alpha$. Since $0 \leq F \leq 1$ we have $1 - 1/\alpha \leq \phi'(t) \leq 1$, so $|\phi'(t)| \leq \max(1, 1/\alpha - 1) \leq 1/\alpha$ for $\alpha \leq 1$. Lipschitzness follows from the subgradient bound; note that no density is assumed.

(iv) For $t \in [0,1]$ and $X \in [0,1]$ we have $(t-X)_+ \leq \mathbf{1}\{X < t\}$, since the left side is zero when $X \geq t$ and at most $t \leq 1$ otherwise. Taking expectations gives $a_t \leq \mathbb{P}(X < t) = F(t^-)$. By

definition $t^\star = \inf\{t : F(t) \geq \alpha\}$, so $F(t) < \alpha$ for every $t < t^\star$ and hence $F(t^{\star-}) \leq \alpha \leq F(t^\star)$; the first inequality gives $a_{t^\star} \leq \alpha$. Finally $W := (t^\star - X)_+ \in [0,1]$, so $W^2 \leq W$ and $\mathrm{Var}(W) \leq \mathbb{E}[W^2] \leq \mathbb{E}[W] = a_{t^\star} \leq \alpha$. □

Property (iv) is what makes the amplitude-aware estimator possible: near $t^\star$ we have $F(t^\star) \approx \alpha$, so the amplitude being estimated is itself of order α.

B CLASSICAL BASELINES

We first supply the two-sided estimator guarantee the classical entries need. Throughout this section we use the loss convention $Z := 1 - X \in [0,1]$, so that the lower-tail reward CVaR of X equals $1 - C_\alpha(Z)$ where $C_\alpha(Z) := \inf_\mu\{\mu + \frac{1}{\alpha}\mathbb{E}[(Z-\mu)_+]\}$ is the upper-tail CVaR of the loss, and $\hat{C}_\alpha$ denotes the corresponding empirical estimator $\inf_\mu\{\mu + \frac{1}{n\alpha}\sum_i (Z_i - \mu)_+\}$.

Proposition 9 (Two-sided classical CVaR estimation)**.** *There are universal constants $c, c' > 0$ such that for every distribution on $[0,1]$, with or without atoms, every $\alpha \in (0,1)$ and every $\delta \in (0,1)$,*

$$\mathbb{P}\left(\left|\hat{C}_\alpha - C_\alpha\right| > c\left[\sqrt{\frac{\ln(c'/\delta)}{\alpha n}} + \frac{\ln(c'/\delta)}{\alpha n}\right]\right) \;\leq\; \delta.$$

In particular $n = O\big(\ln(1/\delta)/(\alpha\varepsilon^2)\big)$ samples suffice for accuracy ε, for $0 < \varepsilon \leq 1$.

Proof. The side $\hat{C}_\alpha - C_\alpha$. Let $\mu^\star$ minimize $\mu + \frac{1}{\alpha}\mathbb{E}[(Z-\mu)_+]$ and put $W := (Z - \mu^\star)_+ \in [0,1]$. Optimality of $\mu^\star$ forces the right derivative $1 - \frac{1}{\alpha}\mathbb{P}(Z > \mu^\star) \geq 0$, that is $\mathbb{P}(Z > \mu^\star) \leq \alpha$; hence $W^2 \leq W \leq \mathbf{1}\{Z > \mu^\star\}$ and $\mathrm{Var}(W) \leq \mathbb{E}[W^2] \leq \mathbb{P}(Z > \mu^\star) \leq \alpha$. Taking $\mu = \mu^\star$ in the empirical infimum,

$$\hat{C}_\alpha - C_\alpha \;\leq\; \Big(\mu^\star + \tfrac{1}{\alpha}\overline{W}\Big) - \Big(\mu^\star + \tfrac{1}{\alpha}\mathbb{E}W\Big) \;=\; \frac{\overline{W} - \mathbb{E}W}{\alpha},$$

and Bernstein's inequality with $\mathrm{Var}(W) \leq \alpha$ and $W \in [0,1]$ gives $\overline{W} - \mathbb{E}W \leq \sqrt{2\alpha\ln(1/\delta)/n} + \frac{2}{3}\ln(1/\delta)/n$, so this side is at most $\sqrt{2\ln(1/\delta)/(\alpha n)} + \frac{2}{3}\ln(1/\delta)/(\alpha n)$. No continuity is used.

The side $C_\alpha - \hat{C}_\alpha$. This is the harder direction, because the empirical minimizer is data dependent. It is exactly the content of Mhammedi et al. (2020, Theorem 11), proved for bounded Z with no continuity assumption, and with α inside the square root: their bound is of the form $\sqrt{12\,C_\alpha\ln(1/\delta)/(5\alpha n)} \vee O(1/(\alpha n)) + C_\alpha\big(\sqrt{\ln(1/\delta)/(\alpha n)} + 3\ln(1/\delta)/(\alpha n)\big)$, which for $Z \in [0,1]$ has the stated form. Combining the two sides and translating back through $Z = 1 - X$ gives the proposition. □

The classical entries of Table 1 are obtained by instantiating standard schemes with the two-sided guarantee of Proposition 9: writing

$$\beta(n,\delta) \;:=\; c\left[\sqrt{\frac{\ln(c'/\delta)}{\alpha n}} + \frac{\ln(c'/\delta)}{\alpha n}\right]$$

for the confidence radius it supplies, $n = O(\ln(1/\delta)/(\alpha\varepsilon^2))$ samples suffice for accuracy ε when $0 < \varepsilon \leq 1$. We state the resulting classical counts asymptotically, because the universal constants c, c' have not been pinned down: the elementary side of Proposition 9 gives $\sqrt{2\ln(1/\delta)/(\alpha n)} + \frac{2}{3}\ln(1/\delta)/(\alpha n)$ explicitly, but the opposite side is imported from Mhammedi et al. (2020, Theorem 11) and its explicit constants would have to be traced through. Note that β carries a linear term in $\ln(1/\delta)/(\alpha n)$ that a pure $\sqrt{\cdot}$ radius does not. Figures 2 and 3 require a numerical classical constant, which Proposition 9 does not supply. We fix one by convention rather than by proof: the radius $3\sqrt{x} + \frac{10}{3}x$ with $x := \ln(3/\delta)/(\alpha n)$ gives $n \geq 15\ln(3/\delta)/(\alpha\varepsilon^2)$ for $0 < \varepsilon \leq 1$; it is consistent with the elementary side proved above and with the shape of Mhammedi et al. (2020, Theorem 11), and we have verified numerically that it implies $|\hat{C}_\alpha - C_\alpha| \leq \varepsilon$. We have not traced the explicit constants of that theorem, so we do not state it as proved. It is used only to fix the horizontal position of the crossover in Figure 3(c); no claim in the paper depends on it.

We record the resulting bounds.

Classical CVaR successive elimination. An arm is eliminated once its confidence radius falls below $\Delta_i/4$, that is once $\beta(n_i, \delta') \le \Delta_i/4$, giving $n_i = O\big(\ln(1/\delta')/(\alpha\Delta_i^2)\big)$ pulls and total $O\big(\sum_{i\neq i^\star} \alpha^{-1}\Delta_i^{-2}\log(K\log(1/\Delta_i)/\delta)\big)$.

Classical CVaR-UCB. With confidence radius $\beta(n, \delta')$ as above and $\delta' = \Theta(1/T)$, arm i ceases to be selected once $\beta(n_i) \le \Delta_i/2$, that is after $n_i = O\big(\ln(T)/(\alpha\Delta_i^2)\big)$ pulls, each incurring regret Δ_i. The total is $O\big(\sum_{i\neq i^\star} \ln(T)/(\alpha\Delta_i)\big)$.

The classical entries above are upper bounds on the cost of specified classical algorithms, obtained from Proposition 9. They are not information-theoretic lower bounds: we do not claim, and do not cite, any matching classical lower bound for CVaR estimation or for CVaR bandits.

C Proof of Lemma 1

Amplitude estimation, stated carefully. We use Brassard et al. (2002, Theorem 12). With M applications of the Grover iterate, and with probability at least $8/\pi^2 \approx 0.8106$, the procedure outputs one of two grid values $\tilde{a} \in \{\tilde{a}_1, \tilde{a}_2\}$ straddling the true amplitude a, both of which satisfy

$$|\tilde{a}_i - a| \;\le\; \frac{2\pi\sqrt{a(1-a)}}{M} + \Big(\frac{\pi}{M}\Big)^2. \tag{12}$$

Boundary cases are exact: if $a = 0$ the output is 0 with certainty, and if $a = 1$ and M is even the output is 1 with certainty. Each run uses at most $2M + 1$ oracle queries, one initial state preparation plus M Grover iterates each invoking A_t and $A_t^\dagger$ once.

Lemma 10 (Median amplification with a two-valued guarantee)**.** *Fix $\beta > 0$ and let $G_\beta := \{a' : |a' - a| \le \beta\}$. Suppose each of k independent runs lands in G_β with probability at least $8/\pi^2$. Then the median of the k outputs lies in G_β except with probability at most $\exp(-0.19291\,k)$, so $k \ge 5.184\ln(1/\delta')$ suffices for failure probability δ'.*

Proof. G_β is an interval. Order the k outputs. If strictly more than $k/2$ of them lie in G_β, then the median, being a middle order statistic, is bracketed by elements of G_β: there are too few bad runs to occupy the median position from either side. Hence the median lies in G_β. This is where the two-valued nature of (12) is absorbed; good runs may split between $\tilde{a}_1$ and $\tilde{a}_2$, but both lie in the same interval. The median therefore fails only if at least $k/2$ runs are bad, each bad with probability at most $1 - 8/\pi^2 = 0.1894$. By Hoeffding's inequality, $\mathbb{P}(\mathrm{Bin}(k, 0.1894) \ge k/2) \le \exp(-2k(0.5 - 0.1894)^2) = \exp(-0.19291k)$. Solving $\exp(-0.19291k) \le \delta'$ gives $k \ge \ln(1/\delta')/0.19291 = 5.184\ln(1/\delta')$. □

Proof of Lemma 1. Set $R = \lceil \ln(1/(\alpha\varepsilon))/\ln(3/2)\rceil$, $\eta = \varepsilon/(2(2R{+}1))$, $M = \lceil 4\pi(2R{+}1)/(\alpha\varepsilon)\rceil$, $k = \lceil 5.184\ln((2R+1)/\delta)\rceil$.

Step 1 (per-evaluation accuracy). Set $\beta = \alpha\eta$. Since $a(1-a) \le 1/4$, (12) gives error at most $\pi/M + \pi^2/M^2$. For $M \ge 2\pi/\beta$ with $\beta \le 2$ we have $\pi/M \le \beta/2$ and $\pi^2/M^2 \le \beta^2/4 \le \beta/2$, so a single run lands in G_β with probability at least $8/\pi^2$. By Lemma 10 with $\delta' = \delta/(2R+1)$, the median estimate satisfies $|\hat{a}_t - a_t| \le \alpha\eta$ except with probability $\delta/(2R{+}1)$. Hence $|\hat{\phi}(t) - \phi(t)| = |\hat{a}_t - a_t|/\alpha \le \eta$. A union bound over the $2R{+}1$ evaluations gives that all evaluations are η-accurate with probability at least $1 - \delta$. Condition on this event.

Step 2 (errant cuts lose at most 2η). Consider a round on $[l, r]$ with trisection points $m_1 < m_2$ and let $\tau \in [l, r]$ attain $V := \max_{[l,r]} \phi$. If τ lies in the kept interval there is nothing to prove. Suppose $\tau > m_2$ but the algorithm keeps $[l, m_2]$, so $\hat{\phi}(m_1) \ge \hat{\phi}(m_2)$ and hence $\phi(m_1) \ge \phi(m_2) - 2\eta$. Write m_2 as a convex combination $m_2 = \lambda m_1 + (1-\lambda)\tau$. Since $m_2 - m_1 = (r-l)/3$ and $\tau - m_1 \le r - m_1 = 2(r-l)/3$, the weight on τ is $1 - \lambda = (m_2 - m_1)/(\tau - m_1) \ge 1/2$. Concavity then gives

$$\phi(m_2) \;\ge\; \lambda\phi(m_1) + (1-\lambda)\phi(\tau) \;\ge\; \tfrac{1}{2}\big(\phi(m_1) + \phi(\tau)\big) \;\ge\; \tfrac{1}{2}\big(\phi(m_2) - 2\eta + \phi(\tau)\big),$$

whence $\phi(m_2) \ge \phi(\tau) - 2\eta = V - 2\eta$. Since m_2 lies in the kept interval, the kept interval's maximum is at least $V - 2\eta$. The case $\tau < m_1$ is symmetric.

Step 3 (accumulation). Let V_j denote the maximum of ϕ over the round-j interval, so $V_0 = \mathrm{CVaR}_\alpha(\nu)$. By Step 2, $V_R \geq \mathrm{CVaR}_\alpha(\nu) - 2R\eta$. The final interval has width $(2/3)^R \leq \alpha\varepsilon$ by the choice of R (exact, since $R \geq \ln(1/(\alpha\varepsilon))/\ln(3/2)$ by definition), and the evaluated midpoint t_f satisfies $|t_f - \tau_R| \leq (2/3)^R/2$, so by Proposition 8(iii), $\phi(t_f) \geq V_R - \frac{1}{\alpha}\cdot\frac{\alpha\varepsilon}{2} = V_R - \varepsilon/2$. Therefore

$$\hat{C} \;\geq\; \hat{\phi}(t_f) \;\geq\; \phi(t_f) - \eta \;\geq\; \mathrm{CVaR}_\alpha(\nu) - (2R+1)\eta - \varepsilon/2 \;=\; \mathrm{CVaR}_\alpha(\nu) - \varepsilon,$$

using $(2R+1)\eta = \varepsilon/2$. For the upper side, $\hat{C} = \max_t \hat{\phi}(t) \leq \max_t \phi(t) + \eta \leq \mathrm{CVaR}_\alpha(\nu) + \varepsilon$.

Step 4 (query count). There are $2R+1$ evaluations, each running k amplitude estimations of at most $2M+1$ queries, giving $Q = (2R+1)k(2M+1)$. Substituting the parameter values and using $\alpha\varepsilon \leq 1/10$, $\delta \leq 1/2$ yields $Q \leq 140(2R+1)^2 \ln((2R+1)/\delta)/(\alpha\varepsilon)$. □

Remark 11 (Honest accounting of the constants). The constant 140 and the $(2R+1)^2$ factor are artifacts of a uniform accuracy schedule and of canonical rather than iterative amplitude estimation. A geometric schedule reduces the squared logarithm to a single logarithm. We have not optimized these, and Section 7 reports the resulting crossover explicitly rather than suppressing it. The constant 140 is nevertheless close to necessary: sweeping the admissible domain and maximizing over δ at the discontinuities of k, the largest prefactor it must cover is 138.1, near $\alpha = 0.119$, $\varepsilon = 0.84$, leaving 1.4% of headroom. Both figures are reproduced by the verification script of Appendix I.

D Proof of Lemma 2

Algorithm 2: AmplitudeAwareCVaREst$(O_\nu, \alpha, \varepsilon, \delta)$

Phase 1 (localize). $l \leftarrow 0$, $r \leftarrow 1$
Function `Test(`m`)`**:**
  $\hat{a} \leftarrow$ median-amplified amplitude estimate of $F(m)$ at accuracy $\alpha\varepsilon/8$
  if $\hat{a} < \alpha - \alpha\varepsilon/8$ **then return** LOW
  else if $\hat{a} > \alpha + \alpha\varepsilon/4$ **then return** HIGH
  else return BAND
$s_0 \leftarrow$ `Test(0)`
if $s_0 \in \{\text{HIGH}, \text{BAND}\}$ **then return** 0 `// CVaR`$_\alpha \leq \varepsilon/4$
for $j = 1, \ldots, B := \lceil \log_2(4/(\alpha\varepsilon)) \rceil$ **do**
  $m \leftarrow (l+r)/2$
  $s \leftarrow$ `Test(`m`)`
  if $s =$ LOW **then** $l \leftarrow m$ `// certifies` $F(m) < \alpha$
  else if $s =$ HIGH **then** $r \leftarrow m$ `// certifies` $F(m) > \alpha$
  else $r_0 \leftarrow m$; **break**
end
if *no* BAND *occurred* **then** $r_0 \leftarrow l$ `// atom exit:` $r - l \leq \alpha\varepsilon/4$
Phase 2 (amplitude-aware search). Run CVaREst restricted to $[0, r_0]$, where $a_t \leq F(r_0) \leq 2\alpha$ for every probe, using $M_2 = \lceil 16\pi\sqrt{2}(2R_2+1)/(\sqrt{\alpha}\,\varepsilon) \rceil$
return the resulting estimate

Set

$$B = \left\lceil \log_2 \tfrac{4}{\alpha\varepsilon} \right\rceil, \qquad R_2 = \left\lceil \tfrac{\ln(1/(\alpha\varepsilon))}{\ln(3/2)} \right\rceil, \qquad \eta = \tfrac{\varepsilon}{4(2R_2+1)},$$

$$M_1 = \left\lceil \tfrac{32\pi\sqrt{3}}{\sqrt{\alpha}\,\varepsilon} \right\rceil, \qquad M_2 = \left\lceil \tfrac{16\pi\sqrt{2}(2R_2+1)}{\sqrt{\alpha}\,\varepsilon} \right\rceil, \qquad k = \left\lceil 5.184 \ln \tfrac{B+2R_2+3}{\delta} \right\rceil.$$

Proof. Step 1 (soundness of the three-way test). The test estimates the amplitude $F(m)$ using M_1 Grover iterates with median amplification. If $F(m) \leq 3\alpha$, then by (12) the error is at most $2\pi\sqrt{3\alpha}/M_1 + (\pi/M_1)^2 \leq \alpha\varepsilon/8$ for the stated M_1. If instead $F(m) > 3\alpha$, then using $\sqrt{F\alpha} \leq F/\sqrt{3}$ the leading term is at most $\varepsilon F/48$, and the second term satisfies $(\pi/M_1)^2 \leq \alpha\varepsilon^2/3072 \leq \varepsilon^2 F/9216$ because $F > 3\alpha$. Hence for $\varepsilon \leq 1$

$$|\hat{a} - F| \;\leq\; F\Big(\tfrac{\varepsilon}{48} + \tfrac{\varepsilon^2}{9216}\Big) \;\leq\; F\Big(\tfrac{1}{48} + \tfrac{1}{9216}\Big) \;<\; \tfrac{F}{47},$$

so the estimate satisfies $\hat{a} \geq \frac{46}{47}F > \frac{138}{47}\alpha > \alpha + \alpha\varepsilon/4$ and the test returns HIGH. Consequently, on the good event: LOW implies $F(m) < \alpha$; HIGH implies $F(m) > \alpha$; and BAND implies $F(m) \in [\alpha(1-\varepsilon/4), \alpha(1+3\varepsilon/8)]$.

Step 2 (invariant). The bisection is entered only when $s_0 =$ LOW, which certifies $F(0) < \alpha$, so the invariant holds at initialization $l = 0$, $r = 1$. The updates preserve $F(l) < \alpha < F(r)$, hence $t^\star = \inf\{t : F(t) \geq \alpha\} \in (l, r]$ throughout. This is exactly the property that a coarse test would fail to provide: a test certifying only $F(m) < 1.5\alpha$ leaves open $F(m) \geq \alpha$, and the bisection may then move past $t^\star$ while ϕ decreases at rate up to $1/2$, incurring unbounded loss.

Step 3 (exits). Three exits are possible. (a) HIGH *or* BAND *at* 0: on HIGH, $F(0) > \alpha$, so the worst α-mass lies entirely at 0 and $\mathrm{CVaR}_\alpha = 0$ exactly. On BAND, $F(0) \geq \alpha(1-\varepsilon/4)$. If $F(0) \geq \alpha$ then again $\mathrm{CVaR}_\alpha = 0$. Otherwise, for $0 \leq t \leq t^\star$ monotonicity of F gives $0 \leq \phi'(t) = 1 - F(t)/\alpha \leq 1 - F(0)/\alpha \leq \varepsilon/4$, and since $\phi(0) = 0$,

$$0 \;\leq\; \mathrm{CVaR}_\alpha \;=\; \phi(t^\star) - \phi(0) \;\leq\; \int_0^{t^\star} \tfrac{\varepsilon}{4}\,\mathrm{d}t \;\leq\; \tfrac{\varepsilon}{4}.$$

Returning 0 is therefore within $\varepsilon/4$ in both cases.

(b) BAND *at* r_0: then $F(r_0) \geq \alpha(1-\varepsilon/4)$. If $t^\star \leq r_0$, Phase 2 contains the optimum. If $t^\star > r_0$, then on $[r_0, t^\star]$ we have $F \in [\alpha(1-\varepsilon/4), \alpha]$, so $|\phi'| = |1 - F/\alpha| \leq \varepsilon/4$ and the missed optimum costs at most $\varepsilon/4$. Also $F(r_0) \leq \alpha(1+3\varepsilon/8) \leq 2\alpha$, so every probe in $[0, r_0]$ has $a_t \leq F(t) \leq F(r_0) \leq 2\alpha$ by monotonicity of F. (c) *Atom exit*: if no BAND occurs in B rounds, the interval has width at most $\alpha\varepsilon/4$ while straddling $t^\star$, so $\phi(t^\star) - \phi(l) \leq \frac{1}{\alpha} \cdot \frac{\alpha\varepsilon}{4} = \varepsilon/4$ by Proposition 8(iii), and evaluating at l suffices.

Step 4 (budget). The total error decomposes into localization drift at most $\varepsilon/4$ (Step 3), Phase-2 search error $2R_2\eta + \eta \leq \varepsilon/4$ by the choice of η, and the Phase-2 width term $|\phi'| \cdot (\text{width}) \leq \frac{1}{\alpha} \cdot \frac{\alpha\varepsilon}{2} = \varepsilon/2$. The three sum to ε. We flag this decomposition as the step most sensitive to off-by-one errors; in particular the width term contains no additional factor of $1/2$, since the two factors of α cancel exactly.

Step 5 (cost). In Phase 2 every amplitude satisfies $a_t \leq 2\alpha$, so achieving accuracy $\alpha\eta$ requires $2\pi\sqrt{2\alpha}/M_2 \leq \alpha\eta/2$, that is $M_2 \geq 4\pi\sqrt{2}/(\sqrt{\alpha}\,\eta)$, which the stated value satisfies; the second term of (12) is dominated. Phase 1 uses $B + 1$ tests at M_1 iterates (the initial TEST(0) together with the B loop tests) and Phase 2 uses $2R_2 + 2$ evaluations at M_2, each with k repetitions, giving $Q' \leq [(B+1)(2M_1+1) + (2R_2+2)(2M_2+1)]k \leq 900(2R_2+1)^2 \ln((B+2R_2+3)/\delta)/(\sqrt{\alpha}\,\varepsilon)$ for $\alpha\varepsilon \leq 10^{-2}$. Write $n = 2R_2 + 1$ and $L = \ln((B + 2R_2 + 3)/\delta)$. Under $\alpha\varepsilon \leq 10^{-2}$, $\varepsilon \leq 1$, $\delta \leq 1/2$ and $\alpha \leq 1/4$ we have $n \geq 25$, $\sqrt{\alpha}\,\varepsilon \leq 0.1$, $B + 1 \leq n$ and $L \geq \ln 72$. Bounding every ceiling directly,

$$(B+1)(2M_1+1)+(2R_2+2)(2M_2+1) \;<\; 161.8\,\frac{n^2}{\sqrt{\alpha}\,\varepsilon}, \qquad k \;\leq\; \Big(5.184+\tfrac{1}{\ln 72}\Big)L \;<\; 5.418\,L,$$

whose product is below $877\,n^2L/(\sqrt{\alpha}\,\varepsilon)$, so the constant 900 is conservative. The margin is real but not large. Sweeping the admissible domain and maximizing over δ at the discontinuities of $k = \lceil 5.184L \rceil$, the largest prefactor actually required is 833.0, attained near $\alpha = 0.011$, $\varepsilon = 0.69$, so 900 holds with about 7% of headroom; the Phase-2 term alone carries $2(16\pi\sqrt{2})(5.184) \approx 737$, so no substantially smaller constant is available. The failure budget is allocated over the same $B+1$ Phase-1 tests and $2R_2 + 2$ Phase-2 evaluations, i.e. over $B + 2R_2 + 3$ estimator calls, which is the union bound already used in the definition of k. □

E PROOFS FOR SECTION 4

Throughout, $\varepsilon_r = 2^{-r}$, $\delta_{i,r} = \delta/(Kr(r+1))$, and E is the event that every estimator call meets its guarantee, so $\mathbb{P}(E) \geq 1 - \delta$ by $\sum_{r \geq 1} 1/(r(r+1)) = 1$. Write Λ_r for the polylogarithmic prefactor of the r-th call. Both algorithms are run at $\alpha \leq 1/4$ and start the schedule at

$$r_{\min} \;:=\; \max\{1,\, \lceil \log_2(100\alpha) \rceil\} \;\leq\; 5,$$

so that $\alpha\varepsilon_r = \alpha 2^{-r} \leq 10^{-2}$ for every $r \geq r_{\min}$ and Lemma 2 is invoked only inside its stated domain. The failure budget is unaffected: restricting to $r \geq r_{\min}$ only shrinks the allocation, since

$\sum_{r \geq r_{\min}} 1/(r(r+1)) = 1/r_{\min} \leq 1$. Write $\varepsilon_{\bar{r}}$ for the accuracy of the terminal call of an arm and set

$$J := \left\lceil \ln \tfrac{1}{\alpha \varepsilon_{\bar{r}}} / \ln \tfrac{3}{2} \right\rceil, \qquad B := \left\lceil \log_2 \tfrac{4}{\alpha \varepsilon_{\bar{r}}} \right\rceil,$$

the round and bisection counts Lemma 2 supplies at that accuracy.

Algorithm 3: QCVaR-SE$(O_1, \ldots, O_K, \alpha, \delta)$

$S \leftarrow [K]$
for $r = r_{\min}, r_{\min}+1, \ldots$ **do**
  foreach $i \in S$ **do** $\hat{C}_i \leftarrow \textsc{Est}(O_i, \alpha, \varepsilon_r, \delta_{i,r})$
  $\hat{C}_{\max} \leftarrow \max_{j \in S} \hat{C}_j$
  $S \leftarrow S \setminus \{i \in S : \hat{C}_i < \hat{C}_{\max} - 2\varepsilon_r\}$
  if $|S| = 1$ **then return** the arm in S
end

Algorithm 4: QCVaR-UCB$(O_1, \ldots, O_K, \alpha, T, \delta)$

foreach i **do** $r_i \leftarrow r_{\min}$; $\hat{C}_i \leftarrow \textsc{Est}(O_i, \alpha, \varepsilon_{r_{\min}}, \delta_{i,r_{\min}})$; $\beta_i \leftarrow \varepsilon_{r_{\min}}$
while ***true*** **do**
  $s \leftarrow \arg\max_i (\hat{C}_i + \beta_i)$ `// optimism`
  $r' \leftarrow r_s + 1$
  if $Q_{\text{spent}} + q(s, r') > T$ **then** play arm s for the remaining $T - Q_{\text{spent}}$ rounds and **return** $\arg\max_i \hat{C}_i$
  $r_s \leftarrow r'$
  $\hat{C}_s \leftarrow \textsc{Est}(O_s, \alpha, \varepsilon_{r_s}, \delta_{s,r_s})$; $\beta_s \leftarrow \varepsilon_{r_s}$
end

Here $q(i, r)$ is the deterministic worst-case query cost of $\textsc{Est}(O_i, \alpha, \varepsilon_r, \delta_{i,r})$, known before the call is made, and Q_{spent} is the number of queries already used; the algorithm is run only when $T \geq Q_{\text{init}}$, the cost of the K initialization calls, and otherwise outputs an arbitrary default. The algorithm therefore uses *exactly* T queries: when the next estimator call does not fit, the remaining budget is spent on the currently selected arm s rather than left idle.

Each estimate is fresh, so no anytime property of the estimator is required; this is the role played by the doubling trick of Wu et al. (2023).

Proof of Theorem 4. The optimal arm survives. For any $j \in S$ at round r, on E we have $\hat{C}_{i^\star} \geq C_{i^\star} - \varepsilon_r \geq C_j - \varepsilon_r \geq \hat{C}_j - 2\varepsilon_r$, so $\hat{C}_{i^\star} \geq \hat{C}_{\max} - 2\varepsilon_r$ and $i^\star$ is never eliminated.

Elimination round. Let $\bar{r}_i := \max\{r_{\min}, \lfloor \log_2(4/\Delta_i) \rfloor + 1\}$, the first round at or after $r_{\min}$ with $\varepsilon_r < \Delta_i/4$ strictly. Then on E,

$$\hat{C}_{i^\star} \;\geq\; C_{i^\star} - \varepsilon_r \;=\; C_i + \Delta_i - \varepsilon_r \;\geq\; (\hat{C}_i - \varepsilon_r) + \Delta_i - \varepsilon_r \;>\; \hat{C}_i + 4\varepsilon_r - 2\varepsilon_r \;=\; \hat{C}_i + 2\varepsilon_r,$$

the strict step using $\Delta_i > 4\varepsilon_r$. Hence $\hat{C}_i < \hat{C}_{\max} - 2\varepsilon_r$ strictly and arm i is eliminated at round $\bar{r}_i$. The strictness matters: at $\varepsilon_r = \Delta_i/4$ exactly, the chain yields only a non-strict inequality, which the elimination rule does not act on.

Cost. Arm i is estimated in rounds $r_{\min}, \ldots, \bar{r}_i$, costing

$$\sum_{r_{\min} \leq r \leq \bar{r}_i} Q(\varepsilon_r, \delta_{i,r}, \alpha) \;\leq\; \Lambda_{\bar{r}_i} \, 2^{\bar{r}_i + 1}/\sqrt{\alpha} \;\leq\; 64 \, \Lambda_{\bar{r}_i}/(\sqrt{\alpha}\, \Delta_i),$$

using $2^{\bar{r}_i} \leq \max\{2^{r_{\min}}, 2^{\lfloor \log_2(4/\Delta_i) \rfloor + 1}\} \leq \max\{32, 8/\Delta_i\} \leq 32/\Delta_i$, which holds because $r_{\min} \leq 5$ and $\Delta_i \leq 1$. The survivor $i^\star$ is estimated through $\bar{r}_{\max} = \max_i \bar{r}_i$, contributing one additional term at $\Delta_{\min}$. Since $900 \cdot 64 = 57600$, the stated bound follows, with the survivor term at $\Delta_{\min}$ verified separately under the Lemma 2 accounting.

□

Proof of Theorem 5. Selection inequality. Suppose arm $i \neq i^\star$ is selected while its radius is $\beta_i = \varepsilon_r$. Selection means $\hat{C}_i + \beta_i \geq \hat{C}_{i^\star} + \beta_{i^\star}$. On E, $\hat{C}_{i^\star} + \beta_{i^\star} \geq C_{i^\star}$ and $\hat{C}_i \leq C_i + \beta_i$, so $C_i + 2\varepsilon_r \geq C_{i^\star}$, that is $\Delta_i \leq 2^{1-r}$ and $r \leq \log_2(4/\Delta_i)$.

Terminal rounds. If the call at round $r_s + 1$ on the selected arm s does not fit in the remaining budget, the last $N_{\mathrm{rem}} := T - Q_{\mathrm{spent}} < q(s, r_s + 1)$ rounds are spent on s. If $s = i^\star$ they incur no regret. Otherwise the selection inequality applies at s, so $\Delta_s \leq 2\varepsilon_{r_s}$, and by Lemma 2 at accuracy $\varepsilon_{r_s+1} = \varepsilon_{r_s}/2$,

$$\Delta_s N_{\mathrm{rem}} \;\leq\; 2\varepsilon_{r_s}\, q(s, r_s+1) \;\leq\; 2\varepsilon_{r_s} \cdot \frac{900\,(2J_s+1)^2 \ln\big((B_s + 2J_s + 3)/\delta_{s,r_s+1}\big)}{\sqrt{\alpha}\,\varepsilon_{r_s}/2} = \frac{3600\,(2J_s+1)^2 \ln\big((B_s + 2J_s + 3)/\delta_{s,r_s+1}\big)}{\sqrt{\alpha}},$$

which is one further additive term of the same $\widetilde{O}(\alpha^{-1/2})$ form as a single arm's contribution, and is the term displayed separately in Theorem 5. The ε_{r_s} cancels exactly, as it does for the per-arm terms.

Call cap. Arm i receives one initialization plus one call per selection, so its last round L_i satisfies $L_i \leq \max\{r_{\min},\ \log_2(4/\Delta_i) + 1\}$.

Query cost and cancellation. Summing the geometric cost as above, $\mathrm{Queries}(i) \leq 64\Lambda_{L_i}/(\sqrt{\alpha}\,\Delta_i)$. Each query to arm i incurs regret Δ_i, so arm i contributes $\Delta_i \cdot 64\Lambda_{L_i}/(\sqrt{\alpha}\,\Delta_i) = 64\Lambda_{L_i}/\sqrt{\alpha}$: the gap cancels exactly. Queries to $i^\star$ incur zero regret. Summing over arms and substituting Λ gives the stated bound. Finally, each call at round r costs at least $\Omega(2^r/\sqrt{\alpha})$ queries, so the budget T caps $L_i \leq \log_2(\sqrt{\alpha}T) + O(1)$ regardless of the gaps, which yields the unconditional polylogarithmic form. □

Proof of Proposition 3. Inspection of the two proofs above shows that the only properties of the objective used are: each C_i is a fixed real number in $[0, 1]$; the estimator guarantee $|\hat{C}_i - C_i| \leq \varepsilon_r$ on E; and the differences $C_{i^\star} - C_i$. No step invokes linearity, additivity, or any distributional decomposition of CVaR_α. This is necessary for soundness, since CVaR_α is not linear in ν, and sufficient for the substitution of Lemma 2 for Lemma 1, which changes only the query-cost function. □

F Proof of Theorem 6

Proof. Instances. Let $m_0 = \alpha/2$ and $m_1 = \alpha/2 + 3\alpha\varepsilon$. Both lie in $(0, \alpha)$ since $6\varepsilon < 1$ for $\varepsilon \leq 1/16$. By (9), $\mathrm{CVaR}_\alpha(\nu_{m_0}) - \mathrm{CVaR}_\alpha(\nu_{m_1}) = 3\varepsilon$. An estimator with additive error ε and success probability $2/3$ therefore distinguishes the two instances with probability at least $2/3$, since the two ε-balls are disjoint.

Oracle completion. The qsample oracle is specified only on $|0\rangle$, and the algorithm must succeed for every valid completion, so the lower bound may select one. Let $|\varphi_m\rangle = \sqrt{m}|0\rangle + \sqrt{1-m}|1\rangle$ be the coin state and let $P = \mathrm{span}\{|\varphi_{m_0}\rangle, |\varphi_{m_1}\rangle\}$. Define $O_{m_1} := WO_{m_0}$ where W acts as the rotation R_θ within P carrying $|\varphi_{m_0}\rangle$ to $|\varphi_{m_1}\rangle$, and as the identity on $P^\perp$. Then

$$\|O_{m_1} - O_{m_0}\| = \|(W - I)O_{m_0}\| = \|W - I\| = 2\sin(\theta/2) \leq \theta, \tag{13}$$

where $\theta = \arccos\langle\varphi_{m_0}|\varphi_{m_1}\rangle$.

Bounding θ. Expanding,

$$1 - \langle\varphi_a|\varphi_b\rangle = \tfrac{1}{2}\Big[(\sqrt{a} - \sqrt{b})^2 + (\sqrt{1-a} - \sqrt{1-b})^2\Big], \tag{14}$$

verified by expanding the right side. With $a = m_0$, $b = m_1$, $|a - b| = 3\alpha\varepsilon$, and $\sqrt{a} + \sqrt{b} \geq 2\sqrt{\alpha/2} = \sqrt{2\alpha}$,

$$(\sqrt{m_0} - \sqrt{m_1})^2 = \frac{(3\alpha\varepsilon)^2}{(\sqrt{m_0} + \sqrt{m_1})^2} \;\leq\; \frac{9\alpha^2\varepsilon^2}{2\alpha} \;=\; \tfrac{9}{2}\,\alpha\varepsilon^2,$$

$$(\sqrt{1-m_0} - \sqrt{1-m_1})^2 \leq \frac{9\alpha^2\varepsilon^2}{2(1-\alpha)} \;\leq\; 6\,\alpha\varepsilon^2,$$

using $1-\alpha \geq 3/4$. Hence $1-\cos\theta \leq \frac{1}{2}(\frac{9}{2}+6)\alpha\varepsilon^2 = \frac{21}{4}\alpha\varepsilon^2$. Jordan's inequality gives $\sin(\theta/2) \geq \theta/\pi$ on $[0,\pi/2]$, so $1-\cos\theta = 2\sin^2(\theta/2) \geq 2\theta^2/\pi^2$ and therefore

$$\theta \;\leq\; \pi\sqrt{\tfrac{1-\cos\theta}{2}} \;\leq\; \pi\sqrt{\tfrac{21}{8}}\,\sqrt{\alpha}\,\varepsilon \;\leq\; 5.09\,\sqrt{\alpha}\,\varepsilon. \tag{15}$$

Hybrid argument. Let $|\psi_T^m\rangle$ be the algorithm's state after T queries against oracle O_m, with all non-oracle unitaries fixed. Telescoping, $\||\psi_T^{m_0}\rangle - |\psi_T^{m_1}\rangle\| \leq T\max\|O_{m_0}^{(\cdot)} - O_{m_1}^{(\cdot)}\|$ over the three query types. For the plain oracle this is $\leq \theta$; for the inverse, $\|O_{m_0}^\dagger - O_{m_1}^\dagger\| = \|(O_{m_0}-O_{m_1})^\dagger\| = \|O_{m_0}-O_{m_1}\|$; for the controlled version, $cO_{m_0} - cO_{m_1} = |1\rangle\langle 1|\otimes(O_{m_0}-O_{m_1})$ is block-diagonal with a zero block, so its operator norm equals $\|O_{m_0}-O_{m_1}\|$. Hence $\||\psi_T^{m_0}\rangle - |\psi_T^{m_1}\rangle\| \leq T\theta$.

Distinguishing. For pure states, success probability p obeys $p \leq \frac{1}{2}(1+D)$ with $D = \sqrt{1-|\langle\psi|\phi\rangle|^2}$. Success $2/3$ forces $D \geq 1/3$, hence $|\langle\psi_T^{m_0}|\psi_T^{m_1}\rangle| \leq \sqrt{8}/3$ and $\||\psi_T^{m_0}\rangle - |\psi_T^{m_1}\rangle\|^2 \geq 2(1-\sqrt{8}/3) = 0.1144$, so the distance is at least 0.338. Combining, $T\theta \geq 0.338$, so $T \geq 0.338/(5.09\sqrt{\alpha}\,\varepsilon) = 0.0665/(\sqrt{\alpha}\,\varepsilon) \geq \frac{1}{16\varepsilon\sqrt{\alpha}}$. □

Remark 12. The constant is not optimized; using $\|O_{m_0}-O_{m_1}\| = 2\sin(\theta/2)$ directly rather than the relaxation $\leq\theta$ would improve it. We report $1/16$ as a safe rounded value of the 0.0665 obtained above.

G Proof of Theorem 7

We follow the weighted adversary argument of Wang et al. (2021, Appendix D), recomputing the oracle-dependent quantities for the mass-parameterized coin.

Proof of Theorem 7. **Hard family.** In the baseline instance, arm i carries ν_{m_i} with masses chosen so that $C_i = 1 - m_i/\alpha$ realizes the target gaps, that is $\Delta_i = (m_i - m_1)/\alpha$, with all masses in $[\alpha/4, \alpha]$. Concretely, set

$$m_1 = \frac{\alpha}{2}, \qquad m_i = \alpha\Big(\frac{1}{2} + \Delta_i\Big)\;\; (i>1), \qquad \eta = \tfrac{1}{2}\Delta_{\min}, \qquad \Delta_{\min} = \min_{i>1}\Delta_i,$$

so that $m_i \in [\alpha/2, \alpha]$ for every $\Delta_i \leq 1/2$. In instance $x \geq 2$, arm x has its mass lowered to $m_x' = m_1 - \alpha\eta = \alpha(\frac{1}{2}-\eta) \in [\alpha/4, \alpha/2]$, making it optimal by margin η; write $\Delta_x' = \Delta_x + \eta \leq \frac{3}{2}\Delta_x$. Baseline and perturbed masses therefore all lie in $[\alpha/4, \alpha]$, which is what the transfer step below requires.

Transfer. The adversary progress measure is $s_k = \sum_{x>1}\Delta_x'^{-2}\langle\psi_x^{(k)}|\psi_1^{(k)}\rangle$. Its boundary values are oracle-independent inner-product facts: $s_0 = \sum_{x>1}\Delta_x'^{-2}$ and $|s_T| \leq 2\sqrt{\delta(1-\delta)}\sum_{x>1}\Delta_x'^{-2}$. The per-query change depends on the oracle family only through the matrix $A(a)^\dagger A(b) - I$ at the single differing arm, whose two entries are

$$u = \sqrt{(1-a)(1-b)} + \sqrt{ab} - 1, \qquad v = \sqrt{(1-a)b} - \sqrt{a(1-b)}, \tag{16}$$

and no coupling across arms. This is what permits the transfer; only u and v must be recomputed. As in Appendix F we must also check the access model of (2). We take the symmetric two-dimensional completion of the coin, so that the oracles for the baseline and the perturbed instance differ only inside the two-dimensional span of the two coin states at the differing arm. An inverse query contributes the same amount, since $\|\mathcal{O}_a^\dagger - \mathcal{O}_b^\dagger\| = \|(\mathcal{O}_a-\mathcal{O}_b)^\dagger\| = \|\mathcal{O}_a - \mathcal{O}_b\|$, and a controlled query contributes the same because $c\mathcal{O}_a - c\mathcal{O}_b = |1\rangle\langle 1|\otimes(\mathcal{O}_a - \mathcal{O}_b)$ is block-diagonal with a zero block. The per-query progress bound below therefore holds for all three query types.

The diagonal entry is quadratic in the gap. By (14), $|u| = \frac{1}{2}[(\sqrt{a}-\sqrt{b})^2 + (\sqrt{1-a}-\sqrt{1-b})^2]$. Here $|a-b| = \alpha\Delta_x'$, the mass shift corresponding to a CVaR gap of Δ_x'. With $a, b \geq \alpha/4$ we get $\sqrt{a}+\sqrt{b} \geq \sqrt{\alpha}$, so $(\sqrt{a}-\sqrt{b})^2 \leq \alpha\Delta_x'^2$, and similarly $(\sqrt{1-a}-\sqrt{1-b})^2 \leq \frac{2}{3}\alpha^2\Delta_x'^2$. Hence

$$|u| \;\leq\; \tfrac{7}{12}\,\alpha\,\Delta_x'^2. \tag{17}$$

The off-diagonal entry is linear in the gap. Rationalizing, $v = (b-a)/(\sqrt{(1-a)b}+\sqrt{a(1-b)})$, since the numerator $(1-a)b-a(1-b) = b-a$. The denominator is a *sum* of two square roots, and by Wang et al. (2021, Lemma 8) it is at least $2\sqrt{p(1-p)}$ for masses in $[p, 1-p]$. Taking $p = \alpha/4$ gives $2\sqrt{(\alpha/4)(1-\alpha/4)} = \sqrt{\alpha(1-\alpha/4)}$, so

$$|v| \;\le\; \frac{\alpha\Delta'_x}{\sqrt{\alpha(1-\alpha/4)}} \;=\; \frac{\sqrt{\alpha}\,\Delta'_x}{\sqrt{1-\alpha/4}} \;\le\; 1.04\,\sqrt{\alpha}\,\Delta'_x \qquad (\alpha \le 1/4). \tag{18}$$

Cauchy-Schwarz. Following Wang et al. (2021, Eqs. 59–60), the diagonal contribution is bounded by $\sqrt{\sum |u|^2\Delta'^{-4}|\alpha_\cdot|^2}$. Since $|u|^2\Delta'^{-4} \le (7/12)^2\alpha^2$ is *constant* in the gap, this term does not scale with $\sqrt{\sum \Delta'^{-2}}$. It is not, however, a flat $O(\alpha)$: the amplitudes entering it belong to the $K-1$ distinct states $|\psi_x^{(k)}\rangle$, whose squared norms sum to $K-1$ and not to 1, so the direct bound over the two (b, b') pairs is $2 \cdot \frac{7}{12}\alpha\sqrt{K-1}$. Write $S = \sum_{x>1} \Delta_x'^{-2}$. Since $\Delta_i \le 1/2$ and $\Delta'_x \le \frac{3}{2}\Delta_x$ give $\Delta'_x \le 3/4$, we have $S \ge \frac{16}{9}(K-1)$, i.e. $\sqrt{K-1} \le \frac{3}{4}\sqrt{S}$, so the diagonal contributes at most $\frac{7}{8}\alpha\sqrt{S} \le \frac{7}{16}\sqrt{\alpha}\,\sqrt{S}$ using $\alpha \le 1/4$, and is absorbed by the off-diagonal term up to a constant. The off-diagonal gives $|v|^2\Delta'^{-4} \le 1.08\,\alpha\Delta_x'^{-2}$, so summing over the two (b, b') pairs yields

$$|s_{k+1} - s_k| \;\le\; \tfrac{7}{16}\sqrt{\alpha}\,\sqrt{S} + 2 \cdot 1.04\,\sqrt{\alpha}\,\sqrt{S} \;\le\; 2.52\,\sqrt{\alpha}\,\sqrt{S}. \tag{19}$$

It is essential that $|v|$ is genuinely linear in Δ'. Were both entries quadratic, both contributions would be flat and the argument would give no gap-dependent bound.

Conclusion. Combining with the boundary values,

$$T \;\ge\; \frac{1-2\sqrt{\delta(1-\delta)}}{2.52\sqrt{\alpha}}\,\sqrt{\textstyle\sum_{x>1}\Delta_x'^{-2}} \;\ge\; \frac{1-2\sqrt{\delta(1-\delta)}}{2.52\sqrt{\alpha}} \cdot \frac{2\sqrt{H}}{3} \;\ge\; \frac{1-2\sqrt{\delta(1-\delta)}}{9}\sqrt{\frac{H}{\alpha}},$$

using $\Delta'_x \le \frac{3}{2}\Delta_x$ in the middle step, valid for $0 < \delta < 1/2$, and $2/(3 \cdot 2.52) = 0.264 \ge 1/9$ in the last. The formal substitution $\alpha \to 1$ would recover the $\Omega(\sqrt{H})$ bound of Wang et al. (2021) up to constants, but $\alpha = 1$ lies outside the standing assumption $\alpha \le 1/4$ of Theorem 7, so this is an analogy only and not a consistency check. □

H THE THREE ROUTES OF SECTION 6

Each route below is a way one might hope to run variable-time amplitude amplification over a coherent CVaR estimator. None currently yields the required bound.

Coherent execution of the adaptive search. VTAA requires the inner routine to run without measurement, but our estimator branches on comparisons between amplitude estimates. This is an open technical question rather than a proved barrier: by deferred measurement the comparison outcomes can be held in ancillas and the interval endpoints updated reversibly, so the 2^R classical branches need not be enumerated. What is missing is a coherent routine with a controlled stopping rule, controlled error accumulation, and clean uncomputation of the amplitude-estimation garbage.

Non-adaptive replacement. The integral representation $\mathrm{CVaR}_\alpha = \frac{1}{\alpha}\int_0^1 (\alpha - F(y))_+ \,\mathrm{d}y$ (Thomas & Learned-Miller, 2019) is coherent by construction, but evaluating it over $\Theta(1/\varepsilon)$ bins at per-bin cost $\Theta(1/(\alpha\varepsilon))$ gives $\Theta(1/(\alpha\varepsilon^2))$, losing the quadratic advantage. The search was what avoided touching every bin.

Nested-expectation machinery. That integral is a nested expectation with a 1-Lipschitz outer positive part and prefactor $1/\alpha$, suggesting composition with Blanchet et al. (2025). The composition is valid, but their cost is controlled by a worst-case bound on the inner second moment rather than by the local variance near the quantile, so the route yields $\widetilde{O}(1/(\alpha\varepsilon))$ and does not recover the $\sqrt{\alpha}$. Blanchet et al. (2025) themselves list CVaR as future work.

I EXPERIMENTAL DETAILS

Setup. All numbers in Figure 2 are query counts, not wall-clock times. Panels (a) and (b) evaluate the proven bounds with the constants exactly as stated in Lemmas 1 and 2 and the classical radius of Proposition 9 at the explicit constant fixed in Appendix B; no fitting is involved. Figure 3(c) locates the crossover by bisection on ε for each α.

Query cost. Algorithm 1's cost is not a random variable: it issues $(2R + 1)$ evaluations, each running k amplitude estimations of at most $2M+1$ queries, with R, M and k all fixed in advance by $(\alpha, \varepsilon, \delta)$. Panel (a) therefore reports the computed value $(2R+1)k(2M+1)$, which at $\varepsilon = 4.5\times10^{-3}$ is 1,338,041,915 queries, against the Lemma 1 bound of 1.41×10^9; the 4.9% gap is ceiling slack, and it never exceeds 6.9% over the range plotted. The third curve is a diagnostic, not a run of Algorithm 1: it sizes each amplitude estimation by choosing the smallest M for which the two terms of (12), evaluated at the *exact* amplitude $a_t = \mathbb{E}[(t - X)_+]$, are each at most half the target. That rule uses the very quantity being estimated and is unavailable to the algorithm, so its cost is a lower envelope rather than an achievable count. We report it because the envelope is stable at 3.06 to 3.21 times below Algorithm 1 across the range, which measures directly how much is lost by sizing M for the worst case $a(1 - a) \le 1/4$ when the amplitudes near the quantile are of order α. That is the premise Lemma 2 turns into a proved $\sqrt{\alpha}$.

Accuracy. Panel (b) runs Algorithm 1 verbatim, at $M = \lceil 4\pi(2R + 1)/(\alpha\varepsilon)\rceil$ and $k = \lceil 5.184 \ln((2R+1)/\delta)\rceil$, on $X \sim \mathrm{Unif}[0, 1]$, for which $\mathrm{CVaR}_\alpha = \alpha/2$ and the maximizer is interior, so the search is non-trivial. Amplitude-estimation outcomes are drawn from the exact measurement distribution induced by phase estimation on the Grover operator, the sum of two Fejér kernels centered at $M\phi_a$ and $M(1 - \phi_a)$ with $a = \sin^2(\pi\phi_a)$. The sampler enumerates a window around each peak; the discarded tail never exceeded 6.8×10^{-4} of the mass and is handled conservatively, by replacing a draw with a uniformly random grid outcome with exactly that probability, so truncation cannot flatter the measured success rate. Over 40 seeded repetitions at each of five accuracies, every run met the ε guarantee; the worst observed error was $1.4 \times 10^{-3}\varepsilon$ and the median was about $6 \times 10^{-4}\varepsilon$. This tests the parts of the proof that are not deterministic: the median amplification of Lemma 10, the errant-cut bound of Step 2, and the accuracy budget of Step 3.

The margin is large and its source is identifiable. Step 3 charges the final interval at the Lipschitz rate $1/\alpha$, contributing $\varepsilon/2$ of the budget, but $\phi'(t^\star) = 1 - F(t^\star)/\alpha$ vanishes whenever F is continuous at the quantile, so on this test distribution the term is second order in the interval width rather than first. The bound cannot simply drop it: on the atomic two-point instances of Section 5, which are the ones that drive the lower bound, ϕ' does not vanish at $t^\star$. A distribution-dependent refinement that exploits continuity where it holds would tighten the constant 140 substantially, and is the most promising direction for the tightening noted in Remark 11.

Constant verification. The numerical constants in this paper are checked against an independent computation rather than by hand, and the check is now applied end to end, including every ceiling and every downstream constant. This includes the amplitude-estimation iterate count, the median-amplification constant 5.184 and its Hoeffding exponent, the elimination threshold of Theorem 4, the Jordan-inequality coefficient 5.09 and the resulting constant $1/16$ of Theorem 6, and the denominator infimum underlying the off-diagonal bound of Theorem 7. We adopted this practice after finding that a hand-derived value for the last of these was incorrect by a factor of two, an error that survived two independent manual checks and was caught only by numerical evaluation. The composed end-to-end bounds of Lemmas 1 and 2 are checked the same way, by evaluating the exact query count against the stated bound over the admissible domain, rather than by composing the individual constants by hand. The verification script is included with the supplementary material.

Reproducibility. Seed 20260720; all figures regenerate from three scripts sharing one style module and one seed. The simulation uses no hardware quantum resources and runs in under a minute on a laptop.