# TIGHT RATES FOR BANDIT FIRST-PRICE AUTOBIDDING

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Automated bidding is widely used to maximize advertising value under return-on-investment (ROI) and budget constraints. We study stochastic first-price auctions with independent values and competing bids under bandit feedback, where the bidder learns only whether it wins each auction. Under an ROI constraint alone, we first show that no algorithm can guarantee both exact expected ROI and sublinear regret. When sublinear constraint violation is allowed, however, our algorithm achieves a $\widetilde{O}(T^{2/3})$ regret bound against the optimal randomized strategy in hindsight, with a violation bound of the same order. This improves the $\widetilde{O}(T^{3/4})$ regret bound of Li et al. (2025) and closes the gap they left open under bandit feedback. We also construct instances that give a matching lower bound for this regret and violation guarantee. When a budget constraint is also imposed, our algorithm achieves $\widetilde{O}(T^{2/3})$ regret and ROI violation bounds for budgets proportional to the horizon. The regret is measured against the optimal randomized strategy in hindsight subject to both constraints, and matches the known lower bound of Aggarwal et al. (2025) in this independent model. The improvement comes from continually updating optimistic estimates of win probabilities and controlling how estimation errors affect regret, without amplification by the constraint multipliers.

## 1 INTRODUCTION

Autobidding has become a standard approach to managing online advertising campaigns. Advertisers specify their objectives and constraints, while an automated bidding algorithm decides how much to bid for each advertising opportunity. In many applications, the goal is to maximize the total value obtained from a campaign subject to a return-on-investment (ROI) target, possibly together with a budget constraint (Aggarwal et al., 2024). The ROI target requires the value generated over the campaign to be sufficiently large relative to total spending, while the budget limits how much the advertiser can spend. The algorithm must therefore choose bids as opportunities arrive, taking into account their contribution to the campaign as a whole. Our focus is on first-price auctions, where the winner pays its own bid. In this setting, a higher bid increases the chance of winning but also the payment upon winning, making the choice of bids central to meeting the advertiser's objectives and constraints.

The bidder must make these decisions with limited information about competing bids. We consider a stochastic setting in which values and competing bids are drawn independently from unknown distributions. The bidder observes its value before each auction, but afterwards learns only whether it won. Under this bandit feedback, each bid affects both the value obtained and the information available for future decisions. Following Li et al. (2025), we measure regret against the optimal randomized strategy in hindsight, which knows the competing-bid distribution and the entire value sequence in advance and satisfies the relevant constraints in expectation. Our goal is to achieve low regret relative to this benchmark while approximately satisfying the ROI constraint over the campaign. When a budget constraint is also imposed, total spending must stay within the budget.

Two recent papers are especially relevant here. Aggarwal et al. (2025) study ROI and budget constraints in nontruthful auctions, allowing correlation between values and competing bids and comparing with a restricted distributional benchmark. Li et al. (2025) study first-price auctions with ROI alone and the unrestricted randomized hindsight benchmark used here. Both obtain $\widetilde{O}(T^{3/4})$ regret



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.

under bandit feedback. An $\Omega(T^{2/3})$ regret lower bound is also known for the setting with a budget constraint (Aggarwal et al., 2025). This leaves open whether the two-thirds rate can be achieved with ROI alone, or with a budget against the randomized hindsight benchmark. We answer both questions in the independent first-price model.

**Our results.** Under an ROI constraint alone, our algorithm achieves a $\widetilde{O}(T^{2/3})$ regret bound against the optimal randomized strategy in hindsight, together with a violation bound of the same order (Theorem 1). Allowing some constraint violation is necessary: Proposition 1 shows that sublinear regret is impossible when the constraint must hold exactly in expectation. We prove a matching lower bound (Theorem 2), establishing a tight regret and ROI violation bound.

With an additional budget constraint, our algorithm keeps total spending within the budget. For budgets proportional to the horizon, Theorem 3 gives a $\widetilde{O}(T^{2/3})$ regret bound against the optimal randomized strategy in hindsight subject to both constraints, together with an ROI violation bound of the same order. Combined with the lower bound of Aggarwal et al. (2025), this establishes a tight regret bound in the independent model. The same regret guarantee also holds against another standard distributional benchmark: the best randomized bidding rule that knows both distributions and chooses bids based on the current value, with both constraints satisfied in expectation (Corollary 1).

**Technique overview.** We begin by discretizing the bid space and maintaining optimistic estimates of the win probabilities at the grid points. Since values and competing bids are independent, the algorithm only needs to estimate one win probability for each bid, without estimating it separately for different values. In each auction, the algorithm uses these estimates and the current value to score candidate bids. Each score combines predicted value with an ROI multiplier times the predicted difference between value and payment. The candidate with the highest score determines the submitted bid. The observed win or loss then updates the probability estimates, while the predicted difference between value and payment updates the multiplier. Both updates enter the next round's scores. The bandit algorithm of Li et al. (2025) estimates win probabilities during an initial exploration phase and uses these fixed estimates for the remaining auctions. Our algorithm continues to learn throughout bidding, allowing the analysis to track estimation error at the bids actually submitted. The key to controlling cumulative estimation error is that frequently selected bids have increasingly accurate win-probability estimates, while rarely selected bids contribute little to the total error.

However, small errors in the estimated win probabilities can still lead to large errors in the bid scores, since these scores also depend on an ROI multiplier that may grow over time. To handle this difficulty, our analysis first bounds the value and constraint violation predicted using the upper confidence bounds on win probabilities. The maximum estimated score allows a comparison with the hindsight benchmark. The multiplier update controls the contribution of the ROI term to this score, giving bounds on predicted value and constraint violation even as the estimates change. We then relate these predictions to actual outcomes by comparing value and payment separately. In this comparison, probability estimation errors are weighted only by values and bids, both of which are bounded. The cumulative error bound from the preceding paragraph therefore enters the regret and violation guarantees without an additional factor from the ROI multiplier.

An additional budget constraint introduces the risk of exhausting the budget early and missing valuable opportunities later. The key observation is that early exhaustion requires spending to run ahead of an even allocation of the budget across auctions. Controlling this excess throughout the horizon therefore limits the loss from early stopping. To obtain this control, we introduce a second multiplier to regulate spending and analyze both multipliers together. A stopping rule based on past spending then ensures that the budget is never exceeded. For budgets proportional to the horizon, the expected loss from stopping early is of the same order as the learning error, yielding the same regret bound as in the ROI-only setting.

### 1.1 Related Work

**ROI-constrained autobidding.** Value maximization under ROI and budget constraints is a central model of autobidding (Aggarwal et al., 2019; 2024). Online algorithms have been developed for truthful auctions (Feng et al., 2023; Vaze & Sinha, 2025) and for nontruthful auctions with finite value and bid spaces (Castiglioni et al., 2024). The closest first-price results are those of Li et al. (2025); Aggarwal et al. (2025), discussed above. Our ROI violation measure takes the positive part of total payment minus total value before averaging. This is stronger than the signed expected

difference used by Li et al. (2025). More recently, Hu et al. (2026) obtain $\widetilde{O}(T^{2/3})$ regret with unknown values and binary feedback in a linear contextual model, assuming a Lipschitz continuous CDF for competing bids and a positive feasibility margin. Our guarantees require neither assumption and control the expected positive part of the realized cumulative ROI deficit.

**Learning in first-price auctions.** Learning to bid in first-price auctions has been studied with stochastic and adversarial inputs, different feedback models, and contextual information (Han et al., 2024; 2020; Badanidiyuru et al., 2023). Other work examines bidding when values are unknown and must also be learned (Cesa-Bianchi et al., 2024; Vijayan et al., 2025). Our use of observations collected at one value to guide bidding at other values is related to contextual bandits with cross-learning (Balseiro et al., 2019; Schneider & Zimmert, 2023).

**Budgets and pacing.** Pacing adjusts bids to manage spending over a campaign. Its regret, equilibrium, and welfare properties have been studied in repeated auctions with budget constraints (Balseiro & Gur, 2019; Conitzer et al., 2022; Gaitonde et al., 2023; Fikioris & Tardos, 2024). For first-price auctions, budget-constrained bidding has also been studied with full or one-sided winning-bid feedback (Wang et al., 2023; Fu et al., 2026). Work on ROI and budget constraints examines how their pacing rules interact, how competing autobidders behave, and how budgets and targets are allocated across channels (Balseiro et al., 2024; Lucier et al., 2024; Deng et al., 2023).

**Online learning with constraints.** Bidding with budget constraints is closely related to bandits with knapsacks and their extensions to contextual decisions, convex constraints, and adversarial environments (Badanidiyuru et al., 2018; 2014; Agrawal & Devanur, 2014; 2016; Immorlica et al., 2019; Kesselheim & Singla, 2020; Castiglioni et al., 2022). ROI constraints also connect to models with replenishable resources, since surplus earned in one auction can cover a deficit in another (Kumar & Kleinberg, 2022; Bernasconi et al., 2024b). More generally, dual methods, adaptive regret, and optimistic estimates have been used for online allocation and learning under long-term constraints (Balseiro et al., 2020; Slivkins et al., 2023; Bernasconi et al., 2024a; 2025; Guo et al., 2025). Guo et al. (2026) develop a modular framework for constrained bidding with a known context distribution. Their oracle assumption uses a common error bound across contexts and bids in each round. Our bandit analysis controls cumulative error at submitted bids, without knowing the value distribution.

## 2 PRELIMINARIES

Consider a bidder participating in repeated first-price auctions with time horizon $T$. In each round $t$, the bidder observes its value $v_t \in [0,1]$ and submits a bid $b_t \in [0,1]$. Let $d_t \in [0,1]$ denote the highest competing bid. The bidder wins if $b_t \geq d_t$, with ties broken in its favor, and pays its own bid upon winning. Writing $X_t = \mathbf{1}\{b_t \geq d_t\}$ for the win indicator, the resulting value and payment are

$$R_t = v_t X_t, \qquad P_t = b_t X_t. \tag{1}$$

The bidder aims to maximize total value subject to an ROI constraint and, when present, a budget constraint.

The values $(v_t)_{t=1}^T$ are drawn i.i.d. from an unknown distribution $H$, and the competing bids $(d_t)_{t=1}^T$ are drawn i.i.d. from an unknown distribution $F$. The two sequences are independent. Thus, $F(b) = \Pr(d_t \leq b)$ is the probability of winning with bid $b$. Under bandit feedback, the bidder observes only the win indicator $X_t$ after each auction. Its bid may depend on the current value, past observations, and its own randomization. Throughout the paper, $\widetilde{O}(\cdot)$ hides logarithmic factors.

**Constraints and offline benchmarks.** An ROI target $\rho > 0$ requires total value to be at least $\rho$ times total payment: $\rho \sum_{t=1}^T P_t \leq \sum_{t=1}^T R_t$. Without loss of generality, we set $\rho = 1$.[1] When a budget $B$ is imposed, the online bidder must also satisfy $\sum_{t=1}^T P_t \leq B$ on every sample path.

The offline benchmark knows $F$ and the entire realized value sequence $V^T = (v_1, \ldots, v_T)$ before bidding, but not the realized competing bids. It chooses a randomized bidding strategy

[1] For a fixed target $\rho > 0$, replacing $v_t$ by $v_t/\rho$ reduces the ROI constraint to target one. The rescaled values remain bounded by a constant, so this only changes constants in the regret and violation bounds and leaves the budget unchanged.

$\pi = (\pi_1, \ldots, \pi_T)$, where each $\pi_t$ is a distribution over bids in $[0, 1]$ and may depend on all of $V^T$. Conditional on $V^T$, its expected total value and payment are

$$\mathcal{R}_F(\pi) = \sum_{t=1}^{T} \int v_t F(b)\, \pi_t(db), \qquad \mathcal{P}_F(\pi) = \sum_{t=1}^{T} \int b F(b)\, \pi_t(db). \tag{2}$$

Under an ROI constraint alone, the benchmark is

$$\mathrm{OPT}(F, V^T) = \sup_{\pi:\, \mathcal{P}_F(\pi) \le \mathcal{R}_F(\pi)} \mathcal{R}_F(\pi). \tag{3}$$

With an additional budget constraint, it becomes

$$\mathrm{OPT}_B(F, V^T) = \sup_{\pi:\, \mathcal{P}_F(\pi) \le \mathcal{R}_F(\pi),\ \mathcal{P}_F(\pi) \le B} \mathcal{R}_F(\pi). \tag{4}$$

The benchmark constraints hold in expectation over randomized bids and competing bids, conditional on the value sequence $V^T$. For the online bidder, we measure ROI violation by the expected positive part of total payment minus total value and require total spending to stay within the budget on every run.

**Regret and constraint violation.** All expectations below include the values, competing bids, and the bidder's randomization. We define regret against the respective benchmarks as

$$\mathrm{Reg}_T = \mathbb{E}\left[\mathrm{OPT}(F, V^T) - \sum_{t=1}^{T} R_t\right], \qquad \mathrm{Reg}_T^B = \mathbb{E}\left[\mathrm{OPT}_B(F, V^T) - \sum_{t=1}^{T} R_t\right]. \tag{5}$$

For $(x)_+ = \max\{x, 0\}$, the expected ROI and budget violations are

$$\mathrm{Viol}_T^{\mathrm{ROI}} = \mathbb{E}\left[\left(\sum_{t=1}^{T} (P_t - R_t)\right)_+\right], \qquad \mathrm{Viol}_T^{\mathrm{B}} = \mathbb{E}\left[\left(\sum_{t=1}^{T} P_t - B\right)_+\right]. \tag{6}$$

We write $\mathrm{Viol}_T = \mathrm{Viol}_T^{\mathrm{ROI}}$ in the ROI-only setting. Taking the positive part after summing over rounds allows surplus from one auction to offset a deficit in another. Taking it before expectation ensures that surplus in one run cannot cancel a violation in another.[2]

## 3 Learning with an ROI constraint

In this section, we study bandit bidding under an ROI constraint alone. We first describe an online algorithm that learns win probabilities throughout bidding, and then prove its $\widetilde{O}(T^{2/3})$ regret and ROI violation bounds. The lower bounds explain why allowing some constraint violation is necessary and establish the tightness of this guarantee.

### 3.1 The bidding algorithm

The algorithm uses estimates of win probabilities to compare bids at the current value, together with an ROI multiplier that controls the tradeoff between value and payment. Each auction provides feedback for improving these estimates, while the predicted ROI surplus determines how the multiplier changes. Both are updated throughout bidding, so the algorithm continues to learn as it collects value. To maintain these estimates, we discretize the bid range $[0, 1]$. For the time horizon $T$, choose the number of grid intervals $K = \max\{2, \lceil T^{1/3} \rceil\}$, and let $c_j = j/K$ denote the grid points for $j = 0, \ldots, K$. A finer grid reduces the change caused by rounding a bid, but also leaves more win probabilities to estimate. We distinguish a candidate bid $b$, used to evaluate bidding choices, from the grid bid actually submitted. The rounding rule $\sigma$ maps a candidate bid to its submitted bid as follows:

$$\sigma(b) = \begin{cases} c_{\lfloor Kb \rfloor + 1}, & b < 1, \\ 1, & b = 1, \end{cases} \qquad b \le \sigma(b) \le b + K^{-1}. \tag{7}$$

[2] An alternative measure is the positive part of the expected deficit, $\left(\mathbb{E}[\sum_{t=1}^{T}(P_t - R_t)]\right)_+$. By Jensen's inequality, our measure is at least as large and can be strictly larger.

Thus, all candidate bids in the interval $[c_{j-1}, c_j)$, including its left endpoint, are submitted as the grid bid $c_j$. Rounding upward preserves the chance of winning, even when a particular competing bid occurs with positive probability. On a win, the actual payment exceeds the candidate bid by at most the grid spacing $1/K$.

Since submitted bids lie on the grid, the algorithm only needs to estimate the win probabilities at the positive grid points. The win probability $F(1) = 1$ at the largest grid bid is already known. Moreover, independence of values and competing bids means that observations collected at the same bid estimate the same win probability, regardless of the value observed in that round. The algorithm can therefore use a single estimate for each bid across all values.

Before round $t$, let $n_{t,j}$ denote the number of previous auctions in which the grid bid $c_j$ was submitted, for $j = 1, \dots, K-1$. When the sample count $n_{t,j}$ is positive, let $\widehat{F}_{t,j}$ denote the fraction of those auctions won. To account for uncertainty in these empirical estimates, we use upper confidence bounds. Set the failure probability parameter $\delta = T^{-2}$ and the logarithmic factor $L_T = \log(2KT^2/\delta)$. Define the upper confidence bound $U_{t,j}$ and its adjusted version $\bar{U}_{t,j}$ by

$$\begin{aligned} U_{t,j} &= \begin{cases} 1, & n_{t,j} = 0, \\ \min\{1, \widehat{F}_{t,j} + \sqrt{L_T/(2n_{t,j})}\}, & n_{t,j} > 0, \end{cases} \qquad j = 1, \dots, K-1, \\ U_{t,K} &= 1, \qquad \bar{U}_{t,j} = \min_{j \le i \le K} U_{t,i}. \end{aligned} \tag{8}$$

The adjustment uses a simple property of first-price auctions: increasing the bid cannot reduce the probability of winning. An upper bound at a larger bid is therefore also an upper bound at every smaller bid. Taking the minimum over larger grid bids tightens the estimates and makes them nondecreasing in the bid. The initial upper bound at an untried bid is one; observations at that bid narrow its confidence interval, and observations at larger bids may also improve its adjusted bound. The confidence bounds cover all grid points and sample counts, allowing bidding decisions to depend on previous feedback.

For a candidate bid $b$, let $Q_t(b)$ denote the adjusted upper estimate of the win probability at its submitted bid: $Q_t(b) = \bar{U}_{t,j}$ whenever $\sigma(b) = c_j$. In particular, the candidate bid zero uses the estimate for the submitted bid $\sigma(0) = 1/K$. Submitting this grid bid ensures that the observed outcome concerns the win probability used in the bidding rule. Submitting zero would instead produce an observation with mean $F(0)$, which need not equal the win probability $F(1/K)$.

With these estimates in place, the algorithm can compare the predicted value and payment of different bids. To account for the ROI constraint, let $\lambda_t \ge 0$ denote the ROI multiplier in round $t$, initialized at $\lambda_1 = 1$. After observing the current value $v_t$, the algorithm selects a candidate bid $\widetilde{b}_t$ according to

$$\widetilde{b}_t \in \arg\max_{b \in [0,1]} Q_t(b)\big[(1+\lambda_t)v_t - \lambda_t b\big]. \tag{9}$$

This score is predicted value plus the ROI multiplier times predicted ROI surplus. A larger multiplier gives more weight to the difference between value and payment. Within each grid interval, the probability estimate $Q_t(b)$ is constant and the score is nonincreasing in the candidate bid. Consequently, the maximization only needs to check the grid endpoints $0, c_1, \dots, c_{K-1}, 1$.

The algorithm submits the rounded bid $b_t = \sigma(\widetilde{b}_t)$. Let $q_t = Q_t(\widetilde{b}_t)$ denote its estimated win probability, and let $r_t^Q$ and $g_t^Q$ denote the predicted value and ROI surplus used by the bidding rule. With update step size $\eta = T^{-2/3}$, the multiplier update is

$$r_t^Q = v_t q_t, \qquad g_t^Q = (v_t - \widetilde{b}_t)q_t, \qquad \lambda_{t+1} = \lambda_t \exp(-\eta g_t^Q). \tag{10}$$

A negative predicted surplus increases the ROI multiplier, giving greater weight to ROI surplus in subsequent auctions; a positive surplus decreases it. The predicted surplus uses the candidate bid $\widetilde{b}_t$, while the actual payment uses the submitted bid $b_t$. Their difference is bounded by the grid spacing and is accounted for in the analysis.

The multiplier update uses the same probability estimate $q_t$ that was used to select the bid. After observing the win indicator $X_t$, the algorithm updates the sample count and empirical win frequency at the submitted grid bid, providing the probability estimates for the following round. Bids selected

more often therefore receive more observations and increasingly narrow confidence intervals. Algorithm 1 summarizes the procedure.

**Algorithm 1** Bandit bidding under an ROI constraint

**Require:** Horizon $T$; $K = \max\{2, \lceil T^{1/3} \rceil\}$, $\eta = T^{-2/3}$, $\delta = T^{-2}$

Set $\lambda_1 = 1$; set sample counts $n_j = 0$ and wins $s_j = 0$ for $j = 1, \dots, K-1$
**for** $t = 1, \dots, T$ **do**
  Form $Q_t$ from the current samples using (8)
  Observe $v_t$ and choose $\widetilde{b}_t$ by (9)
  Set $q_t = Q_t(\widetilde{b}_t)$ and $b_t = \sigma(\widetilde{b}_t)$
  Submit $b_t$ and observe the win bit $X_t$
  $\lambda_{t+1} \leftarrow \lambda_t \exp[-\eta(v_t - \widetilde{b}_t)q_t]$
  **if** $b_t = c_j$ for some $j = 1, \dots, K-1$ **then**
    $n_j \leftarrow n_j + 1$; $s_j \leftarrow s_j + X_t$
  **end if**
**end for**

### 3.2 Regret and ROI violation analysis

The following theorem shows that learning win probabilities throughout bidding, together with the ROI multiplier update, gives the desired regret and violation bounds.

**Theorem 1.** *For any unknown competing-bid distribution $F$ and value distribution $H$, Algorithm 1 under bandit feedback achieves expected regret and ROI violation bounded by*

$$\mathrm{Reg}_T = \widetilde{O}(T^{2/3}), \qquad \mathrm{Viol}_T = \widetilde{O}(T^{2/3}). \tag{11}$$

To compare with the hindsight benchmark, let $f_F(v, \lambda)$ be the maximum bid score for value $v$ and ROI multiplier $\lambda$ under the true win probabilities. Averaging over the value distribution $H$ gives

$$\begin{aligned} f_F(v, \lambda) &= \sup_{b \in [0,1]} F(b)\big[(1+\lambda)v - \lambda b\big], \\ D_F(\lambda) &= \mathbb{E}_{v \sim H}\big[f_F(v, \lambda)\big], \qquad D_F^\star = \inf_{\lambda \geq 0} D_F(\lambda). \end{aligned} \tag{12}$$

Every feasible hindsight strategy has nonnegative expected total ROI surplus. Adding this surplus multiplied by $\lambda$ and maximizing in each round therefore yields $\mathbb{E}[\mathrm{OPT}(F, V^T)] \leq T D_F^\star$.

On the confidence event, $Q_t(b) \geq F(\sigma(b)) \geq F(b)$. Hence the maximum estimated score $f_{Q_t}(v, \lambda)$, defined by replacing $F$ with $Q_t$ above, satisfies $f_{Q_t}(v, \lambda) \geq f_F(v, \lambda)$. For nonnegative coefficients $(1+\lambda)v - \lambda b$, this follows directly from the confidence bounds. For negative coefficients, the true score is nonpositive and is dominated by the nonnegative estimated score of candidate bid zero.

The selected score also includes weighted ROI surplus, so a value guarantee requires controlling this additional term. With the update step size $\eta = T^{-2/3}$, the multiplier update bounds $\sum_{t=1}^T \lambda_t g_t^Q$ from above by $O(T^{2/3})$ and the predicted violation $(-\sum_{t=1}^T g_t^Q)_+$ by $\widetilde{O}(T^{2/3})$. Since the estimates and multiplier are fixed before the current value is observed, averaging the selected score connects it to the dual bound. Subtracting the weighted surplus and accounting for confidence failures then gives $\mathbb{E}[\sum_{t=1}^T r_t^Q] \geq \mathbb{E}[\mathrm{OPT}(F, V^T)] - O(T^{2/3})$.

It remains to relate these predictions to actual outcomes. Let $e_t = |q_t - F(b_t)|$ denote the probability estimation error at the submitted bid $b_t$. Summing confidence widths at the submitted bids gives $\mathbb{E}[\sum_{t=1}^T e_t] = O(K + \sqrt{KTL_T} + T\delta)$: frequently submitted bids become well estimated, while rarely submitted bids contribute few error terms.

Let $\mathcal{H}_t$ be the information available after observing the current value and choosing the bid, but before the auction outcome. Then $\mathbb{E}[R_t \mid \mathcal{H}_t] \geq r_t^Q - e_t$ and $\mathbb{E}[P_t - R_t \mid \mathcal{H}_t] \leq -g_t^Q + K^{-1} + e_t$, where $K^{-1}$ accounts for upward rounding. By establishing predicted value and surplus guarantees before making these comparisons, the analysis introduces estimation errors without multiplying them by the ROI multiplier. The cumulative realized ROI deficit differs from its conditional means by a

martingale sum with expected absolute value at most $\sqrt{T}$, which also controls the positive part in our violation measure.

Finally, the choice $K = \max\{2, \lceil T^{1/3} \rceil\}$ balances cumulative estimation error $\widetilde{O}(\sqrt{KT})$ with rounding cost $T/K$, giving the claimed bounds. The full proof appears in Appendix A.

### 3.3 ROI-ONLY LOWER BOUNDS

For any given bandit bidding algorithm, we construct a competing-bid distribution on which either regret or ROI violation is large. The construction improves win probabilities in a region that the algorithm rarely explores.

**Theorem 2.** *For every possibly randomized bandit bidding algorithm and time horizon $T \geq 216$, there is an instance with i.i.d. competing bids and constant value $v_t = 4/13$ such that*

$$\max\{\mathrm{Reg}_T, \mathrm{Viol}_T\} \geq \frac{1}{520} T^{2/3}. \tag{13}$$

Fix an algorithm and consider its behavior under a base competing-bid distribution. This distribution assigns probability $3/4$ to a competing bid of zero, allowing zero bids to earn value without payment. The remaining probability is placed so that mixing zero with different candidate bids can achieve the same expected value while satisfying ROI. Divide part of the bid range into roughly $T^{1/3}$ disjoint intervals. Since the algorithm makes only one bid per round, at least one interval receives few bids in expectation under the base distribution.

We construct the hard instance by moving probability mass from the right endpoint of this interval to its left endpoint. This increases the win probability inside the interval by an amount of order $T^{-1/3}$, while leaving feedback at all other bids unchanged. The algorithm must therefore bid inside the interval to detect the improvement. Moreover, the increase is small enough that the observations collected there provide little information for distinguishing the two distributions. Although the algorithm can respond to new feedback, its behavior remains sufficiently similar under the modified distribution that it fails to use the improved interval often enough.

The hindsight strategy, by contrast, knows which interval has improved. It mixes a bid at the interval's left endpoint with zero bids, using the ROI surplus from free wins to offset the deficit from bids above value. This feasible mixture increases its expected value by order $T^{-1/3}$ per round, giving an advantage of order $T^{2/3}$ over the horizon. To account for algorithms that obtain more value by violating ROI, the proof compares value minus a fixed multiple of the expected ROI deficit. This comparison shows that the algorithm must incur either regret or ROI violation of the stated order.

The situation changes when the ROI constraint must hold exactly in expectation. A second construction makes useful exploration difficult under this requirement, while giving the hindsight strategy a constant advantage per round.

**Proposition 1.** *Let $\mathfrak{A}_T^{\mathrm{ex}}$ be the class of bandit algorithms $\mathcal{A}$ that, without knowing $(H, F)$, satisfy $\mathbb{E}_{H,F,\mathcal{A}}[\sum_{t=1}^{T}(P_t - R_t)] \leq 0$ for every pair $(H, F)$ in the ROI-only model. Writing $\mathrm{Reg}_T(\mathcal{A}; H, F)$ for their expected regret, for every time horizon $T \geq 1$,*

$$\frac{T}{24} \leq \inf_{\mathcal{A} \in \mathfrak{A}_T^{\mathrm{ex}}} \sup_{H,F} \mathrm{Reg}_T(\mathcal{A}; H, F) \leq T. \tag{14}$$

*The lower bound holds with constant value $v_t = 1/2$ and two-point competing-bid distributions.*

Fix an algorithm satisfying exact expected ROI and set the value to $v = 1/2$. In the base instance, the competing bid is equally likely to be $v$ or 1. Bids below value never win, whereas bids above value incur a positive expected ROI deficit. Since no bid earns surplus to offset this deficit, the algorithm cannot bid above $v$ in this instance. Its expected value is therefore at most $1/4$ per round.

For the alternative instance, let the competing bid be equally likely to be $v-\varepsilon$ or $v+\varepsilon$. The hindsight strategy can now mix these two bids with probabilities $2/3$ and $1/3$, respectively. The surplus from the lower bid exactly offsets the deficit from the higher bid, and the mixture earns expected value $1/3$ per round. Crucially, this value does not decrease as $\varepsilon$ becomes smaller.

For bids at most $v$, the two instances give different feedback only in the narrow interval $[v - \varepsilon, v)$. After fixing the algorithm and horizon, we choose $\varepsilon$ small enough that the algorithm enters this

interval with very small probability in the base instance. Such a choice is possible because the horizon is finite and the interval excludes the bid $v$ itself. The two runs can then produce identical bids and feedback with high probability, keeping the algorithm's expected value close to its value in the base instance. Meanwhile, the hindsight strategy continues to earn $1/3$ per round. The resulting constant gap gives linear regret. Proofs of both lower bounds appear in Appendix B.

## 4 Adding a budget constraint

In this section, we extend the ROI-only algorithm to handle an additional budget constraint. The main difficulty is that spending too quickly may leave the bidder unable to benefit from later auctions, even when ROI is satisfied. The algorithm therefore adds a budget multiplier to control spending throughout the horizon and a stopping rule to ensure that the budget is never exceeded. Let $B = \beta T$ be the total budget, where $\beta \in (0, 1]$ is the average budget per auction. The allowance $\beta$ is an average spending target: the algorithm can spend more in one auction and compensate by spending less in others. The following theorem measures regret against the hindsight benchmark subject to both constraints in (4).

**Theorem 3.** *For any unknown competing-bid distribution $F$ and value distribution $H$, and any budget $B = \beta T$ with $\beta \in (0, 1]$, Algorithm 2 under bandit feedback achieves expected regret and ROI violation bounded by*

$$\mathrm{Reg}_T^B = \widetilde{O}(T^{2/3}/\beta), \qquad \mathrm{Viol}_T^{\mathrm{ROI}} = \widetilde{O}(T^{2/3}), \tag{15}$$

*while satisfying $\sum_{t=1}^{T} P_t \le B$ on every sample path. For any fixed $\beta > 0$, both bounds are $\widetilde{O}(T^{2/3})$.*

The algorithm uses the same win-probability estimates and upward rounding as Algorithm 1. Alongside the ROI multiplier $\lambda_t$, it maintains a budget multiplier $\mu_t$, initialized at one. Before the stopping rule takes effect, each round is called active. In an active round, the algorithm observes the value $v_t$ and chooses a candidate bid to maximize

$$\mu_t \beta + Q_t(b)\big[(1 + \lambda_t)v_t - (\lambda_t + \mu_t)b\big]. \tag{16}$$

The additional term $\mu_t(\beta - bQ_t(b))$ compares predicted payment with the average budget per auction. For the selected candidate bid $\widetilde{b}_t$, let $q_t = Q_t(\widetilde{b}_t)$ be its estimated win probability. Write $g_t = (v_t - \widetilde{b}_t)q_t$ for predicted ROI surplus and $h_t = \beta - \widetilde{b}_t q_t$ for the difference between the per-auction budget and predicted payment. Each multiplier increases when its corresponding quantity is negative. Thus the budget multiplier makes expensive bids less attractive when predicted spending exceeds the allowance $\beta$.

The multiplier controls predicted spending, while budget feasibility requires checking actual payments. Before observing each value, the algorithm checks its remaining budget. If less than one remains, it submits an actual bid of zero in all subsequent rounds and stops updating the estimates and multipliers. These zero bids are not rounded upward. Since every active payment is at most one, total spending cannot exceed the budget. Moreover, deciding whether a round is active before observing its value allows the analysis to continue averaging over the value distribution $H$ without changing that distribution.

The analysis must control spending before the stopping rule takes effect. A bound only at the end of the horizon would not prevent the budget from being exhausted early. The maximum-score comparison from Section 3 continues to apply. However, a bid can generate ROI surplus while its predicted payment exceeds $\beta$. The two constraint terms can therefore offset each other in the selected score. The analysis controls their combined contribution using the potential $\Psi_t = \lambda_t + \mu_t$. Indeed, candidate bid zero has a nonnegative score, so the selected score satisfies $v_t q_t + \lambda_t g_t + \mu_t h_t \ge 0$. Since $v_t q_t \le 1$, the updates give $\Psi_{t+1} \le (1 + \eta^2)\Psi_t + \eta$. This bounds the two multipliers together and controls both predicted deficits up to every active round. In particular, after any number $s$ of active rounds, predicted spending is at most $\beta s + \widetilde{O}(T^{2/3})$.

These predicted guarantees can again be related to actual outcomes using estimation errors only at submitted bids. For an active round, let $e_t = |q_t - F(b_t)|$. The difference between predicted payment and conditional expected payment satisfies $|b_t F(b_t) - \widetilde{b}_t q_t| \le e_t + 1/K$. Thus the cumulative estimation bound from Section 3 also controls spending, without introducing either multiplier into the error term. The corresponding comparisons for value and ROI surplus are unchanged.

**Algorithm 2** Bandit bidding under ROI and budget constraints

**Require:** Time horizon $T$, total budget $B = \beta T$
**Require:** Grid, confidence estimates, parameters $K, \eta, \delta$, and rounding rule $\sigma$ as in Algorithm 1
Initialize $\lambda_1 = \mu_1 = 1$ and total spending $A = 0$
Initialize grid counts and wins as in Algorithm 1
**for** $t = 1, \dots, T$ **do**
  **if** $B - A < 1$ **then** ▷ Check before observing the value
    Observe $v_t$; submit the actual bid $b_t = 0$ and observe $X_t$
    Set $\lambda_{t+1} = \lambda_t$ and $\mu_{t+1} = \mu_t$; leave grid counts and wins unchanged
  **else**
    Form $Q_t$ from past observations and observe $v_t$
    Choose $\widetilde{b}_t \in \{c_0, \dots, c_K\}$ to maximize (16)
    Set $q_t = Q_t(\widetilde{b}_t)$ and $b_t = \sigma(\widetilde{b}_t)$
    Submit $b_t$, observe $X_t$, and update $A \leftarrow A + b_t X_t$
    Set $g_t = (v_t - \widetilde{b}_t)q_t$ and $h_t = \beta - \widetilde{b}_t q_t$
    Update $\lambda_{t+1} = \lambda_t \exp(-\eta g_t)$ and $\mu_{t+1} = \mu_t \exp(-\eta h_t)$
    Update the count and wins at $b_t$ if $b_t < 1$
  **end if**
**end for**

To see why stopping loses little value, let $\tau$ denote the number of active rounds and let $A_\tau = \sum_{t=1}^{\tau} P_t$ be spending by then. If $\tau < T$, the stopping rule implies $A_\tau > B - 1$. Since $B = \beta T$, it follows that

$$\beta(T - \tau) \le 1 + (A_\tau - \beta\tau)_+.$$

This inequality also holds when $\tau = T$. Thus stopping with many auctions left requires spending substantially more than the allowance accumulated over the active rounds.

The bounds at every active prefix now become essential. They control predicted spending at the random stopping time, while estimation error and upward rounding account for the difference from conditional expected payments. Realized payments add a martingale fluctuation with expected absolute value at most $\sqrt{T}$. Together, these bounds give $\mathbb{E}[(A_\tau - \beta\tau)_+] = \widetilde{O}(T^{2/3})$, and hence

$$\mathbb{E}[T - \tau] \le \frac{1 + \mathbb{E}[(A_\tau - \beta\tau)_+]}{\beta} = \widetilde{O}(T^{2/3}/\beta).$$

Each round after active bidding stops can lose at most one unit of value. Dividing the spending bound by $\beta$ therefore explains the additional regret from stopping and the factor $1/\beta$ in Theorem 3. The ROI violation bound follows directly from the surplus analysis on active rounds, and subsequent zero bids contribute nonnegative ROI surplus. It therefore remains $\widetilde{O}(T^{2/3})$. The full proof appears in Appendix C.

Together with the $\Omega(T^{2/3})$ lower bound of Aggarwal et al. (2025), Theorem 3 establishes a tight worst-case regret bound for budgets proportional to the horizon. Their construction uses constant unit values, so ROI holds automatically and values are independent of competing bids. Moreover, their benchmark coincides with ours on these instances. Appendix C.4 gives the comparison.

The analysis also extends to a distributional benchmark that knows both $F$ and $H$ and chooses a randomized bid based on the current value. Let $\mathrm{OPT}^B_{\mathrm{pop}}(F, H)$ be the largest expected value per auction, $\mathbb{E}[vF(b)]$, among such rules satisfying

$$\mathbb{E}[(b - v)F(b)] \le 0, \qquad \mathbb{E}[bF(b)] \le \beta. \tag{17}$$

Here expectations include both the value drawn from $H$ and the randomized bid. The dual argument in Appendix C applies to this benchmark as well, giving the following corollary.

**Corollary 1.** *The expected regret of Algorithm 2 against the optimal feasible randomized bidding rule satisfies*

$$T\,\mathrm{OPT}^B_{\mathrm{pop}}(F, H) - \mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \le \widetilde{O}(T^{2/3}/\beta).$$

*Its ROI violation is $\widetilde{O}(T^{2/3})$, and total payment satisfies $\sum_{t=1}^{T} P_t \le B$ on every sample path.*

### AI USE STATEMENT

The research ideas and initial proof drafts were developed by the authors. Large language models assisted with checking and revising these drafts, translating author-written text into English, polishing the wording, and identifying and checking relevant references. The authors subsequently reviewed and revised the manuscript and take responsibility for its final content, including the mathematical arguments and references.

## REFERENCES

Gagan Aggarwal, Ashwinkumar Badanidiyuru, and Aranyak Mehta. Autobidding with constraints. In *Web and Internet Economics*, volume 11920 of *Lecture Notes in Computer Science*, pp. 17–30. Springer, 2019. doi: 10.1007/978-3-030-35389-6_2. URL https://doi.org/10.1007/978-3-030-35389-6_2.

Gagan Aggarwal, Ashwinkumar Badanidiyuru, Santiago R. Balseiro, Kshipra Bhawalkar, Yuan Deng, Zhe Feng, Gagan Goel, Christopher Liaw, Haihao Lu, Mohammad Mahdian, Jieming Mao, Aranyak Mehta, Vahab Mirrokni, Renato Paes Leme, Andres Perlroth, Georgios Piliouras, Jon Schneider, Ariel Schvartzman, Balasubramanian Sivan, Kelly Spendlove, Yifeng Teng, Di Wang, Hanrui Zhang, Mingfei Zhao, Wennan Zhu, and Song Zuo. Auto-bidding and auctions in online advertising: A survey. *ACM SIGecom Exchanges*, 22(1):159–183, 2024. doi: 10.1145/3699824.3699838. URL https://sigecom.org/exchanges/volume_22/1/AGGARWAL.pdf.

Gagan Aggarwal, Giannis Fikioris, and Mingfei Zhao. No-regret algorithms in non-truthful auctions with budget and ROI constraints. In *Proceedings of the ACM on Web Conference 2025*, pp. 1398–1415. Association for Computing Machinery, 2025. doi: 10.1145/3696410.3714881. URL https://doi.org/10.1145/3696410.3714881.

Shipra Agrawal and Nikhil R. Devanur. Bandits with concave rewards and convex knapsacks. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, pp. 989–1006. Association for Computing Machinery, 2014. doi: 10.1145/2600057.2602844. URL https://doi.org/10.1145/2600057.2602844.

Shipra Agrawal and Nikhil R. Devanur. Linear contextual bandits with knapsacks. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://papers.nips.cc/paper_files/paper/2016/hash/f3144cefe89a60d6a1afaf7859c5076b-Abstract.html.

Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. Resourceful contextual bandits. In *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 1109–1134. PMLR, 2014. URL https://proceedings.mlr.press/v35/badanidiyuru14.html.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *Journal of the ACM*, 65(3):13:1–13:55, 2018. doi: 10.1145/3164539. URL https://arxiv.org/abs/1305.2545.

Ashwinkumar Badanidiyuru, Zhe Feng, and Guru Guruganesh. Learning to bid in contextual first price auctions. In *Proceedings of the ACM Web Conference 2023*, pp. 3489–3497. Association for Computing Machinery, 2023. doi: 10.1145/3543507.3583427. URL https://doi.org/10.1145/3543507.3583427.

Santiago Balseiro, Negin Golrezaei, Mohammad Mahdian, Vahab Mirrokni, and Jon Schneider. Contextual bandits with cross-learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/6aadca7bd86c4743e6724f9607256126-Abstract.html.

Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. Dual mirror descent for online allocation problems. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 613–628. PMLR, 2020. URL https://proceedings.mlr.press/v119/balseiro20a.html.

Santiago R. Balseiro and Yonatan Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019. doi: 10.1287/mnsc.2018.3174. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2018.3174.

Santiago R. Balseiro, Kshipra Bhawalkar, Zhe Feng, Haihao Lu, Vahab Mirrokni, Balasubramanian Sivan, and Di Wang. A field guide for pacing budget and ROS constraints. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 2607–2638. PMLR, 2024. URL https://proceedings.mlr.press/v235/balseiro24a.html.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. Beyond primal-dual methods in bandits with stochastic and adversarial constraints. In *Advances in Neural Information Processing Systems*, volume 37, 2024a. doi: 10.52202/079017-0273. URL https://papers.nips.cc/paper_files/paper/2024/hash/0fd5675f49141c79ad22d7a533c89b12-Abstract-Conference.html.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. Bandits with replenishable knapsacks: the best of both worlds. In *International Conference on Learning Representations*, 2024b. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/54a1495b06c4ee2f07184afb9a37abda-Abstract-Conference.html.

Martino Bernasconi, Matteo Castiglioni, and Andrea Celli. No-regret is not enough! bandits with general constraints through adaptive regret minimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 3877–3898. PMLR, 2025. URL https://proceedings.mlr.press/v267/bernasconi25a.html.

Matteo Castiglioni, Andrea Celli, and Christian Kroer. Online learning with knapsacks: the best of both worlds. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 2767–2783. PMLR, 2022. URL https://proceedings.mlr.press/v162/castiglioni22a.html.

Matteo Castiglioni, Andrea Celli, and Christian Kroer. Online learning under budget and ROI constraints via weak adaptivity. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5792–5816. PMLR, 2024. URL https://proceedings.mlr.press/v235/castiglioni24a.html.

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. The role of transparency in repeated first-price auctions with unknown valuations. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pp. 225–236. Association for Computing Machinery, 2024. doi: 10.1145/3618260.3649658. URL https://cesa-bianchi.di.unimi.it/Pubblicazioni/stoc24.pdf.

Vincent Conitzer, Christian Kroer, Debmalya Panigrahi, Okke Schrijvers, Nicolas E. Stier-Moses, Eric Sodomka, and Christopher A. Wilkens. Pacing equilibrium in first price auction markets. *Management Science*, 68(12):8515–8535, 2022. doi: 10.1287/mnsc.2022.4310. URL https://doi.org/10.1287/mnsc.2022.4310.

Yuan Deng, Negin Golrezaei, Patrick Jaillet, Jason Cheuk Nam Liang, and Vahab Mirrokni. Multi-channel autobidding with budget and ROI constraints. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7617–7644. PMLR, 2023. URL https://proceedings.mlr.press/v202/deng23c.html.

Zhe Feng, Swati Padmanabhan, and Di Wang. Online bidding algorithms for return-on-spend constrained advertisers. In *Proceedings of the ACM Web Conference 2023*, pp. 3550–3560. Association for Computing Machinery, 2023. doi: 10.1145/3543507.3583491. URL https://doi.org/10.1145/3543507.3583491.

Giannis Fikioris and Éva Tardos. Liquid welfare guarantees for no-regret learning in sequential budgeted auctions. *Mathematics of Operations Research*, 50(2):1233–1249, 2024. doi: 10.1287/moor.2023.0274. URL https://doi.org/10.1287/moor.2023.0274.

Zeng Fu, Jiashuo Jiang, and Yuan Zhou. Online bidding for contextual first-price auctions with budgets under one-sided information feedback, 2026. URL https://arxiv.org/abs/2603.07207.

Jason Gaitonde, Yingkai Li, Bar Light, Brendan Lucier, and Aleksandrs Slivkins. Budget pacing in repeated auctions: Regret and efficiency without convergence. In *14th Innovations in Theoretical Computer Science Conference*, volume 251 of *Leibniz International Proceedings in Informatics*, pp. 52:1–52:1. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi: 10.4230/LIPIcs.ITCS.2023.52. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2023.52.

Hengquan Guo, Lingkai Zu, and Xin Liu. Triple-optimistic learning for stochastic contextual bandits with general constraints. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 21252–21276. PMLR, 2025. URL https://proceedings.mlr.press/v267/guo25v.html.

Hengquan Guo, Haobo Zhang, Junwei Pan, Shudong Huang, Nianhua Xie, Lei Xiao, Haijie Gu, Jie Jiang, and Xin Liu. Towards safe and optimal online bidding: A modular look-ahead lyapunov framework. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/pdf?id=AtbRCnvcrZ.

Yanjun Han, Zhengyuan Zhou, Aaron Flores, Erik Ordentlich, and Tsachy Weissman. Learning to bid optimally and efficiently in adversarial first-price auctions, 2020. URL https://arxiv.org/abs/2007.04568.

Yanjun Han, Tsachy Weissman, and Zhengyuan Zhou. Optimal no-regret learning in repeated first-price auctions. *Operations Research*, 73(1):209–238, 2024. doi: 10.1287/opre.2020.0282. URL https://doi.org/10.1287/opre.2020.0282.

Zihao Hu, Yuxiao Wen, Yuan Yao, Jiheng Zhang, and Zhengyuan Zhou. Learning to bid with unknown private values in budget-constrained first-price auctions, 2026. URL https://arxiv.org/abs/2605.09448.

Nicole Immorlica, Karthik Abinav Sankararaman, Robert E. Schapire, and Aleksandrs Slivkins. Adversarial bandits with knapsacks. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science*, pp. 202–219. IEEE, 2019. doi: 10.1109/FOCS.2019.00022. URL https://doi.org/10.1109/FOCS.2019.00022.

Thomas Kesselheim and Sahil Singla. Online learning with vector costs and bandits with knapsacks. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2286–2305. PMLR, 2020. URL https://proceedings.mlr.press/v125/kesselheim20a.html.

Raunak Kumar and Robert D. Kleinberg. Non-monotonic resource utilization in the bandits with knapsacks problem. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/7a62d9a4c03377d1175b8859b4cc16d4-Abstract-Conference.html.

Yilin Li, Yuan Deng, Wei Tang, and Hanrui Zhang. No-regret online autobidding algorithms in first-price auctions. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-2079. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/59c147c7d4fdb732daea3064eab949bf-Abstract-Conference.html.

Brendan Lucier, Sarath Pattathil, Aleksandrs Slivkins, and Mengxiao Zhang. Autobidders with budget and ROI constraints: Efficiency, regret, and pacing dynamics. In *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 3642–3643. PMLR, 2024. URL https://proceedings.mlr.press/v247/lucier24a.html.

Jon Schneider and Julian Zimmert. Optimal cross-learning for contextual bandits with unknown context distributions. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

doi: 10.52202/075280-2259. URL https://papers.neurips.cc/paper_files/paper/2023/hash/a3017a8d202a433be56a3dfdcac6c8eb-Abstract-Conference.html.

Aleksandrs Slivkins, Karthik Abinav Sankararaman, and Dylan J. Foster. Contextual bandits with packing and covering constraints: A modular lagrangian approach via regression. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 4633–4656. PMLR, 2023. URL https://proceedings.mlr.press/v195/foster23c.html.

Rahul Vaze and Abhishek Sinha. Online bidding algorithms with strict return on spend (ROS) constraint, 2025. URL https://arxiv.org/abs/2502.05599.

Sushant Vijayan, Zhe Feng, Swati Padmanabhan, Karthikeyan Shanmugam, Arun Suggala, and Di Wang. Online bidding under RoS constraints without knowing the value. In *Proceedings of the ACM on Web Conference 2025*, pp. 3096–3107. Association for Computing Machinery, 2025. doi: 10.1145/3696410.3714734. URL https://arxiv.org/abs/2503.03195.

Qian Wang, Zongjun Yang, Xiaotie Deng, and Yuqing Kong. Learning to bid in repeated first-price auctions with budgets. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36494–36513. PMLR, 2023. URL https://proceedings.mlr.press/v202/wang23ao.html.

# A PROOF OF THE ROI-ONLY UPPER BOUND

This appendix supplies the details behind Theorem 1. We write $a_t = \sigma(\widetilde{b}_t)$ for the implemented bid; all other notation follows Section 3. All expectations include the values, competing bids, and possible algorithmic randomization.

## A.1 A COMMON DUAL UPPER BOUND

Let $D_F^\star$ be defined in (12). For a fixed value sequence $V^T$, any randomized policy $\pi$ feasible for the ROI constraint in (3) satisfies, for every $\lambda \geq 0$,

$$\mathbb{E}_\pi\left[\sum_{t=1}^{T} R_t\right] \leq \mathbb{E}_\pi\left[\sum_{t=1}^{T}\{R_t + \lambda(R_t - P_t)\}\right] \leq \sum_{t=1}^{T} f_F(v_t, \lambda).$$

The first inequality uses only aggregate expected feasibility. Taking the supremum over feasible policies, then expectation over $V^T$, and finally the infimum over any *fixed* $\lambda$ gives $\mathbb{E}[\mathrm{OPT}(F, V^T)] \leq TD_F^\star$. The same $D_F^\star$ controls the entire horizon; no round or confidence-model segment receives a separate offline comparator.

## A.2 SIMULTANEOUS LOCAL CONFIDENCE

For each grid bid $c_j$, $j < K$, couple the observations to an infinite stack of independent Bernoulli$(F(c_j))$ variables, independent also of the entire value sequence and the algorithm's randomization. Reveal the next variable in that stack whenever $c_j$ is queried. This coupling has the same law as the auction observations: the choice of the current bid is made before the independent competing bid $d_t$ is drawn. In particular, the first $n$ variables in a stack form a fixed-sample prefix even though the round of the $n$-th query is random. Hoeffding's inequality for every such prefix, followed by a union bound over $j < K$ and $n = 1, \ldots, T$, gives an event $\mathcal{E}$ with $\Pr(\mathcal{E}^c) \leq 2(K-1)Te^{-L_T} \leq \delta$. On this event,

$$\left|\widehat{F}_{t,j} - F(c_j)\right| \leq \sqrt{\frac{L_T}{2n_{t,j}}} \quad \text{for all } t, j \text{ with } n_{t,j} > 0. \tag{18}$$

For every $i \geq j$, $U_{t,i} \geq F(c_i) \geq F(c_j)$, so the suffix minimum satisfies $\bar{U}_{t,j} \geq F(c_j)$. Since $i = j$ is present in that minimum, $\bar{U}_{t,j} \leq U_{t,j}$. Thus, on $\mathcal{E}$,

$$0 \leq \bar{U}_{t,j} - F(c_j) \leq \begin{cases} 1, & n_{t,j} = 0, \\ 2\sqrt{L_T/(2n_{t,j})}, & n_{t,j} > 0. \end{cases} \tag{19}$$

At $j = K$, the error is zero because $F(1) = 1$. Applied at the selected coordinate, (19) bounds $e_t = |q_t - F(a_t)|$. The first query to any coordinate costs at most one; subsequent queries cost at most a constant times $\sqrt{L_T/n}$, where $n$ is the number of earlier queries to that coordinate. If coordinate $j$ is queried $N_j$ times,

$$\sum_{t:a_t=c_j} e_t \leq 1 + C\sqrt{N_j L_T} \quad \text{on } \mathcal{E}.$$

Since $\sum_j N_j \leq T$, Cauchy–Schwarz yields

$$\sum_t e_t \leq K + C\sqrt{KTL_T} \quad \text{on } \mathcal{E}.$$

The event $\mathcal{E}$ covers all grid points and sample prefixes, including the ones selected in response to past bits and the current value. Always $e_t \leq 1$, so

$$\mathbb{E}\left[\sum_t e_t\right] \leq K + C\sqrt{KTL_T} + T\delta. \tag{20}$$

This accounting needs no uniform estimate of $F$ between grid points.

### A.3 Optimism and multiplier control

For a candidate bid $b$ with $\sigma(b) = c_j$, the successor rule and monotonicity imply, on $\mathcal{E}$,

$$Q_t(b) = \bar{U}_{t,j} \geq F(\sigma(b)) \geq F(b).$$

For $c = (1+\lambda)v - \lambda b \geq 0$, this implies $Q_t(b)c \geq F(b)c$. For $c < 0$, the true score $F(b)c$ is nonpositive, while the candidate bid 0 has model score $(1+\lambda)vQ_t(0) \geq 0$. Taking maxima shows

$$f_{Q_t}(v,\lambda) \geq f_F(v,\lambda) \quad \text{for all } v, \lambda \geq 0 \text{ on } \mathcal{E}. \tag{21}$$

Whether or not $\mathcal{E}$ holds, $f_{Q_t}(v,\lambda) \geq 0$. Therefore, if $D_{Q_t}(\lambda) = \mathbb{E}_{v\sim H}[f_{Q_t}(v,\lambda)]$,

$$D_{Q_t}(\lambda_t) \geq D_F^\star \quad \text{on } \mathcal{E}, \qquad D_{Q_t}(\lambda_t) \geq 0 \quad \text{always}. \tag{22}$$

Because $0 \leq D_F^\star \leq D_F(0) \leq 1$, the pathwise lower bound in (22) implies

$$\mathbb{E}\left[\sum_{t=1}^{T} D_{Q_t}(\lambda_t)\right] \geq TD_F^\star - T\delta. \tag{23}$$

We have used $\mathcal{E}$ only for this pathwise comparison; we have not conditioned a single-round expectation on a future-dependent event.

The selected model score is $r_t^Q + \lambda_t g_t^Q = f_{Q_t}(v_t, \lambda_t) \geq 0$, and $0 \leq r_t^Q \leq 1$. Thus $\lambda_t g_t^Q \geq -1$. Since $|g_t^Q| \leq 1$ and $\eta \leq 1/2$, the inequality $e^{-x} \leq 1 - x + x^2$ for $|x| \leq 1/2$ yields

$$\lambda_{t+1} \leq (1+\eta^2)\lambda_t - \eta\lambda_t g_t^Q \leq (1+\eta^2)\lambda_t + \eta.$$

Induction gives $\lambda_t \leq M = e^{\eta^2 T}(1+\eta T)$ for $1 \leq t \leq T+1$. More explicitly, for every $0 \leq s \leq T$, $\lambda_{s+1} \leq e^{\eta^2 s}(1+\eta s)$. Summing the first inequality through an arbitrary prefix $s$ and using $\lambda_1 = 1$ yields

$$\sum_{t=1}^{s} \lambda_t g_t^Q \leq \frac{1}{\eta} + \eta s M, \qquad 0 \leq s \leq T. \tag{24}$$

The exact multiplicative update also gives $\lambda_{s+1} = \exp(-\eta \sum_{t=1}^{s} g_t^Q) \leq M$ for every prefix, hence

$$-\sum_{t=1}^{s} g_t^Q \leq \frac{\log M}{\eta}, \qquad 0 \leq s \leq T. \tag{25}$$

Let $\mathcal{F}_{t-1}$ denote the history just before $v_t$ is drawn. Both $Q_t$ and $\lambda_t$ are $\mathcal{F}_{t-1}$-measurable, so

$$\mathbb{E}[r_t^Q + \lambda_t g_t^Q \mid \mathcal{F}_{t-1}] = D_{Q_t}(\lambda_t).$$

Combining this identity with (23) and the case $s = T$ of (24) gives

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t^Q\right] \geq TD_F^\star - T\delta - \eta^{-1} - \eta TM. \tag{26}$$

### A.4 Transfer to realized outcomes

Condition on $\mathcal{H}_t$, the history through the selection of $\widetilde{b}_t, a_t$ but before $d_t$ is drawn. Then $\mathbb{E}[X_t \mid \mathcal{H}_t] = F(a_t)$. As $0 \leq v_t \leq 1$,

$$\mathbb{E}[R_t \mid \mathcal{H}_t] = v_t F(a_t) \geq v_t q_t - e_t = r_t^Q - e_t. \tag{27}$$

The weak dual bound, (26), and (20) now imply

$$\mathrm{Reg}_T \leq \eta^{-1} + \eta TM + K + C\sqrt{KTL_T} + 2T\delta. \tag{28}$$

For violation, let $C_t = P_t - R_t = (a_t - v_t)X_t$, and let $\mu_t = \mathbb{E}[C_t \mid \mathcal{H}_t] = (a_t - v_t)F(a_t)$. The successor rule gives $0 \leq a_t - \widetilde{b}_t \leq 1/K$, hence

$$\begin{aligned} \mu_t + g_t^Q &= (a_t - \widetilde{b}_t)F(a_t) + (\widetilde{b}_t - v_t)\big(F(a_t) - q_t\big) \\ &\leq K^{-1} + e_t. \end{aligned} \tag{29}$$

The differences $C_t - \mu_t = (a_t - v_t)(X_t - F(a_t))$ form a bounded martingale difference sequence with conditional second moment at most one. Thus

$$\mathbb{E}\left[\left|\sum_{t=1}^{T}(C_t - \mu_t)\right|\right] \le \sqrt{T}. \tag{30}$$

By (25) and (29),

$$\sum_t \mu_t \le \eta^{-1}\log M + T/K + \sum_t e_t$$

on every path. Taking the positive part after adding the martingale, then expectation, and using (20) gives

$$\mathrm{Viol}_T \le \eta^{-1}\log M + T/K + K + C\sqrt{KTL_T} + T\delta + \sqrt{T}. \tag{31}$$

With $K = \max\{2, \lceil T^{1/3}\rceil\}$, $\eta = T^{-2/3}$, and $\delta = T^{-2}$, both (28) and (31) are $\widetilde{O}(T^{2/3})$.

## B Proofs of the ROI-only lower bounds

*Proof of Theorem 2.* Fix an arbitrary one-bit algorithm and an integer $T \ge 216$. The algorithm may randomize and adapt to all previously observed bits; its current bid is chosen before the current competing bid is drawn. Set

$$M = \lceil 2T^{1/3}\rceil, \qquad \varepsilon = \frac{1}{3M}, \qquad d_i = \frac{1}{3} + i\varepsilon \quad (0 \le i \le M). \tag{32}$$

Thus $d_M = 2/3$. Fix the value distribution to be a point mass at $v = 4/13$. On the bid interval $[0,1]$, define the base CDF

$$F^\circ(b) = \begin{cases} 3/4, & 0 \le b < d_0, \\ 3/(4-d_i), & d_i \le b < d_{i+1}, \quad 0 \le i < M, \\ 3/(4-d_M), & d_M \le b < 1, \\ 1, & b = 1. \end{cases} \tag{33}$$

For each $j \in \{0, \dots, M-1\}$, put $I_j = [d_j, d_{j+1})$ and define

$$F^{(j)}(b) = \begin{cases} 3/(4-d_{j+1}), & b \in I_j, \\ F^\circ(b), & b \notin I_j. \end{cases} \tag{34}$$

Both are right-continuous, nondecreasing CDFs on $[0,1]$. Indeed, $F^\circ$ places mass at $0, d_0, \dots, d_M, 1$, and $F^{(j)}$ moves the jump at $d_{j+1}$ to $d_j$. This description also covers $j = M-1$, when the source atom is at $d_M$. In particular, the CDFs agree at $d_{j+1}$ and at every bid outside the half-open interval $I_j$. The half-open convention matters because bidding exactly $d_{j+1}$ wins the source atom in both environments.

For any CDF $F$, let $r_F(b) = vF(b)$ and $h_F(b) = (b-v)F(b)$ denote expected reward and expected ROI cost for a deterministic bid $b$. With $\lambda_\star = 1/12$ and $v = 4/13$,

$$r_F(b) - \lambda_\star h_F(b) = \frac{4-b}{12}F(b). \tag{35}$$

At $b = 0$, at each $d_i$, and at $b = 1$, the right side under $F^\circ$ equals $1/4$. Between these points the CDF is constant and $(4-b)/12$ decreases. Therefore the base score is at most $1/4$ for every bid. The alternative differs only on $I_j$. Within that interval the score is maximized at its left endpoint $d_j$, so

$$r_{F^{(j)}}(b) - \lambda_\star h_{F^{(j)}}(b) \le \frac{1}{4} + \gamma_j \mathbf{1}\{b \in I_j\}, \qquad \gamma_j = \frac{\varepsilon}{4(4-d_{j+1})} \le \frac{3\varepsilon}{40}. \tag{36}$$

The last inequality follows from $d_{j+1} \le 2/3$.

We next exhibit an admissible comparator in environment $F^{(j)}$. Bid 0 wins with probability $3/4$, earns expected reward $r_0 = 3/13$, and has expected ROI cost $h_0 = -3/13$. Bid $d_j$ wins with

probability $q_{j+1} = 3/(4 - d_{j+1})$ and has positive expected ROI cost $h_j = (d_j - 4/13)q_{j+1} > 0$. Mix bid $d_j$ with probability

$$p_j = \frac{-h_0}{h_j - h_0} = \frac{4 - d_j - \varepsilon}{12d_j - \varepsilon} \tag{37}$$

and bid 0 otherwise. Since $d_j \geq 1/3 > 4/13$ and $12d_j - \varepsilon > 0$, both $p_j$ and $1 - p_j = (13d_j - 4)/(12d_j - \varepsilon)$ lie in $(0, 1)$. In particular, $(1 - p_j)h_0 + p_j h_j = 0$, so the mixture satisfies the ROI constraint conditional on the (constant) value sequence. Its expected reward per round is

$$r_j^{\mathrm{mix}} = v\big((1 - p_j)3/4 + p_j q_{j+1}\big) = \frac{3d_j}{12d_j - \varepsilon} = \frac{1}{4} + \frac{\varepsilon}{4(12d_j - \varepsilon)} \geq \frac{1}{4} + \frac{\varepsilon}{32}. \tag{38}$$

Using this mixture on every round is feasible for the full hindsight benchmark in (3), which permits a randomized bid in each round; hence $\mathrm{OPT}(F^{(j)}, V^T) \geq T(1/4 + \varepsilon/32)$.

Let $\mathbb{P}_\circ, \mathbb{P}_j$ be the laws of the interaction transcript under $F^\circ, F^{(j)}$, respectively, augmented by the algorithm's private randomization. The transcript contains its bids and win bits, but not the unobserved competing bids. Let $\mathbb{E}_\circ, \mathbb{E}_j$ denote their expectations. Define $N_j = \sum_{t=1}^T \mathbf{1}\{b_t \in I_j\}$. Because each new threshold is independent of the past and of the chosen bid, taking conditional expectations in (36) gives

$$\mathbb{E}_j\left[\sum_{t=1}^T R_t\right] \leq \frac{T}{4} + \lambda_\star \mathbb{E}_j\left[\sum_{t=1}^T (P_t - R_t)\right] + \gamma_j \mathbb{E}_j[N_j]. \tag{39}$$

Write $\mathrm{Reg}_{T,j}$ and $\mathrm{Viol}_{T,j}$ for the regret and ROI violation in environment $F^{(j)}$. Since $x \leq x_+$ for every real $x$, $\mathbb{E}_j[\sum_t (P_t - R_t)] \leq \mathrm{Viol}_{T,j}$. Combining this with (38) and (36) yields

$$\mathrm{Reg}_{T,j} + \frac{1}{12}\,\mathrm{Viol}_{T,j} \geq \frac{\varepsilon T}{32} - \frac{3\varepsilon}{40}\mathbb{E}_j[N_j]. \tag{40}$$

The $M$ intervals $I_j$ are disjoint. Thus $\sum_j \mathbb{E}_\circ[N_j] \leq T$, and some index $j$ satisfies $\mathbb{E}_\circ[N_j] \leq T/M$. We fix such an index. This choice is made under the base law and may depend on the algorithm, as permitted in a minimax lower bound. The only observation-law difference between $F^\circ$ and $F^{(j)}$ occurs after a bid in $I_j$, where the win-bit means are

$$q_j = \frac{3}{4 - d_j}, \qquad q_{j+1} = \frac{3}{4 - d_{j+1}}. \tag{41}$$

They lie in $[9/11, 9/10]$, and $q_{j+1} - q_j = 3\varepsilon/\big((4 - d_j)(4 - d_{j+1})\big) \leq 27\varepsilon/100$. Since $q_{j+1}(1 - q_{j+1}) \geq 9/100$, the Bernoulli KL bound $\mathrm{kl}(p\|q) \leq (p - q)^2/[q(1 - q)]$ gives

$$\mathrm{kl}\big(\mathrm{Bern}(q_j)\|\mathrm{Bern}(q_{j+1})\big) \leq \frac{81}{100}\varepsilon^2 \leq \varepsilon^2. \tag{42}$$

At every history, the algorithm uses the same bid distribution in both environments. Conditional on the bid, its win bit has mean $q_j$ under $F^\circ$ and $q_{j+1}$ under $F^{(j)}$ if the bid lies in $I_j$, and the two conditional laws agree otherwise. Its reward and payment are deterministic functions of the bid, constant value, and win bit. The adaptive KL chain rule therefore gives the exact identity

$$\mathrm{KL}(\mathbb{P}_\circ\|\mathbb{P}_j) = \mathrm{kl}\big(\mathrm{Bern}(q_j)\|\mathrm{Bern}(q_{j+1})\big)\mathbb{E}_\circ[N_j] \leq \varepsilon^2 \mathbb{E}_\circ[N_j] \leq \frac{\varepsilon^2 T}{M}. \tag{43}$$

Pinsker's inequality and $M \geq 2T^{1/3}$ yield

$$\mathrm{TV}(\mathbb{P}_\circ, \mathbb{P}_j) \leq \varepsilon\sqrt{\frac{T}{2M}} \leq \frac{1}{12}. \tag{44}$$

For any random variable in $[0, T]$, its expectations under these two laws differ by at most $T\,\mathrm{TV}$. Since $M \geq 12$ when $T \geq 216$, we conclude that

$$\mathbb{E}_j[N_j] \leq \mathbb{E}_\circ[N_j] + T\,\mathrm{TV} \leq \frac{T}{M} + \frac{T}{12} \leq \frac{T}{6}. \tag{45}$$

Substitution into (40) gives

$$\mathrm{Reg}_{T,j} + \frac{1}{12}\,\mathrm{Viol}_{T,j} \geq \varepsilon T\left(\frac{1}{32} - \frac{1}{80}\right) = \frac{3}{160}\varepsilon T. \tag{46}$$

Let $m_j = \max\{\mathrm{Reg}_{T,j}, \mathrm{Viol}_{T,j}\} \geq 0$. Each term is at most $m_j$, so $\mathrm{Reg}_{T,j} + \mathrm{Viol}_{T,j}/12 \leq (13/12)m_j$. Hence $m_j \geq (9/520)\varepsilon T$. Finally, $M = \lceil 2T^{1/3} \rceil \leq 3T^{1/3}$, so $\varepsilon \geq 1/(9T^{1/3})$, proving $m_j \geq T^{2/3}/520$. □

### B.1 Exact expected ROI feasibility

*Proof of Proposition 1.* Fix $T \geq 1$ and an algorithm $\mathcal{A} \in \mathfrak{A}_T^{\mathrm{ex}}$. Set every value to $v = 1/2$. Consider first the competing-bid distribution $F_0 = (\delta_v + \delta_1)/2$, and write $\mathbb{P}_0$ and $\mathbb{E}_0$ for the resulting law and expectation. Conditional on the bid $b_t$ and the preceding history, the expected ROI deficit is

$$\mathbb{E}_0[P_t - R_t \mid b_t, \text{history}] = (b_t - v)F_0(b_t) \geq \tfrac{1}{2}(b_t - v)_+. \tag{47}$$

Indeed, a bid below $v$ never wins under $F_0$, while a bid above $v$ wins with probability at least $1/2$. Exact expected feasibility on this instance implies

$$0 \geq \mathbb{E}_0\left[\sum_{t=1}^{T}(P_t - R_t)\right] \geq \tfrac{1}{2}\sum_{t=1}^{T}\mathbb{E}_0[(b_t - v)_+].$$

All terms on the right are nonnegative, so $b_t \leq v$ almost surely under $F_0$ for every $t$. The bidder can then win with probability at most $1/2$ in each round, and hence

$$\mathbb{E}_0\left[\sum_{t=1}^{T} R_t\right] \leq \frac{vT}{2} = \frac{T}{4}. \tag{48}$$

For $\varepsilon > 0$, let $E_\varepsilon$ be the event that at least one bid in the $F_0$ run lies in $[v - \varepsilon, v)$. Because there are only finitely many bids, the events $E_{2^{-m}}$ decrease to the empty set as $m \to \infty$. We can therefore choose a dyadic $0 < \varepsilon < 1/4$ such that

$$\mathbb{P}_0(E_\varepsilon) \leq \frac{1}{12}. \tag{49}$$

This choice may depend on $\mathcal{A}$ and $T$, as permitted in a minimax lower bound.

Now let $F_\varepsilon = (\delta_{v-\varepsilon} + \delta_{v+\varepsilon})/2$. Couple the two runs using the same randomness for $\mathcal{A}$ and the same independent fair coins in every round: when the coin is zero, the competing bids are $v$ under $F_0$ and $v - \varepsilon$ under $F_\varepsilon$; when it is one, they are 1 and $v + \varepsilon$, respectively. Almost surely under the base law, all bids in the $F_0$ run are at most $v$. On $E_\varepsilon^c$, each such bid is either below $v - \varepsilon$ or equal to $v$. A bid below $v - \varepsilon$ loses in both runs; a bid equal to $v$ wins in both runs exactly when the common coin is zero. Induction over the rounds shows that the bids and observed outcomes in the two runs coincide on this event. Since rewards and payments are determined by these observations, their total rewards also coincide there. Each total reward lies in $[0, vT]$, so

$$\mathbb{E}_\varepsilon\left[\sum_{t=1}^{T} R_t\right] \leq \mathbb{E}_0\left[\sum_{t=1}^{T} R_t\right] + vT\mathbb{P}_0(E_\varepsilon) \leq \frac{T}{4} + \frac{T}{24} = \frac{7T}{24}. \tag{50}$$

In the $F_\varepsilon$ instance, the hindsight benchmark may bid $v - \varepsilon$ with probability $2/3$ and $v + \varepsilon$ with probability $1/3$ in every round. The low bid wins with probability $1/2$ and earns expected ROI surplus $\varepsilon/2$; the high bid always wins and incurs expected deficit $\varepsilon$. The mixture therefore has zero expected ROI deficit and is feasible for (3). Its expected value per round is $(2/3)(v/2) + (1/3)v = 1/3$, giving $\mathrm{OPT}(F_\varepsilon, (v, \ldots, v)) \geq T/3$. Together with (50), this proves $\mathrm{Reg}_T(\mathcal{A}; \delta_v, F_\varepsilon) \geq T/24$.

For the upper bound, the algorithm that always bids $b_t = v_t$ has $P_t = R_t$ on every sample path, so it belongs to $\mathfrak{A}_T^{\mathrm{ex}}$. Its reward is nonnegative and the benchmark value is at most $T$. Thus its regret is at most $T$ on every instance, proving (14). □

# C PROOFS FOR THE BUDGET EXTENSION

We prove Theorem 3 using the same grid and confidence model as in Section 3. We first analyze an unstopped two-multiplier algorithm, then add the budget stopping rule. Set

$$f_F^B(v,\lambda,\mu) = \mu\beta + \sup_{b\in[0,1]} F(b)\big[(1+\lambda)v - (\lambda+\mu)b\big], \qquad D^\star_{F,B} = \inf_{\lambda,\mu\ge 0} \mathbb{E}_v[f_F^B(v,\lambda,\mu)]. \tag{51}$$

For every value sequence and every policy feasible for both constraints in (4), adding the two constraints with nonnegative multipliers bounds its expected reward by $\sum_t f_F^B(v_t,\lambda,\mu)$. Averaging over $V^T$, then taking the infimum over fixed multipliers, gives

$$\mathbb{E}[\mathrm{OPT}_B(F,V^T)] \le T D^\star_{F,B}. \tag{52}$$

Candidate bid 0 gives $f_F^B(v,\lambda,\mu) \ge \mu\beta \ge 0$; also $f_F^B(v,0,0) \le 1$. Thus $0 \le D^\star_{F,B} \le 1$.

## C.1 THE UNSTOPPED ALGORITHM

**Lemma 1** (Soft-budget intermediate bound)**.** *For every $T \ge 3$ and $\beta \in [0,1]$, the unstopped one-bit algorithm below has*

$$\mathrm{Reg}_T^B = \widetilde{O}(T^{2/3}), \qquad \mathrm{Viol}_T^{\mathrm{ROI}} = \widetilde{O}(T^{2/3}), \qquad \mathrm{Viol}_T^{\mathrm{B}} = \widetilde{O}(T^{2/3}),$$

*with constants independent of $\beta$.*

*Proof of Lemma 1.* The algorithm maintains $\lambda_1 = \mu_1 = 1$ and selects the exact maximizer of (16). Put

$$q_t = Q_t(\widetilde{b}_t), \qquad r_t = v_t q_t, \qquad g_t = (v_t - \widetilde{b}_t)q_t, \qquad h_t = \beta - \widetilde{b}_t q_t,$$

and update

$$\lambda_{t+1} = \lambda_t e^{-\eta g_t}, \qquad \mu_{t+1} = \mu_t e^{-\eta h_t}, \qquad \eta = T^{-2/3}. \tag{53}$$

Write $f_{Q_t}^B$ for the expression in (51) with $F(b)$ replaced by $Q_t(b)$. Since $|g_t|, |h_t| \le 1$, these updates satisfy the same elementary exponential inequality used in Appendix A. Let $S_t = \lambda_t g_t + \mu_t h_t$ and $\Psi_t = \lambda_t + \mu_t$. The selected model score is

$$r_t + S_t = f_{Q_t}^B(v_t,\lambda_t,\mu_t) \ge 0 \tag{54}$$

because candidate bid 0 has score $(1+\lambda_t)v_t Q_t(0) + \mu_t\beta \ge 0$. Therefore $S_t \ge -r_t \ge -1$, and

$$\Psi_{t+1} \le (1+\eta^2)\Psi_t - \eta S_t \le (1+\eta^2)\Psi_t + \eta. \tag{55}$$

With $M_B = e^{\eta^2 T}(2+\eta T)$, induction gives $\Psi_t \le M_B$ for $1 \le t \le T+1$. Telescoping the first inequality in (55) and taking the logarithm of each update give, for every $0 \le s \le T$,

$$\sum_{t=1}^{s} S_t \le \frac{2}{\eta} + \eta s M_B, \tag{56}$$

$$-\sum_{t=1}^{s} g_t \le \frac{\log M_B}{\eta}, \qquad -\sum_{t=1}^{s} h_t \le \frac{\log M_B}{\eta}. \tag{57}$$

These are pathwise statements. In particular, positive surplus in one constraint may offset negative surplus in the other within $S_t$; the argument does not require a separate inequality $-\lambda_t g_t \le 1$ or $-\mu_t h_t \le 1$.

On the confidence event $\mathcal{E}$ of (18), $Q_t(b) \ge F(b)$. For a fixed $v,\lambda,\mu,b$, if $c = (1+\lambda)v - (\lambda+\mu)b \ge 0$, then $Q_t(b)c \ge F(b)c$. If $c < 0$, the true score $\mu\beta + F(b)c \le \mu\beta$, while candidate bid 0 has model score at least $\mu\beta$. Hence

$$f_{Q_t}^B(v,\lambda,\mu) \ge f_F^B(v,\lambda,\mu) \quad \text{on } \mathcal{E}. \tag{58}$$

Off the event, $f_{Q_t}^B \ge 0$. The confidence model and multipliers are fixed before $v_t$ arrives, so the same unconditional argument as (23) yields

$$\mathbb{E}\left[\sum_t (r_t + S_t)\right] \ge T D^\star_{F,B} - T\delta.$$

Using (56),

$$\mathbb{E}\left[\sum_t r_t\right] \geq TD^\star_{F,B} - T\delta - \frac{2}{\eta} - \eta T M_B. \tag{59}$$

For the unstopped algorithm, write $a_t = \sigma(\widetilde{b}_t)$ for the actual submitted bid. The selected-coordinate error $e_t = |q_t - F(a_t)|$ satisfies (20) without change. Condition on the history through the selected bids, before the current competing bid is drawn. Upward rounding gives

$$v_t F(a_t) \geq r_t - e_t, \tag{60}$$

$$(a_t - v_t)F(a_t) \leq -g_t + K^{-1} + e_t, \qquad a_t F(a_t) - \beta \leq -h_t + K^{-1} + e_t. \tag{61}$$

The two realized costs, $P_t - R_t$ and $P_t - \beta$, minus their conditional means are bounded martingale differences. Each cumulative martingale has expected absolute value at most $\sqrt{T}$. Combining (52), (59), (57), (20), and (60)–(61) gives

$$\mathrm{Reg}^B_T \leq \frac{2}{\eta} + \eta T M_B + K + C\sqrt{KTL_T} + 2T\delta, \tag{62}$$

$$\mathrm{Viol}^J_T \leq \frac{\log M_B}{\eta} + \frac{T}{K} + K + C\sqrt{KTL_T} + T\delta + \sqrt{T}, \quad J \in \{\mathrm{ROI}, \mathrm{B}\}. \tag{63}$$

Since $M_B = O(T^{1/3})$ and $K \asymp T^{1/3}$, Lemma 1 follows. □

### C.2 Budget feasibility

*Proof of Theorem 3.* We now assume $\beta > 0$. At the start of each round, before seeing $v_t$, check the remaining realized budget. If it is less than one, permanently bid 0; otherwise run the two-multiplier algorithm above. Once bidding stops, neither the confidence model nor the multipliers are updated. Let $\tau$ be the number of active rounds and $A_\tau = \sum_{t\leq\tau} P_t$. Since each active payment is at most one, an active round starting with $B - A_{t-1} \geq 1$ cannot make $A_t > B$. Subsequent bids of 0 make no payment, so $\sum_t P_t \leq B$ on every path.

Let $I_t = \mathbf{1}\{t \leq \tau\}$. The decision $I_t$ is determined by the history before $v_t$ arrives. In the independent-stack coupling of Appendix A, the confidence event $\mathcal{E}$ depends only on win-bit stacks and is independent of the value sequence. Consequently $v_t$ still has law $H$ conditional on the pre-value history and $\mathcal{E}$. On $\mathcal{E}$, optimism (58) bounds the conditional expected active score below by $I_t D^\star_{F,B}$; off $\mathcal{E}$, the score is nonnegative. Since $D^\star_{F,B} \leq 1$ and $\Pr(\mathcal{E}^c) \leq \delta$,

$$\mathbb{E}\left[\sum_{t\leq\tau}(r_t + S_t)\right] \geq D^\star_{F,B}\mathbb{E}[\tau] - T\delta.$$

The prefix form of (56) bounds $\sum_{t\leq\tau} S_t$ pathwise, yielding

$$\mathbb{E}\left[\sum_{t\leq\tau} r_t\right] \geq D^\star_{F,B}\mathbb{E}[\tau] - T\delta - \frac{2}{\eta} - \eta T M_B. \tag{64}$$

The budget part of (61) gives

$$A_\tau - \beta\tau \leq \frac{\log M_B}{\eta} + \frac{T}{K} + \sum_{t\leq\tau} e_t + M^{\mathrm{pay}}_\tau, \tag{65}$$

where $M^{\mathrm{pay}}_\tau = \sum_{t\leq\tau}\{P_t - a_t F(a_t)\}$ is a stopped martingale and $\mathbb{E}[|M^{\mathrm{pay}}_\tau|] \leq \sqrt{T}$. On paths where $\tau < T$, the stopping rule implies $A_\tau > B - 1 = \beta T - 1$. On paths where $\tau = T$, the following bound is trivial. Thus on all paths,

$$\beta(T - \tau) \leq 1 + \frac{\log M_B}{\eta} + \frac{T}{K} + \sum_{t\leq\tau} e_t + |M^{\mathrm{pay}}_\tau|. \tag{66}$$

Only active rounds query the confidence model, so the selected-error argument in (20) applies to their stopped sequence as well. Taking expectations,

$$\mathbb{E}[T-\tau] \le \frac{1+\eta^{-1}\log M_B + T/K + K + C\sqrt{KTL_T} + T\delta + \sqrt{T}}{\beta}. \tag{67}$$

After stopping, bid 0 contributes nonnegative reward and nonnegative ROI surplus. Applying (60) on active rounds, then (64) and (20), gives

$$\mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \ge D^\star_{F,B}\mathbb{E}[\tau] - T\delta - \frac{2}{\eta} - \eta T M_B - \mathbb{E}\left[\sum_{t\le\tau} e_t\right]. \tag{68}$$

Combining this with (52) and $0 \le D^\star_{F,B} \le 1$ yields

$$\mathrm{Reg}_T^B \le \mathbb{E}[T-\tau] + T\delta + \frac{2}{\eta} + \eta T M_B + \mathbb{E}\left[\sum_{t\le\tau} e_t\right].$$

The selected-error bound and (67) make the right-hand side $\widetilde{O}(T^{2/3}/\beta)$.

To bound ROI violation, define the stopped martingale

$$M_\tau^{\mathrm{ROI}} = \sum_{t\le\tau}\big[(P_t - R_t) - (a_t - v_t)F(a_t)\big].$$

Its increments are bounded by one and $\mathbb{E}[|M_\tau^{\mathrm{ROI}}|] \le \sqrt{T}$. The ROI part of (61) and the prefix form of (57) imply

$$\sum_{t\le\tau}(P_t - R_t) \le \frac{\log M_B}{\eta} + \frac{T}{K} + \sum_{t\le\tau} e_t + M_\tau^{\mathrm{ROI}}.$$

After stopping, each bid 0 has payment zero and nonnegative reward. Taking the positive part of the full-horizon deficit, then expectations, gives

$$\mathrm{Viol}_T^{\mathrm{ROI}} \le \frac{\log M_B}{\eta} + \frac{T}{K} + K + C\sqrt{KTL_T} + T\delta + \sqrt{T} = \widetilde{O}(T^{2/3}).$$

□

### C.3 The unrestricted distributional benchmark

*Proof of Corollary 1.* For any rule feasible in (17) and any $\lambda, \mu \ge 0$,

$$\begin{aligned}\mathbb{E}[vF(b)] &\le \mu\beta + \mathbb{E}\big[F(b)\big((1+\lambda)v - (\lambda+\mu)b\big)\big] \\ &\le \mathbb{E}_v[f_F^B(v,\lambda,\mu)].\end{aligned}$$

Taking the supremum over feasible rules and then the infimum over $\lambda, \mu$ shows $\mathrm{OPT}^B_{\mathrm{pop}}(F,H) \le D^\star_{F,B}$. The stopped reward bound (68), the selected-error estimate (20), and (67) give

$$\mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \ge T D^\star_{F,B} - \widetilde{O}(T^{2/3}/\beta),$$

since $0 \le D^\star_{F,B} \le 1$. This proves the corollary without comparing the two offline benchmark values directly. □

### C.4 Regret lower bound with a budget constraint

We justify the regret-only comparison stated in Section 4. The instance in Aggarwal et al. (2025, Theorem 5.1) is value-maximizing, first-price, has $v_t = 1$ in every round and $B = \beta_0 T$ for a fixed constant $\beta_0 \in (0,1)$, and gives the bidder bandit feedback. For every bid $b \in [0,1]$, $(b - v_t)\mathbf{1}\{b \ge d_t\} \le 0$; hence our ROI constraint holds on every sample path and does not restrict either comparator

or learner on this instance. First-price payment is determined by the bidder's own bid and win bit, so their feedback supplies no additional information here.

Their offline benchmark mixes Lipschitz functions from values to bids. When the only value is 1, each function contributes just its bid at 1, while every constant bid is itself Lipschitz. Mixing such functions is therefore equivalent to an arbitrary distribution over bids. This gives exactly the per-round benchmark $\mathrm{OPT}^{B}_{\mathrm{pop}}(F, H)$ in (17). Moreover, any policy in our sequence-aware benchmark $\mathrm{OPT}_{B}(F, V^{T})$ induces an average bid distribution $\bar{\pi} = T^{-1}\sum_{t}\pi_{t}$. Since the value is constant, this averaging preserves expected reward and payment; conversely one can repeat any bid distribution on every round. Consequently,

$$\mathrm{OPT}_{B}(F, (1, \ldots, 1)) = T\,\mathrm{OPT}^{B}_{\mathrm{pop}}(F, H)$$

on the instance of their theorem. Their $\Omega(T^{2/3})$ regret lower bound thus applies directly to our algorithms that keep total spending within the budget on this fixed-fraction instance. The comparison establishes the worst-case exponent in $T$; it does not determine the optimal dependence on $\beta$.