Bidding and Pricing in Budget and ROI Constrained Markets

Negin Golrezaei
MIT Sloan School of Management

Patrick Jaillet
MIT Electrical Engineering and Computer Science

Jason Cheuk Nam Liang
MIT Operations Research Center

Vahab Mirrokni
Google Research

arXiv:2107.07725v1 [cs.GT] 16 Jul 2021

In online advertising markets, setting budget and *return on investment* (ROI) constraints are two prevalent ways to help advertisers (i.e. buyers) utilize limited monetary resources efficiently. In this work, we provide a holistic view of ROI and budget constrained markets. We first tackle the buyer's bidding problem subject to both budget and ROI constraints in repeated second-price auctions. We show that the optimal buyer hindsight policy admits a "threshold-based" structure that suggests the buyer win all auctions during which her valuation-to-expenditure ratio is greater than some threshold. We further propose a threshold-based bidding framework that aims to mimic the hindsight bidding policy by learning its threshold. We show that when facing stochastic competition, our algorithm guarantees the satisfaction of both budget and ROI constraints and achieves sublinear regret compared to the optimal hindsight policy. Next, we study the seller's pricing problem against an ROI and budget constrained buyer. We establish that the seller's revenue function admits a bell-shaped structure, and then further propose a pricing algorithm that utilizes an episodic binary-search procedure to identify a revenue-optimal selling price. During each binary search episode, our pricing algorithm explores a particular price, allowing the buyer's learning algorithm to adapt and stabilize quickly. This, in turn, allows our seller algorithm to achieve sublinear regret against adaptive buyer algorithms that quickly react to price changes.

Key words: Online advertising, bidding under financial constraints, return-on-investment, pricing

1. Introduction

In online advertising markets, advertisers run ad campaigns in which they bid and compete for ad impressions through various forms of repeated auctions. To efficiently utilize limited monetary resources that are allocated to a certain campaign, advertisers' bidding strategies are typically subject to financial constraints. Such constraints generally include budget and *return-on-investment* (ROI) constraints. Budget constraints primarily reflect advertisers' monetary limits due to organizational planning, while ROI constraints enforces the desired performance/return on the amount of capital spent Kireyev et al. (2016), Golrezaei et al. (2018), Balseiro et al. (2019b). Both constraints are empirically validated in practice (see e.g. Auerbach et al. (2008), Golrezaei et al. (2018)). In this work, we study these financial considerations from two complementary aspects in online advertising markets:

From the perspective of a buyer (advertiser), what is an optimal bidding strategy that achieves a high utility while maintaining both budget and ROI constraints in the long run? From the perspective of a seller (mechanism designer), what is an optimal pricing strategy against a buyer with both budget and ROI constraints who aims to learn her bidding strategy?

Bidding under ROI and budget constraints. We consider a buyer who participates in repeated second price auctions with some predesignated budget and target ROI, and aims to maximize quasi-linear utility. The buyer is subject to a budget constraint which sets a cap on the her total expenditure; and also an ROI constraint that requires the total accumulated valuation divided by total expenditure to be at least the target ROI. The problem of interest is to learn how to bid under both constraints.

The problem of learning how to bid only under a budget constraint has been studied extensively in the literature,[1] and is closely related to the more general area of online resource allocation subject to capacity or packing constraints, which includes but is not limited to online knapsack Vaze (2018), Zhou et al. (2008), packing Seiden (2002), Buchbinder and Naor (2009), Feldman et al. (2010), Kesselheim et al. (2014), secretary problems Babaioff et al. (2007, 2008), Arlotto and Gurvich (2019). One of the common approaches for the problem with only budget constraints is the adaptation of a "pacing" strategy, which is motivated by the primal-dual framework (see Balseiro and Gur (2019),

[1] We refer readers to Balseiro et al. (2021), Balseiro and Gur (2019) for a comprehensive study on budget management strategies and their influences to the overall market. Also see Feldman et al. (2007) for a study on advertiser budget optimization for search-based auctions.

Balseiro et al. (2020a)). Although these primal-dual algorithms for budget constrained problems are shown to have near-optimal performances in their settings, they cannot be generalized to deal with ROI constraints as they cannot guarantee the buyer achieves a predetermined ROI target over all auctions; see Section 3 for an example. Particularly, when there is no ROI constraint and one only needs to deal with budget constraints, primal-dual algorithms would terminate when the budget is depleted, making use of the fact that total buyer expenditure always increases in time. However, this hard-stopping procedure is not valid with ROI constraints as the realized ROI may increase or decrease over time, and can possibly drop below the buyer's target ROI at some point.

Given this crucial observation, rather than designing a bidding strategy based on the primal-dual framework, we develop a bidding strategy that is motivated by the structure of the optimal solution to the primal (hindsight) problem. This solution admits a "threshold-based" structure that suggests the buyer's optimal strategy is to win all auctions during which the "value-to-cost ratio" is greater than some threshold. This structure then inspires our threshold-based bidding algorithm which in every period randomizes over two possible bids. We show that in a stochastic setting, our algorithm can obtain a T-period regret in the order of $O(\sqrt{T})$.

Pricing against single buyer with ROI and budget constraints. We complement our study on learning how to bid with the seller's online pricing problem: we focus on designing seller pricing policies against a budget and ROI constrained buyer. In our pricing problem, as our of our main challenges, both the buyer and seller adopt online algorithms to achieve their respective objectives. Hence, the environment faced by the seller is neither stochastic nor adversarial. A similar pricing problem is studied in Braverman et al. (2018), where they show that when an unconstrained quasi-linear buyer adopts a certain class of learning algorithms, which they refer to as "mean-based" algorithms (e.g. Follow the Perturbed Leader algorithm and EXP3), the seller can extract the buyer's entire surplus; see Deng et al. (2019) for an extension of this work. In our work, due to the existence of the ROI and budget constraints, it is not possible to extract the buyer's entire surplus. Nonetheless, we show that the seller can learn a revenue-optimal selling price by suffering an additive (sublinear) revenue loss.

Designing pricing algorithms that eventually learn revenue-optimal selling price is cumbersome, as the buyer's model primitives (i.e. budget rate, target ROI, valuation distribution) are not known to the seller. We overcome this hurdle by identifying a bell-shaped structure of the seller's per-period revenue function given that the buyer is clairvoyant and acts optimally. We exploit this bell-shaped structure by designing an episodic binary search pricing algorithm. The episodic structure enables the buyer's algorithm to adapt to changes in prices and roughly speaking, allows the seller to obtain regret in the order of how well the buyer reacts to changes in seller's prices. This holds for our proposed buyer bidding algorithm and any other buyer algorithm that is adaptive to changes in prices.

We refer readers to Appendix A for an extended literature review.

2. Preliminaries

Notation. Let $\mathbb{R}_+$ be all strictly positive real numbers. For integer $N \in \mathbb{N}$, denote $[N] = \{1, 2 \dots N\}$ and $\Delta_N = \left\{\boldsymbol{p} \in [0,1]^N : \sum_{n\in[N]} p^n = 1\right\}$ be the N-dimensional probability simplex. For a real number $x \in \mathbb{R}$, denote $(x)_+$ as its positive part. For a vector $\boldsymbol{a}$, denote $\|\boldsymbol{a}\|$ as the Euclidean norm of $\boldsymbol{a}$. For any two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, let $\min\{\boldsymbol{a}, \boldsymbol{b}\} = (\min\{a_i, b_i\})_{i\in[n]}$ be the element-wise minimum. We write $\boldsymbol{a} \preceq \boldsymbol{b}$ if and only if $a_i \leq b_i$ and $\boldsymbol{a} \succeq \boldsymbol{b}$ if and only if $a_i \geq b_i$ for all $i \in [n]$.

Model. Consider a buyer competing in repeated second price auctions over a finite time horizon $T > 2$. During each period $t \in [T]$, the buyer observes her (private) valuation $v_t > 0$ for the auctioned ad impression, and then submits a bid value $b_t > 0$. If b_t is greater than the highest competing bid $d_t > 0$, the buyer wins the auction and pays d_t. Otherwise the buyer pays nothing. At the end of the period, the buyer observes d_t regardless of whether she won the auction or not. We assume that the per-period valuation and competing bid pair (v_t, d_t) is supported on some finite set $\mathcal{W} = \{(v^1, d^1) \dots (v^K, d^K)\}$,[2] which is known to the buyer. We call the occurrence of $(v_t, d_t) = (v^k, d^k)$ a *type-k arrival* during period t. For Sections 3, 4, and 5, where we study bidder's bidding problem,

[2] Continuous supports can be handled by discretization. This will cause additional buyer revenue loss, but by choosing proper discretization sizes, buyer's regret (defined later in Equation (3)) can remain sublinear in T.

we assume there is a time-invariant occurrence distribution $\boldsymbol{p} = (p^1 \dots p^K) \in \Delta_K$, unknown to the buyer, such that that $\mathbb{P}\left((v_t, d_t) = (v^k, d^k)\right) = p^k > 0$ for $\forall\ t \in [T]$.

We assume all valuations and highest competing bids are strictly positive and bounded, and $\left\{(v^k, d^k)\right\}_{k\in[K]}$ are ordered such that $\frac{v^1}{d^1} > \cdots > \frac{v^K}{d^K}$. Denoting $\theta^k = \frac{v^k}{d^k}$, $k \in [K+1]$ as the *value-to-cost* ratio of the type-k arrival, we thus have $\theta^1 > \theta^2 > \ldots > \theta^K > \theta^{K+1} = 0$, where we define $v^{K+1} = 0$ and $d^{K+1} = \infty$. We also write $\bar{d} = \max_{k\in[K]} d^k$ and $\underline{d} = \min_{k\in[K]} d^k$.

Buyer's objective. The buyer aims to maximize her cumulative expected utility $\mathbb{E}\big[\sum_{t\in[T]} (v_t - \alpha d_t)\, z_t\big]$, where $z_t = \mathbb{I}\{b_t \geq d_t\}$ is an indicator of whether the buyer won the auction in period t. The expectation is taken over $\{(v_t, d_t)\}_{t\in[T]}$ and possibly the randomness in bids $\{b_t\}_{t\in[T]}$. Here, $\alpha \geq 0$ can be viewed as the buyer's *private capital cost* that normalizes the buyer's accumulated valuation with total expenditure. For simplicity, we assume $v^k - \alpha d^k \neq 0$ for all $k \in [K]$. The private capital cost utility model, which includes the quasi-linear ($\alpha = 1$) and value-maximizing ($\alpha = 0$) utility models, is studied in various settings (e.g. see Balseiro et al. (2019b)).

Buyer's financial constraints and feasible bidding strategies. They buyer employs bidding strategies that respect both a budget constraint and an ROI constraint. More specifically, a buyer (non-anticipating) bidding strategy $\boldsymbol{\beta}$ (possibly randomized) induces bids $\{b_t^{\boldsymbol{\beta}}\}_{t\in[T]}$ where each bid value $b_t^{\boldsymbol{\beta}}$ can only depend on $\{(v_\tau, d_\tau, b_\tau^{\boldsymbol{\beta}})\}_{\tau\in[t-1]} \cup \{v_t\}$ and the randomness in the strategy. Consequently, the resulting bids $\{b_t^{\boldsymbol{\beta}}\}_t$ of strategy $\boldsymbol{\beta}$ should satisfy the following constraints:

$$\text{ROI} : \textstyle\sum_{t\in[T]} \mathbb{E}\left[(v_t - \gamma d_t)\, \mathbb{I}\big\{b_t^{\boldsymbol{\beta}} \geq d_t\big\}\right] \geq 0,\ \ \text{Budget} : \textstyle\sum_{t\in[T]} \mathbb{E}\left[d_t \mathbb{I}\big\{b_t^{\boldsymbol{\beta}} \geq d_t\big\}\right] \leq \rho T\,. \tag{1}$$

Here, $\rho > 0$ is called the *budget rate*, γ is the buyer's *target ROI* such that $\gamma > \alpha > 0$,[3] and the expectations are taken w.r.t. randomness in $\{(v_t, d_t, b_t^{\boldsymbol{\beta}})\}_{t\in[T]}$. Note that satisfying the ROI constraint guarantees the buyer's returns (measured by total accumulated value divided by total expenditure) are at least the target ROI γ, which can be seen by rewriting the ROI constraint as $\sum_{t\in[T]} \mathbb{E}[v_t \mathbb{I}\{b_t^{\boldsymbol{\beta}} \geq d_t\}] / \sum_{t\in[T]} \mathbb{E}[d_t \mathbb{I}\{b_t^{\boldsymbol{\beta}} \geq d_t\}] \geq \gamma$.

[3] Note that when $\gamma \leq \alpha$, the ROI constraint becomes redundant.

We remark that both budget and ROI constraints are studied in an expected sense. Such "soft" constraints are useful in practice due to the fact that real-world advertisers typically engage in many different online advertising campaigns, so it is reasonable to maintain these financial constraints on an aggregate level. We note that such soft financial constraints are also studied in mechanism design and online learning literature such as Vaze (2018), Golrezaei et al. (2018).

Buyer's regret. We evaluate any bidding strategy by comparing its cumulative utility to the best achievable utility in hindsight, which, for any realization $\boldsymbol{X} = \{(v_t, d_t)\}_{t\in[T]}$ is defined as

$$\begin{aligned} \mathrm{OPT}(\boldsymbol{X};\alpha,\gamma,\rho) = \max_{\boldsymbol{z}\in[0,1]^T} \quad & \textstyle\sum_{t\in[T]} (v_t - \alpha d_t)\, z_t \\ \text{s.t.} \quad & \textstyle\sum_{t\in[T]} (v_t - \gamma d_t)\, z_t \geq 0, \quad \sum_{t\in[T]} d_t z_t \leq \rho T \end{aligned} \tag{2}$$

Note that we considered the LP-relaxation (i.e., $z_t \in [0,1]$ rather than $z_t \in \{0,1\}$) of the hindsight problem as our benchmark. Also note in contrast with the soft ROI and budget constraints that are considered in expectation (see Equation (1), the constraints in the hindsight benchmark are "hard" constraints. Later in Theorem 2 we relate the optimization problems w.r.t. hard and soft constraints.

We quantitatively measure the performance of any bidding strategy $\boldsymbol{\beta}$ that satisfies both ROI and budget constraints in Equation (1) against the aforementioned benchmark using the notion of regret:

$$\mathrm{Reg}^{\boldsymbol{\beta}}(T,\alpha,\gamma,\rho) = \mathbb{E}\left[\mathrm{OPT}(\boldsymbol{X};\alpha,\gamma,\rho) - \textstyle\sum_{t\in[T]} (v_t - \alpha d_t)\, \mathbb{I}\{b_t^{\boldsymbol{\beta}} \geq d_t\}\right], \tag{3}$$

where the expectation is again w.r.t. all randomness over $\boldsymbol{X} = \{(v_t, d_t)\}_{t\in[T]}$ and $\{b_t^{\boldsymbol{\beta}}\}_{t\in[T]}$.

3. A Primal-dual View for Financial Constraints and Failure of Pacing

The primal dual framework coupled with "pacing" has been widely used to resolve the online bidding problem under a a hard budget constraint (see Balseiro et al. (2020b,a)). Nevertheless, in this section we argue that naively applying these approaches to the bidding problem under a single ROI constraint does not necessarily guarantee satisfaction of the ROI constraint in the long run.

Overview of the primal-dual view and a pacing policy for budget constraints. Consider the primal and Lagrangian of the problem where only the budget constraint is present:

$$\begin{aligned} &\mathtt{Primal}^{\mathtt{B}} = \max_{\boldsymbol{z}\in[0,1]^T} \quad \textstyle\sum_{t\in[T]} (v_t - \alpha d_t)\, z_t \quad \text{s.t.} \quad \sum_{t\in[T]} d_t z_t \leq \rho T \\ &L^{\mathtt{B}}(\boldsymbol{z},\lambda) = \lambda\rho T + \textstyle\sum_{t\in[T]} (v_t - (\alpha+\lambda)\, d_t)\, z_t\,, \end{aligned} \tag{4}$$

where $\lambda \geq 0$ is the dual variable w.r.t. the budget constraint. The dual function is defined as

$$\texttt{Dual}^{\texttt{B}}(\lambda) = \max_{\boldsymbol{z}\in[0,1]^T} L^{\texttt{B}}(\boldsymbol{z},\lambda) = L^{\texttt{B}}(\boldsymbol{z}^*(\lambda),\mu) = \lambda\rho T + \sum_{t\in[T]} \left(v_t - (\alpha+\lambda)\, d_t\right)_+ \,, \tag{5}$$

where $z_t^*(\lambda) = \arg\max_{z_t\in[0,1]} L^{\texttt{B}}(\boldsymbol{z},\lambda) = \arg\max_{z_t\in[0,1]} \left(v_t - (\alpha+\lambda)\, d_t\right) z_t$ is referred to as the optimal hindsight auction outcome in period t w.r.t. some fixed $\lambda \geq 0$. It is easy to observe that $z_t^*(\lambda) = \mathbb{I}\left\{v_t - (\alpha+\lambda)\, d_t \geq 0\right\}$. Note that

$$\mathbb{E}\left[\texttt{Primal}^{\texttt{B}}\right] \leq \min_{\lambda\geq 0} \mathbb{E}\left[\texttt{Dual}^{\texttt{B}}(\lambda)\right] = T \min_{\lambda\geq 0}\ \lambda\rho + \sum_{k\in[K]} p^k \left(v^k - (\alpha+\lambda)\, d^k\right)_+ \,, \tag{6}$$

which also induces an optimal dual variable $\lambda^* = \arg\min_{\lambda\geq 0} \mathbb{E}\left[\texttt{Dual}^{\texttt{B}}(\lambda)\right]$. The primal-dual approach aims to achieve this dual function upper bound via approximating the optimal outcome $z_t^*(\lambda^*)$ in Equation (5) through submitting a so-called "paced bid" $b_t = \frac{v_t}{\alpha+\lambda^*}$. This is because by doing so the realized outcome z_t is identical to the optimal outcome $z_t^*(\lambda^*)$, i.e. $z_t = \mathbb{I}\{b_t \geq d_t\} = \mathbb{I}\{v_t \geq (\alpha+\lambda^*)d_t\} = z_t^*(\lambda^*)$. As λ^* is typically unknown, the primal-dual framework is usually coupled with some learning algorithm such as dual Mirror Descent (e.g. see Balseiro et al. (2020b,a)) that maintains an estimate $\hat{\lambda} \geq 0$ over time, and submits a corresponding paced bids $b_t = \frac{v_t}{\alpha+\hat{\lambda}}$. During the run of the learning algorithm, the buyer intentionally stops the algorithm once the budget is depleted to ensure the budget constraint is satisfied. Previous works show that pacing under this primal-dual framework yields sublinear buyer regret against the dual upper bound in Equation (6); e.g. see Balseiro et al. (2021).

Failure of the primal-dual view and pacing policy ROI constraints. Analogously, the primal-dual framework and pacing can also be applied to the online bidding problem with a single ROI constraint:

$$\begin{aligned}
&\texttt{Primal}^{\texttt{R}} = \max_{\boldsymbol{z}\in[0,1]^T}\ \sum_{t\in[T]} \left(v_t - \alpha d_t\right) z_t \quad \text{s.t.} \quad \sum_{t\in[T]} \left(v_t - \gamma d_t\right) z_t \geq 0\,,\\
&L^{\texttt{R}}(\boldsymbol{z},\mu) = \sum_{t\in[T]} \left((1+\mu)\, v_t - (\alpha+\gamma\mu)\, d_t\right) z_t\,,\\
&\texttt{Dual}^{\texttt{R}}(\mu) = \max_{\boldsymbol{z}\in[0,1]^T} L^{\texttt{R}}(\boldsymbol{z},\mu) = \sum_{t\in[T]} \left((1+\mu)\, v_t - (\alpha+\gamma\mu)\, d_t\right)_+ \,,
\end{aligned} \tag{7}$$

where $\mu \geq 0$ is the dual variable corresponding to the ROI constraint, and $\mu^* = \arg\min_{\mu\geq 0} \mathbb{E}\left[\texttt{Dual}^{\texttt{R}}(\mu)\right]$ is the optimal dual variable. In this case, the paced bids become $b_t = \frac{(1+\hat{\mu})v_t}{\alpha+\gamma\hat{\mu}}$, where $\hat{\mu}$ is some estimate of μ^*. However, we present a simple problem instance that demonstrates this pacing strategy motivated by the dual variable does not necessarily guarantee a nonnegative expected cumulative ROI balance, even when the optimal dual variable is known.

Example 1 *Assume* $\alpha = 0$, $v_t = v > 0$ *for all* $t \in [T]$, *and* $d_t = \frac{v}{2\gamma}$ *w.p.* p *and* $\frac{3v}{2\gamma}$ *w.p.* $1-p$ *for some* $p \in (0, 1/2)$. *Then,* $\mathbb{E}[\mathtt{Dual}^R(\mu)] = pvT\left(1+\frac{\mu}{2}\right)_+ + (1-p)vT\left(1-\frac{\mu}{2}\right)_+$ *and hence* $\mu^* = \arg\min_{\mu \geq 0} \mathbb{E}[\mathtt{Dual}^R(\mu)] = 2$. *In a hypothetical ideal world where the buyer knows* μ^*, *the paced bid value during period* $t \in [T]$ *is* $b_t = \frac{(1+\mu^*)v_t}{\gamma\mu^*} = \frac{3v}{2\gamma}$. *However, the cumulative expected ROI of this pacing strategy is* $\sum_{t\in[T]} \mathbb{E}[(v_t - \gamma d_t)\mathbb{I}\{b_t \geq d_t\}] = pT\left(v - \frac{v}{2}\right)\mathbb{I}\{\frac{3v}{2\gamma} \geq \frac{v}{2\gamma}\} + (1-p)T\left(v - \frac{3v}{2}\right)\mathbb{I}\{\frac{3v}{2\gamma} \geq \frac{3v}{2\gamma}\} = Tv\left(p - \frac{1}{2}\right) < 0$. *The final inequality follows from* $p < \frac{1}{2}$.

We remark that although pacing under the primal-dual framework allows the buyer to approximate the dual upper bound with sublinear loss, the pacing strategy alone does not necessarily guarantee primal feasibility (e.g. satisfaction of the budget or ROI constraint), so typically a "stopping mechanic" is imposed. For instance, for online bidding with a hard budget (i.e. $\sum_{t\in[T]} d_t z_t \leq \rho T$ *w.p.*1), the budget constraint is satisfied by terminating the pacing algorithm once budget is depleted. The rationale behind such a hard stopping time procedure is that total expenditure is non-decreasing over time. Nevertheless, such a hard stopping procedure is not applicable for the ROI problem since per period ROI balance $(v_t - \gamma d_t)\, z_t$ may be strictly negative, which means the cummulated ROI balance until period t, namely $\sum_{\tau\in[t]} (v_\tau - \gamma d_\tau)\, z_\tau$, may drop below 0 as time proceeds. This motivates the needs for alternative bidding frameworks to ensure ROI and budget constraint satisfaction.

4. A Reformulated Problem and the Threshold-based Optimal Solution

In this section, we first introduce a reformulation for $\mathrm{OPT}(\{v_t, d_t\}_{t\in[T]}; \alpha, \gamma, \rho)$ whose closed form solution admits what we refer to as a "threshold-based structure". This insight will later motivate our bidding strategy in a very straightforward manner. To begin with, we consider the following optimization problem for any $\boldsymbol{n} \in \mathbb{R}_+^K$ and $c > 0$:

$$\begin{aligned} U(\boldsymbol{n}; \alpha, \gamma, c) = \max_{\boldsymbol{x}\in[0,1]^K} & \sum_{k\in[K]} n^k\,(v^k - \alpha d^k)\, x^k \\ \text{s.t.} \quad & \sum_{k\in[K]} n^k\,(v^k - \gamma d^k)\, x^k \geq 0, \quad \sum_{k\in[K]} n^k d^k x^k \leq c \end{aligned} \tag{8}$$

Here, the decision variables x^k can be interpreted as the proportion of auctions during type-k arrivals won by the buyer. The following proposition shows that $U(\boldsymbol{n};\alpha,\gamma,\rho T)$ and $\mathrm{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)$ are closely related.

Proposition 1 *Define* $N^k=\sum_{t\in[T]}\mathbb{I}\left\{(v_t,d_t)=(v^k,d^k)\right\}$*, and write* $\boldsymbol{N}=(N^k)_{k\in[K]}$*. Note that* N^k *is a random variable, and* $\mathbb{E}\left[N^k\right]=p^kT$*. Then* $OPT(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)=U(\boldsymbol{N};\alpha,\gamma,\rho T)$*.*

The proof of this proposition can be found in Appendix D.1. Before we discuss the solution to the reformulated problem, we introduce the notion of a *threshold vector* to simplify notation.

Definition 1 (Threshold vectors) *We say that an* K*-dimensional vector* $\boldsymbol{x}\in\mathbb{R}^K$ *is a threshold vector if it takes the form of* $\boldsymbol{x}=(1\dots1,q,0\dots0)$*, where the first* $J\in\{0,\dots K\}$ *entries are 1's, followed by some number* $q\in[0,1)$*, and trailing with* $(K-J-1)_+$ *0's.*[4] *Any threshold vector is uniquely characterized by its dimension* K*, as well as, a tuple* $(J,q)\in\{0,\dots K\}\times[0,1)$*, so we denote the vector as* $\psi(J,q)$*. In the special case when* $J=K$*, take* $q=0$*.*

We remark that for any two threshold vectors $\boldsymbol{a},\boldsymbol{b}$ of the same dimension, $\min\{\boldsymbol{a},\boldsymbol{b}\}$ is also a threshold vector. Furthermore, either $\boldsymbol{a}\preceq\boldsymbol{b}$ or $\boldsymbol{a}\succeq\boldsymbol{b}$. Using this definition of threshold vectors, the following Theorem 1 states that the optimal solution to $U(\boldsymbol{n};\alpha,\gamma,c)$ is a threshold vector.

Theorem 1 (Threshold-based solution) *Fix* $\boldsymbol{n}\in\mathbb{R}^K_+$*,* $c>0$*,* $\gamma>0$*, and let* $n^{K+1}=\infty$*. Define* [5]

$$
\begin{aligned}
&r=\max\left\{k\in[K]:\textstyle\sum_{\ell\in[k]}n^\ell\left(v^\ell-\gamma d^\ell\right)\geq0\right\},\quad q^{\mathtt{R}}=\frac{\sum_{k\in[r]}n^kw^k}{n^{r+1}\cdot\left|w^{r+1}\right|},\\
&b=\max\left\{k\in[K]:\textstyle\sum_{\ell\in[k]}n^\ell d^\ell\leq c\right\},\quad \textit{and}\quad q^{\mathtt{B}}=\frac{c-\sum_{k\in[b]}n^kd^k}{n^{b+1}\cdot d^{b+1}},
\end{aligned}
\tag{9}
$$

If we let $\boldsymbol{x}^{\mathtt{R}}=\psi(r,q^{\mathtt{R}})$ *and* $\boldsymbol{x}^{\mathtt{B}}=\psi(b,q^{\mathtt{B}})$ *be two threshold vectors, then* $\boldsymbol{x}^*=\min\{\boldsymbol{x}^{\mathtt{R}},\boldsymbol{x}^{\mathtt{B}},\psi(\kappa_\alpha,0)\}$ *is an optimal solution to* $U(\boldsymbol{n};\alpha,\gamma,c)$*. Here,* $\kappa_\alpha=\max\{k\in[K]:v^k-\alpha d^k\geq0\}$*.*[6] *Furthermore,* $\boldsymbol{x}^*$ *is also a threshold vector characterized by tuple* (J,q) *where*

$$
J=\min\{r,b,\kappa_\alpha\},\quad q=x^{*,J+1}=\min\left\{x^{\mathtt{B},J+1},x^{\mathtt{R},J+1}\right\}\cdot\mathbb{I}\{J+1\leq\kappa_\alpha\}\,.
\tag{10}
$$

[4] For the edge case of $(1,\dots1)\in\mathbb{R}^K$, $J=K$ and hence the number of trailing 0's is $(K-J-1)_+=0$.

[5] In the rest of the paper `B`, `R` will be the shorthand notation for "Budget" and "ROI" respectively.

[6] $\psi(\kappa_\alpha,0)$ is the threshold vector whose first κ_α entries are 1's while the rest are 0.

Remark 1 *We note that the variables* $(b, r, J, q^B, q^R, q, \boldsymbol{x}^B, \boldsymbol{x}^R, \boldsymbol{x}^*)$ *in Theorem 1 depend on the parameters* $\boldsymbol{n}, \alpha, \gamma, c$. *In the rest of the paper, if not stated otherwise, we set* $\boldsymbol{n} = \boldsymbol{p}$ *and* $c = \rho$ *when computing the aforementioned variables.*

For the proof of this theorem, please see Appendix D.2. The structure of the optimal solution $\boldsymbol{x}^*$ that we characterized in Theorem 1 suggests in hindsight the buyer should win all auctions with arrival types $1, 2 \ldots, J$ where the threshold $J = \min\{r, b, \kappa_\alpha\}$, and win a q proportion of the auctions with arrival type $J + 1$ while ignoring all other arrival types $J + 2 \ldots K$.

Threshold-based bidding strategies. Now, we demonstrate how we can transform the idea of having the bidder "win all auctions with arrival types $1, \ldots J$" as illustrated in Theorem 1 into a practical bidding strategy. We point out that instead of considering a threshold for arrival types, we can equivalently study the value-to-cost ratio for each arrival type. Since the value-to-cost ratios (i.e., $\theta^i = v^i/d^i$, $i \in [K]$) are ordered such that $\theta^1 > \ldots \theta^K$, Theorem 1 suggests that the buyer wins the auction during period t if the value-to-cost ratio v_t/d_t is at least θ^J, and win with probability $1 - q$ if v_t/d_t is θ^{J+1}. This value-to-cost ratios viewpoint thus motivates the following bidding strategy which we call a *threshold-based bidding* strategy.

Definition 2 (Threshold-based bidding strategy) *Recall that* $\theta^k = v^k/d^k$ *for any* $k \in [K]$. *For some threshold type* $k \in [K]$ *and remainder probability* $q \in [0, 1)$, *a threshold-based bidding strategy, denoted by* $\beta(v; k, q)$, *maps valuation* v *to bid value* v/θ^k *w.p.* q *and* v/θ^{k+1} *w.p.* $1 - q$.

In the threshold-based strategy, when the buyer submits v_t/θ^k, she wins the auction if $v_t/\theta^k \geq d_t$, which is equivalent to having the value-to-cost ratio during the current period v_t/d_t to be greater than the threshold value-to-cost ratio θ^k. Similarly, submitting bid value v_t/θ^{k+1} allows the buyer to win the auction if current period value-to-cost ratio is greater than θ^{k+1}. In light of this threshold-based bidding strategy, the following theorem states that submitting threshold bids w.r.t. the optimal threshold θ^J and remainder probability q is not only optimal, but also satisfies both budget and ROI constraints. The proof of this theorem is detailed in Appendix D.3.

Theorem 2 (Optimal threshold-based bidding for known $\boldsymbol{p}$) *Let (J,q) be defined as Theorem 1 w.r.t. $\boldsymbol{n}=\boldsymbol{p}$ and $c=\rho$. If for each period $t\in[T]$ the buyer submits threshold bid $b_t=\beta(v_t,J,q)$, then $\mathbb{E}[\sum_{t\in[T]}(v_t-\alpha d_t)\mathbb{I}\{b_t\geq d_t\}]=TU(\boldsymbol{p};\alpha,\gamma,\rho)\geq\mathbb{E}[OPT(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)]$. Furthermore, this bidding strategy satisfies both budget and ROI constraints in Equation (1).*

We remark that when capital cost $\alpha\geq 1$, the optimal threshold $\theta^J=\max\{\theta^b,\theta^r,\alpha\}\geq 1$, suggesting the buyer should always underbid. This means overbidding is possible when $0<\alpha<1$.

5. Online Threshold Bidding Algorithm

In this section, we introduce our bidding framework that harnesses the threshold-based bidding structure described in Theorem 2. We refer to our proposed bidding framework as *Conservative Threshold-based Bidding under Budget and ROI Constraints (*CTBR*)*. The framework consists of three components, namely learning algorithm $\mathcal{A}$; confidence bound ℓ_t, and threshold-based bidding.

Learning algorithm $\mathcal{A}$. The framework takes in any learning algorithm $\mathcal{A}$ that maps the current distribution estimate $\hat{\boldsymbol{p}}_t$ and historical data $\{v_\tau,d_\tau,b_\tau\}_{\tau\in[t-1]}$ to an updated estimate. We point out that a merit of the CTBR framework is that the freedom to choose any learning algorithms enables advertisers to customize their own learning algorithms according to practical considerations, including but not limited to employing non-standard learning algorithms that are robust to corrupted or outlier data that is perhaps originated from behavioral anomalies or market shocks (e.g. see Lykouris et al. (2018), Gupta et al. (2019)). Here, we present two simple learning algorithms as illustrative examples, namely *empirical estimation* (EE) and *Stochastic Gradient Descent* (SGD). Let $\boldsymbol{s}_t=(\mathbb{I}\{(v_t,d_t)=(v^1,d^1)\},\dots,\mathbb{I}\{(v_t,d_t)=(v^K,d^K)\})\in\{0,1\}^K$ characterize the occurrence of each arrival type in period t. Then the two algorithms update estimates for $\boldsymbol{p}$ as followed:

$$\text{EE:}\quad \hat{\boldsymbol{p}}_{t+1}=(t\cdot\hat{\boldsymbol{p}}_t+\boldsymbol{s}_t)/(t+1),\quad \text{SGD:}\quad \hat{\boldsymbol{p}}_{t+1}=\arg\min_{\widetilde{\boldsymbol{p}}\in\Delta_K}\|\widetilde{\boldsymbol{p}}-(\hat{\boldsymbol{p}}_t-\eta_t\hat{g}_t)\|. \tag{11}$$

where for SGD, $\hat{g}_t=\hat{\boldsymbol{p}}_t-\boldsymbol{s}_t$ is a stochastic gradient of the function $f(\widetilde{\boldsymbol{p}})=\frac{1}{2}\|\widetilde{\boldsymbol{p}}-\boldsymbol{p}\|^2$ at the point $\widetilde{\boldsymbol{p}}=\hat{\boldsymbol{p}}_t$, and η_t is called the *step size* at period t. We note that SGD is generally associated with either vanishing step sizes (e.g. $\eta_t=1/\sqrt{t}$ or $1/t$), or constant step sizes that can possibly depend on the total number of periods T. For more details on related descent methods, see e.g. Cohen (1981), Boyd et al. (2004).

Confidence bound ℓ_t. The CTBR algorithm also takes in a sequence of confidence bounds $\{\ell_t\}_{t\in[T]}$ to construct conservative estimates of the optimal arrival-type threshold J and the remainder probability q based off $\mathcal{A}$'s distribution update $\hat{\boldsymbol{p}}_t$; see Equations (12) and (13). Recall that the $1-q$ is the winning probability that the buyer hopes to attain during type-$(J+1)$ arrivals. The confidence bounds in $\hat{r}_t$ and $\hat{b}_t$ allow the buyer to obtain more accurate estimates for the threshold types r and b respectively when t increases (see Lemma 3 of Appendix E.1). On the other hand, the confidence bounds in the estimates $\hat{q}_t^{\mathtt{R}}$ and $\hat{q}_t^{\mathtt{B}}$ make bids more likely to take the smaller value $v_t/\theta^{\hat{J}_t}$ rather than the larger value $v_t/\theta^{\hat{J}_t+1}$ (since $\theta^{\hat{J}_t} < \theta^{\hat{J}_t+1}$). We note that bidding smaller values results in higher ROI and lower expenditure, and hence the bias towards lower values of $\hat{q}_t$ helps the buyer satisfy both budget and ROI constraints. [7] We remark that these confidence bounds can be viewed as the estimation accuracy of the input algorithm $\mathcal{A}$: later in Theorem 3 and 4, we show that when ℓ_t satisfies $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\| \leq \ell_t$ with high probability, CTBR achieves low regret. That being said, as we show via numerical studies in Section 5.1, satisfying the condition $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\| \leq \ell_t$ is not essential in the sense that CTBR algorithm maintains good performance even for naive choices of ℓ_t.

Threshold-based bidding. Motivated by the threshold-based bidding strategy in Theorem 2, CTBR submits a threshold-based bid (see Definition 2) w.r.t. the conservative estimates of the optimal arrival-type threshold J and the remainder probability q.

The bidding framework is detailed in Algorithm 1, in which for notational simplicity, for any $k \in [K+1]$, we define $w^k = v^k - \gamma d^k$ (where $w^{K+1} = -\infty$). We further assume $w^k \neq 0$ so that the buyer's hindsight problem $U(\boldsymbol{p};\alpha,\gamma,\rho)$ admits a unique optimal threshold-based solution. Also, in

[7] Smaller bids result in lower marginal expenditures, and thus higher realized ROI. For instance, assume $d_t = 1$ w.p. 1/3 and $d_t = 2$ w.p. 2/3. If the buyer with fixed valuation 1 submits bid value 1, she only wins the auction and attains value 1 when the highest competing bid is 1. Hence, both her expected accumulated value and expected expenditure are $1 \times 1/3 = 1/3$, so her realized ROI is $\frac{1/3}{1/3} = 1$. If the buyer submits bid value 2, she always wins the auction and attains value 1, but her expected expenditure is $1 \times 1/3 + 2 \times 2/3 = 5/3$, resulting in a lower realized ROI of $\frac{1}{5/3} = 3/5$.

the remaining we will assume $\sum_{k\in[r]} p^k w^k \neq 0$, $\sum_{k\in[J]} p^k w^k \neq 0$, and $\rho - \sum_{k\in[b]} p^k w^k \neq 0$ to rule out edge cases that complicate analysis without providing additional insights.

Algorithm 1 Conservative Threshold-based Bidding under Budget and ROI Constraints: $\text{CTBR}_{\mathcal{A}}$

Input: Learning algorithm $\mathcal{A}$, time-dependent confidence bound ℓ_t (possibly depending on T).

1: Initialize $\hat{J}_1 = 1$, $\hat{\boldsymbol{p}}_1 = (1/K \ldots 1/K) \in \Delta_K$, $\hat{q}_1 = 0$.

2: **for** $t = 1, 2, \ldots$ **do**

3: **Follow threshold-based bidding strategy**: Observe valuation v_t and submit threshold bid $b_t = \beta(v_t, \hat{J}_t, \hat{q}_t)$. After submitting bid b_t, observe highest competing bid d_t.

4: **Update estimate of distribution $\boldsymbol{p}$ by invoking algorithm $\mathcal{A}$:** $\hat{\boldsymbol{p}}_{t+1} = \mathcal{A}\big(\hat{\boldsymbol{p}}_t, \{v_\tau, d_\tau, b_\tau\}_{\tau\in[t]}\big)$.

5: **Update threshold vectors:** $\hat{\boldsymbol{x}}^{\mathtt{B}}_{t+1} = \psi\left(\hat{b}_{t+1}, \left(\hat{q}^{\mathtt{B}}_{t+1}\right)_+\right)$ and $\hat{\boldsymbol{x}}^{\mathtt{R}}_{t+1} = \psi\left(\hat{r}_{t+1}, \left(\hat{q}^{\mathtt{R}}_{t+1}\right)_+\right)$, where

$$\begin{aligned}
&\hat{r}_{t+1} = \max\left\{k \in [K] : \textstyle\sum_{\ell\in[k]} \hat{p}^\ell_{t+1} w^\ell \geq -\sqrt{K}\bar{w}\ell_t\right\} \text{ and } \hat{q}^{\mathtt{R}}_{t+1} = \frac{\sum_{\ell\in\left[\hat{r}_{t+1}\right]} \hat{p}^\ell_t w^\ell - \left(\sqrt{K}+2\right)\bar{w}\ell_t}{\hat{p}_t^{\hat{r}_{t+1}+1}\left|w^{\hat{r}_{t+1}+1}\right|} \\
&\hat{b}_{t+1} = \max\left\{k \in [K] : \textstyle\sum_{\ell\in[k]} \hat{p}^\ell_{t+1} d^\ell \leq \rho + \sqrt{K}\bar{d}\ell_t\right\} \text{ and } \hat{q}^{\mathtt{B}}_{t+1} = \frac{\rho - \sum_{\ell\in\left[\hat{b}_{t+1}\right]} \hat{p}^\ell_t d^\ell - \left(\sqrt{K}+2\right)\bar{d}\ell_t}{\hat{p}_t^{\hat{b}_{t+1}+1}\left|d^{\hat{b}_{t+1}+1}\right|}.
\end{aligned} \tag{12}$$

6: **Update the threshold-based bidding strategy:** Calculate $\hat{\boldsymbol{x}}_t = \min\left\{\hat{\boldsymbol{x}}^{\mathtt{B}}_{t+1}, \hat{\boldsymbol{x}}^{\mathtt{R}}_{t+1}, \psi(\kappa_\alpha, 0)\right\} = \psi(\hat{J}_{t+1}, \hat{q}_{t+1})$ where $\hat{J}_{t+1}$ and $\hat{q}_{t+1}$ are:

$$\hat{J}_{t+1} = \min\left\{\hat{r}_{t+1}, \hat{b}_{t+1}, \kappa_\alpha\right\} \quad \text{and} \quad \hat{q}_{t+1} = \min\left\{\hat{x}^{\mathtt{R},\hat{J}_{t+1}}_{t+1}, \hat{x}^{\mathtt{B},\hat{J}_{t+1}}_{t+1}\right\} \cdot \mathbb{I}\left\{\hat{J}_{t+1} + 1 \leq \kappa_\alpha\right\}. \tag{13}$$

7: **end for**

Remark 2 *In Equation (12), we always have $\hat{q}^{\mathtt{R}}_{t+1}, \hat{q}^{\mathtt{B}}_{t+1} < 1$. This is easy to see via combining two observations: (1) $\hat{p}_t^{\hat{r}_{t+1}+1} > 0$, and $w^{\hat{r}_{t+1}+1} < 0$. These inequalities hold because if $\hat{p}_t^{\hat{r}_{t+1}+1} = 0$ or $w^{\hat{r}_{t+1}+1} \geq 0$, $\sum_{\ell\in\left[\hat{r}_{t+1}+1\right]} \hat{p}^\ell_{t+1} w^\ell \geq \sum_{\ell\in\left[\hat{r}_{t+1}\right]} \hat{p}^\ell_{t+1} w^\ell \geq -\sqrt{K}\bar{w}\ell^{\mathcal{A}}(t,\delta)$, which contradicts the maximality of $\hat{r}_{t+1}$; (2) By the maximality of $\hat{r}_{t+1}$, we know that $\sum_{\ell\in[\hat{r}_{t+1}+1]} \hat{p}^\ell_{t+1} w^\ell < -\sqrt{K}\bar{w}\ell_t < 0$ so $\sum_{\ell\in[\hat{r}_{t+1}]} \hat{p}^\ell_{t+1} w^\ell < -\hat{p}_{t+1}^{\hat{r}_{t+1}+1} w^{\hat{r}_{t+1}+1}$. Dividing both sides by $-\hat{p}_{t+1}^{\hat{r}_{t+1}+1} w^{\hat{r}_{t+1}+1} > 0$ concludes $\hat{q}^{\mathtt{R}}_{t+1} < 1$. A similar argument also implies $\hat{q}^{\mathtt{B}}_{t+1} < 1$.*

As of our main result for this section, we theoretically show performance guarantees for our proposed CTBR framework. Due to space limitations, in the following theorem, we present an informal statement on the regret of CTBR. For a more detailed statement, we refer readers to Theorem 8 in Appendix B where we include additional results and discussion for Section 5.

Theorem 3 (Regret of CTBR) *Let* $\{\hat{\boldsymbol{p}}_t\}_{t\in[T]}$ *be the estimates of learning algorithm* $\mathcal{A}$ *and assume the input confidence bound* $\ell_t:(0,1)\to\mathbb{R}^+$*is decreasing in* t*,* $\lim_{t,T\to\infty}\ell_t=0$*,* [8] *and satisfies* $\mathbb{P}(\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\|\leq\ell_t)\geq 1-\frac{1}{T}$*. Then for large* T*, bidding according to* $\mathrm{CTBR}_{\mathcal{A}}$ *satisfies both ROI and budget constraints in Equation (1), and the regret is* $Reg(\boldsymbol{p},T,\alpha,\gamma,\rho)=\widetilde{\mathcal{O}}(\sum_{t\in[T]}\ell_t)$*.*

To put the performance of the CTBR framework in more detailed context, in the following theorem, we present $\mathrm{CTBR}_{\mathcal{A}}$'s regret when the learning algorithm $\mathcal{A}$ is EE and SGD respectively; see Equation (11). We again refer readers to Theorem 9 in Appendix B for a formal version of the statement.

Theorem 4 ($\mathrm{CTBR}_{\mathcal{A}}$ with EE and SGD) *(i) When* $\mathcal{A}$ *is EE, or SGD with vanishing step size* $\eta_t=1/t$*,* $\mathrm{CTBR}_{\mathcal{A}}$ *with input confidence bound* $\ell_t=\Theta(1/\sqrt{t})$ *incurs a regret of* $\widetilde{\mathcal{O}}(\sqrt{T})$*. (ii) When* $\mathcal{A}$ *is SGD with constant step size* $\eta_t=\eta=T^{-2/3}$*,* $\mathrm{CTBR}_{\mathcal{A}}$ *with input confidence bound* $\ell_t=\Theta((1-2\eta)^{t/2}+\sqrt{\eta})$ *incurs a regret of* $\widetilde{\mathcal{O}}\left(T^{2/3}\right)$*. Finally, in the aforementioned scenarios (i) and (ii),* $\mathrm{CTBR}_{\mathcal{A}}$ *satisfies both budget and ROI constraints in Equation (1).*

5.1. Numerical Study on Learning How to Bid

This section presents a numerical study regarding the CTBR framework. We consider three regimes, namely *ROI dominant*, *Budget dominant*, *α dominant*, each corresponding to model primitives $(\alpha,\gamma,\rho,\boldsymbol{p})$ such that in the optimal solution of the reformulated problem $U(\boldsymbol{p};\alpha,\gamma,\rho)$ in Equation (8), the ROI constraint is binding, the budget constraint is binding, and both constraints are non-binding, respectively. With a slight abuse of notation, we use $\mathtt{R},\mathtt{B}$, and $\mathtt{A}$ to denote the *ROI dominant*, *Budget dominant*, and *α dominant* regimes respectively.

Our experimental setup is described as followed. We fix capital cost $\alpha=1$ and consider three sets of parameters $(\gamma^{\mathtt{B}},\rho^{\mathtt{B}})=(1.2,0.05)$, $(\gamma^{\mathtt{R}},\rho^{\mathtt{R}})=(2.1,0.4)$ and $(\gamma^{\mathtt{A}},\rho^{\mathtt{A}})=(1.2,0.4)$, associated with each regime respectively. For each regime $y=\mathtt{R},\mathtt{B},\mathtt{A}$ we generate $N=100$ probability distributions

[8] Note that the confidence bound ℓ_t may depend on both the current period t and total horizon length T; e.g. see SGD with constant step sizes in Theorem 4

$\boldsymbol{P}^{y,i} \in \Delta_K$ $(i = 1, \ldots N)$ uniformly at random such that $\{(\alpha, \gamma^y, \rho^y, \boldsymbol{P}^{y,i})\}_{i\in[N]}$ belong to the respective regimes. Here, each of the K entries in $\boldsymbol{P}^{y,i}$ are sampled from Uniform$(0,1)$, and then rescaled by the sum of all entries to form a probability distribution. Then, for each set of model primitives $(\alpha, \gamma^y, \rho^y, \boldsymbol{P}^{y,i})$ $(y \in \{\texttt{R}, \texttt{B}, \texttt{A}\}, i \in [N])$, we sample $T = 10{,}000$ independent pairs $\{(v_t, d_t)\}_{t\in[T]}$ from $\boldsymbol{P}^{y,i} \in \Delta_K$ supported on $\mathcal{W} \subseteq \{0.2, 0.4, 0.6, 0.8, 1\}^2$ where $K = |\mathcal{W}| = 19$.

CTBR implementation robust to ℓ_t. Although buyers may not know the exact ℓ_t that satisfies the high probability bound condition in Theorem 3, here we demonstrate that from a practical viewpoint CTBR is robust to simple choices of ℓ_t. For $y \in \{\texttt{R}, \texttt{B}, \texttt{A}\}$ and each set of model primitives in $\{(\alpha, \gamma^y, \rho^y, \boldsymbol{P}^{y,i})\}_{i\in[N]}$, we run CTBR$_{\mathrm{EE}}$ with confidence bound $\ell_t = \frac{t^{-s}}{\max\{\bar{d}, \bar{w}\}\sqrt{K}}$ for $s = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1$ respectively over the T periods. Here $\max\{\bar{d}, \bar{w}\}\sqrt{K}$ is a normalization factor for illustrative purposes. In Figure 1, we observe that for all s, CTBR$_{\mathrm{EE}}$ achieves high utility and a realized target ROI greater than the target most of the time, demonstrating that CTBR with naively chosen confidence bounds produces fairly robust outcomes in general scenarios. An interesting insight from Figure 1 is the tradeoff between utility and realized ROI as s varies: at fixed t, larger s corresponds to the buyer being more aggressive and thus willing to win auctions with smaller value-to-cost ratios[9] via submitting larger bids. This would result in higher utility but reduce realized ROI as the marginal cost increases. Finally, we comment that the worst case realized ROI occurs in the ROI dominant regime (where the buyer aims for a 210% return) for $s = 1/2$. Here, ROI target is achieved in only $\sim 65\%$ of instances, yet we note that this is mainly because as s decreases, CTBR$_{\mathrm{EE}}$ converges more slowly, so our $T = 10{,}000$ is relatively small and CTBR$_{\mathrm{EE}}$ did not yet converge for $s = 1/2$. For more details on convergence see Figure 4 in Appendix C.

Comparison with benchmark bidding algorithms. We also compare CTBR with four benchmark bidding algorithms, namely `Conserv` , `Budget-Pacing`, `ROI-Pacing` and `Pacing`. Fixing buyer parameters (α, γ, ρ), submitted bids for each benchmark are as followed. `Conserv` simply bids v_t/γ

[9] Equation (12) states that larger ℓ_t yields larger $\hat{r}_{t+1}$ and $\hat{b}_{t+1}$. This may result in a larger estimate for the threshold arrival type J, corresponding to smaller values of $\theta^{\hat{J}}$ and thus higher thresholded bids $b_t = v_t/\theta^{\hat{J}}$.

in each period to guarantee realized ROI is greater than the target ROI γ with probability 1. For the other pacing-type algorithms, denote $\hat{\lambda}_t$ and $\hat{\mu}_t$ as the estimated dual variables in period t corresponding to the budget constraint and ROI constraints respectively for the optimization problem in Equation (2). Then, `Budget-Pacing` bids $\frac{v_t}{\alpha+\hat{\lambda}_t}$, `ROI-Pacing` bids $\frac{(1+\hat{\mu}_t)v_t}{\alpha+\gamma\hat{\mu}_t}$, and `Pacing` bids $\frac{(1+\hat{\mu}_t)v_t}{\alpha+\gamma\hat{\mu}_t+\hat{\lambda}_t}$. Note that the dual variables $\hat{\lambda}_t$ and $\hat{\mu}_t$ are updated via projected stochastic sub-gradient descent (see Balseiro et al. (2021)). We include the pseudocode of each benchmark in Appendix C. For CTBR we run CTBR_{EE} with confidence bound $\ell_t = \frac{t^{-1}}{\max\{\bar{d},\bar{w}\}\sqrt{K}}$. In Figure 2 (Left), we observe CTBR_{EE} achieves nearly the largest per period utility $\sum_{t\in[T]}(v_t - d_t)z_t/T$ in all regimes, and significantly outperforms pacing-type strategies. This is primarily because the benchmark algorithms either deplete their budget too slowly (e.g. see the budget depletion trajectory for these benchmarks in the α-dominant regime in Figure 2 (Right)), resulting in less auction wins and hence lower total utility; or they deplete their budget too quickly (e.g. see the budget depletion trajectory for `Conserv`, `ROI-Pacing` and `Pacing` in the Budget-dominant regime), and miss out opportunities to learn the best bidding strategy, leading to low overall utility. In contrast, CTBR strikes a balance between "learning" and "budget depletion" so that the improvement of bidding decisions occurs at a moderate rate as expenditure increases.

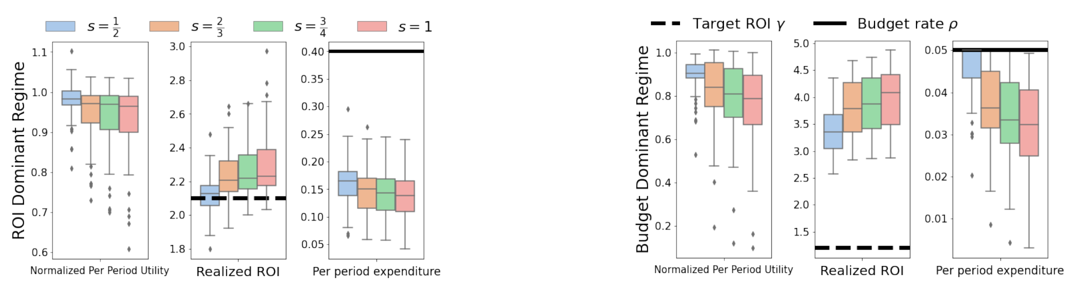

Figure 1 **CTBR robustness. The figure shows performance metrics for CTBR_{EE} run with $\ell_t = \frac{t^{-s}}{\max\{\bar{d},\bar{w}\}\sqrt{K}}$ for $s = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1$ in ROI dominant (Left) and budget dominant regimes (Right). For α dominant regimes, all metrics are nearly the same across s and due to space limits we omit the results. Per period utility is normalized by Optimal utility $U(\boldsymbol{p};\alpha,\gamma,\rho)$. Each box plot is w.r.t. $N = 100$ probability instances. Results suggests even for simple choices of ℓ_t, CTBR can achieve high utility while maintaining ROI constraints in most scenarios.**

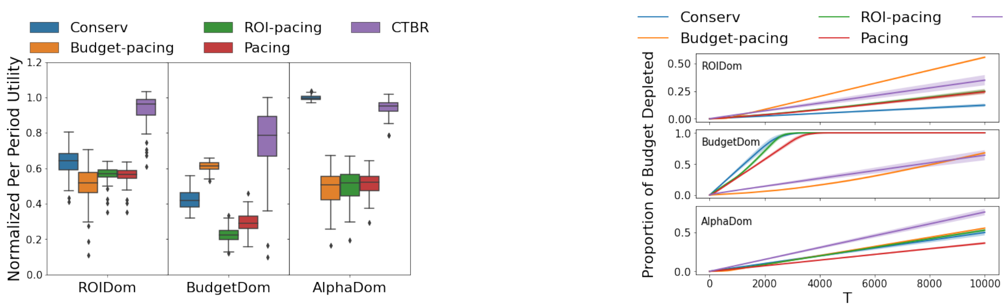

Figure 2 Performance comparison with benchmarks. (Left): Normalized per-period utility. Each box plot is constructed with $N = 100$ probability instances. (Right): Trajectory for proportion of budget depleted. The shaded area represents the standard deviation of budget depletion over $N = 100$ probability instances. These results suggest CTBR outforms other benchmarks because, universally across all regimes, CTBR maintains a better balance between budget depletion and learning optimal bidding decisions.

6. Pricing Against an ROI and Budget Constrained Buyer

Here, we take the alternative perspective of the seller and consider the pricing problem against a single ROI and budget constrained buyer. In this scenario, the second price auction effectively becomes a posted price auction, where the seller's price takes the role of highest competing bid value d_t. More specifically, the buyer still decides on some bid value b_t, but instead of having the buyer actually submit the bid, the buyer decides to take the item at period t and pay price d_t if $b_t \geq d_t$, or leave the item and pay nothing otherwise. That being said, any bidding strategy (e.g. $\text{CTBR}_{\mathcal{A}}$) for the second price auction can be effectively transformed into a "take or leave" strategy for the posted price auction. We note that the seller does not get to observe the buyer's "bid" but instead only knows whether the buyer took the item or not.

Seller's pricing problem. Consider the seller committing to a finite price set $\mathcal{D} = \{D^m\}_{m\in[M]}$ where $D^1 > \cdots > D^M$. The seller sets a price $d_t \in \mathcal{D}$ during each period against a *value-maximizing* buyer (whose capital cost $\alpha = 0$). [10] Then, the buyer observes both d_t and its valuation v_t and

[10] In the posted price setting, assuming the buyer's capital cost $\alpha = 0$ is without loss of generality. This is because before the buyer makes a take/leave decision, she gets to observe both her valuation v_t and price d_t. Hence, in the buyer's hindsight problem $\text{OPT}(\{v_t, d_t\}_{t\in[T]}; \alpha, \gamma, \rho)$ defined in Equation (2), the buyer can rescale her valuations by setting $\widetilde{v}_t = v_t - \alpha d_t$ as well as her target ROI $\widetilde{\gamma} = \gamma - \alpha$. This will result in the equivalent hindsight problem $\widetilde{\text{OPT}}(\{\widetilde{v}_t, d_t\}_{t\in[T]}; 0, \widetilde{\gamma}, \rho) := \max_{\boldsymbol{z}\in[0,1]^T} \sum_{t\in[T]} \widetilde{v}_t z_t \quad \text{s.t.} \sum_{t\in[T]} (\widetilde{v}_t - \widetilde{\gamma} d_t)\, z_t \geq 0, \ \ \sum_{t\in[T]} d_t z_t \leq \rho T$, which implies that we can simply assume $\alpha = 0$.

makes a take/leave decision $z_t \in \{0,1\}$ based on some learning algorithm that satisfies both ROI and budget constraint in Equation (1). Here, we assume that the buyer's valuations are supported on the set $\mathcal{V} = \{V^n\}_{n\in[N]}$ where $V^1 > \cdots > V^N$ and associated with some probability distribution $\boldsymbol{g} \in \Delta_N$ such that $\mathbb{P}(v_t = V^n) = g^n$ for any period $t \in [T]$. On a separate note, $\{D^m\}_{m\in[M]}$ and $\{V^n\}_{n\in[N]}$ can be thought of as the unique values of highest competing bids and valuations in the set $\mathcal{W} = \{(v^k, d^k)\}_{k\in[K]}$ of the previous sections, i.e. $\mathcal{W} = \mathcal{V} \times \mathcal{D}$, where $K = MN$. Furthermore, imposing any distribution $\widetilde{\boldsymbol{g}} \in \Delta_M$ on support $\mathcal{D}$, combined with distribution $\boldsymbol{g}$ on $\mathcal{V}$, induces a product distribution $\boldsymbol{p} = \boldsymbol{g} \times \widetilde{\boldsymbol{g}}$ over $\mathcal{W}$.

The buyer's target ROI γ and budget rate ρ are private to the buyer and is unknown to the seller. Both the seller and the buyer do not know the valuation distribution $\boldsymbol{g}$, and hence learn how to price and make take/leave decisions respectively as time proceeds. In other words, the posted price auction involves *a two-sided learning paradigm* where both the seller and buyer adopt different online decision-making strategies that interact with one another. Our goal is to better understand whether we can design an effective pricing algorithm facing an environment that is driven by the buyer's unknown algorithm.

Seller's regret. We evaluate the performance of any sequence of pricing decision $\{d_t\}_{t\in[T]} \in \mathcal{D}^T$ by benchmarking its realized revenue, namely $\sum_{t\in[T]} d_t z_t$, to the maximum revenue that could have been obtained if the seller had set a fixed price over all T periods assuming the buyer makes optimal decisions. Mathematically, we first rewrite the buyer's problem in Equation (8) as followed for any fixed seller price $d \in \mathcal{D}$ as

$$U(d) = \max_{\boldsymbol{x}\in[0,1]^N} \sum_{n\in[N]} g^n V^n x^n \quad \text{s.t.} \sum_{n\in[N]} g^n \left(V^n - \gamma d\right) x^n \geq 0, \text{ and } d \sum_{n\in[N]} g^n x^n \leq \rho. \quad (14)$$

In light of Theorem 1, we let $\boldsymbol{x}_d \in [0,1]^N$ be the unique optimal threshold solution to $U(d)$ for any $d \in \mathcal{D}$, and note that the optimal solution induces a per-period expenditure of

$$\pi(d) := d \sum_{n\in[N]} g^n x_d^n\,, \quad (15)$$

which is also the expected per-period revenue of the seller when the buyer knows $\boldsymbol{g}$ and acts optimally. We will refer to $\pi(\cdot)$ as the seller's *revenue function*. Here, we remark that $\sum_{n\in[N]} g^n x_d^n$ is the probability that the buyer decides to take price d. Then, for any sequence of pricing decisions $\{d_t\}_{t\in[T]} \in \mathcal{D}^T$, we define the seller's regret over T periods to be

$$\mathrm{Reg}_{\mathrm{sell}} = T \max_{d\in\mathcal{D}} \pi(d) - \sum_{t\in[T]} \mathbb{E}\left[d_t z_t\right] , \tag{16}$$

where the expectation is taken w.r.t. $\{(v_t, d_t)\}$ and randomness in the buyer's strategy.

The benchmark $U(d)$ is in fact equivalent to our definition of $U(\boldsymbol{p}; 0, \gamma, \rho)$ in Section 4. Assume the seller sets price $d_t = D^m \in \mathcal{D}$ for some $m \in [M]$ and for all $t \in [T]$. In this case the buyer's hindsight benchmark is given by $\mathrm{OPT}(\{v_t, D^m\}_{t\in[T]}; 0, \gamma, \rho)$. Under this fixed price scenario, d_t can be viewed as being drawn from the distribution $\boldsymbol{e}^m \in \Delta_M$ (m'th unit vector in $\mathbb{R}^M$), which then induces the product distribution $\boldsymbol{p} = \boldsymbol{g} \times \boldsymbol{e}^m \in \Delta_K$ over $\mathcal{W}$. Then, it is not difficult to see for any $m \in [M]$, $U(D^m)$ is equivalent to $U(\boldsymbol{g} \times \boldsymbol{e}^m; 0, \gamma, \rho)$.[11] This equivalency between $U(d)$ and $U(\boldsymbol{p}; 0, \gamma, \rho)$ suggests that the benchmark $U(d)$ indeed implies optimal buyer actions, as for any $m \in [M]$ such that $D^m \in \mathcal{D}$, we have $U(D^m) = U(\boldsymbol{g} \times \boldsymbol{e}^m; 0, \gamma, \rho) \geq \mathbb{E}\left[\mathrm{OPT}(\{v_t, D^m\}_{t\in[T]}; 0, \gamma, \rho)\right]$, where the final inequality follows from Theorem 2.

Finally, we remark the seller's regret resembles that of an M-arm multi-arm bandit (MAB) problem (see Lattimore and Szepesvári (2020) for a detailed introduction), where we can view each price $D^m \in \mathcal{D}$ as an arm and $D^m z_t$ to be the reward by pulling arm m. Nevertheless, we point out that our problem is more complex as the seller's reward $D^m z_t$ for setting price D^m during period t is related to the buyer-specific algorithm, which likely depends on the buyer's past decisions as well as past prices set by the seller. Although our pricing problem is more difficult than MAB which

[11] Consider any decision vector $\widetilde{\boldsymbol{x}} \in [0,1]^K$ that is feasible to $U(\boldsymbol{g} \times \boldsymbol{e}^m; 0, \gamma, \rho)$. Using the definition that $\boldsymbol{p} = \boldsymbol{g} \times \boldsymbol{e}^m$, we have $p^k = g^n$ is $(v^k, d^k) = (V^n, D^m)$ and $p^k = 0$ otherwise. Thus, the objective of $U(\boldsymbol{g} \times \boldsymbol{e}^m; 0, \gamma, \rho)$, namely $\sum_{k\in[K]} p^k v^k \widetilde{x}^k = \sum_{n\in[N]} \sum_{k: v^k = V^N} p^k v^k \widetilde{x}^k = \sum_{n\in[N]} g^n V^n x^n$ where x^n is the change of variable for $\widetilde{x}^k$ when $k \in [N]$ satisfies $(v^k, d^k) = (V^n, D^m)$. A similar argument shows that the two constraints are also equivalent, i.e. $\sum_{k\in[K]} p^k d^k \widetilde{x}^k = D^m \sum_{n\in[N]} g^n x^n$ and $\sum_{k\in[K]} p^k \left(v^k - \gamma d^k\right) \widetilde{x}^k = \sum_{n\in[N]} g^n \left(V^n - \gamma D^m\right) x^n$.

typically requires the decision maker to explore all M arms, in the next section we demonstrate that by exploiting the special structures of our problem, we only need to explore $\mathcal{O}(\log(M))$ arms (prices).

6.1. Bell-shaped Structure of the Revenue Function

In this section, we first motivate our pricing algorithm by analyzing some underlying structures of the seller revenue $\pi(d)$ defined in Equation (15). The goal of this section is to develop efficient ways to identify $\arg\max_{d\in\mathcal{D}} \pi(d)$ by avoiding exploring each possible price in $\mathcal{D}$ which will result in a regret that scales linearly in the number of prices M. In the rest of the paper, we make the following assumption to rule out trivial problem instances (e.g. cases when the optimal solution $\boldsymbol{x}_d$ corresponding to some $d \in \mathcal{D}$ has all 0 entries or when one of the constraints are redundant):

Assumption 1 *For any $d \in \mathcal{D}$, assume $V^N - \gamma d < 0$, $V^1 - \gamma d > 0$ and $\sum_{n\in[N]}(V^n - \gamma d)g^n \neq 0$. Furthermore, assume $\bar{d} > \rho$ and $\underline{d} < \rho$.*

To begin with, we categorize all prices $d \in \mathcal{D}$ according to whether constraints are binding under the corresponding optimal solution $\boldsymbol{x}_d$.

Definition 3 *Fix target ROI γ, budget rate ρ, valuation distribution $\boldsymbol{g} \in \Delta_K$ and selling price $d \in \mathcal{D}$. Recall $\boldsymbol{x}_d$ is the optimal threshold-based solution to $U(d)$ in Equation (14). Then we say d is*

- ***Non-binding**, if under $\boldsymbol{x}_d$, both constraints are non binding, i.e., $d\sum_{n\in[N]} g^n x_d^n < \rho$ and $\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n > 0$;*
- ***Budget binding** if under $\boldsymbol{x}_d$, the budget constraints is binding, i.e. $d\sum_{n\in[N]} g^n x_d^n = \rho$ and $\sum_{n\in[N]}(V^n - \gamma d) g^n x_d^n \geq 0$;*
- ***ROI binding** if under $\boldsymbol{x}_d$, the ROI constraint is binding, i.e. $\sum_{n\in[N]}(V^k - \gamma d) g^n x_d^n = 0$ and $d\sum_{n\in[N]} g^n x_d^n \leq \rho$.*

It is apparent that any price $d \in \mathcal{D}$ must belong to at least one of these categories. Also, if a price is non-binding, it cannot be budget binding or ROI binding. However, it may be possible that a price d is both budget binding and ROI binding. This can only occur for certain model primitives $\rho, \gamma, \boldsymbol{g}$. We also point out that for any budget binding price $d \in \mathcal{D}$, the seller would like to extract the entire budget from the buyer since the per-period revenue under optimal buyer action is $\pi(d) = d \sum_{n \in [N]} g^n x_d^n = \rho$.

Our main result of this subsection is the following Theorem 5, which states that as we traverse $\mathcal{D}$ in increasing price order, prices are first non-binding and the revenue $\pi(d)$ increases in d; then prices become budget binding, where revenue remains constant at $\pi(d) = \rho$; finally prices become ROI binding, where $\pi(d)$ decreases in d. The proof can be found in Appendix F.1.

Theorem 5 (Bell-shaped Structure of the Revenue Function) *Suppose that Assumption* 1 *holds. Then, the following hold*

1. *For any non-binding prices* $d, \widetilde{d}$, *if* $d < \widetilde{d}$ *then* $\pi(d) < \pi(\widetilde{d})$.
2. *If* d *is budget binding, any price* $\widetilde{d} > d$ *cannot be non-binding, which means* $\widetilde{d}$ *is budget binding or ROI binding.*
3. *If* d *is ROI binding, then any* $\widetilde{d} > d$ *must also be ROI binding. Furthermore,* $\pi(d) > \pi(\widetilde{d})$.

We provide an illustration of Theorem 5 in Figure 3 that depicts the "non-binding → budget binding → ROI binding" transition phenomenon, as well as a corresponding revenue "increase → plateau → decrease", as we traverse prices in increasing order. We note that for specific model primitives $\boldsymbol{g}, \gamma, \rho$, there may exists no budget binding prices (as shown in right subfigure in Figure 3), meaning that there are scenarios in which it is impossible for the buyer to extract the entire buyer budget. Nevertheless, this transition phenomena suggests that we can efficiently identify the maximizing revenue $\arg\max_{d \in \mathcal{D}} \pi(d)$ by utilizing a simple binary search approach. Hence, we utilize this structure of $\pi(d)$ to motivate our pricing algorithm.

6.2. Pricing Algorithm against an ROI and Budget Constrained Buyer

The main challenge the seller faces is her lack of knowledge on the buyer's model primitives, namely the buyer's valuation distribution $\boldsymbol{g}$, target ROI γ and budget rate ρ. Furthermore, the seller has limited information feedback as she only observes whether the buyer took the price or not, i.e., the seller only observes the outcome $z_t \in \{0, 1\}$. This lack of information makes it very difficult for the seller to estimate the buyer's model primitives. Nevertheless, we propose a simple pricing algorithm that bypasses this lack of knowledge via exploiting the price transition phenomenon as characterized in Theorem 5 and Figure 3. We demonstrate that this algorithm achieves good performance when facing a general class of algorithms that is adaptive to nonstationary environments.

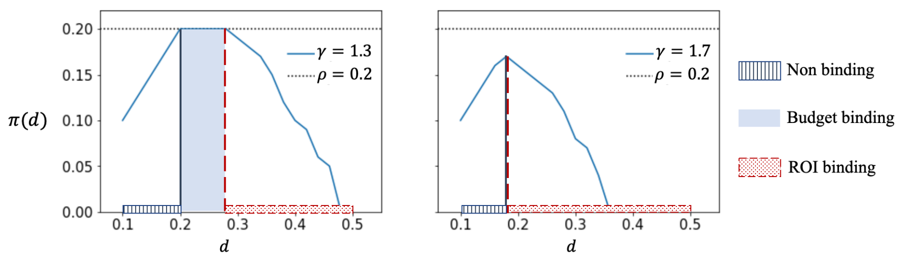

Figure 3 **Seller revenue function bell-shape structure. Model primitives: number of unique buyer valuations $N = 6$, valuation set $\mathcal{V} = (0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$, valuation distribution $g = (0.1, 0.1, 0.2, 0.1, 0.2, 0.3)$, number of unique selling prices $M = 21$, seller price set $\mathcal{D} = (0.5, 0.48 \ldots 0.1)$, buyer budget rate $\rho = 0.2$, capital cost $\alpha = 0$. The left and right subfigures correspond to target ROI $\gamma = 1.3$ and 1.7 respectively. In both cases, prices transition from non-binding to budget binding, and finally to ROI binnding. Revenue $\pi(d)$ increases as in d when prices are non-binding, decreases in d when prices are ROI binding, and remains at ρ when prices are budget binding. Note that when $\gamma = 1.7$, there are no budget binding prices.**

Our proposed pricing algorithm consists of an exploration phase and an exploitation phase. During the exploration phase, the algorithm searches for a revenue maximizing price $D^* \in \arg\max_{d \in \mathcal{D}} \pi(d)$ through an episodic structure: the seller initiates the first episode $\mathcal{E}_1$, and fixes the price chosen in this episode D_1 for E consecutive periods. At the end of the episode (i.e. after E periods since the beginning of the episode), the seller records the average per-period revenue $\hat{\pi}(D_1) = \frac{D_1}{E} \sum_{t \in \mathcal{E}_1} z_t$, where $z_t \in \{0, 1\}$ indicates whether the buyer takes the price at time $t \in \mathcal{E}_1$. The process then repeats as the seller moves on to episodes $\mathcal{E}_2, \ldots$ This exploration phase eventually terminates when the seller has explored enough prices. The seller's pricing decision in each episode is governed by a binary search procedure over the set $\mathcal{D}$, such that every price is chosen at most once across all episodes, and the exploration phase will have $\mathcal{O}(\log(M))$ episodes. Our pricing algorithm is shown in Algorithm 2.

We note that our proposed algorithm does not try to learn the buyer's model primitives. We further point out that such a binary-search approach is a natural choice to identify revenue-optimal prices in the simplest monopolistic pricing setting under a typical unimodal assumption, [12] and one may wonder whether this approach can have good performances against a much more complex setting where the buyer is ROI and budget constrained and aims to learn her optimal bidding strategy. Surprisingly, in the next section we are in fact able to show this simple approach achieves good performances against buyers who are adaptive to price changes.

Algorithm 2 Binary Search Exploration Exploitation

Input: Episode length E.

Initialize iteration index $\text{iter} = 1$.

[Exploration]:

Set D^1 for E consecutive periods, and record per-period revenue $\hat{\pi}\left(D^1\right)$. Then set D^M for E consecutive periods, and record per-period revenue $\hat{\pi}\left(D^M\right)$.

Set $m^* \leftarrow \arg\max_{m \in \{1,M\}} \hat{\pi}\left(D^m\right)$ $\text{L} = 1$, $\text{R} = M$, $\text{med} = \lfloor \frac{\text{L}+\text{R}}{2} \rfloor$.

while $\text{L} < \text{R}$ **do**

    $\text{iter} \leftarrow \text{iter} + 1$.

    if per-period revenue $\hat{\pi}\left(D^k\right)$ is not recorded for $k = \text{med}, \text{med}+1$ **then**

        Set price D^k for E consecutive periods and record per-period revenue $\hat{\pi}\left(D^k\right)$ for $k = \text{med}, \text{med}+1$

    end if

    if $\hat{\pi}\left(D^{\text{med}}\right) < \hat{\pi}\left(D^{\text{med}+1}\right)$ **then**

        Set $m^* \leftarrow \arg\max_{m \in \{m^*, \text{med}+1\}} \hat{\pi}\left(D^m\right)$, $\text{L} \leftarrow \text{med} + 1$, $\text{med} \leftarrow \lfloor \frac{\text{L}+\text{R}}{2} \rfloor$

    else

        Set $m^* \leftarrow \arg\max_{m \in \{m^*, \text{med}\}} \hat{\pi}\left(D^m\right)$, $\text{R} \leftarrow \text{med} - 1$, $\text{med} \leftarrow \lfloor \frac{\text{L}+\text{R}}{2} \rfloor$

    end if

end while

[Exploitation]: Set price D^{m^*} for the remaining periods.

[12] In monopolistic pricing, the revenue-optimal price p^* is charachterized by $p^* = \arg\max_p pF(p)$, where F is the cdf of buyer valuations. A typical assumption is such that the function $pF(p)$ is unimodal.

6.3. Regret Analysis of Our Pricing Algorithm

In this section, we provide theoretical guarantees for our proposed pricing algorithm against a class of buyer's bidding strategies. Recall that the seller is pricing against a buyer who is subject to both budget and ROI constraints, and adopts some algorithm that satisfies both ROI and budget constraint in Equation (1) to maximize total cumulative value. The class of strategies that we consider imposes some notion of adaptiveness to non-stationary environments, as discussed in the following definition.

Definition 4 (ξ-Adaptive Bidding Strategies) *Assume the T-period horizon is divided into H consecutive episodes $\mathcal{E}_1 \dots \mathcal{E}_H$, i.e.$\sum_{h\in[H]} |\mathcal{E}_h| = T$. In each episode $h \in [H]$, the seller sets a fixed price $D_h \in \mathcal{D}$, and the buyer decides on a binary sequence of take-or-leave actions $\{z_t\}_{t\in[T]} \in \{0,1\}^T$. Then we say a buyer's strategy is ξ-adaptive for $\xi \in (0,1)$ if*

1. ***Adaptivity:*** *there exists a universal error function $\phi : \mathbb{N}\times\mathbb{N} \to [0,1]$ decreasing in the first argument such that for all episodes $h = 1,2,\dots H$, $\left|\frac{1}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t - \frac{\pi(D_h)}{D_h}\right| \le \phi(|\mathcal{E}_h|,T)$ w.p. at least $1-\frac{1}{T}$;*[13]
2. ***Stability:*** *there exists a minimum episode length $E_0 = \Omega(T^{1-\xi})$ such that $\phi(|\mathcal{E}_h|,T) < \frac{G}{2\bar{d}}$ for any episode h whose length $|\mathcal{E}_h| \ge E_0$. Here $G := \min_{d,\widetilde{d}\in\mathcal{D}:\pi(d)\neq\pi(\widetilde{d})} \left|\pi(d) - \pi(\widetilde{d})\right|$ is the minimum gap between any two non equal revenues corresponding to prices in $\mathcal{D}$.*

Note that the error function $\phi(|\mathcal{E}_h|,T)$ does not depend on the actual price D_h set in episode $h \in [H]$. The first adaptivity condition characterizes the buyer algorithm's ability to adapt and optimally react to prices across different episodes. The term $\left|\frac{D_h}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t - \pi(D_h)\right|$ is the seller's average revenue loss, relative to the revenue from optimal buyers, over a certain period under fixed price D_h. However, the term can alternatively be viewed as the buyer's deviation from optimal behavior induced by the optimal threshold solution $\boldsymbol{x}_{D_h}$ because $\frac{\pi(D_h)}{D_h} = \sum_{n\in[N]} g^n x^n_{D_h}$ is the optimal probability of which the buyer should take price D_h. The first adaptivity condition hence states that the buyer's deviation

[13] Note that z_t for any $t \in \mathcal{E}_h$ would also depend on the price D_h.

from optimality under a fixed price is bounded. The second stability condition states that for long enough episodes, the deviation of buyer behavior from optimality stabilizes and reaches a certain low point, namely $\frac{G}{2\bar{d}}$. Additionally, the stability condition states that a ξ-adaptive algorithm will require an order of $\Omega(T^{1-\xi})$ periods to stabilize. Thus, the larger ξ, the more stable the algorithm is. Finally, we remark that if error function ϕ for some buyer algorithm is independent of T, then the stability condition would be easily satisfied.

The main result of this subsection is presented in Theorem 6, which characterizes the performance of our pricing algorithm against any ξ-adaptive buyer algorithm. The proof of Theorem 6 can be found in Appendix F.2.

Theorem 6 (Pricing against ξ-adaptive buyer strategies) *Assume the buyer runs some ξ-adaptive algorithm with error function ϕ. Fix $\epsilon \in (0,\xi)$ and let $T_\epsilon > 0$ satisfy $\phi(T^{1-\xi+\epsilon}, T) < \frac{G}{2\bar{d}}$ for all $T > T_\epsilon$.* [14] *Then, if the seller adopts the pricing strategy in Algorithm 2 with episode length $E = T^{1-\xi+\epsilon}$ over a time horizon $T > \max\left\{T_\epsilon, \left(4\bar{d}\lfloor\log_2(M)\rfloor + 4\bar{d}\right)^{\frac{1}{\xi-\epsilon}}\right\}$, under Assumption 1 the seller's regret is bounded as*

$$Reg_{sell} \leq 2\bar{d}\left(\lfloor\log_2(M)\rfloor + 1\right) \cdot T^{1-\xi+\epsilon} + 2\bar{d}T \cdot \phi\left(\tfrac{T}{2}, T\right) + \bar{d}\left(\lfloor\log_2(M)\rfloor + 1\right)^2/2\,. \tag{17}$$

Theorem 6 delineates how a ξ-adaptive algorithm's adaptivity and stability properties factor into seller regret (see discussion after Definition 4 for details on adaptivity and stability). The first term $T^{1-\xi+\epsilon}$ in the seller's regret corresponds to the buyer algorithm's stability property, which characterize its length of periods needed to stabilize in each episode; the second term $\phi\left(\frac{T}{2}, T\right)$ corresponds to the adaptivity property, which represents the buyer's deviation from the optimal threshold-based strategy characterized by the optimal solution $\boldsymbol{x}_d$ to $U(d)$ for any price $d \in \mathcal{D}$. Finally, we remark that the seller does not need to know the exact value of ξ, as some lower bound ξ would be sufficient.

[14] T_ϵ exists because the minimum episode length from the stability condition can be taken as $E_0 = T^{1-\xi+\epsilon}$ for large enough T so that $\phi(T^{1-\xi+\epsilon}, T) < \frac{G}{2\bar{d}}$. Here, we also used the fact that ϕ is decreasing in its first argument.

We note that ξ-adaptivity is not at all restrictive. We will show that our $\mathrm{CTBR}_{\mathcal{A}}$ coupled with the simple SGD learning algorithm and constant step sizes is $\frac{1}{3}$-adaptive.

Theorem 7 (CTBR with constant step size SGD is $\frac{1}{3}$-adaptive) *There exists some $T_0 \in \mathbb{N}$ such that for all $T > T_0$,* CTBR *equipped with SGD with constant step size $\eta = T^{-\frac{2}{3}}$ is $\frac{1}{3}$-adaptive. The corresponding minimum episode length can be taken as $E_0 = T^{\frac{2}{3}+\epsilon}$ for some $\epsilon = \Theta(1/\log(T))$ and $\epsilon < 1/3$.*

The proof for Theorem 7 is provided in Appendix F.3. We remark that although SGD with constant step size is adaptive, SGD with vanishing step size is not adaptive. To see this intuitively, consider the entire horizon T being split equally into two halves. Suppose that the difference between the prices in the two halves is large. From Theorem 9, we know that SGD with vanishing step size adjusts to the fixed price quickly at a $\frac{1}{\sqrt{t}}$ rate in the first half. But when the algorithm enters the second half, all step sizes are less than $\frac{2}{T}$, which does not provide enough flexibility for the algorithm to adapt to the new price.

Corollary 1 (Seller's regret against $\mathrm{CTBR}_{\mathcal{A}}$ with constant step size SGD) *Let $T > T_0$ where T_0 is defined in Theorem 7. Assume the buyer runs* $\mathrm{CTBR}_{\mathcal{A}}$ *with SGD and constant step size $\eta = T^{-\frac{2}{3}}$. Then for a fixed $\epsilon \in (0, \frac{1}{3})$ and $\epsilon = \Theta(1/\log(T))$, if the seller sets prices with episode length $E = T^{\frac{2}{3}+\epsilon}$ using Algorithm 2, then for all $T > \max\left\{T_0, \left(4\bar{d}\lfloor \log_2(M)\rfloor + 4\bar{d}\right)^{\frac{1}{\frac{1}{3}-\epsilon}}\right\}$, the seller's regret is bounded as*

$$Reg_{sell} \leq 2\bar{d}\widetilde{H}\cdot T^{\frac{2}{3}+\epsilon} + 4\bar{d}\sqrt{T\log\left(2T^2\right)} + \frac{\mathcal{T}}{2} + C\left(\widetilde{A}+\widetilde{B}\right)T^{\frac{2}{3}} + \bar{d}\widetilde{H}^2/2 = \Theta(T^{\frac{2}{3}+\epsilon}),$$

where $\widetilde{H} = \lfloor \log_2(M)\rfloor + 1$, C and S are defined as in Theorem 9, $\widetilde{A} = \sqrt{2 + 16\sqrt{\log\left(T^2\log\left(T\right)\right)}}$, $\widetilde{B} = 2\sqrt{\left(1 + 72\log\left(T^2\log\left(T\right)\right)\right)}$, and $\mathcal{T} = \min\left\{t\in[T] : \widetilde{A}\left(1-2T^{-\frac{2}{3}}\right)^t + \widetilde{B}T^{-\frac{1}{3}} < S\right\} = \Theta(T^{\frac{1}{3}})$.

The proof for Corollary 1 can be found in Appendix F.4.

References

Agrawal S, Devanur NR, Li L (2016) An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. *Conference on Learning Theory*, 4–18 (PMLR).

Agrawal S, Wang Z, Ye Y (2014) A dynamic near-optimal algorithm for online linear programming. *Operations Research* 62(4):876–890.

Alon N, Awerbuch B, Azar Y (2003) The online set cover problem. *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, 100–105.

Amin K, Rostamizadeh A, Syed U (2013) Learning prices for repeated auctions with strategic buyers. *arXiv preprint arXiv:1311.6838* .

Arlotto A, Gurvich I (2019) Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* 9(3):231–260.

Auerbach J, Galenson J, Sundararajan M (2008) An empirical analysis of return on investment maximization in sponsored search auctions. *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising*, 1–9.

Azar Y, Bhaskar U, Fleischer L, Panigrahi D (2013) Online mixed packing and covering. *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, 85–100 (SIAM).

Azar Y, Cohen IR, Panigrahi D (2014) Online covering with convex objectives and applications. *arXiv preprint arXiv:1412.3507* .

Babaioff M, Immorlica N, Kempe D, Kleinberg R (2007) A knapsack secretary problem with applications. *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, 16–28 (Springer).

Babaioff M, Immorlica N, Kempe D, Kleinberg R (2008) Online auctions and generalized secretary problems. *ACM SIGecom Exchanges* 7(2):1–11.

Balseiro S, Golrezaei N, Mahdian M, Mirrokni V, Schneider J (2019a) Contextual bandits with cross-learning. *Advances in Neural Information Processing Systems* 32:9679–9688.

Balseiro S, Golrezaei N, Mirrokni V, Yazdanbod S (2019b) A black-box reduction in mechanism design with private cost of capital. *Available at SSRN 3341782* .

Balseiro S, Kim A, Mahdian M, Mirrokni V (2021) Budget management strategies in repeated auctions. *Operations Research* Forthcoming.

Balseiro S, Kim A, Russo DJ (2019c) On the futility of dynamics in robust mechanism design. *Columbia Business School Research Paper Forthcoming* .

Balseiro S, Lu H, Mirrokni V (2020a) The best of many worlds: Dual mirror descent for online allocation problems. *arXiv preprint arXiv:2011.10124* .

Balseiro S, Lu H, Mirrokni V (2020b) Dual mirror descent for online allocation problems. *International Conference on Machine Learning*, 613–628 (PMLR).

Balseiro SR, Gur Y (2019) Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* 65(9):3952–3968.

Bergemann D, Said M (2010) Dynamic auctions. *Wiley Encyclopedia of Operations Research and Management Science* .

Boyd S, Boyd SP, Vandenberghe L (2004) *Convex optimization* (Cambridge university press).

Braverman M, Mao J, Schneider J, Weinberg M (2018) Selling to a no-regret buyer. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 523–538.

Buchbinder N, Naor J (2009) Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research* 34(2):270–286.

Cohen A (1981) Stepsize analysis for descent methods. *Journal of Optimization Theory and Applications* 33(2):187–205.

Dantzig GB (1957) Discrete-variable extremum problems. *Operations research* 5(2):266–288.

Deng Y, Schneider J, Sivan B (2019) Prior-free dynamic auctions with low regret buyers .

Devanur NR, Hayes TP (2009) The adwords problem: online keyword matching with budgeted bidders under random permutations. *Proceedings of the 10th ACM conference on Electronic commerce*, 71–78.

Feldman J, Henzinger M, Korula N, Mirrokni VS, Stein C (2010) Online stochastic packing applied to display ad allocation. *European Symposium on Algorithms*, 182–194 (Springer).

Feldman J, Muthukrishnan S, Pal M, Stein C (2007) Budget optimization in search-based advertising auctions. *Proceedings of the 8th ACM conference on Electronic commerce*, 40–49.

Golrezaei N, Jaillet P, Liang JCN (2019) Incentive-aware contextual pricing with non-parametric market noise. *arXiv preprint arXiv:1911.03508* .

Golrezaei N, Javanmard A, Mirrokni V (2020) Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Operations Research* .

Golrezaei N, Lobel I, Paes Leme R (2018) Auction design for roi-constrained buyers. *Available at SSRN 3124929* .

Gupta A, Koren T, Talwar K (2019) Better algorithms for stochastic bandits with adversarial corruptions. *Conference on Learning Theory*, 1562–1578 (PMLR).

Han Y, Zhou Z, Flores A, Ordentlich E, Weissman T (2020a) Learning to bid optimally and efficiently in adversarial first-price auctions. *arXiv preprint arXiv:2007.04568* .

Han Y, Zhou Z, Weissman T (2020b) Optimal no-regret learning in repeated first-price auctions. *arXiv preprint arXiv:2003.09795* .

Kesselheim T, Tönnis A, Radke K, Vöcking B (2014) Primal beats dual on online packing lps in the random-order model. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 303–312.

Kireyev P, Pauwels K, Gupta S (2016) Do display ads influence search? attribution and dynamics in online advertising. *International Journal of Research in Marketing* 33(3):475–490.

Kleinberg R (2005) A multiple-choice secretary algorithm with applications to online auctions. *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, 630–631 (Citeseer).

Kleinberg R, Leighton T (2003) The value of knowing a demand curve: Bounds on regret for online posted-price auctions. *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, 594–605 (IEEE).

Laffont JJ, Robert J (1996) Optimal auction with financially constrained buyers. *Economics Letters* 52(2):181–186.

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Lykouris T, Mirrokni V, Paes Leme R (2018) Stochastic bandits robust to adversarial corruptions. *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 114–122.

Pai MM, Vohra R (2014) Optimal auctions with financially constrained buyers. *Journal of Economic Theory* 150:383–425.

Qian J, Fruit R, Pirotta M, Lazaric A (2020) Concentration inequalities for multinoulli random variables. *arXiv preprint arXiv:2001.11595* .

Rakhlin A, Shamir O, Sridharan K (2011) Making gradient descent optimal for strongly convex stochastic optimization. *arXiv preprint arXiv:1109.5647* .

Seiden SS (2002) On the online bin packing problem. *Journal of the ACM (JACM)* 49(5):640–671.

Vaze R (2018) Online knapsack problem under expected capacity constraint. *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, 2159–2167 (IEEE).

Weed J, Perchet V, Rigollet P (2016) Online learning in repeated auctions. *Conference on Learning Theory*, 1562–1583 (PMLR).

Weissman T, Ordentlich E, Seroussi G, Verdu S, Weinberger MJ (2003) Inequalities for the l1 deviation of the empirical distribution. *Hewlett-Packard Labs, Tech. Rep* .

Zhou Y, Chakrabarty D, Lukose R (2008) Budget constrained bidding in keyword auctions and online knapsack problems. *International Workshop on Internet and Network Economics*, 566–576 (Springer).

Appendix. Appendices for Bidding and Pricing in Budget and ROI Constrained Markets

Appendices are organized as followed. In Appendix A, we present an extended literature review and discuss some broader related works. In Appendices B and C, we include additional material/results/discussions to Sections 5 and 5.1, respectively. All proofs of our theoretical results are included in Appendices D, E and F. Specifically, in Appendix D we provide proofs for our theoretical results in Section 4, while in appendices E and F we present the proofs for our results in Sections 5 and 6, respectively.

A. Extended Literature Review

As the most closely related works have been discussed in the introduction section, here we only further discuss broader related works.

Other related work in online resource allocation There has been extensive research on online resource allocation with budget/capacity constraints (see e.g. Kleinberg (2005), Devanur and Hayes (2009), Agrawal et al. (2016)) and here we briefly discuss those that are the most relevant.[15] Zhou et al. (2008) studies the budget-constrained bidding problem for sponsored search in an adversarial setting and present an algorithm with competitive ratio that depends on upper and lower bounds on the value-to-cost ratios; Babaioff et al. (2007), Arlotto and Gurvich (2019) study variants of the knapsack and secretary problems under the random order arrival model and stochastic arrival model, respectively, both presenting near optimal algorithms in their respective settings. Our work differs from this line of research as we incorporate an ROI constraint while also considering the problem of how to price against budget and ROI constrained buyers. Finally, Agrawal et al. (2014) utilizes a primal-dual framework to study online linear programming (LP) with packing constraints, where the positive-valued constraint matrix is revealed column by column (each column corresponds to a highest competing bid d_t) along with the corresponding objective coefficient (corresponding to utility $v_t - \alpha d_t$).

[15] The buyer's online bidding problem can be viewed as an online resource allocation problem. However, a key difference is that in bidding, the buyer does not observe the highest competing bid d_t (equivalently the amount of resource depleted) before making a decision; as in the resource allocation problem, both the reward and resource depletion are revealed before decision making. Therefore, to apply a resource allocation algorithm in the bidding problem, one must additionally impose some bidding mechanic that indirectly achieves the desired allocation through constructing appropriate bid values.

Their algorithm determines the decision variable corresponding to the arriving column based on the dual variables of past revealed columns.

Mechanism design and market equilibrium for budget and ROI constrained buyers One relevant line of research addresses the mechanism design problem for budget or ROI constrained buyers. As one of the pioneering works regarding mechanism for financially constrained buyers, Laffont and Robert (1996) derives the optimal mechanism for symmetric buyers and public budget information. On the contrary, a more recent paper Pai and Vohra (2014) studies the general multidimensional mechanism design setting against buyers with private budgets. Regarding ROI constrained buyers, Golrezaei et al. (2018) shows that the optimal mechanism for symmetric ROI-constrained buyers is either second-price auctions with reduced reserve prices or subsidized second-price auctions. The work also derives an optimal mechanism for asymmetric ROI buyers. There is also a wide range of work that study dynamic mechanism design for budget constrained buyers, and we refer the reader to the survey Bergemann and Said (2010) and references therein.

Online bidding in repeated auctions under feedback constraints Other than budget capacities and ROI targets, buyers are also typically constrained in terms of the amount information available as they participate in auctions. For example, Balseiro et al. (2019a) studies bidding problem in first price auctions under different feedback structures where an unconstrained quasi-linear buyer only observes whether or not she wins the auction, and Han et al. (2020b,a) study a similar problem where the buyer also gets to observe the highest competing bid if she did not win the auction. As another related work, Weed et al. (2016) studies the bidding problem where the buyer does not know her valuation before submitting her bid, and only observes her valuation if she wins the auction. The work considers the stochastic and adversarial highest competing bid settings, and presents algorithms that build on the UCB and EXP3 algorithms, respectively.

Selling to truthful and strategic buyers Kleinberg and Leighton (2003) studies the scenario where the seller sell items through a repeated posted price mechanism to a single truthful buyer who simply accepts the price if her valuation is greater than the offered price. the work presents optimal algorithms in the settings where the buyer's valuations are fixed, stochastic and adversarial, respectively. Amin et al. (2013) also concerns selling through a posted price mechanism, but to a strategic buyer who may choose not to accept a price bellow her valuation (or accept a price above her valuation). The work presents learning algorithms in both the fixed valuation and stochastic valuation settings under the assumption that discount their utilities over time. Other related works include Golrezaei et al. (2020) which studies the dynamic pricing problem for repeated contextual second price auctions facing multiple strategic buyers. The work proposes learning algorithms that are robust to buyers' strategic behavior under various seller information structures and provides corresponding performance guarantees. Golrezaei et al. (2019) relaxes several assumptions for one of the settings in Golrezaei et al. (2020), and presents an algorithm with improved performance guarantees. Finally, Balseiro et al. (2019c) considers the dynamic mechanism design problem against strategic buyers, and further identifies a class of problems in which the optimal mechanism is to simply repeat some static mechanism over time.

Online optimization with covering constraints The buyer's ROI constraint takes the form of a long-term covering constraint. The related problem of optimization under online covering constraints have been studied in Alon et al. (2003), Azar et al. (2013, 2014). However, the setting in these works differ from ours: Instead of making irrevocable online decisions, these works focus on updating a decision vector upon the arrival of a covering constraint each period such that this constraint is satisfied. In other words, they consider the decision problem where covering constraints are satisfied in each period, while our buyers of interest only need to satisfy the covering (ROI) constraint in the long run. Another key difference is that in these works the covering constraints are all positive, which means these constraints can be easily satisfied (per period) by increasing each entry of the decision vector. On the contrary, in our problem the ROI balance per period $(v_t - \gamma d_t)z_t$ may be negative, and hence makes constraint satisfaction more difficult.

B. Additional Material for Section 5 Online Threshold Bidding Algorithm

The following Theorem 8 is a more detailed version of Theorem 3 in Section 5, which provides a general regret upper bound for our CTBR framework w.r.t. any input learning algorithm $\mathcal{A}$.

Theorem 8 (Regret of CTBR) *Let $\{\hat{\boldsymbol{p}}_t\}_{t\in[T]}$ be the estimates of learning algorithm $\mathcal{A}$. Assume there exists estimation error function $\ell_t : (0,1) \to \mathbb{R}^+$ decreasing in t so that $\mathbb{P}(\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \le \ell_t) \ge 1 - \frac{1}{T}$, and $\lim_{t\to\infty} \ell_t < S$, where* [16]

$$S := \frac{1}{2}\min\left\{p^{b+1}, p^{r+1}, \frac{-\sum_{k\in[r+1]} p^k w^k}{\sqrt{K}\bar{w}}, \frac{\sum_{k\in[r]} p^k w^k}{\left(\sqrt{K}+1\right)\bar{w}}, \frac{\sum_{k\in[b+1]} p^k d^k - \rho}{\sqrt{K}\bar{d}}, \frac{\rho - \sum_{k\in[b]} p^k d^k}{\left(\sqrt{K}+1\right)\bar{d}}\right\}.$$

[16] The definition $r = \max\left\{k \in [K] : \sum_{\ell\in[k]} p^\ell \left(v^\ell - \gamma d^\ell\right) \ge 0\right\}$ implies $p^{r+1} > 0$ and $\sum_{k\in[r+1]} p^k w^k < 0$ always hold. In the edge case where $r = K$, we defined $d^{K+1} = \infty$, so $w^{K+1} = -\infty$. Similarly, the definition of b always implies $p^b > 0$ and $\sum_{k\in[b+1]} p^k d^k > \rho$.

Define $\mathcal{T}_\mathcal{A} = \min\{t \in [T] : \ell_t < S\}$ *to be the earliest period* t *under which the* ℓ_t *falls bellow* S*. Then for large enough* T *such that*

$$T > \max\left\{2\mathcal{T}_\mathcal{A}, \frac{4\bar{w}(\mathcal{T}_\mathcal{A}+1)}{\sum_{k\in[J]} p^k w^k}\right\}, \text{ and } \sum_{t=\mathcal{T}_\mathcal{A}+1}^{T} \ell_t > 2\left(\mathcal{T}_\mathcal{A}+1\right), \tag{18}$$

bidding according to $\mathrm{CTBR}_\mathcal{A}$ *with confidence bound* ℓ_t *satisfies both ROI and budget constraints in Equation (1). Furthermore,* $Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) \leq \max_{k\in[K]} |v^k - \alpha d^k| \cdot \left(2\mathcal{T}_\mathcal{A} + 1 + C\sum_{t\in[T]} \ell_t\right)$*, where* $C = \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)$.

The proof for Theorem 8 can be found in Appendix E.1. Here, we provide some intuition for the conditions and results of Theorem 8. The variable $\mathcal{T}_\mathcal{A}$ can be viewed as the number of periods required for the input algorithm $\mathcal{A}$ to produce sufficiently accurate estimates. Hence, the first condition in Equation (18), namely $T > \max\left\{2\mathcal{T}_\mathcal{A}, \frac{4\bar{w}(\mathcal{T}_\mathcal{A}+1)}{\sum_{k\in[J]} p^k w^k}\right\}$ simply states the horizon length should be large enough for the algorithm to stabilize. We also note that $\mathcal{T}_\mathcal{A}$ is possibly a function of T, so as long as it is sublinear in T, the first condition is automatically fulfilled for large enough T. Regarding the second condition in Equation (18), the term $\sum_{t=\mathcal{T}_\mathcal{A}+1}^{T} \ell_t$ is required to be large. This is because ℓ_t represents how conservative the buyer is in terms of her estimations for the remainder probabilities $\hat{q}^{\mathtt{R}}_{t+1}$ and $\hat{q}^{\mathtt{B}}_{t+1}$ (see Equation (12)): the larger ℓ_t, the more conservative the buyer is, and hence the more likely the buyer can satisfy both budget and ROI constraints. On the contrary, the term $\sum_{t=\mathcal{T}_\mathcal{A}+1}^{T} \ell_t$ also appears in the regret, which highlights a very natural trade-off between constraint satisfaction and overall utility loss.

The following Theorem 9 is a more detailed version of Theorem 4 in Section 5, which characterizes regret upper bounds for our CTBR framework when the input learning algorithm $\mathcal{A}$ is EE, SGD with vanishing step size, and SGD with constant step size, respectively.

Theorem 9 ($\mathbf{CTBR}_\mathcal{A}$ **with EE and SGD)** *Recall the EE and SGD algorithms described in Equation (11), and let* $\{\hat{\boldsymbol{p}}_t\}_{t\in[T]}$ *be the corresponding estimates for* $\boldsymbol{p}$*. Then the following hold:*

1. *When* $\mathcal{A}$ *is EE, the corresponding error function in Theorem 8 is* $\ell_t = \sqrt{\frac{2K\log(2T)}{t}}$*. Then, there exists some* $T_0 < \infty$ *such that for all* $T > T_0$*, the regret of* $\mathrm{CTBR}_\mathcal{A}$ *is*

$$Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) \leq \max_{k\in[K]} \left|v^k - \alpha d^k\right| \cdot \left(2\mathcal{T}_\mathcal{A} + 1 + C\sqrt{2KT\log(T)}\right), \tag{19}$$

where $\mathcal{T}_\mathcal{A}$ *and* C *are defined in Theorem 8. Here* $\mathcal{T}_\mathcal{A} = \Theta(\log(T))$*, so* $Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) = \widetilde{\mathcal{O}}(\sqrt{T})$.

2. *When $\mathcal{A}$ is SGD with vanishing step size $\eta_t = \frac{1}{t}$, then $\ell_t = \sqrt{\frac{600\log(T\log(T))+12}{t}}$. Then, there exists some $T_1 < \infty$ such that for all $T > T_1$, the regret of* $\text{CTBR}_{\mathcal{A}}$ *is*

$$Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) \leq \max_{k\in[K]} \left|v^k - \alpha d^k\right| \cdot \left(2\mathcal{T}_{\mathcal{A}} + 1 + 2C\sqrt{T}\sqrt{600\log\left(T\log\left(T\right)\right)+12}\right), \tag{20}$$

Here $\mathcal{T}_{\mathcal{A}} = \Theta(\log(T))$, so $Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) = \widetilde{\mathcal{O}}(\sqrt{T})$.

3. *If SGD is run with constant step size $\eta_t = \eta \in (0,1)$, the corresponding error function in Theorem 8 is $\ell_t = A\left(1-2\eta\right)^{\frac{t-1}{2}} + B\sqrt{\eta}$, where*

$$A = \sqrt{2 + 16\sqrt{\log\left(T\log\left(T\right)\right)}}, \quad B = 2\sqrt{\left(1 + 72\log\left(T\log\left(T\right)\right)\right)}. \tag{21}$$

Then, by taking $\eta = T^{-\frac{2}{3}}$, there exists some $T_2 < \infty$ such that for all $T > T_2$, the regret of $\text{CTBR}_{\mathcal{A}}$ *is*

$$Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) \leq \max_{k\in[K]} \left|v^k - \alpha d^k\right| \cdot \left(2\mathcal{T}_{\mathcal{A}} + 1 + C(A/2 + B)T^{\frac{2}{3}}\right). \tag{22}$$

Here $\mathcal{T}_{\mathcal{A}} = \Theta(T^{\frac{1}{3}})$, so $Reg(\boldsymbol{p}, T, \alpha, \gamma, \rho) = \widetilde{\mathcal{O}}(T^{\frac{2}{3}})$.

Finally, in each of the above scenarios, $\text{CTBR}_{\mathcal{A}}$ *satisfies both budget and ROI constraints in Equation (1).*

Here, the high probability bounds for the event $\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$ when $\mathcal{A}$ is EE directly follows from Proposition 1 of Qian et al. (2020) which is a restatement of the concentration inequalities for multinomial random variables developed in Weissman et al. (2003).[17] Hence the proof for the case where $\mathcal{A}$ is EE directly follows from Theorem 8 and hence will be omitted. The high probability bound for SGD with vanishing step sizes follows exactly from Proposition 1 in Rakhlin et al. (2011), but for completeness we include it in our Appendix E.2. The high probability bound for constant step size is based on a modification of that proof, and also provided in the same appendix.

C. Additional Material for Section 5.1 Empirical Study on Learning How to Bid

CTBP implementation robust to ℓ_t. In the following Figure 4, we illustrate the rate of convergence of CTBR_{EE}, with confidence bound $\ell_t = \frac{t^{-s}}{\max\{\bar{d},\bar{w}\}\sqrt{K}}$ for $s \in \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1\}$.

[17] The high probability bounds in Qian et al. (2020) or Weissman et al. (2003) are w.r.t. the ℓ_1 norm. That is, the resuls therein imply that for $\mathcal{A} = \text{EE}$, $\mathbb{P}\left(\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|_1 \leq \ell_t\right) \geq 1 - \frac{1}{T}$. Nevertheless, since $\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|_1 \geq \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|$, we have $\mathbb{P}\left(\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t\right) \geq \mathbb{P}\left(\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|_1 \leq \ell_t\right) \geq 1 - \frac{1}{T}$.

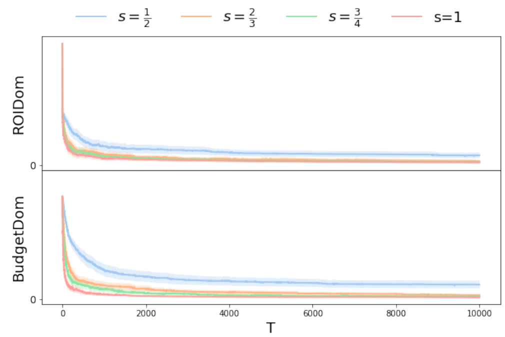

Figure 4 **Convergence of CTBR$_{\mathsf{EE}}$. Recalling $\hat{\boldsymbol{x}}_t$ defined in Algorithm 1 to be the estimate of the optimal threshold-based solution $\boldsymbol{x}^*$ w.r.t. a set of model primitives $(\alpha, \gamma^y, \rho^y, \boldsymbol{P}^{y,i})$, this figure shows the estimation error $\|\hat{\boldsymbol{x}}_t - \boldsymbol{x}^*\|$ over time for CTBR$_{\mathsf{EE}}$ with confidence bound $\ell_t = \frac{t^{-s}}{\max\{\bar{d},\bar{w}\}\sqrt{K}}$ for $s \in \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1\}$. The shaded area delineates the standard deviation of the estimation error over $N = 100$ probability instances. We omit the α-dominant regime as all CTBR$_{\mathsf{EE}}$ converge to $\boldsymbol{x}^*$ very quickly and all lines nearly overlap. This figure suggests that larger s yields faster convergence of CTBR$_{\mathsf{EE}}$ to the optimal threshold-based solution.**

Comparison with benchmark bidding algorithms. Here, we present the pseudocode for our considered benchmark bidding algorithms `Conserv`, `Budget-Pacing`, `ROI-Pacing` and `Pacing`, respectively. First, recall $\hat{\lambda}_t$ and $\hat{\mu}_t$ are estimates of the optimal dual variables w.r.t. the budget constraint and ROI constraint, respectively, in the buyer's hindsight optimization problem $\text{OPT}(\{(v_t, d_t)\}_{t\in[T]}; \alpha, \gamma, \rho)$ defined in Equation (2). Also recall $\rho > 0$ is the budget rate, $\gamma > 0$ is the buyer's target ROI, $\alpha \in (0, \gamma)$ is the buyer's private capital cost, and $z_t \in \{0,1\}$ is the indicator variable that denotes whether the buyer won the item in period t. The benchmark bidding algorithms proceed as followed:

- `Conserv`: for $t \in [T]$, observe v_t and bid $b_t = v_t/\gamma$.
- `Budget-Pacing`: for $t \in [T]$, observe v_t and bid $b_t = \frac{v_t}{\alpha + \hat{\lambda}_t}$. Then observe payment $d_t z_t$ and update

$$\hat{\lambda}_{t+1} = \Pi_{[0,\bar{\lambda}]}\left(\hat{\lambda}_t - \bar{\eta}_t\left(\rho - d_t z_t\right)\right). \tag{23}$$

- `ROI-Pacing`: For $t \in [T]$, observe v_t and bid $b_t = \frac{(1+\hat{\mu}_t)v_t}{\alpha + \gamma\hat{\mu}_t}$. Then, observe payment $d_t z_t$ and auction outcome z_t. Update

$$\hat{\mu}_{t+1} = \Pi_{[0,\bar{\mu}]}\left(\hat{\mu}_t - \bar{\eta}_t\left(v_t z_t - \gamma d_t z_t\right)\right). \tag{24}$$

- `Pacing`: Observe v_t and bid $b_t = \frac{(1+\hat{\mu}_t)v_t}{\alpha + \gamma\hat{\mu}_t + \hat{\lambda}_t}$. Then, observe payment $d_t z_t$ and auction outcome z_t. Update $\hat{\lambda}_{t+1}$ and $\hat{\mu}_{t+1}$ according to Equations (23) and (24), respectively.

In the above algorithms, $\Pi_{\mathcal{C}}$ is the projection onto a set $\mathcal{C}$, and $\bar{\mu}$, $\bar{\lambda}$ are upper bounds for the corresponding optimal dual variables. Note that subgradient descent step-sizes $\bar{\eta}_t$ are typically chosen to be in the order of $\mathcal{O}(1/\sqrt{T})$ to yield optimal bidding performance (see Balseiro et al. (2021)).

D. Proofs for Section 4

Definition 5 *Recall in Definition 1, we defined $\psi(J,q)$ to be a K-dimensional threshold vector. For any other dimension $K' \neq K$, we will add a subscript and use $\psi_{K'}$ to denote threshold vectors in $\mathbb{R}^{K'}$.*

D.1. Proof of Proposition 1

We rewrite $\text{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)$ defined in Equation (2) by grouping all periods t during which a type k arrival occurs:

$$\begin{aligned}
\text{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho) = \max_{\boldsymbol{z}\in[0,1]^T} \quad & \textstyle\sum_{k\in[K]} (v^k - \alpha d^k)\left(\sum_{t:(v_t,d_t)=(v^k,d^k)} z_t\right)\\
\text{s.t.} \quad & \textstyle\sum_{k\in[K]} (v^k - \gamma d^k)\left(\sum_{t:(v_t,d_t)=(v^k,d^k)} z_t\right) \geq 0\\
& \textstyle\sum_{k\in[K]} d^k\left(\sum_{t:(v_t,d_t)=(v^k,d^k)} z_t\right) \leq \rho T
\end{aligned}$$

Since $\sum_{t:(v_t,d_t)=(v^k,d^k)} z_t \in [0,N^k]$, applying the change of variables $\sum_{t:(v_t,d_t)=(v^k,d^k)} z_t = N^k x^k$ for all $k\in[K]$ with some decision variable $x^k\in[0,1]$ yields the desired result. ■

D.2. Proof for Theorem 1

Our proof relies on the following lemma, whose proof can be found in Appendix D.4.1.

Lemma 1 *Consider $\{(a^i,b^i)\}_{i\in[m]}$ where $(a^i,b^i)\in\mathbb{R}_+\times\mathbb{R}_+$ for all $i\in[m]$. Assume $\frac{a^1}{b^1} > \frac{a^2}{b^2}\cdots > \frac{a^m}{b^m}$, and denote $b^{m+1}=\infty$. Then, for some $c>0$, (i) the unique optimal solution to*

$$\texttt{Knapsack} = \max_{\boldsymbol{y}\in[0,1]^m} \sum_{i\in[m]} a^i y^i \quad s.t. \sum_{i\in[m]} b^i y^i \leq c\,,$$

is the m-dimensional threshold vector[18] $\boldsymbol{y}^* = \psi_m(J,q)\in[0,1]^m$ *where $J=\max\{i\in[m]:\sum_{j\in[i]} b^j \leq c\}$, and* $q = \begin{cases} \frac{c-\sum_{i\in[J]} b^i}{b^{J+1}} & J\leq m-1\\ 0 & J=m\end{cases}$*. (ii) for any $e\geq 0$, the following optimization problem*

$$\texttt{Negval-Knapsack} = \max_{\boldsymbol{y}\in[0,1]^m} \sum_{i\in[m]} \left(a^i - e b^i\right) y^i \quad s.t. \sum_{i\in[m]} b^i y^i \leq c\,,$$

[18] Here, we recall ψ_m denotes m-dimensional threshold vectors for any $m\in\mathbb{N}$, and for simplicity we omit the subscript if we are working with K-dimensional threshold vectors

admits a unique optimal solution which is the threshold vector $\min\{\boldsymbol{y}^*, \psi(\kappa_e, 0)\}$*, where* $\kappa_e = \max\{i \in [m] : a^i \geq eb^i\}$*.*

We now return to our proof for Theorem 1, which consists of 3 steps:

1. We show that $\boldsymbol{X}^{\mathtt{B}} := \min\{\boldsymbol{x}^{\mathtt{B}}, \psi(\kappa_\alpha, 0)\}$ is the unique optimal solution to the "budget only" problem:

$$\texttt{P-Budget} = \max_{\boldsymbol{x}\in[0,1]^K} \sum_{k\in[K]} n^k \left(v^k - \alpha d^k\right) x^k \text{ s.t. } \sum_{k\in[K]} n^k d^k x^k \leq c\,, \tag{25}$$

   where we recall $\boldsymbol{x}^{\mathtt{B}} = \psi(b, q^{\mathtt{B}}) \in [0,1]^K$ is the threshold vector defined in the statement of the lemma.

2. We show that $\boldsymbol{X}^{\mathtt{R}} := \min\{\boldsymbol{x}^{\mathtt{R}}, \psi(\kappa_\alpha, 0)\}$ is the unique optimal solution the "ROI constraint only" problem:

$$\texttt{P-ROI} = \max_{\boldsymbol{x}\in[0,1]^K} \sum_{k\in[K]} n^k \left(v^k - \alpha d^k\right) x^k \text{ s.t. } \sum_{k\in[K]} n^k \left(v^k - \gamma d^k\right) x^k \geq 0\,, \tag{26}$$

   where we recall $\boldsymbol{x}^{\mathtt{R}} = \psi(b, q^{\mathtt{R}}) \in [0,1]^K$ is the threshold vector defined in the statement of the lemma.

3. We show that $\boldsymbol{x}^* = \min\{\boldsymbol{X}^{\mathtt{B}}, \boldsymbol{X}^{\mathtt{R}}\} = \min\{\boldsymbol{x}^{\mathtt{B}}, \boldsymbol{x}^{\mathtt{R}}, \psi(\kappa_\alpha, 0)\}$ is feasible to $U(\boldsymbol{n};\alpha,\gamma,c)$. In other words, we show $\boldsymbol{x}^*$ is feasible to both `P-Budget` and `P-ROI`. The rest of the proof is almost trivial: `P-Budget`, `P-ROI` and $U(\boldsymbol{n};\alpha,\gamma,c)$ have the same objective functions, while each of `P-Budget` and `P-ROI` has one less constraint than $U(\boldsymbol{n};\alpha,\gamma,c)$, respectively. So $\texttt{P-Budget} \geq U(\boldsymbol{n};\alpha,\gamma,c)$ and $\texttt{P-ROI} \geq U(\boldsymbol{n};\alpha,\gamma,c)$. If $\boldsymbol{x}^* = \boldsymbol{X}^{\mathtt{B}}$ and $\boldsymbol{x}^*$ is feasible to $U(\boldsymbol{n};\alpha,\gamma,c)$, then $\texttt{P-Budget} = U(\boldsymbol{n};\alpha,\gamma,c)$ and $\boldsymbol{X}^{\mathtt{B}}$ is the unique optimal solution to both `P-Budget` and $U(\boldsymbol{n};\alpha,\gamma,c)$. A similar argument holds for the case when $\boldsymbol{x}^* = \boldsymbol{X}^{\mathtt{R}}$.

Proof for (1) Since $\frac{n^k(v^k - \alpha d^k)}{n^k d^k} = \theta^k - \alpha$ and $\theta^1 > \cdots > \theta^K$, applying Lemma 1 (ii) with $m = K$, $a^k = n^k v^k$, $b^k = n^k d^k$, $c = c$ and $e = \alpha$ allows us to directly conclude that $\min\{\boldsymbol{x}^{\mathtt{B}}, \psi(\kappa_\alpha, 0)\}$ is the unique optimal solution to `P-Budget`.

Proof for (2) Let $\widetilde{\boldsymbol{x}} \in [0,1]^K$ be any optimal solution to `P-ROI`. Recall $\kappa_\gamma = \max\{k \in [K] : v^k \geq \gamma d^k\}$ so that $v^k \geq \gamma d^k$ for all $k \leq \kappa_\gamma$. [19] Then it is easy to see for any $k \in [\kappa_\gamma]$, $\widetilde{x}^k = 1$. This is because if there exists some $j \leq \kappa_\gamma$ such that $\widetilde{x}^j < 1$, then the solution $\boldsymbol{x} = (\widetilde{x}^1 \ldots \widetilde{x}^{j-1}, 1, \widetilde{x}^{j+1}, \ldots \widetilde{x}^K)$ is feasible and yields a strictly larger objective than $\widetilde{\boldsymbol{x}}$:

$$\sum_{k\in[K]} n^k \left(v^k - \alpha d^k\right) x^k - \sum_{k\in[K]} n^k \left(v^k - \alpha d^k\right) \widetilde{x}^k = \left(v^j - \alpha d^j\right)(1 - \widetilde{x}^j) > 0\,. \tag{27}$$

[19] Recall $\kappa_\alpha = \max\{k \in [K] : v^k \geq \alpha d^k\}$, so $\kappa_\gamma < \kappa_\alpha$ because $\alpha < \gamma$ and $\frac{v^1}{d^1} > \frac{v^2}{d^2} > \ldots \frac{v^K}{d^K}$.

Note that the final inequality cannot be equal because we assumed $\widetilde{x}^j < 1$ and $v^k \neq \alpha d^k$ for all $k \in [K]$. Hence, the optimal solution to P-ROI takes the form of $\widetilde{\boldsymbol{x}} = (\underbrace{1 \dots 1}_{\kappa_\gamma \text{ 1's}}, \widetilde{y}^{\kappa_\gamma+1}, \dots \widetilde{y}^K) \in [0,1]^K$. Hence, we know that $\widetilde{\boldsymbol{y}} := (\widetilde{y}^{\kappa_\gamma}, \dots \widetilde{y}^K)$ must satisfy

$$\widetilde{\boldsymbol{y}} \in \arg\max_{\boldsymbol{x} \in [0,1]^{K-\kappa_\gamma}} \sum_{k=\kappa_\gamma+1}^{K} n^k \left(v^k - \alpha d^k\right) x^k \text{ s.t. } \sum_{k=\kappa_\gamma+1}^{K} n^k \left(\gamma d^k - v^k\right) x^k \leq \widetilde{c}, \tag{28}$$

where we defined $\widetilde{c} = \sum_{k \in [\kappa_\gamma]} n^k \left(v^k - \gamma d^k\right) > 0$. Note that we have $\gamma d^k - v^k > 0$ for all $k = \kappa_\gamma + 1 \dots K$. By simple calculations it is easy to see that for any $i, j \in \{\kappa_\gamma + 1 \dots K\}$, we have

$$\frac{v^i}{d^i} > \frac{v^j}{d^j} \iff \frac{\left(1 - \frac{\alpha}{\gamma}\right) n^i v^i}{n^i \left(\gamma d^i - v^i\right)} > \frac{\left(1 - \frac{\alpha}{\gamma}\right) n^j v^j}{n^j \left(\gamma d^j - v^j\right)}.$$

Hence $\frac{n^k v^k}{n^k \left(\gamma d^k - v^k\right)}$ decreases in k for $k \in \{\kappa_\gamma + 1 \dots K\}$. Therefore, in the context of Lemma 1 (ii), if we let with $a^i = \left(1 - \frac{\alpha}{\gamma}\right) n^i v^i$, $b^i = n^i \left(\gamma d^i - v^i\right)$, and $e = \frac{\alpha}{\gamma}$, we have $a^i - e b^i = n^i \left(v^i - \alpha d^i\right)$. So further setting $c = \widetilde{c}$ in Lemma 1 (ii), we get

$$\widetilde{\boldsymbol{y}} = \min \left\{ (\underbrace{1 \quad \dots \quad 1}_{\text{entries } y^{\kappa_\gamma+1}, \dots y^{\widetilde{r}}}, \widetilde{q}^r, 0 \dots, 0), \psi_{K-\kappa_\gamma}(\kappa_\alpha - \kappa_\gamma, 0) \right\} \in [0,1]^{K-\kappa_\gamma},$$

where

$$\widetilde{r} = \max \left\{ k \in \{\kappa_\gamma + 1 \dots K\} : \sum_{i=\kappa_\gamma+1}^{k} n^i \left(\gamma d^i - v^i\right) \leq \widetilde{c} \right\} \overset{(i)}{=} \max \left\{ k \in [K] : \sum_{i \in [k]} n^i \left(v^i - \gamma d^i\right) \geq 0 \right\} = r$$

$$\widetilde{q}^{\mathtt{R}} = \frac{\widetilde{c} - \sum_{k=\kappa_\gamma+1}^{\widetilde{r}} n^k \left(\gamma d^k - v^k\right)}{\left(\gamma d^{\widetilde{r}+1} - v^{\widetilde{r}+1}\right) n^{\widetilde{r}+1}} \overset{(ii)}{=} \frac{\sum_{k \in [\widetilde{r}]} \left(v^k - \gamma d^k\right) n^k}{\left(\gamma d^{\widetilde{r}+1} - v^{\widetilde{r}+1}\right) n^{\widetilde{r}+1}} = q^{\mathtt{R}}.$$

Here, in (i) and (ii) we used the definition of $\widetilde{c} = \sum_{k \in [\kappa_\gamma]} n^k \left(v^k - \gamma d^k\right) > 0$ and rearranged terms. Combining the fact that the optimal solution to P-ROI is $\widetilde{\boldsymbol{x}} = (\underbrace{1 \dots 1}_{\kappa_\gamma \text{ 1's}}, \widetilde{y}^{\kappa_\gamma+1}, \dots \widetilde{y}^K) \in [0,1]^K$, and $\widetilde{\boldsymbol{y}}$ is uniquely determined by Equation (28), we can conclude that $\widetilde{\boldsymbol{x}} = \min\{\psi(r, q^{\mathtt{R}}), \psi(\kappa_\alpha, 0)\} = \min\{\boldsymbol{x}^{\mathtt{R}}, \psi(\kappa_\alpha, 0)\}$ is the unique optimal solution to P-ROI.

Proof for (3) We use the following lemma whose proof can be found in Appendix D.4.2.

Lemma 2 (Ordering property for threshold vectors) *Consider $\{a^i\}_{i \in [m]} \in \mathbb{R}_+$ and $\{b^i\}_{i \in [m]} \in \mathbb{R}$ where there exists some $j \in [m]$ such that $b^i > 0$ for all $i = 1 \dots j$ and $b^i < 0$ for all $i = j+1, \dots m$. Let $\boldsymbol{Z}, \boldsymbol{Y} \in [0,1]^m$ be two threshold vectors such that $\boldsymbol{Y} = \psi_m(J_Y, q_Y)$, $\boldsymbol{Z} = \psi_m(J_Z, q_Z)$, and $\boldsymbol{Z} \succeq \boldsymbol{Y}$. Then the following hold:*

(i) $\sum_{i \in [m]} a^i Z^i \geq \sum_{i \in [m]} a^i Y^i$.

(ii) *If $\sum_{i \in [m]} b^i Z^i \geq 0$ then $\sum_{i \in [m]} b^i Y^i \geq 0$. Furthermore, if $b^{J_Y+1} < 0$, then $\sum_{i \in [m]} b^i Y^i \geq \sum_{i \in [m]} b^i Z^i \geq 0$.*

(iii) If $\sum_{i\in[m]} b^i Y^i < 0$ *then* $\sum_{i\in[m]} b^i Z^i < 0$.

Since $n^k d^k > 0$ for all $k \in [K]$ and $\boldsymbol{x}^* = \min\{\boldsymbol{X}^{\text{B}}, \boldsymbol{X}^{\text{R}}\} \preceq \boldsymbol{X}^{\text{B}}$, we can apply Lemma 2 (i) with $m = K$, $a^k = n^k d^k$, $\boldsymbol{Z} = \boldsymbol{X}^{\text{B}}$ and $\boldsymbol{Y} = \boldsymbol{x}^*$, which yields $\sum_{k\in[K]} n^k d^k x^{*,k} \le \sum_{k\in[K]} n^k d^k X^{\text{B},k} \le c$, where the last inequality is due to the fact that $\boldsymbol{X}^{\text{B}}$ is feasible to `P-Budget`. This implies $\boldsymbol{x}^*$ is also feasible to `P-Budget`.

On the other hand, $\boldsymbol{x}^* = \min\{\boldsymbol{X}^{\text{B}}, \boldsymbol{X}^{\text{R}}\} \preceq \boldsymbol{X}^{\text{R}}$. Since $n^k(v^k - \gamma d^k) > 0$ for $k = 1 \ldots \kappa_\gamma$ and $n^k(v^k - \gamma d^k) < 0$ for $k = \kappa_\gamma + 1 \ldots K$, we apply Lemma 2 (ii) with $m = K$, $b^k = n^k(v^k - \gamma d^k)$, $\boldsymbol{Z} = \boldsymbol{X}^{\text{R}}$ and $\boldsymbol{Y} = \boldsymbol{x}^*$, which shows

$$\sum_{k\in[K]} n^k\left(v^k - \gamma d^k\right) X^{\text{R},k} \overset{(i)}{\ge} 0 \overset{(ii)}{\Longrightarrow} \sum_{k\in[K]} n^k\left(v^k - \gamma d^k\right) x^{*,k} \ge 0\,,$$

where (i) follows from the fact that $\boldsymbol{X}^{\text{R}}$ is feasible to `P-ROI` and (ii) follows from Lemma 2 (ii). Hence $\boldsymbol{x}^*$ is also feasible to `P-ROI`. ■

D.3. Proof of Theorem 2

Let (J, q) be defined as Theorem 1 w.r.t. $\boldsymbol{n} = \boldsymbol{p}$ and $c = \rho$. If for each period $t \in [T]$ the buyer submits the threshold bid $b_t = \beta(v_t, J, q)$, then

$$\mathbb{E}\Big[\sum_{t\in[T]} (v_t - \alpha d_t)\,\mathbb{I}\{b_t \ge d_t\}\Big] = TU(\boldsymbol{p}; \alpha, \gamma, \rho) \ge \mathbb{E}\left[\text{OPT}(\{v_t, d_t\}_{t\in[T]}; \alpha, \gamma, \rho)\right].$$

Let $\boldsymbol{x}^*$ be the optimal solution to $U(\boldsymbol{p}; \alpha, \gamma, \rho)$. According to Theorem 1 $\boldsymbol{x}^*$ is the threshold vector $\psi(J, q)$ where $J \in [K]$ and $q \in [0, 1)$ are the optimal threshold type and remained probability as defined in Theorem 1. To show the proposed bidding strategy is $\mathcal{B}$-feasible and achieves a utility equal to $TU(\boldsymbol{p}; \alpha, \gamma, \rho)$, we show the following:

1. **Budget constraint satisfied**: $\sum_{t\in[T]} \mathbb{E}[d_t \mathbb{I}\{b_t \ge d_t\}] = T \sum_{k\in[K]} p^k d^k x^{*,k} \le \rho T$.
2. **ROI constraint satisfied**: $\sum_{t\in[T]} \mathbb{E}[w_t \mathbb{I}\{b_t \ge d_t\}] = T \sum_{k\in[K]} w^k d^k x^{*,k} \ge 0$.
3. **Optimal utility**: $\sum_{t\in[T]} \mathbb{E}[(v_t - \alpha d_t)\mathbb{I}\{b_t \ge d_t\}] = T \sum_{k\in[K]} (v^k - \alpha d^k) x^{*,k} = TU(\boldsymbol{p}; \alpha, \gamma, \rho)$.

To prove (1), consider the following

$$\begin{aligned}\mathbb{E}[d_t \mathbb{I}\{b_t \ge d_t\}] \;&=\; (1-q)\mathbb{E}\left[d_t \mathbb{I}\left\{\frac{v_t}{\theta^J} \ge d_t\right\}\right] + q\mathbb{E}\left[d_t \mathbb{I}\left\{\frac{v_t}{\theta^{J+1}} \ge d_t\right\}\right] \\ =\; (1-q)\sum_{k\in[K]} d^k p^k \mathbb{I}\left\{\theta^k \ge \theta^J\right\} + q\sum_{k\in[K]} d^k p^k \mathbb{I}\left\{\theta^k \ge \theta^{J+1}\right\} \;&=\; \sum_{k\in[J]} d^k p^k + q d^{J+1} p^{J+1} \overset{(i)}{=} \sum_{k\in[K]} d^k p^k x^{*,k}\,,\end{aligned}$$

where in the last equality we used the fact that $\boldsymbol{x}^* = \psi(J, q)$. Multiplying both sides by T concludes (1). The proofs for (2) and (3) are identical to that of (1) simply by replacing d_t with w_t and $v_t - \alpha d_t$ respectively.

To show $TU(\boldsymbol{p};\alpha,\gamma,\rho) \geq \mathbb{E}\left[\mathrm{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)\right]$, we first dualize $\mathrm{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)$. Let $\mu,\lambda \geq 0$ be the dual variables associated with the ROI and budget constraint, respectively. We then have

$$\begin{aligned}\mathbb{E}\left[\mathrm{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)\right] \;&\leq\; \mathbb{E}\left[\max_{\boldsymbol{z}\in[0,1]^T}\sum_{t\in[T]}\left((1+\mu v_t)-(\alpha+\gamma\mu+\lambda)\,d_t\right)z_t\right]+\lambda\rho T \\ \leq \sum_{t\in[T]}\mathbb{E}\left[\left((1+\mu v_t)-(\alpha+\gamma\mu+\lambda)\,d_t\right)_+\right]+\lambda\rho T \;&=\; T\sum_{k\in[K]}p^k\left((1+\mu v^k)-(\alpha+\gamma\mu+\lambda)\,d^k\right)_+ +\lambda\rho T\,.\end{aligned} \tag{29}$$

Similarly, if we dualize $U(\boldsymbol{p};\alpha,\gamma,\rho)$, again with dual variables $\mu,\lambda\geq 0$ that corresponds to the ROI and budget constraint, respectively:

$$\begin{aligned}U(\boldsymbol{p};\alpha,\gamma,\rho) \;&\overset{(i)}{\leq}\; \max_{\boldsymbol{x}\in[0,1]^K}\sum_{k\in[K]}\left((1+\mu v^k)-(\alpha+\gamma\mu+\lambda)\,d^k\right)p^k x^k+\lambda\rho \\ &=\; \sum_{k\in[K]}p^k\left((1+\mu v^k)-(\alpha+\gamma\mu+\lambda)\,d^k\right)_+ +\lambda\rho\,.\end{aligned}$$

Note that if we define $\widetilde{\mu},\widetilde{\lambda}=\arg\min_{\mu,\lambda\geq 0}\sum_{k\in[K]}\left((1+\mu v^k)-(\alpha+\gamma\mu+\lambda)\,d^k\right)_+ +\lambda\rho$, then by strong duality (i) becomes an equality w.r.t. $\widetilde{\mu},\widetilde{\lambda}$, and hence

$$U(\boldsymbol{p};\alpha,\gamma,\rho)=\sum_{k\in[K]}p^k\left((1+\widetilde{\mu}v^k)-\left(\alpha+\gamma\widetilde{\mu}+\widetilde{\lambda}\right)d^k\right)_+ +\widetilde{\lambda}\rho.$$

Since Equation (29) holds for all $\mu,\lambda\geq 0$, we can conclude

$$\mathbb{E}\left[\mathrm{OPT}(\{v_t,d_t\}_{t\in[T]};\alpha,\gamma,\rho)\right] \;\leq\; T\sum_{k\in[K]}p^k\left((1+\widetilde{\mu}v^k)-\left(\alpha+\gamma\widetilde{\mu}+\widetilde{\lambda}\right)d^k\right)_+ +\widetilde{\lambda}\rho T \;=\; TU(\boldsymbol{p};\alpha,\gamma,\rho)\,.$$

■

D.4. Additional Proofs for Appendix D

D.4.1. Proof for Lemma 1 The problem in (i) is exactly the well-studied 0-1 knapsack problem with arbitrary item sizes (see e.g. Dantzig (1957)), and we will omit the proof here. For (ii), let $\widetilde{\boldsymbol{y}}$ be any optimal solution to `Negval-Knapsack`. We claim that for any $i>\kappa_e$, $\widetilde{y}^i=0$. This is easy to see because if there exists some $j>\kappa_e$ s.t. $\widetilde{y}^j>0$, then the solution $\boldsymbol{y}=(\widetilde{y}^1\ldots\widetilde{y}^{j-1},0,\widetilde{y}^{j-1},\ldots\widetilde{y}^m)$ is feasible to `Negval-Knapsack`, and also yields a strictly larger objective than that of $\widetilde{\boldsymbol{y}}$ since

$$\sum_{i\in[m]}\left(a^i-eb^i\right)\widetilde{y}^i-\sum_{i\in[m]}\left(a^i-eb^i\right)y^i=\left(a^j-eb^j\right)\widetilde{y}^j<0,$$

where we used the fact that $a^j<eb^j$ by the definition of κ_e. Hence, the optimal solution to problem to `Negval-Knapsack`takes the form of $\widetilde{\boldsymbol{y}}=(\widetilde{y}^1,\widetilde{y}^2\ldots\widetilde{y}^{\kappa_e},0,\ldots 0)\in[0,1]^m$. Furthermore, we observe that $\widetilde{\boldsymbol{y}}_e:=(\widetilde{y}^1,y^2\ldots\widetilde{y}^{\kappa_e})$ must be the optimal solution to

$$\max_{\boldsymbol{y}\in[0,1]^{\kappa_e}}\sum_{i\in[\kappa_e]}\left(a^i-eb^i\right)y^i \quad \text{s.t.} \sum_{i\in[\kappa_e]}b^iy^i\leq c\,.$$

Now, since $a^i - eb^i \geq 0$ for all $i \in [\kappa_e]$, and $\frac{a^1-eb^1}{b^1} > \cdots > \frac{a^{\kappa_e}-eb^{\kappa_e}}{b^{\kappa_e}}$ (due to the fact that $\frac{a^i-eb^i}{b^i} = \frac{a^i}{b^i} - e$ and the decreasing ordering of $\frac{a^i}{b^i}$'s in i), we can apply (i) and conclude that $\widetilde{\boldsymbol{y}}_e$ is uniquely determined by the threshold vector $\psi_{\kappa_e}(J_e, q_e) \in \mathbb{R}^{\kappa_e}$, where $J_e = \max\{i \in [\kappa_e] : \sum_{j\in[i]} b^j \leq c\}$ and $q_e = \begin{cases} \frac{c-\sum_{i\in[J_e]} b^i}{b^{J_e+1}} & J_e \leq \kappa_e - 1 \\ 0 & J_e = \kappa_e \end{cases}$. Note that $J \geq J_e$ always holds, and recall $\boldsymbol{y}^* = \psi_m(J,q)$. It remains to consider two scenarios

- If $J_e = \kappa_e$, then $J \geq J_e = \kappa_e$, so $\boldsymbol{y}^* \succeq \psi_m(\kappa_e, 0) \in \mathbb{R}^m$, and hence $\min\{\boldsymbol{y}^*, \psi(\kappa_e, 0)\} = \psi_m(\kappa_e, 0)$. On the other hand, $\widetilde{\boldsymbol{y}} = (\widetilde{y}^1, \widetilde{y}^2 \ldots \widetilde{y}^{\kappa_e}, 0, \ldots 0) = (\underbrace{1\ldots1}_{J_e \text{ 1's}}, 0, \ldots 0) = \psi(\kappa_e, 0)$, so $\widetilde{\boldsymbol{y}} = \min\{\boldsymbol{y}^*, \psi_m(\kappa_e, 0)\}$.
- If $J_e \leq \kappa_e - 1$, then $J_e = J$ and $q_e = q$. So $\min\{\boldsymbol{y}^*, \psi_m(\kappa_e, 0)\} = \boldsymbol{y}^*$. On the other hand, $\widetilde{\boldsymbol{y}} = (\widetilde{y}^1, \widetilde{y}^2 \ldots \widetilde{y}^{\kappa_e}, 0, \ldots 0) = (\underbrace{1\ldots1}_{J_e \text{ 1's}}, q_e, 0, \ldots 0)$, so $\widetilde{\boldsymbol{y}} = \boldsymbol{y}^* = \min\{\boldsymbol{y}^*, \psi_m(\kappa_e, 0)\}$.

■

D.4.2. Proof for Lemma 2

(i) Since $a^i > 0$ for all $i \in [m]$, and $\boldsymbol{Z} \succeq \boldsymbol{Y}$ (i.e. $Z^i \geq Y^i$ for all $i \in [m]$), it is easy to see $\sum_{i\in[m]} a^i Z^i \geq \sum_{i\in[m]} a^i Y^i$.

(ii) Note that $Y^{J_Y+1} = q_Y$ while $Y^i = 0$ for all $i > J_Y + 1$. We prove the claim by contradiction. Assume $\sum_{i\in[m]} b^i Y^i < 0$, then it is easy to see $b^{J_Y+1} < 0$. This is because if $b^{J_Y+1} > 0$, then $b^i > 0$ for all $i = 1 \ldots J_Y + 1$ by the definition of $\{b^i\}_{i\in[m]}$, and hence $\sum_{i\in[m]} b^i Y^i = \sum_{i\in[J_Y+1]} b^i Y^i \geq 0$ contradicting our assumption.

Next, since $\sum_{i\in[m]} b^i Y^i < 0 \leq \sum_{i\in[m]} b^i Z^i$, we have $\sum_{i\in[m]} b^i (Z^i - Y^i) \geq 0$. On the other hand,

$$\sum_{i\in[m]} b^i (Z^i - Y^i) \overset{(i)}{=} \sum_{i=J_Y+1}^m b^i (Z^i - Y^i) \overset{(ii)}{<} 0\,.$$

Here, (i) follows from the definition of a threshold vector so that $Y^i = 1$ for all $i = 1 \ldots J_Y$ and also $Z^i = 1$ for all $i = 1 \ldots J_Y$ due to $\boldsymbol{Z} \succeq \boldsymbol{Y}$. (ii) follows from the fact that $b^{J_Y+1} < 0$ so $b^i < 0$ for all $i \geq J_Y + 1$ due to the definition of $\{b^i\}_{i\in[m]}$. Hence, we arrive at a contradiction, which allows us to conclude the first half of the claim, i.e. $\sum_{i\in[m]} b^i Z^i \geq 0$ implies $\sum_{i\in[m]} b^i Y^i \geq 0$.

We now show the second half of the claim i.e. $b^{J_Y+1} < 0$ implies $\sum_{i\in[m]} b^i Y^i \geq \sum_{i\in[m]} b^i Z^i \geq 0$. If $b^{J_Y+1} < 0$, then $b^i < 0$ for all $i = J_Y + 1 + \ldots J_Z + 1$, and hence

$$\sum_{i\in[m]} b^i (Z^i - Y^i) = b^{J_Y+1}(Z^{J_Y+1} - Y^{J_Y+1}) + \sum_{i=J_Y+2}^{J_Z+1} b^i Z^i \overset{(i)}{<} 0\,.$$

Note that in the above inequality the summand $\sum_{i=J_Y+2}^{J_Z+1} b^i Z^i$ does not exist if $J_Y = J_Z$, and in (i) we also used the fact that $Y^i = 0$ for all $i > J_Y + 1$ using the definition of a threshold veector.

(iii) We again use a contradiction argument by assuming $\sum_{i\in[m]} b^i Z^i \geq 0$, and the rest of the proof is almost identical to that of (ii) so we will omit it here. ■

E. Proofs for Section 5

E.1. Proof for Theorem 8

Define the event $\mathcal{G}_t = \{\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| < \ell_t\}$, and recall that $\mathbb{P}(\mathcal{G}_t) \geq 1 - \frac{1}{T}$, $w^k = v^k - \gamma d^k$, $\bar{w} = \max_{k\in[K]} |v^k - \gamma d^k|$, and $\bar{d} = \max_{k\in[K]} d^k$. As in the proof of Theorem 2, we recognize that for any $t \in [T]$,

$$\mathbb{E}\left[w_t \mathbb{I}\{b_t \geq d_t\} \,\Big|\, \hat{\boldsymbol{x}}_t\right] = \sum_{k\in[K]} w^k p^k \hat{x}_t^k \quad \text{and} \quad \mathbb{E}\left[d_t \mathbb{I}\{b_t \geq d_t\} \,\Big|\, \hat{\boldsymbol{x}}_t\right] = \sum_{k\in[K]} d^k p^k \hat{x}_t^k\,,$$

where $\hat{\boldsymbol{x}}_t = \psi(\hat{J}_{t+1}, \hat{q}_{t+1})$ as defined in Equation (13) in Algorithm 1, and b_t is the threshold bid $\beta(v_t, \hat{J}_t, \hat{q}_t)$. Hence, to show that Algorithm 1 satisfies both ROI and budget constraint in Equation (1), it suffices to show

$$\sum_{t\in[T]}\sum_{k\in[K]} \mathbb{E}\left[w^k p^k \hat{x}_t^k\right] \geq 0, \quad \text{and} \quad \sum_{t\in[T]}\sum_{k\in[K]} \mathbb{E}\left[d^k p^k \hat{x}_t^k\right] \leq \rho T\,.$$

Our results rely on the following lemma which mainly states that when the estimate $\hat{\boldsymbol{p}}_t$ for $\boldsymbol{p}$ is accurate for large t, then the corresponding estimates $\hat{b}_t, \hat{r}_t, \hat{q}_t^{\mathrm{R}}, \hat{q}_t^{\mathrm{B}}$ are all accurate. See proof in Appendix E.3.1.

Lemma 3 *Assume event $\mathcal{G}_t = \{\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| < \ell_t\}$ holds for $t \geq \mathcal{T}_{\mathcal{A}} = \min\{t \in [T] : \ell_t < S\}$, where S is defined in the statement of Theorem 8. Then, the following conditions hold*

Condition (i): $\hat{b}_t = b$ and hence $\hat{J}_t = J$. Condition (ii): $\hat{q}_t^{R}$, $\hat{q}_t^{B} > 0$.

Condition (iii): $\bar{w} \cdot \ell_t \leq \sum_{k\in[K]} \hat{x}_t^{R,k} p^k w^k \leq \left(3\sqrt{K}+5\right)\bar{w}\ell_t$ and $\bar{d} \cdot \ell_t \leq \rho - \sum_{k\in[K]} \hat{x}_t^{B,k} p^k d^k \leq \left(3\sqrt{K}+5\right)\bar{d}\ell_t$.

Condition (iv): $|(q^R - \hat{q}_t^R)\, p^{r+1}| \leq \frac{\bar{w}}{\underline{w}}\left(3\sqrt{K}+5\right)\ell_t$ and $|(q^B - \hat{q}_t^B)\, p^{b+1}| \leq \frac{\bar{d}}{\underline{d}}\left(3\sqrt{K}+5\right)\ell_t$.

The remaining proof consists of 3 parts: (1) proving ROI constraint is satisfied; (2) proving budget constraint is satisfied; (3) upper bounding regret.

(1) Proving ROI constraint is satisfied. i.e., $\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[k]} \hat{x}_t^k p^k w^k\right] \geq 0$. We lower bound the realized ROI balance as followed:

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[K]} \hat{x}_t^k p^k w^k\right] \;\geq\; -\bar{w}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}}\sum_{k\in[K]} \left(\mathbb{E}\left[\hat{x}_t^k p^k w^k \mathbb{I}\{\mathcal{G}_t\}\right] + \mathbb{E}\left[\hat{x}_t^k p^k w^k \mathbb{I}\{\mathcal{G}_t^c\}\right]\right) \tag{30}$$

$$\geq\; -\bar{w}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{E}\left[\sum_{k\in[K]} \hat{x}_t^k p^k w^k \mathbb{I}\{\mathcal{G}_t\}\right] - \bar{w}\sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{P}\left(\mathcal{G}_t^c\right) \overset{(i)}{>} -\bar{w}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{E}\left[\sum_{k\in[K]} \hat{x}_t^k p^k w^k \mathbb{I}\{\mathcal{G}_t\}\right] - \bar{w}$$

$$\overset{(ii)}{\geq} \begin{cases} -\bar{w}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}} \left(\sum_{k\in[J]} p^k w^k\right)\mathbb{P}\left(\mathcal{G}_t\right) - \bar{w}, & \text{if } w^{J+1} > 0 \\ -\bar{w}\mathcal{T}_{\mathcal{A}} + \bar{w}\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t \mathbb{P}\left(\mathcal{G}_t\right) - \bar{w}, & \text{if } w^{J+1} < 0 \end{cases},$$

where (i) follows from $\mathbb{P}(\mathcal{G}_t^c) = \mathbb{P}(\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| > \ell_t) < \frac{1}{T}$; For (ii) we have two scenarios: when $w^{J+1} > 0$, we know that $w^k > 0$ for all $k \le J+1$, so $\sum_{k\in[K]} \hat{x}_t^k p^k w^k = \sum_{k\in[J]} p^k w^k + \hat{q}_t p^{J+1} w^{J+1} \ge \sum_{k\in[J]} p^k w^k$ where we used the definition of the threshold vector $\hat{\boldsymbol{x}}_t = \psi(\hat{J}_t, \hat{q}_t)$ and $\hat{J}_t = J$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$ according to Lemma 3 (i). For the scenario when $w^{J+1} < 0$, we have $\mathrm{Reg}(\boldsymbol{p}, T, \alpha, \gamma, \rho) \le \max_{k\in[K]} |v^k - \alpha d^k| \cdot \left(2\mathcal{T}_{\mathcal{A}} + 1 + C\sum_{t\in[T]} \ell_t\right)$, where for (iii) we applied the ordering property for threshold vectors in Lemma 2 (ii) where we take $\boldsymbol{Y} = \hat{\boldsymbol{x}}_t$, $\boldsymbol{Z} = \hat{\boldsymbol{x}}_t^{\mathtt{R}}$ using the fact that $\psi(\hat{J}_t, \hat{q}_t) = \hat{\boldsymbol{x}}_t = \min\{\hat{\boldsymbol{x}}_t^{\mathtt{R}}, \hat{\boldsymbol{x}}_t^{\mathtt{B}}, \psi(\kappa_\alpha, 0)\} \preceq \hat{\boldsymbol{x}}_t^{\mathtt{R}} = \psi(\hat{J}_t^{\mathtt{R}}, \hat{q}_t^{\mathtt{R}})$ and $w^{\hat{J}_t+1} = w^{J+1} < 0$, and for (iv) we directly applied Lemma 3 (iii) as we assumed $t > \mathcal{T}_{\mathcal{A}}$ and the event $\mathcal{G}_t$ holds.

It remains to further lower bound Equation (30) under the two scenarios $w^{J+1} > 0$ and $w^{J+1} < 0$. When $w^{J+1} > 0$, we have

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[K]} \hat{x}_t^k p^k w^k\right] \ \ge\ -\bar{w}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}}\left(\sum_{k\in[J]} p^k w^k\right)\mathbb{P}(\mathcal{G}_t) - \bar{w}$$
$$\ge\ -\bar{w}\mathcal{T}_{\mathcal{A}} + \left(1 - \frac{1}{T}\right)(T - \mathcal{T}_{\mathcal{A}})\left(\sum_{k\in[J]} p^k w^k\right) - \bar{w} \ \overset{(i)}{\ge}\ \frac{T}{4}\left(\sum_{k\in[J]} p^k w^k\right) - \bar{w}\,(\mathcal{T}_{\mathcal{A}} + 1) \ \overset{(ii)}{\ge}\ 0\,,$$

where in (i) we used the condition $\frac{1}{T} < \frac{1}{2}$ and $T > 2\mathcal{T}_{\mathcal{A}}$. Furthermore, (ii) follows from our assumption that $T > \frac{4\bar{w}(\mathcal{T}_{\mathcal{A}}+1)}{\sum_{k\in[J]} p^k w^k}$. When $w^{J+1} < 0$, we have

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[K]} \hat{x}_t^k p^k w^k\right] \ \ge\ -\bar{w}\mathcal{T}_{\mathcal{A}} + \bar{w}\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t \mathbb{P}(\mathcal{G}_t) - \bar{w} \ \ge\ -\bar{w}\mathcal{T}_{\mathcal{A}} + \left(1 - \frac{1}{T}\right)\bar{w}\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t - \bar{w}$$
$$\overset{(i)}{\ge}\ \frac{\bar{w}}{2}\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t - \bar{w}\,(\mathcal{T}_{\mathcal{A}} + 1) \ \overset{(ii)}{\ge}\ 0\,,$$

where in (i) we used the condition $\frac{1}{T} < \frac{1}{2}$. Moreover, (ii) follows from the condition $\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t > 2\,(\mathcal{T}_{\mathcal{A}} + 1)$.

Proving budget constraint is satisfied. i.e., $\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[k]} \hat{x}_t^k p^k d^k\right] \le \rho T$. Using the ordering property for threshold vectors in Lemma 2 (i), since $p^k d^k \ge 0$ for all k and $\hat{\boldsymbol{x}}_t = \min\{\hat{\boldsymbol{x}}_t^{\mathtt{B}}, \hat{\boldsymbol{x}}_t^{\mathtt{R}}, \psi(\kappa_\alpha, 0)\} \preceq \hat{\boldsymbol{x}}_t^{\mathtt{B}}$, we have $\sum_{k\in[K]} \hat{x}_t^k p^k d^k \le \sum_{k\in[K]} \hat{x}_t^{\mathtt{B},k} p^k d^k$ for all $t \in [T]$.Hence,

$$\mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[K]} \hat{x}_t^k p^k d^k\right] \ \le\ \mathbb{E}\left[\sum_{t\in[T]}\sum_{k\in[K]} \hat{x}_t^{\mathtt{B},k} p^k d^k\right]$$
$$\le\ \bar{d}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}}\mathbb{E}\left[\sum_{k\in[K]} \hat{x}_t^{\mathtt{B},k} p^k d^k \mathbb{I}\{\mathcal{G}_t\}\right] + \sum_{t>\mathcal{T}_{\mathcal{A}}}\mathbb{E}\left[\sum_{k\in[K]} \hat{x}_t^{\mathtt{B},k} p^k d^k \mathbb{I}\{\mathcal{G}_t^c\}\right]$$
$$\overset{(i)}{\le}\ \bar{d}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}}\left(\rho - \bar{d}\ell_t\right)\mathbb{P}(\mathcal{G}_t) + \bar{d}\,(T - \mathcal{T}_{\mathcal{A}})\,\frac{1}{T}$$
$$<\ \bar{d}\mathcal{T}_{\mathcal{A}} + \sum_{t>\mathcal{T}_{\mathcal{A}}}\left(\rho - \bar{d}\ell_t\right) + \bar{d} \ <\ \rho T - \bar{d}\left(\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t - \mathcal{T}_{\mathcal{A}} - 1\right) \ \overset{(ii)}{<}\ \rho T\,.$$

where (i) follows directly from Lemma 3 (iii) since $t > \mathcal{T}_{\mathcal{A}}$ and we assume event $\mathcal{G}_t$ holds; (ii) follows from $\sum_{t>\mathcal{T}_{\mathcal{A}}} \ell_t > 2\,(\mathcal{T}_{\mathcal{A}} + 1)$.

Bounding regret. As in the proof of Theorem 2, it is easy to see for any $t \in [T]$, $\mathbb{E}\left[(v_t - \alpha d_t)\,\mathbb{I}\{b_t \geq d_t\}\,\middle|\,\hat{\boldsymbol{x}}_t\right] = \sum_{k\in[K]}(v^k - \alpha d^k)\,p^k \hat{x}_t^k$. Then, we can bound the regret as follows:

$$\begin{aligned}
&\operatorname{Reg}(\boldsymbol{p}, T, \alpha, \gamma, \rho) \\
&= \mathbb{E}\left[\operatorname{OPT}(\{v_t, d_t\}_{t\in[T]}; \alpha, \gamma, \rho)\right] - \sum_{t\in[T]} \mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \mathbb{I}\{b_t \geq d_t\}\right] \\
&\overset{(i)}{\leq} T \cdot U(\boldsymbol{p}; \alpha, \gamma, \rho) - \sum_{t\in[T]} \mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \mathbb{I}\{b_t \geq d_t\}\right] \\
&\overset{(ii)}{\leq} 2\mathcal{T}_{\mathcal{A}} \max_{k\in[K]} \left|v^k - \alpha d^k\right| + \sum_{t>\mathcal{T}_{\mathcal{A}}} \left(u\left(\boldsymbol{p}; \alpha, \gamma, \rho\right) - \mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \mathbb{I}\{b_t \geq d_t\}\right)\right] \\
&= 2\mathcal{T}_{\mathcal{A}} \max_{k\in[K]} \left|v^k - \alpha d^k\right| + \sum_{t>\mathcal{T}_{\mathcal{A}}} \left(u\left(\boldsymbol{p}; \alpha, \gamma, \rho\right) - \mathbb{E}\left[\mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \mathbb{I}\{b_t \geq d_t\}\right) \middle| \hat{\boldsymbol{x}}_t\right]\right] \\
&= 2\mathcal{T}_{\mathcal{A}} \max_{k\in[K]} \left|v^k - \alpha d^k\right| + \sum_{t>\mathcal{T}_{\mathcal{A}}} \left(u\left(\boldsymbol{p}; \alpha, \gamma, \rho\right) - \mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \mathbb{I}\{b_t \geq d_t\}\right) p^k \hat{x}_t^k\right] \\
&\overset{(iii)}{=} 2\mathcal{T}_{\mathcal{A}} \max_{k\in[K]} \left|v^k - \alpha d^k\right| + \sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{E}\left[\sum_{k\in[K]} \left(v^k - \alpha d^k\right) \left(x^{*,k} - \hat{x}_t^k\right) p^k\right] \\
&\leq 2\mathcal{T}_{\mathcal{A}} \max_{k\in[K]} \left|v^k - \alpha d^k\right| + \max_{k\in[K]} \left|v^k - \alpha d^k\right| \sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{E}\left[\sum_{k\in[K]} \left|x^{*,k} - \hat{x}_t^k\right| p^k\right].
\end{aligned} \tag{31}$$

Here, (i) follows from upper bounding $\operatorname{OPT}(\{v_t, d_t\}_{t\in[T]}; \alpha, \gamma, \rho)$ with $T \cdot U(\boldsymbol{p}; \alpha, \gamma, \rho)$ as shown in Lemma 2; (ii) follows from $U(\boldsymbol{p}; \alpha, \gamma, \rho) \leq \max_{k\in[K]} |v^k - \alpha d^k|$ and $\left|\sum_{k\in[K]} (v^k - \alpha d^k)\,\mathbb{I}\{b_t \geq d_t\}\right| \leq \max_{k\in[K]} |v^k - \alpha d^k|$ for all $t \in [T]$; (iii) follows from the definition of $\boldsymbol{x}^* = \psi(J, q)$ being the optimal solution of $u(\boldsymbol{p}; \alpha, \gamma, \rho)$.

Now, considering $\mathbb{E}\left[\sum_{k\in[K]} |x^{*,k} - \hat{x}_t^k|\, p^k\right]$ for all $t > \mathcal{T}_{\mathcal{A}}$, we have

$$\mathbb{E}\left[\sum_{k\in[K]} \left|x^{*,k} - \hat{x}_t^k\right| p^k\right] \overset{(i)}{<} \mathbb{E}\left[\sum_{k\in[K]} \left|x^{*,k} - \hat{x}_t^k\right| p^k \mathbb{I}\{\mathcal{G}_t\}\right] + \frac{1}{T} \overset{(ii)}{=} \mathbb{E}\left[\left|q - \hat{q}_t\right| p^{J+1} \mathbb{I}\{\mathcal{G}_t\}\right] + \frac{1}{T}.$$

In (i) we used the fact that $|x^{*,k} - \hat{x}_t^k| \leq 1$ for all $k \in [K]$ since $x^{*,k}, \hat{x}_t^k \in [0,1]$ and $\mathbb{P}(\mathcal{G}_t^c) < \frac{1}{T}$. In (ii), we first evoked Theorem 1 such that $\boldsymbol{x}^* = \psi(J, q)$; then we used the definition $\hat{\boldsymbol{x}}_t = \psi(\hat{J}_t, \hat{q}_t)$ and according to Lemma 3 (i) we have $\hat{J}_t = J$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$. Plugging this back into Equation (31), we get

$$\operatorname{Reg}(\boldsymbol{p}, T, \alpha, \gamma, \rho) \leq \max_{k\in[K]} \left|v^k - \alpha d^k\right| \left(2\mathcal{T}_{\mathcal{A}} + 1 + \sum_{t>\mathcal{T}_{\mathcal{A}}} \mathbb{E}\left[\left|q - \hat{q}_t\right| p^{J+1} \mathbb{I}\{\mathcal{G}_t\}\right]\right).$$

Therefore, it now remains to show that under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$, we have

$$\left|(\hat{q}_t - q)\, p^{J+1}\right| \;\leq\; \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\} \left(3\sqrt{K} + 5\right) \ell_t. \tag{32}$$

First, recall the definitions $J=\min\{b,r,\kappa_\alpha\}$,

$$\hat{q}_t \overset{(i)}{=} \min\left\{\hat{x}_t^{\mathtt{R},J+1},\hat{x}_t^{\mathtt{B},J+1},\mathbb{I}\left\{\kappa_\alpha\geq J+1\right\}\right\} \quad \text{and} \quad q=\min\left\{x^{\mathtt{R},J+1},x^{\mathtt{B},J+1},\mathbb{I}\left\{\kappa_\alpha\geq J+1\right\}\right\}.$$

The definition in (i) should be $\hat{q}_t=\min\left\{\hat{x}_t^{\mathtt{R},\hat{J}_t+1},\hat{x}_t^{\mathtt{B},\hat{J}_t+1},\mathbb{I}\left\{\kappa_\alpha\geq \hat{J}_t\right\}\right\}$, but again under event $\mathcal{G}_t$ for $t>\mathcal{T}_{\mathcal{A}}$, Lemma 3 states that $\hat{J}_t=J$.

If $J+1>\kappa_\alpha$, then we know that $\hat{q}_t=q=0$ so the inequality in Equation (32) trivially holds. If $J+1\leq\kappa_\alpha$, then either $J=r$ or $J=b$. Furthermore, $\hat{q}_t=\min\left\{\hat{x}_t^{\mathtt{R},J+1},\hat{x}_t^{\mathtt{B},J+1}\right\}$ and $q=\min\left\{x^{\mathtt{R},J+1},x^{\mathtt{B},J+1}\right\}$. Hence, in the following we consider each of the four events:

- $\{x^{\mathtt{R},J+1}\leq x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}\leq\hat{x}_t^{\mathtt{B},J+1}\right\}$
- $\{x^{\mathtt{R},J+1}> x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}>\hat{x}_t^{\mathtt{B},J+1}\right\}$
- $\{x^{\mathtt{R},J+1}\leq x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}>\hat{x}_t^{\mathtt{B},J+1}\right\}$
- $\{x^{\mathtt{R},J+1}> x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}\leq\hat{x}_t^{\mathtt{B},J+1}\right\}$

We observe that if the event $\{x^{\mathtt{R},J+1}\leq x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}\leq\hat{x}_t^{\mathtt{B},J+1}\right\}$ holds, then $J=r$, so $p^{J+1}=p^{r+1}$, as well as $q=q^{\mathtt{R}}$ and $\hat{q}_t=\hat{x}_t^{\mathtt{R},r+1}=\hat{q}_t^{\mathtt{R}}$. In this case Equation (32) holds directly as a result of Lemma 3 (iv). Similarly, if the event $\{x^{J+1,\mathtt{R}}>x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}>\hat{x}_t^{\mathtt{B},J+1}\right\}$ holds, then $J=b$, so $p^{J+1}=p^{b+1}$ as well as $q=q^{\mathtt{B}}$ and $\hat{q}_t=\hat{x}_t^{\mathtt{B},b+1}=\hat{q}_t^{\mathtt{B}}$, so Equation (32) again holds directly as a result of Lemma 3 (iv).

Now, consider the scenario where the event $\{x^{\mathtt{R},J+1}\leq x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}>\hat{x}_t^{\mathtt{B},J+1}\right\}$ holds, which implies $J=r$ and $\hat{J}_t=b$, as well as $\hat{q}_t=\hat{x}_t^{\mathtt{B},J+1}=\hat{q}_t^{\mathtt{B}}$, and $q=x^{\mathtt{R},J+1}=q^{\mathtt{R}}$. Since $\hat{b}_t=b$ and $\hat{r}_t=r$ under event $\mathcal{G}_t$, we know that $J=r=b$, which further implies $p^{J+1}=p^{b+1}=p^{r+1}$. Therefore,

$$\begin{aligned}\hat{q}_t p^{J+1}=\hat{x}_t^{\mathtt{B},J+1}p^{J+1} &\overset{(i)}{<} \hat{x}^{\mathtt{R},J+1}p^{J+1}=\hat{q}_t^{\mathtt{R}}p^{r+1}\overset{(iii)}{\leq}\underbrace{q^{\mathtt{R}}p^{r+1}}_{=qp^{J+1}}+\max\left\{\frac{\bar{w}}{\underline{w}},\frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\ell_t\\ \underbrace{q^{\mathtt{R}}p^{r+1}}_{=qp^{J+1}}=x^{\mathtt{R},J+1}p^{J+1} &\overset{(ii)}{\leq} x^{\mathtt{B},J+1}p^{J+1}=q^{\mathtt{B}}p^{b+1}\overset{(iv)}{\leq}\underbrace{\hat{q}_t^{\mathtt{B}}p^{b+1}}_{=\hat{q}_tp^{J+1}}+\max\left\{\frac{\bar{w}}{\underline{w}},\frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\ell_t\,.\end{aligned} \tag{33}$$

Here (i) and (ii) are valid because the event$\{x^{\mathtt{R},J+1}\leq x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}>\hat{x}_t^{\mathtt{B},J+1}\right\}$ holds; both (iii) and (iv) follow from Lemma 3 (iv). Hence combining the above inequalities we can conclude

$$|q-\hat{q}_t|\,p^{J+1}\leq\max\left\{\frac{\bar{w}}{\underline{w}},\frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\ell_t\,. \tag{34}$$

Following the same analysis, when the event $\{x^{\mathtt{R},J+1}> x^{\mathtt{B},J+1}\}\cap\left\{\hat{x}^{\mathtt{R},J+1}\leq\hat{x}_t^{\mathtt{B},J+1}\right\}$ holds, we can again show $J=r=b$, so $p^{J+1}=p^{b+1}=p^{r+1}$, and also $\hat{q}_t=\hat{x}_t^{\mathtt{R},J+1}=\hat{q}_t^{\mathtt{R}}$, and $q=x^{\mathtt{B},J+1}=q^{\mathtt{B}}$. Using a similar argument as in Equation (33), we can conclude $|q-\hat{q}_t|\,p^{J+1}\leq\max\left\{\frac{\bar{w}}{\underline{w}},\frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\ell_t$. ■

E.2. Proof for Theorem 9

For both SGD with vanishing and constant step sizes, we only need to show $\|\boldsymbol{p}^* - \hat{\boldsymbol{p}}_t\| \leq \ell_t$ w.p. at least $1 - \frac{1}{T}$, and check that there exists some $\widetilde{T} < \infty$ such that for all $T > \widetilde{T}$, the following conditions hold

$$T > \max\left\{2\mathcal{T}_{\mathcal{A}}, \frac{4\bar{w}\left(\mathcal{T}_{\mathcal{A}}+1\right)}{\sum_{k\in[J]} p^k w^k}\right\} \text{ and } \sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \ell_t > 2\left(\mathcal{T}_{\mathcal{A}}+1\right),$$

where $\mathcal{T}_{\mathcal{A}} = \min\{t \in [T] : \ell_t < S\}$ for S defined in Theorem 8. Note that $\mathcal{T}_{\mathcal{A}}$ can possibly depend on T since $\ell_t^{\mathcal{A}}$ may depend on T. Then, we can further bound the regret by applying Theorem 8.

In the following proof, recall

$$\boldsymbol{s}_t = \left(\mathbb{I}\left\{(v_t, d_t) = (v^1, d^1)\right\}, \ldots, \mathbb{I}\left\{(v_t, d_t) = (v^K, d^K)\right\}\right).$$

Furthermore, let $\mathcal{F}_t$ be the sigma algebra generated by $\{(v_\tau, d_\tau)\}_{\tau\in[t]}$.

Proof for SGD with vanishing step size

Here, we utilize the following Lemma which is equivalent to Proposition 1 in Rakhlin et al. (2011). The proof is exactly the same as that in Rakhlin et al. (2011), but for completeness we will include it in Appendix E.3.2

Lemma 4 *Assume $(v_t, d_t) \sim \boldsymbol{p}^* \in \Delta_k$ for all $t \in [T]$, and let $\{\hat{\boldsymbol{p}}_t\}_{t\in[T]}$ be the estimates for $\boldsymbol{p}^*$ generated by SGD with vanishing step size $\eta_t = \frac{1}{t}$ for all t. Then w.p. at least $1-\delta$ for some $\delta \in (0, 1/e)$, we have $\|\boldsymbol{p}^* - \hat{\boldsymbol{p}}_t\| \leq \sqrt{\frac{600\log\left(\frac{\log(T)}{\delta}\right)+12}{t}}$.*

Now, returning to the proof for SGD with vanishing step sizes, we can simply take $\delta = \frac{1}{T}$ and set $\ell_t = \sqrt{\frac{600\log(T\log(T))+12}{t}}$, which yields the desired high probability bound in the Theorem statement.

We now show that there exists some $T_1 < \infty$ such that for all $T > T_1$ the following conditions always hold.

$$T > \max\left\{2\mathcal{T}_{\mathcal{A}}, \frac{4\bar{w}\left(\mathcal{T}_{\mathcal{A}}+1\right)}{\sum_{k\in[J]} p^k w^k}\right\} \text{ and } \sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \ell_t > 2\left(\mathcal{T}_{\mathcal{A}}+1\right),$$

where $\mathcal{T}_{\mathcal{A}} = \min\{t \in [T] : \ell_t < S\}$. It is easy to see that $\mathcal{T}_{\mathcal{A}} = \left\lceil \sqrt{\frac{600\log(T\log(T))+12}{S}} \right\rceil = \Theta(\log(T))$. Therefore, for any large enough T, we must have $T > \max\left\{2\mathcal{T}_{\mathcal{A}}, \frac{4\bar{w}(\mathcal{T}_{\mathcal{A}}+1)}{\sum_{k\in[J]} p^k w^k}\right\} = \Theta(\log(T))$.

Next, we show that we can satisfy this condition $\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \ell_t > 2\left(\mathcal{T}_{\mathcal{A}}+1\right)$ for large enough T. Recall that $\ell_t = \sqrt{\frac{600\log(T\log(T))+12}{t}}$ and note that

$$\begin{aligned}\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \frac{1}{\sqrt{t}} &= \sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \int_t^{t+1} \frac{1}{\sqrt{t}} d\tau \geq \sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \int_t^{t+1} \frac{1}{\sqrt{\tau}} d\tau = \int_{\mathcal{T}_{\mathcal{A}}+1}^{T+1} \frac{1}{\sqrt{\tau}} d\tau \\ &= 2\left(\sqrt{T+1} - \sqrt{\mathcal{T}_{\mathcal{A}}+1}\right) = \Theta(\sqrt{T}).\end{aligned} \tag{35}$$

Therefore, for any large enough T, we must have $\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \ell_t \geq \Theta(\sqrt{T}) > 2(\mathcal{T}_{\mathcal{A}}+1) = \Theta(\log(T))$.

To bound the regret of the CTBR with SGD and vanishing step size $\eta = \frac{1}{t}$, by applying Theorem 8 we only need to bound $\sum_{t\in[T]} \ell_t$. Using a similar integration argument as in Equation (35), we have $\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \frac{1}{\sqrt{t}} \leq \int_0^T \frac{1}{\sqrt{t}} = 2\sqrt{T}$. This allows us to bound the regret as followed:

$$
\begin{aligned}
\operatorname{Reg}(\boldsymbol{p}, T, \alpha, \gamma, \rho) &\overset{(i)}{\leq} \max_{k\in[K]} \left|v^k - \alpha d^k\right| \cdot \left(2\mathcal{T}_{\mathcal{A}} + 1 + C \sum_{t\in[T]} \ell_t\right) \\
&\leq \max_{k\in[K]} \left|v^k - \alpha d^k\right| \cdot \left(2\mathcal{T}_{\mathcal{A}} + 1 + 2C\sqrt{T} \cdot \sqrt{600 \log\left(T \log\left(T\right)\right) + 12}\right)
\end{aligned}
$$

where (i) follows from Theorem 8. ■

Proof for SGD with constant step size

For the high probability bound for SGD with constant step sizes, we prove a slightly more general result, as described in the following Lemma. The proof can be found in Appendix E.3.3

Lemma 5 *Assume $(v_t, d_t) \sim \boldsymbol{p}^* \in \Delta_k$ for all $t = s, s+1, \ldots$ for some starting point $s \in [T]$, and let $\{\hat{\boldsymbol{p}}_t\}_{t\in[T]}$ be the estimates for $\boldsymbol{p}^*$ generated by SGD with constant step size $\eta_t = \eta \in (0, 1/4)$ for all t. Then for any distribution $\boldsymbol{p}_s$ at the starting point, w.p. at least $1-\delta$ for some $\delta \in (0, 1/e)$, we have*

$$
\|\boldsymbol{p}^* - \hat{\boldsymbol{p}}_t\| \leq \sqrt{2 + 16\sqrt{\log\left(\frac{\log\left(T\right)}{\delta}\right)}} \left(1-2\eta\right)^{\frac{t-s}{2}} + 2\sqrt{\left(1 + 72 \log\left(\frac{\log\left(T\right)}{\delta}\right)\right)} \cdot \sqrt{\eta}.
$$

Returning to the proof for Theorem 9 with constant SGD step sizes, we can set the starting point $s = 1$ and $\delta = \frac{1}{T}$ in Lemma 5, which concludes that loss function ℓ_t can be taken as

$$
\ell_t = \underbrace{\sqrt{2 + 16\sqrt{\log\left(T \log\left(T\right)\right)}}}_{A} \left(1-2\eta\right)^{\frac{t-1}{2}} + \underbrace{2\sqrt{\left(1 + 72 \log\left(T \log\left(T\right)\right)\right)}}_{B} \sqrt{\eta}.
$$

We now show that when we take $\eta = T^{-\frac{2}{3}}$, there exists some $T_2 < \infty$ such that for all $T > T_2$ the following conditions always hold.

$$
T > \max\left\{2\mathcal{T}_{\mathcal{A}}, \frac{4\bar{w}\left(\mathcal{T}_{\mathcal{A}}+1\right)}{\sum_{k\in[J]} p^k w^k}\right\} \text{ and } \sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T} \ell_t > 2\left(\mathcal{T}_{\mathcal{A}}+1\right),
$$

where $\mathcal{T}_{\mathcal{A}} = \min\{t \in [T] : \ell_t < S\}$. It is easy to see that

$$
\mathcal{T}_{\mathcal{A}} = 1 + 2\left\lceil \frac{\log\left(S - B\sqrt{\eta}\right)}{\log(i) + \log\left(1-2\eta\right)} \right\rceil = 1 + 2\left\lceil \frac{\log(S) + \log\left(1 - \frac{B}{S}T^{-\frac{1}{3}}\right)}{\log(i) + \log\left(1 - 2T^{-\frac{2}{3}}\right)} \right\rceil. \tag{36}
$$

Since $\lim_{T\to\infty} \frac{B}{S}T^{-\frac{1}{3}} = 0$ and $\lim_{T\to\infty} = T^{-\frac{2}{3}} = 0$, we know that for large T, $\log\left(1 - \frac{B}{S}T^{-\frac{1}{3}}\right) = \Theta(BT^{-\frac{1}{3}}) = \Theta(T^{-\frac{1}{3}})$ and $\log\left(1 - 2T^{-\frac{2}{3}}\right) = \Theta(T^{-\frac{2}{3}})$, so $\mathcal{T}_{\mathcal{A}} = \Theta(T^{\frac{1}{3}})$. Therefore, for any large enough T, we must have $T > \max\left\{2\mathcal{T}_{\mathcal{A}}, \frac{4\bar{w}(\mathcal{T}_{\mathcal{A}}+1)}{\sum_{k\in[J]} p^k w^k}\right\} = \Theta(T^{\frac{1}{3}})$.

On the other hand, note that $\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T}\ell_t > B(T-\mathcal{T}_{\mathcal{A}})\sqrt{\eta} = B(T-\mathcal{T}_{\mathcal{A}})T^{-\frac{1}{3}} = \Theta(T^{\frac{2}{3}})$, where in the last equality we used the fact that $\mathcal{T}_{\mathcal{A}} = \Theta(T^{\frac{1}{3}})$. Therefore, for any large enough T, we must have $\sum_{t=\mathcal{T}_{\mathcal{A}}+1}^{T}\ell_t \geq \Theta(T^{\frac{2}{3}}) > 2\,(\mathcal{T}_{\mathcal{A}}+1) = \Theta(T^{\frac{1}{3}})$.

Finally, to bound the regret for the CTBR with SGD and constant step size $\eta = T^{-\frac{2}{3}}$, by applying Theorem 8 we only need to bound $\sum_{t\in[T]}\ell_t$:

$$\begin{aligned}
\sum_{t\in[T]}\ell_t &= \frac{A}{\sqrt{1-2\eta}}\sum_{t\in[T]}\left(\sqrt{1-2\eta}\right)^t + TB\sqrt{\eta} \;=\; \frac{A}{\sqrt{1-2\eta}}\sum_{t\in[T]}\left(\sqrt{1-2\eta}\right)^t + BT^{\frac{2}{3}} \\
&\leq \frac{A}{\sqrt{1-2\eta}}\cdot\frac{1}{1-\sqrt{1-2\eta}} + BT^{\frac{2}{3}} \;=\; \frac{A}{\sqrt{1-2\eta}}\frac{1+\sqrt{1-2\eta}}{2\eta} + BT^{\frac{2}{3}} \\
&< \frac{A}{2\eta} + BT^{\frac{2}{3}} \;=\; \left(\frac{A}{2}+B\right)T^{\frac{2}{3}}\,.
\end{aligned} \tag{37}$$

In the first equation, we recall $\ell_t = A(1-2\eta)^{\frac{t-1}{2}} + B\sqrt{\eta} = \frac{A}{\sqrt{1-2\eta}}\left(\sqrt{1-2\eta}\right)^t + B\sqrt{\eta}$. ■

E.3. Additional Proofs for Appendix E

E.3.1. Proof for Lemma 3

Proof of part (i) Recall the definitions in Algorithm 1: $\hat{r}_t = \max\left\{k\in[K] : \sum_{\ell\in[k]}\hat{p}_t^\ell w^\ell \geq -\sqrt{K}\bar{w}\ell_t\right\}$ and $\hat{b}_t = \max\left\{k\in[K] : \sum_{\ell\in[k]}\hat{p}_t^\ell d^\ell \leq \rho + \sqrt{K}\bar{d}\ell_t\right\}$. Then to show $\hat{r}_{t+1} = r$ and $\hat{b}_{t+1} = b$, it suffices to show

$$\sum_{\ell\in[r]}\hat{p}_t^\ell w^\ell \geq -\sqrt{K}\bar{w}\ell_t \quad\text{and}\quad \sum_{\ell\in[r+1]}\hat{p}_t^\ell w^\ell < -\sqrt{K}\bar{w}\ell_t \tag{38}$$

$$\sum_{\ell\in[b]}\hat{p}_t^\ell w^\ell \leq \rho + \sqrt{K}\bar{d}\ell_t \quad\text{and}\quad \sum_{\ell\in[b+1]}\hat{p}_t^\ell w^\ell > \rho + \sqrt{K}\bar{d}\ell_t\,. \tag{39}$$

We first show Equation (38). Under the event $\mathcal{G}_t = \{\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| < \ell_t\}$, we have

$$\Big|\sum_{\ell\in[r]}\hat{p}_t^\ell w^\ell - \sum_{\ell\in[r]}p^\ell w^\ell\Big| \leq \|\boldsymbol{w}\|\cdot\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \overset{(i)}{\leq} \sqrt{K}\bar{w}\ell_t \tag{40}$$

$$\Big|\sum_{\ell\in[r+1]}\hat{p}_t^\ell w^\ell - \sum_{\ell\in[r+1]}p^\ell w^\ell\Big| \leq \|\boldsymbol{w}\|\cdot\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \overset{(ii)}{\leq} \sqrt{K}\bar{w}\ell_t\,, \tag{41}$$

where both (i) and (ii) follow from the fact that event $\mathcal{G}_t$ holds. From Equation (40), we have

$$\sum_{\ell\in[r]}\hat{p}_t^\ell w^\ell \geq \sum_{\ell\in[r]}p^\ell w^\ell - \sqrt{K}\bar{w}\ell_t \overset{(i)}{\geq} -\sqrt{K}\bar{w}\ell_t\,, \tag{42}$$

where (i) follows from the fact that $\sum_{\ell\in[r]}p^\ell w^\ell \geq 0$ by the definition of $r = \max\left\{k\in[K] : \sum_{\ell\in[k]}p^\ell w^\ell \geq 0\right\}$. On the other hand, Equation (41), implies:

$$\sum_{\ell\in[r+1]}\hat{p}_t^\ell w^\ell \leq \sum_{\ell\in[r+1]}p^\ell w^\ell + \sqrt{K}\bar{w}\ell_t \overset{(i)}{\leq} -2\sqrt{K}\bar{w}\ell_t + \sqrt{K}\bar{w}\ell_t = -\sqrt{K}\bar{w}\ell_t\,,$$

where (i) holds due to the definition of $\mathcal{T}_{\mathcal{A}}$ such that for $t > \mathcal{T}_{\mathcal{A}}$, we have $\ell_t < -\frac{\sum_{\ell\in[r+1]} p^\ell w^\ell}{2\sqrt{K}\bar{w}} \Longrightarrow \sum_{\ell\in[r+1]} p^\ell w^\ell < -2\sqrt{K}\bar{w}\ell_t$. Hence we have shown Equation (38) which implies $\hat{r}_t = r$.

The proof for Equation (39) is basically identical: we first recognize

$$\Big|\sum_{\ell\in[b]} \hat{p}_t^\ell d^\ell - \sum_{\ell\in[b]} p^\ell d^\ell\Big| \leq \|\boldsymbol{d}\| \cdot \|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \leq \sqrt{K}\bar{d}\ell_t \tag{43}$$

$$\Big|\sum_{\ell\in[b+1]} \hat{p}_t^\ell d^\ell - \sum_{\ell\in[b+1]} p^\ell d^\ell\Big| \leq \|\boldsymbol{d}\| \cdot \|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \leq \sqrt{K}\bar{d}\ell_t\,. \tag{44}$$

Then Equation (43) implies

$$\sum_{\ell\in[b]} \hat{p}_t^\ell d^\ell \leq \sum_{\ell\in[b]} p^\ell d^\ell + \sqrt{K}\bar{d}\ell_t \overset{(i)}{\leq} \rho + \sqrt{K}\bar{d}\ell_t\,, \tag{45}$$

where (i) follows from $b = \max\left\{k \in [K] : \sum_{\ell\in[k]} p^\ell d^\ell \leq \rho\right\}$. On the other hand, Equation (44) implies

$$\sum_{\ell\in[b+1]} \hat{p}_t^\ell d^\ell \geq \sum_{\ell\in[b+1]} p^\ell d^\ell - \sqrt{K}\bar{d}\ell_t \overset{(i)}{\geq} \rho + 2\sqrt{K}\bar{d}\ell_t - \sqrt{K}\bar{d}\ell_t = \rho + \sqrt{K}\bar{d}\ell_t\,,$$

where (i) holds due to the definition of $\mathcal{T}_{\mathcal{A}}$ such that for $t > \mathcal{T}_{\mathcal{A}}$, we have $\ell_t < \frac{\sum_{\ell\in[b+1]} p^\ell d^\ell - \rho}{2\sqrt{K}\bar{d}} \Longrightarrow \sum_{\ell\in[b+1]} p^\ell d^\ell > \rho + 2\sqrt{K}\bar{d}\ell_t$. Hence we have shown Equation (39) which implies $\hat{b}_t = b$. Finally, we can conclude that

$$\hat{J}_t = \min\left\{\hat{b}_t, \hat{r}_t, \kappa_\alpha\right\} = \min\{r, b, \kappa_\alpha\} = J\,.$$

Proof of part (ii) Here, we want to show that $\hat{q}_t^{\mathtt{R}}$, $\hat{q}_t^{\mathtt{B}} > 0$ for any $t \geq \mathcal{T}_{\mathcal{A}}$ when event $\mathcal{G}_t$ holds. When event $\mathcal{G}_t$ holds, we have

$$\sum_{k\in[\hat{r}_t]} \hat{p}_t^k w^k \overset{(i)}{=} \sum_{k\in[r]} \hat{p}_t^k w^k \overset{(ii)}{\geq} \sum_{k\in[r]} p^k w^k - \sqrt{K}\bar{w}\ell_t \tag{46}$$

where in (i) we used the result in part (i) of the lemma, and (ii) follows from Equation (42). Hence, from the definition of $\hat{q}_{t+1}^{\mathtt{R}}$ we get

$$\hat{q}_{t+1}^{\mathtt{R}} = \frac{\sum_{k\in[\hat{r}_t]} \hat{p}_t^k w^k - \left(\sqrt{K}+2\right)\bar{w}\ell_t}{\hat{p}_t^{\hat{r}_{t+1}+1}\,|w^{\hat{r}_{t+1}+1}|} \overset{(i)}{\geq} \frac{\sum_{k\in[r]} p^k w^k - 2\left(\sqrt{K}+1\right)\bar{w}\ell_t}{\hat{p}_t^{\hat{r}_{t+1}+1}\,|w^{\hat{r}_{t+1}+1}|} \overset{(ii)}{>} 0\,,$$

Here, (i) follows from Equation (46); (ii) follows from $\ell_t < S \leq \frac{\sum_{k\in[r]} p^k w^k}{2(\sqrt{K}+1)\bar{w}}$. Recall the definition of

$$S := \frac{1}{2}\min\left\{p^{b+1}, p^{r+1}, \frac{-\sum_{k\in[r+1]} p^k w^k}{\sqrt{K}\bar{w}}, \frac{\sum_{k\in[r]} p^k w^k}{\left(\sqrt{K}+1\right)\bar{w}}, \frac{\sum_{k\in[b+1]} p^k d^k - \rho}{\sqrt{K}\bar{d}}, \frac{\rho - \sum_{k\in[b]} p^k d^k}{\left(\sqrt{K}+1\right)\bar{d}}\right\},$$

and $\mathcal{T}_{\mathcal{A}} = \min\{t \in [T] : \ell_t < S\}$. A similar argument also shows that under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$, $\hat{q}_{t+1}^{\mathtt{B}} > 0$;

Proof of part (iii) We first show the following result w.r.t. the ROI constraint for $t > \mathcal{T}_{\mathcal{A}}$ under event $\mathcal{G}_t$:

$$0 < \bar{w}\ell_t \leq \sum_{k\in[K]} \hat{x}_t^{\mathtt{R},k} p^k w^k \leq \left(3\sqrt{K}+5\right)\bar{w}\ell_t\,. \tag{47}$$

The lower bound holds because

$$\begin{aligned}\sum_{k\in[K]} \hat{x}_t^{\mathtt{R},k} p^k w^k &= \sum_{k\in[\hat{r}_t]} p^k w^k + \left(\hat{q}_t^{\mathtt{R}}\right)_+ p^{\hat{r}_t+1} w^{\hat{r}_t+1} \overset{(i)}{=} \sum_{k\in[r]} p^k w^k + \left(\hat{q}_t^{\mathtt{R}}\right)_+ p^{r+1} w^{r+1}\\ &\overset{(ii)}{=} \sum_{k\in[r]} p^k w^k + \hat{q}_t^{\mathtt{R}} p^{r+1} w^{r+1}\,,\end{aligned} \tag{48}$$

where in (i) we evoked Lemma 3 (i) such that for $t > \mathcal{T}_{\mathcal{A}}$, $\hat{r}_t = r$, and (ii) follows from Lemma 3 (ii) such that $\hat{q}_t^{\mathtt{R}} > 0$ for all $t > \mathcal{T}_{\mathcal{A}}$. Now, recalling the definition

$$\hat{q}_t^{\mathtt{R}} = \frac{\sum_{k\in[r]} \hat{p}_t^k w^k - (\sqrt{K}+2)\bar{w}\ell_t}{\hat{p}_t^{r+1}|w^{r+1}|} \overset{(i)}{=} -\frac{\sum_{k\in[r]} \hat{p}_t^k w^k - (\sqrt{K}+2)\bar{w}\ell_t}{\hat{p}_t^{r+1} w^{r+1}}\,,$$

where in (i) we used the fact that $r = \hat{r}$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$, so $w^{r+1} = w^{\hat{r}_t+1} < 0$ where the inequality follows from Remark 2. Hence we have

$$\begin{aligned}\hat{q}_t^{\mathtt{R}} p^{r+1} w^{r+1} &= \hat{q}_t^{\mathtt{R}} \hat{p}_t^{r+1} w^{r+1} + \hat{q}_t^{\mathtt{R}}\left(p^{r+1} - \hat{p}_t^{r+1}\right) w^{r+1}\\ \overset{(i)}{=} &-\left(\sum_{k\in[r]} \hat{p}_t^k w^k\right) + \left(\sqrt{K}\bar{w}\ell_t + 2\bar{w}\ell_t\right) + \hat{q}_t^{\mathtt{R}}\left(p^{r+1} - \hat{p}_t^{r+1}\right) w^{r+1}\\ \overset{(ii)}{\geq} &-\left(\sum_{k\in[r]} \hat{p}_t^k w^k\right) + \left(\sqrt{K}\bar{w}\ell_t + 2\bar{w}\ell_t\right) - \bar{w}\ell_t \overset{(iii)}{\geq} -\left(\sum_{k\in[r]} p^k w^k + \sqrt{K}\bar{w}\ell_t\right) + \sqrt{K}\bar{w}\ell_t + \bar{w}\ell_t\\ = &-\left(\sum_{k\in[r]} p^k w^k\right) + \bar{w}\ell_t\,.\end{aligned}$$

Equality (i) follows from the definition of $\hat{q}_t^{\mathtt{R}} (> 0)$. Both (ii) and (iii) follow from $\|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \leq \ell_t$ under the event $\mathcal{G}_t$ which implies $\sum_{k\in[r]}\left(p^k - \hat{p}_t^k\right) w^k \leq \|\boldsymbol{w}\| \cdot \|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \leq \sqrt{K}\bar{w}\ell_t$ by Cauchy-schwarz, and $p^{r+1} - \hat{p}_t^{r+1} \leq \|\hat{\boldsymbol{p}}_t - \boldsymbol{p}\| \leq \ell_t$. Plugging this back into Equation (48) yields our desired inequality in Equation (47) that lower bounds the single period ROI balance. To show the upper bound, first note that

$$\sum_{k\in[K]} \hat{x}_t^{\mathtt{R},k} p^k w^k \overset{(i)}{=} \sum_{k\in[r]} p^k w^k + \hat{q}_t^{\mathtt{R}} p^{r+1} w^{r+1} \overset{(ii)}{=} \left(\hat{q}_t^{\mathtt{R}} - q^{\mathtt{R}}\right) p^{r+1} w^{r+1}\,,$$

where (i) follows from Equation (48) and (ii) follows from the definition of $q^{\mathtt{R}}$ such that $p^{r+1}w^{r+1}q^{\mathtt{R}} + \sum_{k\in[r]} p^k w^k = 0$. In the following we show $|(\hat{q}_t^{\mathtt{R}} - q^{\mathtt{R}})\, p^{r+1}w^{r+1}| \leq \left(3\sqrt{K}+5\right)\bar{w}\ell_t$. Consider the following

$$
\begin{aligned}
&\left|(\hat{q}_t^{\mathtt{R}} - q^{\mathtt{R}})\, p^{r+1}w^{r+1}\right| \\
=\;& \left|p^{r+1}\Big(\frac{\sum_{k\in[r]}\hat{p}_t^k w^k - \left(\sqrt{K}+2\right)\bar{w}\ell_t}{\hat{p}_t^{r+1}} - \frac{\sum_{k\in[r]} p^k w^k}{p^{r+1}}\Big)\right| \\
=\;& \left|\Big(\frac{p^{r+1}}{\hat{p}_t^{r+1}} - 1\Big)\cdot\Big(\sum_{k\in[r]}\hat{p}_t^k w^k\Big) + \sum_{k\in[r]}\hat{p}_t^k w^k - \sum_{k\in[r]} p_t^k w^k\right| + \left(\sqrt{K}+2\right)\bar{w}\ell_t\cdot\frac{p^{r+1}}{\hat{p}_t^{r+1}} \\
\leq\;& \left|p^{r+1} - \hat{p}_t^{r+1}\right|\cdot\frac{\left|\sum_{k\in[r]}\hat{p}_t^k w^k\right|}{\hat{p}_t^{r+1}} + \left|\sum_{k\in[r]}\left(\hat{p}_t^k - p_t^k\right) w^k\right| + \left(\sqrt{K}+2\right)\bar{w}\ell_t\cdot\frac{p^{r+1}}{\hat{p}_t^{r+1}} \\
\overset{(i)}{\leq}\;& \ell_t\cdot\frac{\left|\sum_{k\in[r]}\hat{p}_t^k w^k\right|}{\hat{p}_t^{r+1}} + \sqrt{K}\bar{w}\ell_t + \left(\sqrt{K}+2\right)\bar{w}\ell_t\cdot\frac{p^{r+1}}{\hat{p}_t^{r+1}} \\
\overset{(ii)}{\leq}\;& \bar{w}\ell_t + \sqrt{K}\bar{w}\ell_t + \left(\sqrt{K}+2\right)\bar{w}\ell_t\cdot\frac{p^{r+1}}{\hat{p}_t^{r+1}} \overset{(iii)}{\leq} \left(3\sqrt{K}+5\right)\bar{w}\ell_t\,.
\end{aligned}
\tag{49}
$$

Here, in (i) we utilized the fact that under event $\mathcal{G}_t$, $\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$, so $\left|p^{r+1} - \hat{p}_t^{r+1}\right| \leq \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$, and $\left|\sum_{k\in[r]}\left(\hat{p}_t^k - p_t^k\right) w^k\right| \leq \|\boldsymbol{w}\|\cdot\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \sqrt{K}\bar{w}\ell_t$. In (ii) we used the fact that $\sum_{k\in[\hat{r}_t+1]}\hat{p}_t^k w^k < 0$ according to Algorithm 1, so

$$
\sum_{k\in[\hat{r}_t+1]}\hat{p}_t^k w^k < -\hat{p}^{\hat{r}_t+1}w^{\hat{r}_t+1} \Longrightarrow \frac{\sum_{k\in[r+1]}\hat{p}_t^k w^k}{\hat{p}^{r+1}} \overset{(iv)}{=} \frac{\sum_{k\in[\hat{r}_t+1]}\hat{p}_t^k w^k}{\hat{p}^{\hat{r}_t+1}} < -w^{\hat{r}_t+1} \leq \bar{w}\,,
$$

where the equality in (iv) follows from $r = \hat{r}_t$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$. Finally, (iii) in Equation (49) holds because for $t > \mathcal{T}_{\mathcal{A}}$ we have $\ell_t \leq S \leq \frac{p^{r+1}}{2}$ and hence $\hat{p}_t^{r+1} \geq p^{r+1} - \ell_t \geq \frac{p^{r+1}}{2}$.

We now turn to show the following upper and lower bounds w.r.t. the budget constraint in a similar fashion:

$$
0 < \bar{d}\ell_t \leq \rho - \sum_{k\in[K]}\hat{x}_t^{\mathtt{B},k}p^k d^k \leq \left(3\sqrt{K}+5\right)\bar{d}\ell_t\,.
$$

For the lower bound, we start off with the relation

$$
\begin{aligned}
\rho - \sum_{k\in[K]}\hat{x}_t^{\mathtt{B},k}p^k d^k &\overset{(i)}{=} \sum_{k\in[b+1]} x^{\mathtt{B},k}p^k d^k - \sum_{k\in[\hat{b}_t+1]}\hat{x}_t^{\mathtt{B},k}p^k d^k \overset{(ii)}{=} \left(q^{\mathtt{B}+1} - \left(\hat{q}_t^{\mathtt{B}+1}\right)_+\right)p^{b+1}d^{b+1} \\
&\overset{(iii)}{=} \left(q^{\mathtt{B}+1} - \hat{q}_t^{\mathtt{B}+1}\right)p^{b+1}d^{b+1}\,,
\end{aligned}
\tag{50}
$$

where (i) follows from the definition of $\boldsymbol{x}^{\mathtt{B}}$; (ii) follows from the fact that $b = \hat{b}_t$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$, and (iii) follows from $\hat{q}_t^{\mathtt{B}+1} > 0$ for $t > \mathcal{T}_{\mathcal{A}}$ according to Lemma 3 (ii). Then, we have

$$
\begin{aligned}
\left(q^{\mathtt{B}+1} - \hat{q}_t^{\mathtt{B}+1}\right)p^{b+1}d^{b+1} &\overset{(i)}{=} \rho - \sum_{k\in[b]} p^k d^k - \hat{q}_t^{\mathtt{B}+1}p^{b+1}d^{b+1} \\
&= \Big(\rho - \sum_{k\in[b]} p^k d^k\Big) - \hat{q}_t^{\mathtt{B}+1}\hat{p}_t^{b+1}d^{b+1} + \hat{q}_t^{\mathtt{B}+1}\left(\hat{p}_t^{b+1} - p^{b+1}\right)d^{b+1}
\end{aligned}
$$

$$\stackrel{(ii)}{=} \left(\rho - \sum_{k\in[b]} p^k d^k\right) - \left(\rho - \sum_{k\in[b]} \hat{p}_t^k d^k - \left(\sqrt{K}+2\right)\bar{d}\ell_t\right) + \hat{q}_t^{\mathtt{B}+1}\left(\hat{p}_t^{b+1} - p^{b+1}\right) d^{b+1}$$

$$= \left(\sqrt{K}+2\right)\bar{d}\ell_t + \sum_{k\in[b]} \left(\hat{p}_t^k - p^k\right) d^k + \hat{q}_t^{\mathtt{B}+1}\left(\hat{p}_t^{b+1} - p^{b+1}\right) d^{b+1}$$

$$\stackrel{(iii)}{\geq} \left(\sqrt{K}+2\right)\bar{d}\ell_t - \sqrt{K}\bar{d}\ell_t - \bar{d}\ell_t \;=\; \bar{d}\ell_t\,.$$

Here (i) follows from the definition of $q^{\mathtt{B}+1}$ such that $\sum_{k\in[b]} p^k d^k + q^{\mathtt{B}+1}p^{b+1}d^{b+1} = \rho$; (ii) follows from the definition of $\hat{q}_t^{\mathtt{B}+1}$ and the fact that $b = \hat{b}_t$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$; (iii) follows from $|p^{b+1} - \hat{p}^{b+1}| \leq \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$ and $\left|\sum_{k\in[b]} \left(\hat{p}_t^k - p^k\right) d^k\right| \leq \sqrt{K}\bar{d}\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \bar{d}\ell_t$. Combining this with Equation (50) yields the desire lower bound.

On the other hand for the upper bound w.r.t. the budget constraint, we start from Equation (50),

$$\begin{aligned}
\left|\rho - \sum_{k\in[K]} \hat{x}_t^{\mathtt{B},k} p^k d^k\right| &= \left|\left(q^{\mathtt{B}+1} - \hat{q}^{\mathtt{B}+1}\right)\right| p^{b+1} d^{b+1} \\
&= \left|p^{b+1}\left(\frac{\rho - \sum_{k\in[b]} \hat{p}_t^k d^k - \left(\sqrt{K}+2\right)\bar{d}\ell_t}{\hat{p}^{b+1}} - \frac{\rho - \sum_{k\in[r]} p^k d^k}{p^{b+1}}\right)\right| \\
&= \left|\left(\frac{p^{b+1}}{\hat{p}^{b+1}} - 1\right)\cdot\left(\rho - \sum_{k\in[b]} \hat{p}_t^k d^k\right) - \sum_{k\in[b]} \hat{p}_t^k d^k + \sum_{k\in[b]} p_t^k d^k\right| + \left(\sqrt{K}+2\right)\bar{d}\ell_t\cdot\frac{p^{b+1}}{\hat{p}^{b+1}} \\
&\leq \left|p^{b+1} - \hat{p}^{b+1}\right|\cdot\frac{\left|\rho - \sum_{k\in[b]} \hat{p}_t^k d^k\right|}{\hat{p}^{b+1}} + \left|\sum_{k\in[b]} \left(\hat{p}_t^k - p_t^k\right) d^k\right| + \left(\sqrt{K}+2\right)\bar{d}\ell_t\cdot\frac{p^{b+1}}{\hat{p}^{b+1}} \\
&\stackrel{(i)}{\leq} \ell_t\cdot\frac{\left|\rho - \sum_{k\in[b]} \hat{p}_t^k d^k\right|}{\hat{p}^{b+1}} + \sqrt{K}\bar{d}\ell_t + \left(\sqrt{K}+2\right)\bar{d}\ell_t\cdot\frac{p^{b+1}}{\hat{p}^{b+1}} \\
&\stackrel{(ii)}{\leq} \bar{d}\ell_t + \sqrt{K}\bar{d}\ell_t + \left(\sqrt{K}+2\right)\bar{d}\ell_t\cdot\frac{p^{b+1}}{\hat{p}^{b+1}} \\
&\stackrel{(iii)}{\leq} \left(3\sqrt{K}+5\right)\bar{d}\ell_t\,,
\end{aligned} \tag{51}$$

Here, in (i) we utilized the fact that under event $\mathcal{G}_t$, $\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$, so $|p^{b+1} - \hat{p}^{b+1}| \leq \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \ell_t$, and $\left|\sum_{k\in[b]} \left(\hat{p}_t^k - p_t^k\right) d^k\right| \leq \|\boldsymbol{d}\|\cdot\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \leq \sqrt{K}\bar{d}\ell_t$. In (ii) we used the fact that $\sum_{k\in[\hat{b}_t]} \hat{p}_t^k d^k \leq \rho < \sum_{k\in[\hat{b}_t+1]} \hat{p}_t^k d^k$ according to Algorithm 1, so

$$0 \leq \rho - \sum_{k\in[\hat{b}_t]} \hat{p}_t^k d^k < \hat{p}^{\hat{b}_t+1} d^{\hat{b}_t+1} \Longrightarrow 0 \leq \frac{\rho - \sum_{k\in[b+1]} \hat{p}_t^k d^k}{\hat{p}^{b+1}} \stackrel{(iv)}{=} \frac{\rho - \sum_{k\in[\hat{b}_t+1]} \hat{p}_t^k d^k}{\hat{p}^{\hat{b}_t+1}} < d^{\hat{b}_t+1} \leq \bar{d}\,,$$

where the equality in (iv) follows from $b = \hat{b}_t$ under event $\mathcal{G}_t$ for $t > \mathcal{T}_{\mathcal{A}}$. Finally, (iii) in Equation (51) holds because for $t > \mathcal{T}_{\mathcal{A}}$ we have $\ell_t \leq S \leq \frac{p^{b+1}}{2}$, and hence $\hat{p}^{b+1} \geq p^{b+1} - \ell_t \geq \frac{p^{b+1}}{2}$.

Proof for Lemma 3 (iv) In Equation (49) within the proof of (iii),we showed

$$\left|\left(\hat{q}_t^{\mathtt{R}} - q^{\mathtt{R}}\right) p^{r+1} w^{r+1}\right| \leq \left(3\sqrt{K}+5\right)\bar{w}\ell_t\,.$$

Hence using the fact that $|w^{r+1}| \geq \underline{w}$, we get $|(\hat{q}_t^{\mathrm{R}} - q^{\mathrm{R}})\, p^{r+1}| \leq \left(3\sqrt{K}+5\right) \frac{\bar{w}}{\underline{w}} \ell_t$. Similarly, in Equation (51) within the proof of (iii), we showed $|(\hat{q}_t^{\mathrm{B}} - q^{\mathrm{B}})\, p^{b+1} d^{b+1}| \leq \left(3\sqrt{K}+5\right) \bar{d}\ell_t$, so using the fact that $|d^{r+1}| \geq \underline{d}$, we get $|(\hat{q}_t^{\mathrm{B}} - q^{\mathrm{B}})\, p^{b+1}| \leq \left(3\sqrt{K}+5\right) \frac{\bar{d}}{\underline{d}} \ell_t$. ■

E.3.2. Proof for Lemma 4 Recall the updates step of SGD: $\hat{\boldsymbol{p}}_{t+1} = \arg\min_{\widetilde{\boldsymbol{p}} \in \Delta_K} \|\widetilde{\boldsymbol{p}} - (\hat{\boldsymbol{p}}_t - \eta_t (\hat{\boldsymbol{p}}_t - \boldsymbol{s}_t))\|$. Then, the contraction property for projections imply

$$\|\boldsymbol{p} - \hat{\boldsymbol{p}}_{t+1}\|^2 \leq \|\boldsymbol{p} - (\hat{\boldsymbol{p}}_t - \eta_t (\hat{\boldsymbol{p}}_t - \boldsymbol{s}_t))\|^2 = \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 + 2\eta_t \langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \hat{\boldsymbol{p}}_t - \boldsymbol{s}_t \rangle + \eta_t^2 \|\hat{\boldsymbol{p}}_t - \boldsymbol{s}_t\|^2 .$$

Hence, we have

$$\begin{aligned} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_{t+1}\|^2 &\leq \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 + 2\eta_t \langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \hat{\boldsymbol{p}}_t - \boldsymbol{s}_t \rangle + \eta_t^2 \|\hat{\boldsymbol{p}}_t - \boldsymbol{s}_t\|^2 \\ &\overset{(i)}{\leq} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 + 2\eta_t \langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \hat{\boldsymbol{p}}_t - \boldsymbol{s}_t \rangle + 4\eta_t^2 = (1 - 2\eta_t)\, \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 + 2\eta_t \underbrace{\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \boldsymbol{p} - \boldsymbol{s}_t \rangle}_{:= Z_t} + 4\eta_t^2 \\ &= \left(1 - \frac{2}{t}\right) \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 + \frac{2}{t} Z_t + \frac{4}{t^2}\,, \end{aligned} \tag{52}$$

where in (i) we used the fact that $\|\hat{\boldsymbol{p}}_t - \boldsymbol{s}_t\| \leq \|\hat{\boldsymbol{p}}_t\| + \|\boldsymbol{s}_t\| \leq 2$.

Now, telescoping the above recursive inequality until $t = 2$ we get

$$\begin{aligned} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_{t+1}\|^2 &\leq 2\sum_{\tau=2}^{t} \frac{1}{\tau} \left(\prod_{j=\tau+1}^{t} \left(1 - \frac{2}{j}\right) \right) Z_\tau + 4 \sum_{\tau=2}^{t} \frac{1}{\tau^2} \left(\prod_{j=\tau+1}^{t} \left(1 - \frac{2}{j}\right) \right) \\ &\overset{(i)}{=} 2\sum_{\tau=2}^{t} \frac{1}{\tau} \left(\frac{\tau(\tau-1)}{t(t-1)} \right) Z_\tau + 4 \sum_{\tau=2}^{t} \frac{1}{\tau^2} \left(\frac{\tau(\tau-1)}{t(t-1)} \right) \\ &\overset{(ii)}{\leq} \frac{2}{t(t-1)} \sum_{\tau=2}^{t} (\tau - 1)\, Z_\tau + \frac{4}{t}\,. \end{aligned} \tag{53}$$

Here, (i) and (ii) follow from:

$$\prod_{j=\tau+1}^{t} \left(1 - \frac{2}{j}\right) = \prod_{j=\tau+1}^{t} \frac{j-2}{j} = \frac{\tau(\tau-1)}{t(t-1)}, \text{ and } \sum_{\tau=2}^{t} \frac{1}{\tau^2} \left(\frac{\tau(\tau-1)}{t(t-1)} \right) = \sum_{\tau=2}^{t} \frac{\tau - 1}{t(t-1)\tau} < \frac{1}{t}\,.$$

Then it is easy to see $\mathbb{E}\left[(t-1)\, Z_t \middle| \mathcal{F}_{t-1}\right] = (t-1)\, \mathbb{E}\left[\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \boldsymbol{p} - \boldsymbol{s}_t \rangle \middle| \mathcal{F}_{t-1}\right] = 0$ for all $t \in [T]$, so $\{(\tau - 1)\, Z_\tau\}_\tau$ is a martingale difference sequence w.r.t. the filtration $\{\mathcal{F}_t\}_t$. Furthermore, $(\tau - 1)\, Z_\tau$ for all $\tau \leq t$ is bounded uniformly by

$$|(\tau-1)\, Z_\tau| = (\tau - 1)\, |\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_\tau, \boldsymbol{p} - \boldsymbol{s}_\tau \rangle| \leq (\tau - 1)\, \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \cdot \|\boldsymbol{p} - \boldsymbol{s}_\tau\| \leq 2\,(\tau-1) \leq 2\,(t-1)$$

and the conditional variance of $2\,(\tau - 1)\, Z_\tau$ is bounded as followed w.p. 1:

$$\mathrm{Var}\left((\tau-1)\, Z_\tau \middle| \mathcal{F}_{\tau-1}\right) = (\tau-1)^2\, \mathbb{E}\left[\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_\tau, \boldsymbol{p} - \boldsymbol{s}_\tau \rangle^2 \middle| \mathcal{F}_{\tau-1}\right] \overset{(i)}{\leq} 4\,(\tau-1)^2\, \|\boldsymbol{p} - \hat{\boldsymbol{p}}_\tau\|^2\,,$$

where in the final inequality we used Cauchy-Schwarz and $\|\boldsymbol{p}-\hat{\boldsymbol{s}}_t\| \le 2$. Hence, using Lemma 7, w.p. at least $1-\delta$ for any $\delta \in (0,1/e)$ the following holds for any $t = 2 \dots T$:

$$\begin{aligned}\sum_{\tau=2}^{t} (\tau-1) Z_\tau &\le \sqrt{\log\left(\frac{\log(T)}{\delta}\right)} \max\left\{8\sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2}, 4(t-1)\sqrt{\log\left(\frac{\log(T)}{\delta}\right)}\right\} \\ &\le \sqrt{\log\left(\frac{\log(T)}{\delta}\right)} \left(8\sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2} + 4(t-1)\sqrt{\log\left(\frac{\log(T)}{\delta}\right)}\right).\end{aligned}$$

Plugging this back into Equation (53), w.p. at least $1-\delta$ we have

$$\begin{aligned}\|\boldsymbol{p}-\hat{\boldsymbol{p}}_{t+1}\|^2 &\le \frac{2}{t(t-1)}\sqrt{\log\left(\frac{\log(T)}{\delta}\right)} \left(8\sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2} + 4(t-1)\sqrt{\log\left(\frac{\log(T)}{\delta}\right)}\right) + \frac{4}{t} \\ &= \underbrace{\frac{16\sqrt{\log\left(\frac{\log(T)}{\delta}\right)}}{t(t-1)}}_{:=\frac{b}{t(t-1)}} \sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2} + \underbrace{\frac{8\log\left(\frac{\log(T)}{\delta}\right)+4}{t}}_{:=\frac{c}{t}}. \qquad (54)\end{aligned}$$

The remaining is an induction argument, where we find some constant $a > 0$ independent of t such that $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\|^2 \le \frac{a}{t}$ for all $t \in [T]$ (induction hypothesis). [20] Equation (54) and the induction hypothesis imply

$$\begin{aligned}&\frac{b}{t(t-1)}\sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2} + \frac{c}{t} \le \frac{b}{t(t-1)}\sqrt{\sum_{\tau=2}^{t} (\tau-1)^2 \frac{a}{\tau}} + \frac{c}{t} \\ &\le \frac{b}{t(t-1)}\sqrt{a\sum_{\tau=2}^{t} (\tau-1)} + \frac{c}{t} = \frac{b}{\sqrt{2t(t-1)}}\sqrt{a} + \frac{c}{t} = \frac{1}{t+1}\left(\frac{b(t+1)}{\sqrt{2t(t-1)}}\sqrt{a} + \frac{c(t+1)}{t}\right) \\ &\le \frac{1}{t+1}\cdot\frac{3}{2}\left(b\sqrt{a}+c\right).\end{aligned}$$

where in the last inequality we used that fact that for all $t \ge 2$, $\frac{t+1}{\sqrt{2t(t-1)}} \le \frac{3}{2}$ and $\frac{t+1}{t} \le \frac{3}{2}$. Hence, it suffices to have $a > 0$ such that $\frac{3}{2}(b\sqrt{a}+c) \le a \Longrightarrow \sqrt{a} \ge \frac{1}{2}\left(\frac{3b}{2} + \sqrt{\frac{9b^2}{4}+6c}\right)$. Using the basic inequality $2(x^2+y^2) \ge (x+y)^2$, we can take

$$a = 600\log\left(\frac{\log(T)}{\delta}\right) + 12 = \frac{9b^2}{4} + 3c \ge \frac{1}{2}\left(\frac{3b}{2} + \sqrt{\frac{9b^2}{4}+6c}\right)^2$$

This concludes that the loss function ℓ_t can be taken as $\ell_t = \sqrt{\frac{600\log\left(\frac{\log(T)}{\delta}\right)+12}{t}}$. ■

E.3.3. Proof for Lemma 5 Following the same proof as in Equation (52), we have

$$\begin{aligned}\|\boldsymbol{p}^*-\hat{\boldsymbol{p}}_{t+1}\|^2 &\le (1-2\eta_t)\|\boldsymbol{p}^*-\hat{\boldsymbol{p}}_t\|^2 + 2\eta_t \underbrace{\langle \boldsymbol{p}^*-\hat{\boldsymbol{p}}_t, \boldsymbol{p}-\boldsymbol{s}_t\rangle}_{:=Z_t} + 4\eta_t^2 \\ &= (1-2\eta)\|\boldsymbol{p}^*-\hat{\boldsymbol{p}}_t\|^2 + 2\eta Z_t + 4\eta^2.\end{aligned}$$

[20] Note that the variable b here is local to this lemma and different from the budget index.

Now, telescoping the above recursive inequality until the starting point $s \in [T]$, we get

$$\begin{aligned} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_{t+1}\|^2 &\leq (1-2\eta)^{t+1-s} \|\boldsymbol{p}^* - \hat{\boldsymbol{p}}_s\|^2 + 2\eta \sum_{\tau=s}^{t} (1-2\eta)^{t-\tau} Z_\tau + 4\eta^2 \sum_{\tau=s}^{t} (1-2\eta)^{t-\tau} \\ &\overset{(i)}{\leq} 2(1-2\eta)^{t+1-s} + 2\eta (1-2\eta)^t \sum_{\tau=s}^{t} (1-2\eta)^{-\tau} Z_\tau + 2\eta \,. \end{aligned} \tag{55}$$

In (i) we used the fact that $\sum_{\tau=s}^{t} (1-2\eta)^{t-\tau} \leq \sum_{\tau=-\infty}^{t} (1-2\eta)^{t-\tau} = \frac{1}{2\eta}$. We now describe a high probability bound for the summand $\sum_{\tau=s}^{t} (1-2\eta)^{-\tau} Z_\tau$. It is easy to see $\mathbb{E}\left[(1-2\eta)^{-\tau} Z_\tau \middle| \mathcal{F}_{\tau-1}\right] = (1-2\eta)^{-\tau} \mathbb{E}\left[\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_\tau, \boldsymbol{p} - \boldsymbol{s}_\tau \rangle \middle| \mathcal{F}_{t-1}\right] = 0$ so $\left\{(1-2\eta)^{-\tau} Z_\tau\right\}_\tau$ is a martingale difference sequence w.r.t. the filtration $\{\mathcal{F}_t\}_t$. Furthermore, $(1-2\eta)^{-\tau} Z_\tau$ is bounded uniformly by

$$\begin{aligned} \left|(1-2\eta)^{-\tau} Z_\tau\right| &= (1-2\eta)^{-\tau} \left|\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_t, \boldsymbol{p} - \boldsymbol{s}_t \rangle\right| \leq (1-2\eta)^{-\tau} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\| \cdot \|\boldsymbol{p} - \boldsymbol{s}_t\| \\ &\leq 2(1-2\eta)^{-\tau} \leq 2(1-2\eta)^{-t} \,. \end{aligned}$$

The conditional variance of $(1-2\eta)^{-\tau} Z_\tau$ is bounded as followed w.p. 1:

$$\mathrm{Var}\left((1-2\eta)^{-\tau} Z_\tau \middle| \mathcal{F}_{\tau-1}\right) = (1-2\eta)^{-2\tau} \mathbb{E}\left[\langle \boldsymbol{p} - \hat{\boldsymbol{p}}_\tau, \boldsymbol{p} - \boldsymbol{s}_\tau \rangle^2 \middle| \mathcal{F}_{\tau-1}\right] \overset{(i)}{\leq} 2(1-2\eta)^{-2\tau} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_\tau\|^2 \,,$$

where in the final inequality we used Cauchy-Schwarz and $\|\boldsymbol{p} - \hat{\boldsymbol{s}}_t\|^2 \leq 2$. Hence, using Lemma 7, w.p. at least $1-\delta$ for some $\delta \in (0, 1/e)$ the following holds:

$$\begin{aligned} \sum_{\tau=s}^{t} (1-2\eta)^{-\tau} Z_\tau &\leq \sqrt{\log\left(\frac{\log(T)}{\delta}\right)} \max\left\{8\sqrt{\sum_{\tau=s}^{t} (1-2\eta)^{-2\tau} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_\tau\|^2}, 4(1-2\eta)^{-t} \sqrt{\log\left(\frac{\log(T)}{\delta}\right)}\right\} \\ &\leq \sqrt{\log\left(\frac{\log(T)}{\delta}\right)} \left(8\sqrt{\sum_{\tau=s}^{t} (1-2\eta)^{-2\tau} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_\tau\|^2} + 4(1-2\eta)^{-t} \sqrt{\log\left(\frac{\log(T)}{\delta}\right)}\right) . \end{aligned}$$

Plugging this back into Equation (55) and denoting $c = \sqrt{\log\left(\frac{\log(T)}{\delta}\right)}$, w.p. at least $1-\delta$ we have

$$\|\boldsymbol{p} - \hat{\boldsymbol{p}}_{t+1}\|^2 \leq 2(1-2\eta)^{t+1-s} + 16\eta c (1-2\eta)^t \sqrt{\sum_{\tau=s}^{t} (1-2\eta)^{-2\tau} \|\boldsymbol{p} - \hat{\boldsymbol{p}}_\tau\|^2} + \left(2 + 8c^2\right)\eta \,. \tag{56}$$

The remaining is again an induction argument, where we find some constants $a, b > 0$ [21] independent of t (but possibly dependent on T) such that

$$\|\boldsymbol{p} - \hat{\boldsymbol{p}}_t\|^2 \leq a(1-2\eta)^{t+1-s} + b\eta \quad \forall t = s, s+1 \ldots .$$

[21] Note that the variable b here is local to this lemma and different from the budget index.

We do so by considering the induction step, i.e. $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_{t'}\|^2 \le a\,(1-2\eta)^{t'-s} + b\eta$ for all $t' = s, s+1 \dots t$, and we aim to show $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_{t+1}\|^2 \le a\,(1-2\eta)^{t+1-s} + b\eta$. Using the induction hypothesis we have

$$\sum_{\tau=s}^{t} (1-2\eta)^{-2\tau} \|\boldsymbol{p}-\hat{\boldsymbol{p}}_\tau\|^2 \;\le\; \sum_{\tau=s}^{t} (1-2\eta)^{-2\tau}\left(a\,(1-2\eta)^{\tau-s} + b\eta\right)$$
$$=\; a\,(1-2\eta)^{-s}\sum_{\tau=s}^{t}(1-2\eta)^{-\tau} + b\eta\sum_{\tau=s}^{t}(1-2\eta)^{-2\tau} \;\le\; \frac{a}{2\eta}(1-2\eta)^{-s-t} + \frac{b}{2}(1-2\eta)^{-2t}\,.$$

In the last inequality, we used the following

$$\sum_{\tau=s}^{t}(1-2\eta)^{-\tau} \;=\; \frac{1}{2\eta}\left((1-2\eta)^{-t} - (1-2\eta)^{-(s-1)}\right) \;\le\; \frac{1}{2\eta}(1-2\eta)^{-t}$$
$$\sum_{\tau=s}^{t}(1-2\eta)^{-2\tau} \;=\; \frac{1}{1-(1-2\eta)^2}\left((1-2\eta)^{-2t} - (1-2\eta)^{-2(s-1)}\right) \;\le\; \frac{1}{4\eta-4\eta^2}(1-2\eta)^{-2t} \le \frac{1}{2\eta}(1-2\eta)^{-2t}\,.$$

where we used the fact that $4\eta - 4\eta^2 \ge 2\eta$ for any $\eta \le \frac{1}{2}$. Hence, combining this with Equation (56), we get

$$\begin{aligned}
\|\boldsymbol{p}-\hat{\boldsymbol{p}}_{t+1}\|^2 \;&\le\; 2\,(1-2\eta)^{t+1-s} + 16\eta c\sqrt{a\,(1-2\eta)^{t-s}\cdot\frac{1}{2\eta} + \frac{b}{2}} + \left(2+8c^2\right)\eta \\
&\overset{(i)}{\le}\; 2\,(1-2\eta)^{t+1-s} + 16c\sqrt{(1-2\eta)^{t-s}\cdot\frac{a\eta}{2}} + \left(2+8c^2+16c\sqrt{b}\right)\eta \\
&\overset{(ii)}{\le}\; \left(2+\frac{8c}{1-2\eta}\right)(1-2\eta)^{t+1-s} + \left(2+8c^2+16c\sqrt{b}+4ac\right)\eta \\
&\overset{(iii)}{\le}\; (2+16c)\,(1-2\eta)^{t+1-s} + \left(2+8c^2+16c\sqrt{b}+4ac\right)\eta\,,
\end{aligned} \tag{57}$$

where in (i) we used the inequality $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$; in (ii) we used the inequality $\sqrt{xy} \le \frac{x+y}{2}$; in (iii) we recall that $\eta < \frac{1}{4}$.

Now, if we take $a := 2+16c = 2+16\sqrt{\log\left(\frac{\log(T)}{\delta}\right)}$, and b such that

$$b \ge 2+8c^2+4ac+16c\sqrt{b} \Longrightarrow \sqrt{b} \ge \sqrt{2+8c^2+4ac} + 8c\,,$$

then following Equation (57), we can conclude

$$\|\boldsymbol{p}-\hat{\boldsymbol{p}}_{t+1}\|^2 \le (2+16c)\,(1-2\eta)^{t+1-s} + b\eta = a\,(1-2\eta)^{t+1-s} + b\eta\,.$$

To find such a constant b, we use the basic inequality $(x+y)^2 \le 2\,(x^2+y^2)$, and thus can take

$$\begin{aligned}
b := 4\left(1+72\log\left(\frac{\log(T)}{\delta}\right)\right) &> 4\left(1+4c+68c^2\right) \;=\; 2\left(2+8c^2+4ac+64c^2\right) \\
&\ge \left(\sqrt{2+8c^2+4ac}+8c\right)^2\,.
\end{aligned}$$

In the equality, we used the definition $a = 2+16c$. We conclude the proof by realizing for the a, b that we chose above, $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\|^2 \le a\,(1-2\eta)^{t-s} + b\eta$ holds for the base case $t = s$ because trivially $\|\boldsymbol{p}-\hat{\boldsymbol{p}}_t\|^2 \le 2 < a + b\eta$. Finally, using the inequality $\sqrt{x+y} \le \sqrt{x}+\sqrt{y}$ concludes the proof. ■

F. Proofs for Section 6

F.1. Proof of Theorem 5

Our proof relies on the following fact

Fact 1 *If price $d \in \mathcal{D}$ is nonbinding, then the corresponding optimal solution $\boldsymbol{x}_d$ to $U(d)$ is $\boldsymbol{x}_d = (1\dots1) \in \mathbb{R}^n_+$.*

Proof. We prove the claim via contradiction. Assume there is some index $k \in [N]$ such that $\boldsymbol{x}_d^k < 1$. Then consider the solution $\boldsymbol{x} = (x_d^1 \dots x_d^{k-1}, y, x_d^{k+1}, \dots x_d^n)$ where we replaced the k'th entry of $\boldsymbol{x}_d$ with

$$y = x_d^k + \epsilon, \qquad \text{where } \epsilon := \min\left\{\frac{\rho - \sum_{n\in[N]} g^n x_d^n}{dg^k}, \frac{\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n}{|V^k - \gamma d|\, g^k}\right\} \overset{(i)}{>} 0\,,$$

where (i) follows from $\boldsymbol{x}_d$ is nonbinding, i.e. $\rho > \sum_{n\in[N]} g^n x_d^n$ and $\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n > 0$. Then

$$d \sum_{n\in[N]} g^n x^n \; = \; d \sum_{n\in[N]} g^n x_d^n + dg^k \epsilon \; \le \; d \sum_{n\in[N]} g^n x^n + \left(\rho - \sum_{n\in[N]} g^n x_d^n\right) = \rho\,.$$

On the other hand, if $V^k - \gamma d > 0$, then

$$\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n = \sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n + \left(V^k - \gamma d\right) g^k \epsilon > \sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n > 0\,.$$

If $V^k - \gamma d < 0$, then

$$\begin{aligned}
\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n \; &= \; \sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n + \left(V^k - \gamma d\right) g^k x_d^k \epsilon \\
&\ge \; \sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n + \left(V^k - \gamma d\right) \cdot \frac{\sum_{n\in[N]} (V^n - \gamma d)\, g^n x_d^n}{|V^k - \gamma d|} = 0
\end{aligned}$$

where in the last equality we used $|V^k - \gamma d| = -(V^k - \gamma d)$ since $V^k - \gamma d < 0$.

The above shows $\boldsymbol{x}$ is feasible to $U(d)$. On the other hand, $\sum_{n\in[N]} V^n g^n x_d^n < \sum_{n\in[N]} V^n g^n x^n$, so $\boldsymbol{x}$ yields a strictly larger objective than $\boldsymbol{x}_d$, contradicting the optimality of $\boldsymbol{x}_d$. ■

We now return to our proof for Theorem 5.

(1). When both $d, \tilde{d}$ are non-binding, Fact 1 implies $\boldsymbol{x}_d = \boldsymbol{x}_{\tilde{d}} = (1\dots1)$.

$$\pi(d) = d \sum_{n\in[N]} g^n x_d^n = d \sum_{n\in[N]} g^n < \tilde{d} \sum_{n\in[N]} g^n = \tilde{d} \sum_{n\in[N]} g^n x_{\tilde{d}}^n = \pi(\tilde{d})\,.$$

(2). We prove this claim by contradiction. Assume $\tilde{d}$ is non-binding and $\tilde{d} > d$ where d is budget binding. Fact 1 states that $\boldsymbol{x}_{\tilde{d}} = (1\dots1)$. Hence

$$\rho = \pi(d) = d \sum_{n\in[N]} g^n x_d^n \le d \sum_{n\in[N]} g^n x_{\tilde{d}}^n < \tilde{d} \sum_{n\in[N]} g^n x_{\tilde{d}}^n \overset{(i)}{<} \rho\,,$$

where (i) follows from the definition that $\widetilde{d}$ is non-binding. Hence we obtain a contradiction, and $\widetilde{d}$ cannot be non-binding. This means $\widetilde{d}$ must be budget or ROI binding.

(3). Here we show that if some price $d \in \mathcal{D}$ is ROI binding so that $\sum_{n\in[N]}(V^n - \gamma d)g^n x_d^n = 0$, any price $\widetilde{d} > d$ must also be ROI binding. We first claim that $\boldsymbol{x}_{\widetilde{d}} \preceq \boldsymbol{x}_d$. To show this, we use a contradiction argument. Assume $\boldsymbol{x}_{\widetilde{d}} \succeq \boldsymbol{x}_d$, and let the threshold vector $\boldsymbol{x}_d$ be characterized by $\boldsymbol{x}_d = \psi(J', q')$. Under Assumption 1, we note that $\boldsymbol{x}_d$ cannot have all 0 entries and hence $x_d^1 > 0$. However, since $\sum_{n\in[N]}(V^n - \gamma d)g^n x_d^d = 0$, it must be the case that $x_d^{J'+1} < 0$. Now, applying the ordering property for threshold vectors in Lemma 2 (ii) by taking $\boldsymbol{Z} = \boldsymbol{x}_{\widetilde{d}}$ and $\boldsymbol{Y} = \boldsymbol{x}_d$, we have

$$0 = \sum_{n\in[N]}(V^n - \gamma d)g^n x_d^n \geq \sum_{n\in[N]}(V^n - \gamma d)g^n x_{\widetilde{d}}^n > \sum_{n\in[N]}(V^n - \gamma \widetilde{d})g^n x_{\widetilde{d}}^n \,.$$

In the last inequality we used the fact that $\widetilde{d} > d$. Hence, this contradicts the feasibility of $\boldsymbol{x}_{\widetilde{d}}$, so we conclude that $\boldsymbol{x}_{\widetilde{d}} \preceq \boldsymbol{x}_d$. This further implies

$$\rho \geq \underbrace{d \sum_{n\in[N]} g^n x_d^k}_{=\pi(d)} \overset{(i)}{=} \frac{1}{\gamma}\sum_{n\in[N]} V^n g^n x_d^k \overset{(ii)}{>} \frac{1}{\gamma}\sum_{n\in[N]} V^n g^n x_{\widetilde{d}}^k \overset{(iii)}{\geq} \underbrace{\widetilde{d}\sum_{n\in[N]} g^n x_{\widetilde{d}}^k}_{=\pi(\widetilde{d})},$$

where (i) follows from d being ROI binding, i.e. $\sum_{n\in[N]}(V^n - \gamma d)g^n x_d^n = 0$; (ii) follows from $\boldsymbol{x}_{\widetilde{d}} \preceq \boldsymbol{x}_d$; (iii) follows from feasibility of $\widetilde{d}$ so that $\sum_{n\in[N]}(V^n - \gamma\widetilde{d})g^n x_{\widetilde{d}}^n \geq 0$. Therefore, $\rho \geq \pi(d) > \pi(\widetilde{d})$.

Finally, $\rho > \pi(\widetilde{d})$ implies that $\widetilde{d}$ is either non-binding or ROI binding. We note that it is not possible for $\widetilde{d}$ to be non-binding, because $\widetilde{d}$ non-binding implies $\boldsymbol{x}_{\widetilde{d}} = (1\dots 1)$ according Fact 1, contradicting $\boldsymbol{x}_{\widetilde{d}} \preceq \boldsymbol{x}_d$ which we showed earlier. Here we used the fact that $\boldsymbol{x}_d \neq (1\dots 1)$ because $\boldsymbol{x}_d$ is ROI binding and Assumption 1 states for any $d \in \mathcal{D}$, $\sum_{n\in[N]}(V^n - \gamma d)g^n \neq 0$. ■

F.2. Proof of Theorem 6

For a fixed T we use the shorthand notation $\phi = \phi(E, T)$ in this proof. Because $\phi(T^{1-\xi+\epsilon}, T) < \frac{G}{2\bar{d}}$ for all $T > T_\epsilon$, and because the exploration episode length is $E = T^{1-\xi+\epsilon}$, we know that $\phi < \frac{G}{2\bar{d}}$.

Recall $\pi(d) := d\sum_{n\in[N]} g^n x_d^n$ for any $d \in \mathcal{D}$ is the per-period seller revenue function defined in Equation (15), and $\hat{\pi}(D_h) = \frac{D_h}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t$ the estimate of $\pi(D_h)$ for episode $h \in [H]$ (with fixed price D_h). Since our the binary search procedure in our proposed pricing Algorithm 2 has exactly $H = \lfloor \log_2(M) \rfloor + 1$ iterations , the boundedness condition for a ξ-adaptive algorithm can be restated as: w.p. at least $1 - \frac{1}{T}$,

$$\left|\frac{\hat{\pi}(D_h)}{D_h} - \frac{\pi(D_h)}{D_h}\right| = \left|\frac{1}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t - \frac{\pi(D_h)}{D_h}\right| \leq \phi \Longrightarrow |\hat{\pi}(D_h) - \pi(D_h)| \leq D_h\phi \leq \bar{d}\phi \,. \tag{58}$$

Our proof relies on the following lemma:

Lemma 6 *If $\hat{\pi}(D_i) \geq \hat{\pi}(D_j)$ for some episodes $i, j \in [H]$ $(i \neq j)$, then w.p. at least $1 - \frac{1}{T}$, $\pi(D_i) \geq \pi(D_j)$. Furthermore, this implies the event $\mathcal{G} = \{\hat{\pi}(D_i) \geq \hat{\pi}(D_j) \Longrightarrow \pi(D_i) \geq \pi(D_j) \;\; \text{for all } i, j \in [H]\}$ holds with probability at least $1 - \frac{H(H-1)}{2T}$, where $H = \lfloor \log_2(M) \rfloor + 1$ is the total number of binary search iterations (i.e. number of episodes in the exploration phase).*

Proof. Because $\hat{\pi}(D_i) \geq \hat{\pi}(D_j)$, applying Equation (58) for episodes i, j yields

$$\pi(D_i) + \bar{d}\phi \geq \hat{\pi}(D_i) \geq \hat{\pi}(D_j) \geq \pi(D_j) - \bar{d}\phi \Longrightarrow 2\bar{d}\phi \geq \pi(D_j) - \pi(D_i)\,,$$

Now, contrary to our claim, suppose that $\pi(D^i) < \pi(D^j)$.We then have

$$2\bar{d}\phi \geq \pi(D_j) - \pi(D_i) \geq G = \min_{d, \widetilde{d} \in \mathcal{D}: \pi(d) \neq \pi(\widetilde{d})} \left| \pi(d) - \pi(\widetilde{d}) \right|\,,$$

which contradicts the definition of ξ-adaptivity such that $2\bar{d}\phi < G$ for episode lengths $\Omega(T^{1-\xi})$. As there are $H(H-1)/2$ pairs (i, j) such that $i \neq j$, a simple union bound shows event $\mathcal{G}$ holds with probability at least $1 - \frac{H(H-1)}{2T}$. ■

We now return to our proof of Theorem 6. We first show that under event $\mathcal{G} = \{\hat{\pi}(D_i) \geq \hat{\pi}(D_j) \Longrightarrow \pi(D_i) \geq \pi(D_j) \;\; \text{for all } i, j \in [H]\}$, we have $\max_{d \in \mathcal{D}} \pi(d) = \pi\left(D^{m^*}\right)$ where we recall that $m^* = \arg\max_{m \in [M]} \hat{\pi}(D^m)$.

We use an induction argument that shows after each iteration of the binary search procedure in the exploration phase of Algorithm 2, $\pi(D^m) \leq \pi(D^{m^*})$ for all $m \leq \mathrm{L}$ and $m \geq \mathrm{R}$. The base case is the first iteration, where we have $\mathrm{L} = 1$, $\mathrm{R} = M$. If $m^* = \mathrm{L} = 1$, then under event $\mathcal{G}$ we get

$$\hat{\pi}(D^1) \geq \hat{\pi}(D^M) \overset{(i)}{\Longrightarrow} \pi(D^1) \geq \pi(D^M)\,.$$

Hence after the first iteration $\pi(D^m) \leq \pi(D^{m^*})$ for any $m \leq \mathrm{L}$ and $m \geq \mathrm{R}$. The case for $m^* = \mathrm{R}$ follows from the same argument.

Now assume that the induction hypothesis holds, i.e. at the beginning of some iteration with the tuple $(\mathrm{L}, \mathrm{R}, m^*)$, we have $\pi(D^m) \leq \pi(D^{m^*})$ $m \leq \mathrm{L}$ and $m \geq \mathrm{R}$. According to Algorithm 2, we only need to show two cases in order to validate the induction procedure.

- **Case 1.** If $\hat{\pi}(D^{\mathrm{med}}) < \hat{\pi}(D^{\mathrm{med}+1})$, then we show $\pi(D^m) \leq \pi(D^{\mathrm{med}+1})$ for all $m = 1 \ldots \mathrm{med} + 1$
- **Case 2.** If $\hat{\pi}(D^{\mathrm{med}}) \geq \hat{\pi}(D^{\mathrm{med}+1})$, then we show $\pi(D^m) \geq \pi(D^{\mathrm{med}})$ for all $m = \mathrm{med} + 1 \ldots M$

Note that under Case 1., $\text{med}+1$ will be the new value of m^* in the next iteration (i.e. the next induction step). So by showing $\pi(D^m) \le \pi(D^{\text{med}+1})$ for all $m = 1\ldots\text{med}+1$, we validate the induction hypothesis for the next induction step. A similar argument holds for Case 2.

Case 1. When $\hat{\pi}(D^{\text{med}}) < \hat{\pi}(D^{\text{med}+1})$, under event $\mathcal{G}$ we have $\pi(D^{\text{med}}) \le \pi(D^{\text{med}+1})$. We claim that D^{med} cannot be an ROI binding price. Assume the contrary that D^{med} is ROI binding. Then, part (3) of Theorem 5 states $\pi(D^{\text{med}+1}) < \pi(D^{\text{med}})$, leading to a contradiction. Hence D^{med} must be either a nonbinding price or a budget binding price. Applying part (1) of Theorem 5, we can then conclude that for any $m \le \text{med}$, $\pi(D^m) \le \pi(D^{\text{med}})$, so

$$\pi(D^m) \le \pi(D^{\text{med}}) \le \pi(D^{\text{med}+1}) \quad \forall m = 1\ldots\text{med}\,.$$

At the end of the iteration, as we update $m^{*+} = \text{med}+1$ (here we denote m^{*+} as the updated value to distinguish from its initial value at the start of the iteration), we have $\pi(D^{m^{*+}}) \ge \pi(D^{\text{med}+1}) \ge \pi(D^{\text{med}})\ldots\pi(D^1)$. On the other hand, since $\hat{\pi}(D^{m^{*+}}) = \max_{m\in\{m^*,\text{med}+1\}} \hat{\pi}(D^m) \ge \hat{\pi}(D^{m^*})$, event $\mathcal{G}$ implies

$$\pi(D^{m^{*+}}) \ge \pi(D^{m^*}) \overset{(i)}{\ge} \pi(D^m) \quad \forall m = \text{R}\ldots M\,,$$

where (i) follows from the induction hypothesis. Therefore, we have

$$\pi(D^{m^{*+}}) \ge \pi(D^m) \quad \forall m = \text{R}\ldots M \text{ and } m = 1\ldots\text{med}+1\,,$$

and by realizing $(\text{med}+1, \text{R}, m^{*+})$ is the initial tuple for the next iteration concludes the induction step.

Case 2. The case when $\hat{\pi}(D^{\text{med}}) \ge \hat{\pi}(D^{\text{med}+1})$ follows from an identical argument, and we will omit the details. This concludes the induction proof.

The above implies that when the event $\mathcal{G} = \{\hat{\pi}(D_i) \ge \hat{\pi}(D_j) \Longrightarrow \pi(D_i) \ge \pi(D_j) \text{ for all } i,j \in [H]\}$ holds throughout the exploration phase, the above induction argument implies we have $\pi(D^{m^*}) \ge \pi(D^m)$for all $m \in [M]$. Hence $\pi(D^{m^*}) = \max_{d\in\mathcal{D}} \pi(d)$ w.p. at least $1 - \frac{H(H-1)}{2T}$ according to Lemma 6 where $H = \lfloor \log_2(M) \rfloor + 1$.

Furthermore, we point out that in each iteration of the binary search procedure the seller explores at most two prices. Hence exploration phase, which we denote as $\mathcal{E}$, has length at most $2E\left(\lfloor \log_2(M) \rfloor + 1\right) = 2T^{1-\xi+\epsilon}\left(\lfloor \log_2(M) \rfloor + 1\right)$ periods. Therefore, the seller's regret can be upper bounded as

$$\begin{aligned} \text{Reg}_{\text{sell}} \;&=\; T\max_{d\in\mathcal{D}} \pi(d) - \sum_{t\in[T]} \mathbb{E}\left[d_t z_t\right] \;\le\; \bar{d}|\mathcal{E}| + \sum_{t=|\mathcal{E}|+1}^{T} \max_{d\in\mathcal{D}} \pi(d) - \mathbb{E}\left[d_t z_t\right] \\ &\overset{(i)}{\le}\; \bar{d}|\mathcal{E}| + \sum_{t\in[T]/\mathcal{E}} \mathbb{E}\left[\left(\pi(D^{m^*}) - D^{m^*} z_t\right)\mathbb{I}\left\{\mathcal{G}\right\}\right] + \bar{d}(T - |\mathcal{E}|)\mathbb{P}\left(\mathcal{G}^c\right) \end{aligned}$$

$$\leq \bar{d}|\mathcal{E}| + D^{m^*}(T-|\mathcal{E}|)\cdot\mathbb{E}\left[\left|\frac{\pi(D^{m^*})}{D^{m^*}} - \frac{1}{T-|\mathcal{E}|}\sum_{t\in[T]/\mathcal{E}} z_t\right|\mathbb{I}\{\mathcal{G}\}\right] + \bar{d}(T-|\mathcal{E}|)\mathbb{P}(\mathcal{G}^c)$$
$$\overset{(ii)}{\leq} \bar{d}|\mathcal{E}| + 2\bar{d}T\cdot\phi(T-|\mathcal{E}|,T) + \bar{d}T\mathbb{P}(\mathcal{G}^c)$$
$$\overset{(iii)}{\leq} 2\bar{d}(\lfloor\log_2(M)\rfloor+1)\cdot T^{1-\xi+\epsilon} + 2\bar{d}T\phi\left(\frac{T}{2},T\right) + \bar{d}(\lfloor\log_2(M)\rfloor+1)^2/2\,.$$

In (i) we used the fact that $\max_{d\in\mathcal{D}}\pi(d) = \pi(D^{m^*})$ under event $\mathcal{G}$ and $d_t = D^{m^*}$ for all exploitation periods $t\in[T]/\mathcal{E}$; (ii) directly follows from the definition of event $\mathcal{G}$. In (iii), we used the fact that $T > \left(4\bar{d}\lfloor\log_2(M)\rfloor + 4\bar{d}\right)^{\frac{1}{\xi-\epsilon}}$, which implies $T^{\xi-\epsilon} > 4\bar{d}\lfloor\log_2(M)\rfloor + 4\bar{d}$ and hence $\frac{T}{2} > 2\bar{d}(\lfloor\log_2(M)\rfloor+1)T^{1-\xi+\epsilon} > |\mathcal{E}|$. As ϕ is decreasing in the first argument, we then have $\phi\left(\frac{T}{2},T\right) \geq \phi(T-|\mathcal{E}|,T)$. ■

F.3. Proof for Theorem 7

In this proof, we show a more detailed statement by constructively checking the boundedness and stability conditions for ξ-adaptive algorithms. In particular, let C and S be defined as in Theorems 8 and 9. Define $\widetilde{A},\widetilde{B}$ and $\mathcal{T}$ as the following:

$$\widetilde{A} = \sqrt{2+16\sqrt{\log(T^2\log(T))}},\quad \widetilde{B} = 2\sqrt{(1+72\log(T^2\log(T)))}\,,$$

and

$$\mathcal{T} = \min\left\{t\in[T] : \widetilde{A}\left(1-2T^{-\frac{2}{3}}\right)^t + \widetilde{B}T^{-\frac{1}{3}} < S\right\} = \Theta(T^{\frac{1}{3}})\,.$$

To see why , please see Equation (36) and the discussion following Equation (36). Define

$$\phi(\tau,T) = \sqrt{\frac{2\log(2T^2)}{\tau}} + \frac{\mathcal{T}}{\tau} + C\left(\frac{\widetilde{A}T^{\frac{2}{3}}}{\tau} + \widetilde{B}T^{-\frac{1}{3}}\right) = \Theta\left(\frac{1}{\sqrt{\tau}} + \frac{T^{\frac{2}{3}}}{\tau} + T^{-\frac{1}{3}}\right)\,, \tag{59}$$

and for any fixed $\epsilon\in(0,\frac{1}{3})$, assume that entire time horizon T is large such that $\phi(T^{\frac{2}{3}+\epsilon},T) < \frac{G}{2\bar{d}}$. [22] Let the entire T periods is divided into H consecutive episodes $\mathcal{E}_1\ldots\mathcal{E}_H$ where the price in episode $h\in[H]$ is fixed to be $D_h\in\mathcal{D}$, and the buyer runs $\mathrm{CTBR}_{\mathcal{A}}$ with SGD and constant step size $\eta = T^{-\frac{2}{3}}$. We will show that

(1) $\left|\frac{1}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t - \frac{\pi(D_h)}{D_h}\right| \leq \phi(|\mathcal{E}_h|,T)$ for any episode $h\in[H]$ w.p. at least $1-\frac{1}{T}$;

(2) $\phi(|\mathcal{E}_h|,T) < \frac{G}{2\bar{d}}$ for any episode h whose length $|\mathcal{E}_h| \geq T^{\frac{2}{3}+\epsilon}$.

Note that showing (1) and (2) would imply that $\mathrm{CTBR}_{\mathcal{A}}$ with SGD and constant step size is $\frac{1}{3}$-adaptive. We show (1) first, and then deduct (2) through a simple argument.

[22] Note that $\phi(T^{\frac{2}{3}+\epsilon},T) = \Theta(T^{-\epsilon})$ so there always exist large enough T such that $\phi(T^{\frac{2}{3}+\epsilon},T) < \frac{G}{2\bar{d}}$.

Our proof for showing (1) consists of 2 main steps, namely Step 1. providing a reformulation for $\pi(D_h)$; Step 2. showing a high probability bound for $\sum_{t\in\mathcal{E}_h} z_t$; and Step 3. bounding $\pi(D_h)-\sum_{t\in\mathcal{E}_h} z_t$.

Step 1. Reformulation for $\pi(D_h)$.

As discussed in Section 6, the sets $\{D^m\}_{m\in[M]}$ and $\{V^n\}_{n\in[N]}$ can be thought of as the unique values of highest competing bids and valuations, respectively, in the set $\mathcal{W}=\{(v^k,d^k)\}_{k\in[K]}$ studied in Sections 4 and 5. In other words, $\mathcal{W}=\mathcal{V}\times\mathcal{D}$, where $K=MN$. Furthermore, if we impose any distribution $\widetilde{\boldsymbol{g}}\in\Delta_M$ on support $\mathcal{D}$, then $\widetilde{\boldsymbol{g}}$ combined with the valuation distribution $\boldsymbol{g}$ on $\mathcal{V}$ induces a product distribution $\boldsymbol{p}=\boldsymbol{g}\times\widetilde{\boldsymbol{g}}$ over $\mathcal{W}$. Hence, in each episode where the prices is fixed to D^m for some $m\in[M]$, the imposed distribution $\widetilde{\boldsymbol{g}}\in\Delta_M$ on the support of prices $\mathcal{D}$ is the m'th unit vector $\boldsymbol{e}^m\in\Delta_M$. Hence, the valuation-price pairs (v_t,d_t) can be viewed as being drawn from the induces product distribution $\boldsymbol{p}=\boldsymbol{g}\times\boldsymbol{e}^m\in\Delta_K$ over $\mathcal{W}$.

Following this argument, we can denote the induced distributions for the valuation-price pairs (v_t,d_t) in each episode $h\in[H]$ as $\boldsymbol{p}_h$. Hence, we can directly apply Lemma 5 (that we used in the proof of Theorem 9) with $\delta=\frac{1}{T^2}$, and conclude for any $t\in\mathcal{E}_h$, we have w.p. at least $1-\frac{1}{T^2}$

$$\|\boldsymbol{p}_h-\hat{\boldsymbol{p}}_t\|\leq\underbrace{\sqrt{2+16\sqrt{\log\left(T\log\left(T\right)\right)}}}_{=\widetilde{A}}\left(1-2\eta\right)^{\frac{t-s_h}{2}}+\underbrace{2\sqrt{\left(1+72\log\left(T\log\left(T\right)\right)\right)}}_{=\widetilde{B}}\cdot\sqrt{\eta}\,, \tag{60}$$

where we defined s_h to be the first period in episode h, i.e. $s_h=\min\{t:t\in\mathcal{E}_h\}$, and recall the constant step size $\eta=T^{-\frac{2}{3}}$. Furthermore, as argued in Section 6, the problem $U(D_h)$, defined in Equation (14), and $U(\boldsymbol{p}_h;0,\gamma,\rho)$ which is defined in Section 4 are equivalent. Hence, the seller's per-period revenue is

$$\pi(D_h)=D_h\sum_{n\in[N]}x^n_{D_h}g^n=D_h\sum_{k\in[K]}x^{*,k}_h p^k_h\,,$$

where $\boldsymbol{x}_d\in[0,1]^N$ is the unique optimal threshold vector solution to $U(d)$ and $\boldsymbol{x}^*_h\in[0,1]^K$ is the unique optimal threshold vector solution to $U(\boldsymbol{p}_h;0,\gamma,\rho)$. We let the threshold vector $\boldsymbol{x}^*_h$ be characterized by $(J_h,q_h)\in[K]\times[0,1)$. Then we can further write the seller's revenue for episode h as

$$\pi(D_h)=D_h\sum_{k\in[K]}x^{*,k}_h p^k_h=D_h\left(\sum_{k\in[J_h]}p^k_h+q_h p^{J_h+1}_h\right)\,.$$

Note that $p^k_h=0$ for all $k\in[K]$ such that $d^k\neq D_h$, and also it must be the case that $p^{J_h+1}_h>0$.

Step 2. High probability bound for $\sum_{t\in\mathcal{E}_h} z_t$.

Let $\mathcal{F}_t$ be the sigma algebra generated by $\{(v_\tau, d_\tau, b_\tau)\}_{\tau\in[t]}$, which characterizes all randomness in the buyer and seller's behavior up to period t. According to the posted price version of CTBR$_\mathcal{A}$, the take/leave indicator $z_t = \mathbb{I}\{b_t \geq d_t\}$ where b_t is the virtual bid that is not submitted and only used to decide whether to accept price d_t or not during period t. Further, recall the threshold vector $\hat{\boldsymbol{x}}_t = \psi(\hat{J}_t, \hat{q}_t)$ in CTBR$_\mathcal{A}$ for $t \in \mathcal{E}_h$ is the estimate of $\boldsymbol{x}_h^*$, and note that $\hat{\boldsymbol{x}}_t$ is $\mathcal{F}_{t-1}$-measurable. Then, for $t \in \mathcal{E}_h$ we have

$$\begin{aligned}
\mathbb{E}\left[D_h z_t \middle| \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[d_t z_t \middle| \hat{\boldsymbol{x}}_t\right] = \mathbb{E}\left[d_t \mathbb{I}\{b_t \geq d_t\} \middle| \hat{\boldsymbol{x}}_t\right] \\
&= (1-\hat{q}_t)\mathbb{E}\left[d_t \mathbb{I}\left\{\frac{v_t}{\theta^{\hat{J}_t}} \geq d_t\right\} \middle| \hat{\boldsymbol{x}}_t\right] + \hat{q}_t \mathbb{E}\left[d_t \mathbb{I}\left\{\frac{v_t}{\theta^{\hat{J}_t+1}} \geq d_t\right\} \middle| \hat{\boldsymbol{x}}_t\right] \\
&\overset{(i)}{=} (1-\hat{q}_t) \sum_{k\in[K]} d^k p_h^k \mathbb{I}\left\{\theta^k \geq \theta^{\hat{J}_t}\right\} + \hat{q}_t \sum_{k\in[\hat{J}_t]} d^k p_h^k \mathbb{I}\left\{\theta^k \geq \theta^{\hat{J}_t+1}\right\} \\
&= \sum_{k\in[\hat{J}_t]} d^k p_h^k + \hat{q}_t d^{\hat{J}_t+1} p_h^{\hat{J}_t+1} \overset{(ii)}{=} D_h \left(\sum_{k\in[\hat{J}_t]} p_h^k + \hat{q}_t p_h^{\hat{J}_t+1}\right),
\end{aligned}$$

where in (i) we used the fact that when period t belongs in episode $\mathcal{E}_h$, $(v_t, d_t) \sim \boldsymbol{p}_h$; in (ii) we used the fact that $p_h^k = 0$ for all $k \in [K]$ such that $d^k \neq D_h$. This implies that by defining $Y_t = \sum_{k\in[\hat{J}_t]} p_h^k + \hat{q}_t p_h^{\hat{J}_t+1} - z_t$, the sequence $\{Y_t\}_{t\in\mathcal{E}_h}$ is a martingale difference sequence. Also, since it is easy to see that $|Y_t| \leq 1$, by Azuma Hoeffding's inequality we have for any $\delta \in (0,1)$

$$\mathbb{P}\left(\widetilde{\mathcal{G}}\right) \geq 1-\delta \text{ where } \widetilde{\mathcal{G}} := \left\{\left|\sum_{t\in\mathcal{E}_h}\left(\sum_{k\in[\hat{J}_t]} p_h^k + \hat{q}_t p_h^{\hat{J}_t+1} - z_t\right)\right| \leq \sqrt{2|\mathcal{E}_h|\log\left(2T^2\right)}\right\}. \tag{61}$$

In the following we will take $\delta = \frac{1}{T^2}$ so event $\widetilde{\mathcal{G}}$ holds w.p. at least $\frac{1}{T^2}$.

Finally, by applying Lemma 3 (i) for $t - s_h > \mathcal{T}$, under event $\mathcal{G}_t = \left\{\|\boldsymbol{p}_h - \hat{\boldsymbol{p}}_t\| \leq \widetilde{A}\,(1-2\eta)^{\frac{t-s_h}{2}} + \widetilde{B}\sqrt{\eta}\right\}$ we have $\hat{J}_t = J_h$. Furthermore, following the exact proof of Theorem 8 (where we bound the regret), we can recover Equation (34), which in the episodic pricing setting states that under event

$$\mathcal{G}_t = \left\{\|\boldsymbol{p}_h - \hat{\boldsymbol{p}}_t\| \leq \widetilde{A}\,(1-2\eta)^{\frac{t-s_h}{2}} + \widetilde{B}\sqrt{\eta}\right\}$$

for $t - s_h > \mathcal{T}$ we have

$$|q_h - \hat{q}_t|\, p^{J_h+1} \leq \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\left(\widetilde{A}\,(1-2\eta)^{\frac{t-s_h}{2}} + \widetilde{B}\sqrt{\eta}\right). \tag{62}$$

Recall Equation (60) implies event $\mathcal{G}_t$ holds w.p. at least $1 - \frac{1}{T^2}$.

Step 3. Bounding $\frac{\pi(D_h)}{D_h} - \sum_{t\in\mathcal{E}_h} z_t$.

We now combine the results of Step 1 and Step 2 in the above Equations (61) and (62), when the event

$\widetilde{\mathcal{G}} \cap (\cap_{t>\mathcal{T}} \mathcal{G}_t)$ holds, we have $\hat{J}_t = J_h$ where $(J_h, q_h) \in [K] \times [0,1)$ characterized the threshold vector $\boldsymbol{x}_h^* \in [0,1]^K$ which is the unique optimal threshold vector solution to $U(\boldsymbol{p}_h; 0, \gamma, \rho)$. Hence, we have

$$\begin{aligned}
\left|\sum_{t\in\mathcal{E}_h}\left(\frac{\pi(D_h)}{D_h} - z_t\right)\right| &= \left|\sum_{t\in\mathcal{E}_h}\left(\sum_{k\in[J_h]} p_h^k + q p_h^{J_h+1} - z_t\right)\right| \\
&\overset{(i)}{\leq} \left|\sum_{t\in\mathcal{E}_h}\left(\sum_{k\in[\hat{J}_t+1]} p_h^k + \hat{q}_t p_h^{\hat{J}_t+1+1} - z_t\right)\right| + \sum_{t\in\mathcal{E}_h}\left|\left(\sum_{k\in[J_h]} p_h^k + q p_h^{J_h+1}\right) - \left(\sum_{k\in[\hat{J}_t]} p_h^k + q p_h^{\hat{J}_t+1}\right)\right| \\
&\leq \left|\sum_{t\in\mathcal{E}_h}\left(\sum_{k\in[\hat{J}_t+1]} p_h^k + \hat{q}_t p_h^{\hat{J}_{h\,t}+1+1} - z_t\right)\right| + \mathcal{T} + \sum_{t-s_h>\mathcal{T}} |q_h - \hat{q}_t|\, p^{J_h+1} \\
&\overset{(ii)}{\leq} \sqrt{2|\mathcal{E}_h|\log(2T^2)} + \mathcal{T} + \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\left(\widetilde{A}\sum_{t-s_h>\mathcal{T}}(1-2\eta)^{\frac{t-s_h}{2}} + \widetilde{B}|\mathcal{E}_h|\sqrt{\eta}\right),
\end{aligned}$$

where in (i) we used the fact that $J_h = \hat{J}_t$

(ii) we plugged in the Azuma-Hoeffding inequality result showed in Equation (61) with $\delta = \frac{1}{T^2}$.

Since $\mathcal{G}_t$ holds with probability at least $1 - \frac{1}{T^2}$, it is easy to see the event $\widetilde{\mathcal{G}} \cap (\cap_{t>\mathcal{T}} \mathcal{G}_t)$ holds with probability at least $1 - T \cdot \frac{1}{T^2} = 1 - \frac{1}{T}$ via applying a union bound. Hence, we have w.p. at least $1 - \frac{1}{T}$

$$\begin{aligned}
&\left|\frac{\pi(D_h)}{D_h} - \frac{1}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h} z_t\right| \\
&\leq \sqrt{\frac{2\log(2T^2)}{|\mathcal{E}_h|}} + \frac{\mathcal{T}}{|\mathcal{E}_h|} + \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\left(\frac{\widetilde{A}}{|\mathcal{E}_h|}\sum_{t\in\mathcal{E}_h: t-s_h>\mathcal{T}}(1-2\eta)^{\frac{t-s_h}{2}} + \widetilde{B}\sqrt{\eta}\right) \\
&\overset{(i)}{<} \sqrt{\frac{2\log(2T^2)}{|\mathcal{E}_h|}} + \frac{\mathcal{T}}{|\mathcal{E}_h|} + \max\left\{\frac{\bar{w}}{\underline{w}}, \frac{\bar{d}}{\underline{d}}\right\}\left(3\sqrt{K}+5\right)\left(\frac{\widetilde{A}}{|\mathcal{E}_h|}T^{\frac{2}{3}} + \widetilde{B}T^{-\frac{1}{3}}\right) = \phi(|\mathcal{E}_h|, T).
\end{aligned}$$

In (i) we used the fact that $\sum_{t\in\mathcal{E}_h: t-s_h>\mathcal{T}}(1-2\eta)^{\frac{t-s_h}{2}} < \sum_{t=0}^{\infty}(1-2\eta)^{\frac{t}{2}} = \frac{1}{1-\sqrt{1-2\eta}} = \frac{1+\sqrt{1-2\eta}}{2\eta} < \frac{1}{\eta}$. The final equality follows from the definition of $\phi(\tau, T)$ in Equation (59).

Finally, from the definition of $\phi(\tau, T) = \Theta\left(\frac{1}{\sqrt{\tau}} + \frac{T^{\frac{2}{3}}}{\tau} + T^{-\frac{1}{3}}\right)$, we have $\phi(T^{\frac{2}{3}+\epsilon}, T) = \Theta\left(\max\left\{T^{-\epsilon}, T^{-\frac{1}{3}}\right\}\right)$. Therefore, there exists $\epsilon = \Theta(1/\log(T))$ and $T_0 > 0$ such that $\phi(T^{\frac{2}{3}+\epsilon}, T) = \phi(|\mathcal{E}_h|, T) < \frac{G}{2\bar{d}}$ for all $T > T_0$. ■

F.4. Proof of Corollary 1

In Equation (59) of the proof of Theorem 7 (see Appendix F.3), we characterized the universal error function ϕ for constant step size SGD. Simply plugging this error function into the general regret of our proposed seller algorithm in Theorem 6 yields the desired result. ■

G. Supplementary Lemmas

Lemma 7 (Lemma 3 in Rakhlin et al. (2011)) *Let $Y_1 \ldots Y_T$ be a martingale difference sequence with a uniform bound $|Y_t| \le b$ for all t. Let $V_s = \sum_{t \in [s]} Var(Y_t)$ be the sum of conditional variances of Y_t's up to index s. Further, let $\sigma_s = \sqrt{V_s}$. Then we have, for any $\delta \in (0, 1/e)$ and $T \ge 4$,*

$$\mathbb{P}\left(\sum_{t \in [s]} Y_t > 2\sqrt{\log(\log(T)/\delta)} \cdot \max\left\{ 2\sigma_s, b\sqrt{\log(\log(T)/\delta)} \right\} \text{ for some } s < T \right) \le \delta \, .$$