Leveraging The Hints: Adaptive Bidding in Repeated First-Price Auctions

Wei Zhang[1] **Yanjun Han**[2*] **Zhengyuan Zhou**[3,4*] **Aaron Flores**[5] **Tsachy Weissman**[6]
[1]MIT EECS [2]MIT IDSS [3]Arena Technologies [4]NYU Stern [5]Yahoo! Research [6]Stanford EE
{w_zhang, yjhan}@mit.edu z@arena-ai.com aaron.flores@yahooinc.com tsachy@stanford.edu

Abstract

With the advent and increasing consolidation of e-commerce, digital advertising has very recently replaced traditional advertising as the main marketing force in the economy. In the past four years, a particularly important development in the digital advertising industry is the shift from second-price auctions to first-price auctions for online display ads. This shift immediately motivated the intellectually challenging question of how to bid in first-price auctions, because unlike in second-price auctions, bidding one's private value truthfully is no longer optimal. Following a series of recent works in this area, we consider a differentiated setup: we do not make any assumption about other bidders' maximum bid (i.e. it can be adversarial over time), and instead assume that we have access to a hint that serves as a prediction of other bidders' maximum bid, where the prediction is learned through some blackbox machine learning model. We consider two types of hints: a single point-prediction, and a hint interval (representing a type of confidence region into which others' maximum bid falls). We establish minimax near-optimal regret bounds for both cases and highlight the quantitatively different behavior between them. We also provide improved regret bounds when the others' maximum bid exhibits the further structure of sparsity. Finally, we complement the theoretical results with demonstrations using real bidding data.

1 Introduction

As e-commerce proliferates across the industries, digital advertising has become the predominant marketing force in the economy: in 2019, businesses in the US alone [Wag19] have spent more than 129 billion dollars on digital advertising, surpassing for the first time the combined amount spent via traditional advertising channels by 20 billion dollars. Since then, this number has been growing and continues to outpace traditional advertising spending [Wag19]. Within digital advertising, the key step that generates such revenue for digital advertising is online ads auctions. In the past, second-price auctions, as a result of its truthful nature, have been the standard format for online ads auctions [LR00, Kle04, LRBPR07]. However, the industry has recently witnessed a shift from second-price auctions to first-price auctions in display ads auctions, which currently account for 54% of the digital advertising market* share [DRS19].

This industry-wide shift to first-price auctions has occurred for several reasons, including larger revenue (the exchange charges a percentage of the winning bid) [Ben18a] and no last-look advantage

*Correspondence to Yanjun Han (yjhan@mit.edu) and Zhengyuan Zhou (z@arena-ai.com).

*Search auctions dominate the remaining market, which are still conducted using second-price (or generalize second-price) auctions.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

for exchanges: under second-price auctions, an ad exchange can examine all the submitted bids and *then* raise the floor price to above the second-highest bid and obtain a larger[*] revenue [Ben18b].

Motivated by these considerations, several ad exchanges, including AppNexus (now Xandr), Index Exchange and OpenX, started to roll out first-price auctions in 2017 and completed the transition by 2018 [Slu17, App18]. Google Ad Manager (previously Adx) followed suit and also completed the move to first-price auctions at the end of 2019 [Dav19] and incorporated additional transparency[*] in their new first-price auction platform: bidders would be able to see the minimum-bid-to-win (i.e. full information) after each auction on Google Ad Manager, whereas in many other ad exchanges, a bidder only knows whether he/she wins the bid (i.e. binary feedback) [Dav19]. Situated in this background, an important question arises: how should a bidder adaptively bid in repeated online first-price auctions to maximize the cumulative payoffs?

Prior to the shift, bidding is straightforward: the optimal bidding strategy is simply truthfully bid one's private value (regardless of what the other bidders do). However, this truthful property no longer holds in first-price auctions. As such, bidding in first-price auctions–and the various inference/learning problems arising from it–quickly become complicated. In response, an online decision-making approach has emerged recently, where a bidder at each auction t needs to decide the amount b_t to bid with a given valuation v_t, whereas others' maximum bid m_t is either assumed to be iid drawn from a distribution (unrelated to anything else) or fully adversarial. More specifically, [BGM+19] (and its follow-up journal version [BGM+21]) studied the binary feedback setting and show that: 1) if m_t is drawn iid from an underlying distribution (with a generic CDF), then one achieves the minimax optimal regret of $\widetilde{\Theta}(T^{\frac{2}{3}})$; 2) if m_t is adversarial, then one achieves the minimax optimal regret of $\widetilde{\Theta}(T^{\frac{3}{4}})$. Subsequently, [HZW20] considered the winning-bid only feedback (i.e. a bidder can observe the winning bid) and established that if m_t is drawn iid from an underlying distribution (with a generic CDF), one can achieve the minimax optimal regret of $\widetilde{\Theta}(T^{\frac{1}{2}})$. While it remains unknown what the result would be when m_t is adversarial under winning-bid only feedback, [HZF+20] studied the full-information feedback setting and showed that the minimax optimal regret of $\widetilde{\Theta}(T^{\frac{1}{2}})$ can be achieved when m_t is adversarial[*]. [ZKH+21] also studied the full-information feedback setting where it designed and implemented a space-efficient variant of the algorithm proposed in [HZF+20] and through empirical evaluations, showed that the algorithmic variant is quite effective.

In practice, others' highest bid m_t is often neither stochastic nor adversarial, and contextual information is often available to gain some knowledge of m_t. We aim to make inroads into this more practical setup by considering a differentiated setup from the existing and growing adaptive-bidding-in-first-price auctions literature: we do not make any assumption about other bidders' maximum bid (i.e. it can be adversarial over time), or model the contexts directly [BFG21]; instead we assume an access to a hint that serves as a prediction of other bidders' maximum bid, where the prediction is learned through some blackbox machine learning model which could be much more powerful than simple linear models. We consider two types of hints: one where a single point-prediction is available, and the other where a hint interval (representing a type of confidence region into which others' maximum bid falls) is available. We establish minimax optimal regret bounds for both cases and highlight the quantitatively different behavior between the two settings. We also provide improved regret bounds when the others' maximum bid exhibits the further structure of sparsity. Finally, we complement the theoretical results with demonstrations using real bidding data.

1.1 Additional Application: Personalized Hospitality Pricing

Another important application that shares similar elements to adaptive bidding in first-price auctions and that hence is also amenable to the methodological framework we develop in this paper is personalized hospitality pricing. In personalized hospitality pricing, a travel distribution platform applies a markup to a given hotel room provided by a supplier (either the hotel itself or some travel

[*]At the extreme, raising the floor price to barely under the highest bid effectively turns it into a first-price auction without the bidders being aware.

[*]Google was under sustained criticism of leveraging last-look advantage in second-price auctions. This is likely an effort to offset the previous negative image, although there was no such mention in Google's official language.

[*]Note that under both full-information feedback and iid m_t, a pure exploitation algorithm already achieves the minimax optimal regret $\Theta(\sqrt{T})$.

aggregator such as Expedia) and presents the final price along with the hotel room whenever a user (either a consumer or a travel agency) searches for a hotel when booking travel. To see the parallel with first-price auctions, the supplier's cost corresponds to the private value, and the final price (which is supplier's cost plus the markup) corresponds to the "bid". Note that, the user can access many other competing travel distribution platforms, each of which may provide a different price for the same hotel room (type). Consequently, there is a bidding element because the user will take the lowest price. Further, in this problem, the platform would want to provision personalized markups, where the markup is decided based on search features (destination city, number of nights, days until first check-in), hotel features (ratings, room types) and other generic features (holiday season, time of the year etc). Note that as of this writing, although almost all existing markup provisioning schemes are fixed business rules that are handcrafted (many "if this feature then that markup" logic statements), an adaptive learning approach – such as the one proposed in this paper, has great applicability in practice. In particular, Arena Technologies[*], a leading enterprise AI solution provider, provides reinforcement learning enabled personalized hospitality pricing. Using publicly available travel pricing data, Arena builds price prediction models that serve as hints, which are in turn used in its real-time adaptive personalization engine. Such a differentiated infrastructure – both in terms of engineering sophistication and learning flexibility – makes it easy for the deployment and testing of our (and any future improved) algorithm, thereby broadening its potential impact.

2 Problem Formulation

We study the problem of repeated first-price auction with hint as follows. Consider a time horizon with total length T, and there is one round of first-price auction taking place at each time. At the beginning of each round, a bidder observes a particular item and has a private value $v_t \in [0, 1]$ for it. Then, based on her past observations of others' bid and v_t for that round, she bids $b_t \in [0, v_t]$ for that item. The bidder wins the round if and only if b_t is larger than others' highest bid, defined as m_t. Under the above settings one could write the instantaneous reward at time t for that bidder:

$$r(b_t; v_t, m_t) = (v_t - b_t) \cdot \mathbb{1}\{b_t \geq m_t\}. \tag{1}$$

Define policy π as the overall bidding strategy, which is a sequence of bidding prices $(b_1, b_2, \cdots, b_T)$ under corresponding private value sequence $\{v_t\}$ and others' highest bid sequence $\{m_t\}$, and we use $\mathbb{E}[r(b_t; m_t, v_t)]$ to denote the expected reward under policy π while the expectation is taken over the randomness inside randomized policy π. Then we define the regret under policy π:

$$\text{Reg}(\pi) = \max_{a \in \mathcal{F}_{\text{L-M}}} \sum_{t=1}^{T} r(a(v_t); m_t, v_t) - \sum_{t=1}^{T} \mathbb{E}[r(b_t; m_t, v_t)], \tag{2}$$

where a denotes a bidding oracle - a map from private values v_t to bidding prices b_t, and $a(v_t)$ is the bidding price under oracle a.[*] Here $\mathcal{F}_{\text{L-M}}$ is the set of oracles we compete with, which is the set of all 1-Lipschitz and increasing functions from $[0, 1] \to [0, 1]$.

The above settings are similar to that in [HZW20] [HZF+20], and in this work, we include additional information provided to the bidder at each round. The goal is to analyze how the performance of hints may influence the regret bounds in theory. We consider two forms of hint, both of which contain a point estimate h_t of the minimum-bid-to-win m_t at time t, and the difference lies in whether a bidder observes a single hint or a hint interval, with the latter defined as a pair (h_t, σ_t) satisfying:

$$\mathbb{E}\left[|h_t - m_t|^q\right] \leq \sigma_t^q, \quad t = 1, 2, \cdots, T, \tag{3}$$

where q measures how accurate the error estimation is, namely, as q becomes larger the bidder is more confident that h_t and m_t's difference is smaller than σ_t. If we consider the extreme case of $q \to \infty$, then m_t is almost surely inside $[h_t - \sigma_t, h_t + \sigma_t]$. Note that (3) always holds regardless of whether the learner observes a hint or a hint interval, and the main difference is that σ_t is only revealed in the latter scenario. The bidder's goal is to maximize the cumulative reward for the whole time horizon, and equivalently, to minimize the overall regret. Let $L := \sum_{t=1}^{T} \sigma_t$ be the total error of hints, the learner aims to achieve a small regret adaptive to the unknown quantity L: the regret is never larger than the no-hint case even for large L, but becomes significantly smaller if L is small.

[*] See https://www.arena-ai.com/ for more information.

[*] In the following we may use $r_{t,a}$ to abbreviate $r(a(v_t); m_t, v_t)$.

3 Regret Gap Between Single Hint and Hint Interval

In this section we identify two unique features of leveraging the hints in first-price auctions. First, the best way to use hints is different from the ones in the literature of online learning: instead of adding an optimistic term based on the hints in the multiplicative weights algorithm, in first-price auctions we should manually add the hints as a new expert. Second, first-price auctions exhibit a provable gap between the regrets under single hints and hint intervals, a phenomenon which does not occur in the hint literature.

3.1 Online Learning with Hints

There is a rich line of literature related to online learning with hints where one aims to achieve data-dependent regret bounds in terms of the variation in the environment [AAGO06, HK11, CYL+12, RS13, SL14, WL18, BLLW19], by taking part of the revealed past losses implicitly as the hint. The algorithm to leverage these hints typically falls into the category of optimistic online mirror descent, where the hint is used to form an optimistic estimate of the current loss. A piece of work that explicitly formulated the hint similar to ours is [WLA20], which used an optimistic EXP4 algorithm to achieve a regret bound of $O(\min\{\sqrt{T}, \sqrt{L}T^{1/4}\})$ when the total error L of the loss predictors is known.

The nature of first-price auction problems, however, is different from the above. The crucial feature is the discontinuity in the reward function (1) when b_t is close to the minimum-bid-to-win m_t, which means that even an accurate prediction of m_t does not imply an accurate reward prediction under every bid b_t. This distinction not only leads to different optimal regrets, but also results in different algorithms for regret minimization, as well as a curious gap between the optimal regrets under single hints and hint intervals. The concept of the hint interval does not offer additional help over single hints in many other online learning problems, and to the best of our knowledge, is new in the literature.

3.2 Regret Gap Between a Single Hint and a Hint Interval

In this section, we show that the above distinctions are already present in a toy example where all private values v_t are the same for $t = 1, 2, \cdots, T$ (say $v_t \equiv 1$). Our first result states that even under this very simple case, there is a strict separation between the regret bounds of knowing single hints and that of knowing hint intervals.

Theorem 1. *For $L \in [1, T]$, $q \in [1, \infty)$, if $v_t \equiv 1$ and the bidder observes a hint interval at each time t, the policy π_1 in Algorithm 1 satisfies*

$$\mathrm{Reg}(\pi_1) \leq C\sqrt{\log T \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}},$$

for a numerical constant C independent of $\{m_t, h_t, \sigma_t\}$. Moreover, the following minimax lower bound holds:

$$\inf_{\pi} \sup_{\{m_t, h_t, \sigma_t\}} \mathrm{Reg}(\pi) \geq c\sqrt{T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}.$$

Here $c > 0$ is a numerical constant independent of (T, L), the supremum is taken over all $\{m_t, h_t, \sigma_t\}$ sequences that satisfy (3), and the infimum is taken over all possible policies π.

Theorem 2. *If $v_t \equiv 1$ and the bidder observes a single hint at each time t, then for every $q \in [1, \infty)$, the policy π_2 in Algorithm 1 with T extra experts bidding $h_t + \{0, 1/T, \cdots, 1 - 1/T\}$ achieves*

$$\mathrm{Reg}(\pi_2) \leq C\,(\log T)^{\frac{1}{2}}\,(T \cdot L)^{\frac{1}{4}},$$

for a numerical constant C independent of $\{m_t, h_t\}$. Moreover, the following minimax lower bound holds:

$$\inf_{\pi} \sup_{\{m_t, h_t\}} \mathrm{Reg}(\pi) \geq c\,(T \cdot L)^{\frac{1}{4}}.$$

Here $c > 0$ is a numerical constant independent of (T, L), the supremum is taken over all $\{m_t, h_t\}$ sequences that satisfy (3), and the infimum is taken over all possible policies π.

The above result shows that there is a strict separation in regret bounds compared to the previous subsection when additional information is given. Observe that in Theorem 1, when q becomes larger, the minimax regret becomes smaller because the error estimation is more accurate. In Theorem 2, however, the lower bound stays the same order as q changes, and is strictly larger than the case of knowing the hint interval as long as $q > 1$. When $q \to \infty$, the upper and lower bound in Theorem 1 gives an optimal $\widetilde{O}(\sqrt{L})$ magnitude for the minimax regret.

The intuition behind this separation can be explained as follows: hint intervals can be considered as single hints plus an additional information of its accuracy. For example, if the hint interval has length $\epsilon \to 0$, one strategy would be to bid exactly as the hint suggests; however, if the bidder do not observe an interval, it is hard for her to wisely arrange the weight she put in the hints given and thus leading to a smaller reward compared to previous case. This distinction turns out to be crucial because of the discontinuity of the reward in (1).

Also note that the upper and lower bounds in Theorems 1 and 2 are tight within logarithmic factors, and exhibit the desired adaptive regret in the hint performance L: when L is as large as T, an $\widetilde{O}(\sqrt{T})$ regret is attainable which is optimal without the hints; as the quality of the hints becomes better, the regret dependence on T is greatly improved.

3.3 Algorithm for Section 3.2

It is a classical result in online learning that the multiplicative weights algorithm in [LW94] leads to a classical regret upper bound of $O\left(\sqrt{T \cdot \log K}\right)$, for K experts and time horizon T under the setting of prediction with expert advice. With hints in first-price auctions, instead of using optimistic online mirror descent in the literature, we modify the multiplicative weights in another way: we run the same multiplicative weights algorithm with an additional expert in the existing set of experts, whose bid depends on the hint h_t.

Specifically, the algorithm to achieve the upper bound in Theorem 1 is Algorithm 1. Construct T base experts, and let the i-th ($i = 1, 2, \ldots, T$) base expert bid $\frac{i}{T}$ at each time t (since $v_t \equiv 1$ the Lipschitz oracle will bid a constant value). It is easy to see that the discretization error incurs an additional regret at most $T \cdot (1/T) = \Theta(1)$. Now with the hint interval, we include an extra expert a^* who bids $h_t + \sigma_t^{q/(q+1)}$ at each time t. Consequently, we have a set of $K = T + 1$ experts in total, containing T base experts and one hint expert. The multiplicative weights algorithm is then applied as follows:

$$p_{t,a} = \frac{\exp\left(\eta_t \cdot \sum_{s<t} r_{s,a}\right)}{\sum_{a' \in [K]} \exp\left(\eta_t \cdot \sum_{s<t} r_{s,a'}\right)}, \quad a \in [K], \quad t = 1, \ldots, T,$$

The main ideas of our algorithm are:

- We discretize the bids b_t (and possibly the private values v_t when we do not assume that $v_t \equiv 1$ later) when constructing oracles without large loss of cumulative rewards compared to continuous ones, since an enumeration of all 1-Lipschitz functions is unrealistic.
- Imagine if the bidder knew the possible range of m_t at the beginning of time t: $[\underline{m}_t, \overline{m}_t]$, then she could bid the upper point $\overline{m}_t$, and this is a $(\overline{m}_t - \underline{m}_t)$-good expert in the sense of [HZF+20, Lemma 1], which is an appropriate tool to handle the discontinuity in the rewards.

We show in Appendix B.1.1 that Algorithm 1 achieves the regret upper bound shown in Theorem 1. As for Theorem 2 we include N hint experts, each of which bids a constant gap Δ_i above h_t at all time, i.e. the first one bids $b_t = h_t$ for all t, and the second one bids $b_t = h_t + \frac{1}{T}$ for all t, ect. We now have a "dense" set of hint experts covering all strategies that bid $b_t = h_t + c$, $c \in [0, 1]$, $t = 1, 2, \cdots, T$, with total loss at most $\Theta(1)$. (See details of the proof in Appendix B.1.2.)

Notice that in both cases we required knowledge of L to decide learning rate η_t at each time t, but this problem can be addressed by existing techniques [ACBG02, YEYS04]. Indeed, the upper bound in Theorem 2 still holds even if L is not known in advance.

Algorithm 1: Multiplicative Weights with Hint Intervals

Input: Time horizon T; Hints accuracy L; q.
Output: A bidding policy π.
Initialization: Construct $K = T + 1$ experts, with the i-th ($i = 1, 2, \ldots, T$) base expert bidding $\frac{i}{T}$ at each time t and an extra expert a^* bidding $h_t + \sigma_t^{q/(q+1)}$ at each time t;
for $a \in [K]$ **do**
  Initialize $r_{s,a} \leftarrow 0$;
end
for $t \in \{1, 2, \cdots, T\}$ **do**
  The bidder receives a private value $v_t \equiv 1$;
  The bidder observes hint $h_t \in [0, 1]$, along with its accuracy σ_t;
  Set $b_{t,a^*} \leftarrow h_t + \sigma_t^{\frac{q}{q+1}}$;
  Set learning rate $\eta_t \leftarrow \min\left\{\frac{1}{4}, \sqrt{\frac{\log K}{\sum_{s \le t} \sigma_s^{\frac{q}{q+1}}}}\right\}$;
  Let $b_t \leftarrow b_{t,a}$ with probability

$$p_{t,a} = \frac{\exp\left(\eta_t \cdot \sum_{s<t} r_{s,a}\right)}{\sum_{a' \in [K]} \exp\left(\eta_t \cdot \sum_{s<t} r_{s,a'}\right)}.$$

  The bidder receives others' highest bid m_t;
  for $a \in [K]$ **do**

$$r_{t,a} \leftarrow r_{t,a} + r(b_{t,a}; v_t, m_t).$$

  end
  Update $R_t \leftarrow R_{t-1} + r(b_t; v_t, m_t)$;
end

3.4 Varying Private Prices

When private values could vary, oracles cannot bid the same price anymore. The following theorem holds, again showing a strict separation and furthermore indicates that for varying v_t, hints does not help much for continuous bidding value.

Theorem 3. *Let $L \in [1, T]$, $q \in [1, \infty)$, and $v_t \in [0, 1]$ for $t = 1, 2, \ldots, T$. If the bidder observes hint intervals, the following characterization of the minimax regret holds:*

$$\inf_{\pi} \sup_{\{v_t, m_t, h_t, \sigma_t\}} \text{Reg}(\pi) = \widetilde{\Theta}\left(\min\left\{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}, \sqrt{T}\right\}\right),$$

where the supremum is taken over rewards under all possible $\{v_t, m_t, h_t, \sigma_t\}$ sequences and hints that satisfy (3), and the infimum is taken over all possible policies π.

Theorem 4. *Let $L \in [1, T]$, and $v_t \in [0, 1]$ for $t = 1, 2, \ldots, T$. If the bidder observes single hints, then $\forall q \in [1, \infty)$, the following characterization of the minimax regret holds:*

$$\inf_{\pi} \sup_{\{v_t, m_t, h_t, \sigma_t\}} \text{Reg}(\pi) = \widetilde{\Theta}\left(\sqrt{T}\right),$$

where the supremum is taken over rewards under all possible $\{v_t, m_t, h_t, \sigma_t\}$ sequences and hints that satisfy (3), and the infimum is taken over all possible policies π.

Although both results in Theorems 3 and 4 are tight within logarithmic factors, they are pessimistic results. When the hint intervals are observed, Theorem 3 shows that the help from the hint exhibits a thresholding phenomenon: either bidding without hints or only bidding the hints is optimal. For example, when $q = \infty$, the minimax regret is simply a tedious quantity $\widetilde{\Theta}(\min\{L, \sqrt{T}\})$. When there are only single hints, Theorem 4 even shows that the hint is of no help unless its quality is very high, i.e. $L \le 1$. This pessimistic situation is alleviated in the next section, by imposing an additional assumption that others' bids m_t are only supported on a few locations.

4 Exploiting the Sparsity of Others' Bids

The previous section gives a pessimistic result that hints help only to the same extent of bidding the hint itself. To mitigate this drawback, we identify a useful structure in practical online first-price auctions: the maximum competing bids m_t are supported on only a few locations. See Figure 1 for a typical example in real data. The reason why sparsity arises in practice is partially due to the scenario where the maximum competing bid is the reserve price set by the seller.

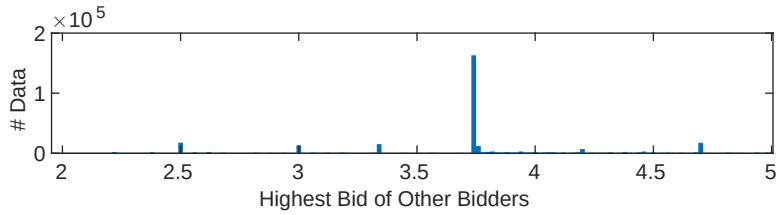


Figure 1: Histogram of others' highest bid in real data.

Assuming that m_t is only supported on K locations, the central question in this section is as follows:

How does the sparsity improve the minimax regret? Can we devise a learning algorithm that is adaptive to the parameter K (and other parameters such as L)?

4.1 Minimax Regret with Sparsity

The central result of this section is the following characterization of the minimax regret with sparsity:

Theorem 5. *For $q \in [1, \infty)$ and varying private prices, suppose that the minimum-bid-to-win m_t only takes K support values and the hint intervals are available. The regret is upper bounded by:*

$$\inf_{\pi} \sup_{\{v_t, m_t, h_t, \sigma_t\}} \mathrm{Reg}(\pi) = O\left(\min\left\{\sqrt{\log T \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}} \cdot K}, T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}, \sqrt{T}\right\}\right),$$

where the supremum is taken over all $\{v_t, m_t, h_t, \sigma_t\}$ sequences and hints that satisfies (3), and the infimum is taken over all possible policies π. In addition, the following minimax lower bound for the regret holds:

$$\inf_{\pi} \sup_{\{v_t, m_t, h_t, \sigma_t\}} \mathrm{Reg}(\pi) = \Omega\left(\min\left\{\sqrt{T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}} \cdot K}, T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}, \sqrt{T}\right\}\right).$$

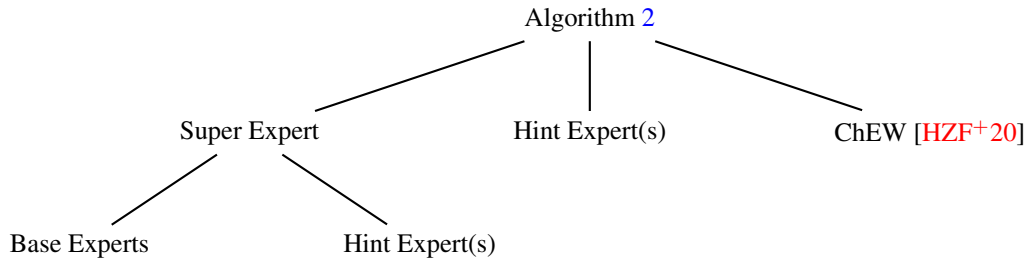
Theorem 5 shows that the minimax regret exhibits an elbow with respect to the sparsity: the regret grows sublinearly with K for small K, but reduces to the fixed value $\Theta\left(\min\left\{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}, \sqrt{T}\right\}\right)$ of Theorem 3 when K is large. Moreover, the better the hint quality is, the more helpful the sparsity will be. For example, if $L = \Theta(1)$, sparsity helps reduce the regret whenever $K < T^{1/(q+1)}$; as the other extreme, if $L = \Theta(T)$, the classical $\Theta(\sqrt{T})$ regret is always unavoidable, no matter how small K is.

4.2 Meta Algorithm Adaptive to Unknown L and K

In the proof of Theorem 5 (see appendix C.1) we prove the three upper bounds can indeed be achieved separately. If a bidder observes the overall quality of hints L at the beginning, she can easily choose the best among those three bounds and achieve an optimal. It is not straightforward though, to achieve them simultaneously without this knowledge. For instance, to achieve each single upper bound, there is a data-driven algorithm agnostic to L (such as using the idea in [ACBG02]); however, the choice of which algorithm to use still depends on the knowledge of L.

Algorithm 2 (pseudocode in Appendix A) with meta-experts addresses this problem and ensures to achieve the optimal one among the three upper bounds for any $L = \Omega(1)$. The intuition is to give hints a higher priority if the error is small. Meanwhile, with this meta structure, it is possible to set different fine tuning strategy for the two layers. By including three "Meta Experts" in the upper layer,

whose strategies are indeed output of algorithms rather than pre-designed oracles, we can achieve the minimum of the three regret bounds listed in Theorem 5.



Theorem 6. *Algorithm 2 achieves the regret upper bound of Theorem 5, while it is agnostic to the knowledge of L.*

Theorem 6 is also true when the support size K and support locations are unknown together with L.* Using similar ideas of the doubling trick, we can deal with unknown K. At the beginning assume a constant value K_0 (e.g. $K_0 = 8$), and run the above algorithm until the number of observed supports up to time t exceeds the current value. If that happens, we double the value of K: $K_{i+1} = 2 \cdot K_i$ ($i \in \mathbb{N}$), and restart the learning algorithm. Since assuming a larger K than reality only makes the oracle stronger, substituting the regret upper bound (4) below to this strategy leads to regret upper bound (let $M := \left\lceil \log\left(\frac{K}{8}\right)\right\rceil + 1$):

$$O\left(\sqrt{\log T} \cdot \sum_{i=0}^{M-1} \sqrt{K_i \cdot \sum_{t=T_i}^{T_{i+1}} \sigma_t^{\frac{q}{q+1}}}\right) \leq O\left(\sqrt{\log T} \cdot \sqrt{(K_0 + \ldots + K_M) \cdot \left(\sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}\right)}\right)$$
$$\leq O\left(\sqrt{\log T \cdot 2K \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}\right),$$

where $K_i := K_0 \cdot 2^i$ and T_i are the dividing points when current number of supports exceeds current K ($T_{M+1} = T$).

For unknown support values construct an expert set as follows. We divide y-axis (bidding price) into T small intervals, with the i-th interval being $[\frac{i}{T}, \frac{i+1}{T}]$ ($i = 0, 1, \cdots, T-1$). Our goal is to allocate the K supports into these T intervals, with each interval containing at most one support (two supports in the same interval could be merged with an additional $O(1)$ regret). The total number of allocations is $\binom{T}{K}$, and for each allocation, the oracle can only choose from T^K bidding functions (cf. the proof of Theorem 5), thus the size of the complete expert set is:

$$|\text{expert set}| \leq \binom{T}{K} \cdot T^K \leq T^{2K}.$$

Then the regret upper bound can be written as (using similar analysis to Appendix B.1.1)

$$O\left(\sqrt{\log(T^{2K}) \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}\right) = O\left(\sqrt{K \cdot \log T \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}\right). \tag{4}$$

4.2.1 Single Hints Case

In Section 3 we mainly showed that there is a strict difference in the regret bound for single hints and hint intervals case. Here we also provide results for single hints case for completeness.

Theorem 7. *For $q \in [1, \infty)$ and varying private prices, if the minimum bid to win takes K support values and the bidder only observes a single hint at each time then the regret bounds hold:*

$$\inf_{\pi} \sup_{\{v_t, m_t, h_t, \sigma_t\}} \text{Reg}(\pi) = \widetilde{\Theta}\left(\min\left\{\sqrt{T}, \sqrt{\sqrt{LT} \cdot K}\right\}\right),$$

where the supremum is taken over all K-sparse m_t sequence and hints that satisfies (3), and the infimum is taken over all possible policies π.

*Since our algorithm is adaptive to T as discussed in the previous section, it is equivalent to say no parameter about the whole game is known.

5 Real-data Experiments

This section presents several real-data experiments in repeated first-price auctions based on practical bidding data, where the hint is the context-based prediction provided by blackbox machine learning models. For business confidentiality we do not disclose further information about the datasets.

Our experiments are run on auction datasets from the first-price auctions on real-world sites, with around 0.38 million data points, where each data point is a quadruple of scalars (v_t, m_t, h_t, σ_t). Here v_t is the private value and m_t is the minimum bid to win for each auction. The private value v_t is computed by Verizon Media based on an independent learning scheme not relying on the auction, and is therefore taken as given. The quantity m_t is the minimum bid to win and is returned by the platform after each auction, which is by definition the other bidders' highest bid (possibly including the seller's reserve price and measured up to 1 cent). These datasets have already been pruned to only contain data points with $v_t > m_t$, for otherwise the bidder never wins regardless of her bids. Hints h_t and its accuracy σ_t are provided by fitting a lognormal model using other contextual information.

We did two parts of experiments. In the first part, we applied the main insight in Section 3 that for fixed private value, given extra information of hints helps reduce the regret for optimal policy, and hint intervals outperforms single hints, as long as $q > 1$. We allocate all data points to separate bins according to the private value v_t and each bin is a subproblem described in Section 3. Figure 2(a) compares the results of binned exponential weighting (without hint) and whether hint intervals or single hints are given. We can observe although the performance of hints itself is obviously bad, by incorporating hint into the learning method, we could improve the cumulative reward by 2.96% if knowing point estimation only, and by 3.55% if an interval is provided in each round.

We also provide a polynomial time algorithm for conducting the meta algorithm we proposed before. Even with T^K experts, we may use dynamic programming to achieve a space complexity $O(DK)$ and time complexity $O(DKT)$, where D is parameter for the discretization on private values. We implemented Algorithm 3 on dataset with support size $K = 5$.

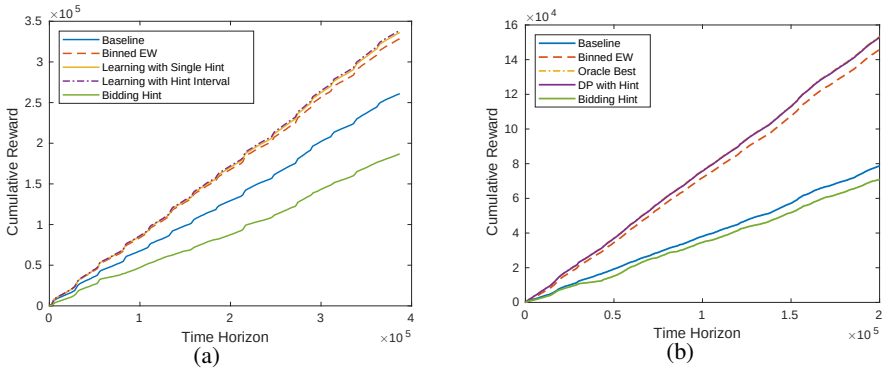


Figure 2: Cumulative rewards as a function of time, based on Algorithms 1 and 3. Panel (a) uses a binning method to show the improvement by incorporating hints. The yellow solid line corresponds to learning with single hints, while the dashdot line corresponds to hint intervals. As comparisons, we also include a baseline algorithm (blue) provided by Verizon Media, the simple exponential weights algorithm without hints (red), and just bidding the hint (green). Panel (b) illustrates the performance of our Algorithm 3 (purple solid line), which almost coincides with the performance of the best increasing and 1-Lipschitz function for the oracle (yellow dashed line). The other algorithms are the same as Panel (a).

6 Conclusion

In this work we study the overall regret for a particular bidder without further assumption for other bidders in repeated first-price auction. We target at the case when an additional information is given to only this bidder at each round. We show that even in a simple setting, there is a strict gap between two different forms of hints, where in either one the upper bound and lower bound matches with regard to log factor. We further consider the case when others' highest bid lies in finite support,

and provide modified algorithm as well as matching regret upper and lower bounds. While not knowing the critical parameters for the whole game in the beginning, our adaptive algorithm always achieves the best among three different upper bounds. Finally, we appreciate data from company that corresponds with our framework, and carry out two parts of experiments on it. The first one is mainly to show the performance improvement with the information given, and in the second part, we provide a way to implement our algorithm in polynomial time and show corresponding results.

7 Acknowledgements

This work was supported in part by NSF awards CCF-2106467 and CCF-2106508. Zhengyuan Zhou acknowledges the generous support from New York University's Center for Global Economy and Business faculty research grant. We would like to thank the NeurIPS 2022 reviewers for their constructive feedback and Eric Ordentlich for helpful discussions in the early stage of this work.

References

[AAGO06] Chamy Allenberg, Peter Auer, László Györfi, and György Ottucsák. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In *International Conference on Algorithmic Learning Theory*, pages 229–243. Springer, 2006.

[ACBG02] Peter Auer, Nicolo Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

[App18] AppNexus. Demystifying auction dynamics for digital buyers and sellers. AppNexus white paper, 2018.

[Ben18a] Ross Benes. First-price auctions are driving up ad prices. https://www.emarketer.com/content/first-price-auctions-are-driving-up-ad-prices, 2018. Published: October 17, 2018.

[Ben18b] Ross Benes. How ssps use deceptive price floors to squeeze ad buyers. https://digiday.com/marketing/ssps-use-deceptive-price-floors-squeeze-ad-buyers, 2018. Published: September 13, 2017.

[BFG21] Ashwinkumar Badanidiyuru, Zhe Feng, and Guru Guruganesh. Learning to bid in contextual first price auctions. *arXiv preprint arXiv:2109.03173*, 2021.

[BGM+19] Santiago Balseiro, Negin Golrezaei, Mohammad Mahdian, Vahab Mirrokni, and Jon Schneider. Contextual bandits with cross-learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[BGM+21] Santiago Balseiro, Negin Golrezaei, Mohammad Mahdian, Vahab Mirrokni, and Jon Schneider. Contextual bandits with cross-learning, 2021.

[BLLW19] Sébastien Bubeck, Yuanzhi Li, Haipeng Luo, and Chen-Yu Wei. Improved path-length regret bounds for bandits. In *Conference On Learning Theory*, pages 508–528. PMLR, 2019.

[CBL06] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[CYL+12] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, pages 6–1. JMLR Workshop and Conference Proceedings, 2012.

[Dav19] Jessica Davies. What to know about google's implementation of first-price ad auctions. https://digiday.com/media/buyers-welcome-auction-standardization-as-google-finally-goes-all-in-on-first-price/, 2019. Published: September 6, 2019.

[DRS19] Stylianos Despotakis, R Ravi, and Amin Sayedi. First-price auctions in online display advertising. *Available at SSRN 3485410*, 2019.

[HK11] Elad Hazan and Satyen Kale. Better algorithms for benign bandits. *Journal of Machine Learning Research*, 12(4), 2011.

[HZF+20] Yanjun Han, Zhengyuan Zhou, Aaron Flores, Erik Ordentlich, and Tsachy Weissman. Learning to bid optimally and efficiently in adversarial first-price auctions. *arXiv preprint arXiv:2007.04568*, 2020.

[HZW20] Yanjun Han, Zhengyuan Zhou, and Tsachy Weissman. Optimal no-regret learning in repeated first-price auctions. *arXiv preprint arXiv:2003.09795*, 2020.

[Kle04] Paul Klemperer. *Auctions: theory and practice*. Princeton University Press, 2004.

[LR00] David Lucking-Reiley. Vickrey auctions in practice: From nineteenth-century philately to twenty-first-century e-commerce. *Journal of economic perspectives*, 14(3):183–192, 2000.

[LRBPR07] David Lucking-Reiley, Doug Bryan, Naghi Prasad, and Daniel Reeves. Pennies from ebay: The determinants of price in online auctions. *The journal of industrial economics*, 55(2):223–233, 2007.

[LW94] N Littlestone and MK Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[RS13] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013.

[SL14] Jacob Steinhardt and Percy Liang. Adaptivity and optimism: An improved exponentiated gradient algorithm. In *International Conference on Machine Learning*, pages 1593–1601. PMLR, 2014.

[Slu17] Sarah Sluis. Big changes coming to auctions, as exchanges roll the dice on first-price. https://www.adexchanger.com/platforms/big-changes-coming-auctions-exchanges-roll-dice-first-price/, 2017. Published: September 5, 2017.

[Wag19] Kurt Wagner. Digital advertising in the us is finally bigger than print and television. https://www.vox.com/2019/2/20/18232433/digital-advertising-facebook-google-growth-tv-print-emarketer-2019, 2019. Published: Feburary 20, 2019.

[WL18] Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory*, pages 1263–1291. PMLR, 2018.

[WLA20] Chen-Yu Wei, Haipeng Luo, and Alekh Agarwal. Taking a hint: How to leverage loss predictors in contextual bandits? In *Conference on Learning Theory*, pages 3583–3634. PMLR, 2020.

[YEYS04] Rani Yaroshinsky, Ran El-Yaniv, and Steven S Seiden. How to better use expert advice. *Machine Learning*, 55(3):271–309, 2004.

[ZKH+21] Wei Zhang, Brendan Kitts, Yanjun Han, Zhengyuan Zhou, Tingyu Mao, Hao He, Shengjun Pan, Aaron Flores, San Gultekin, and Tsachy Weissman. Meow: A space-efficient nonparametric bid shading algorithm. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3928–3936, 2021.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Results are stated and properly qualified.
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] All theorem's have clearly stated assumptions
   (b) Did you include complete proofs of all theoretical results? [Yes] All proofs are detailed in the Appendix
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Pseudocode is included but data is confidential
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Didou report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Pseudocode of Algorithm 2

Algorithm 2: Meta-Expert Learning Algorithm

Input: Time horizon T; support size K; accuracy of error information $(\sigma_1, \cdots, \sigma_T)$; norm parameter $q \in [1, \infty]$.
Output: A bidding policy π.
Initialization: Construct T^K base experts $\{f_i\}$ $(i = 1, 2, \cdots, T^K)$ that cover the oracle with cumulative reward difference at most $O(1)$ (as in the proof of Theorem 5);
for $i = 1, 2, \cdots T^K$ **do**
  Initialize $R_{0,f_i} \leftarrow 0$;
end
Initialize $R_{0,h} \leftarrow 0$, $R_{0,g} \leftarrow 0$, $R_{0,f} \leftarrow 0$;
Initialize $L_0 \leftarrow 0$;
for $t \in \{1, 2, \cdots, T\}$ **do**
  The bidder receives private value $v_t \in [0, 1]$;
  Set learning rate $\eta_{t,1} \leftarrow \min\left\{\frac{1}{4}, \sqrt{\frac{K \log T}{L_t}}\right\}$;
  The bidder observes hint $h_t \in [0, 1]$, along with its accuracy σ_t;
  $L_t \leftarrow L_{t-1} + \sigma_t^{\frac{q}{q+1}}$;
  Set $b_{t,h} \leftarrow h_t + \sigma_t^{\frac{q}{q+1}}$;
  Sample $b_{t,g}$ according to ChEW policy;
  for $i = 1, 2, \cdots, T^K$ **do**
    Let $b_{t,f} \leftarrow f_i(v_t)$ with probability

$$p_{t,i} := \frac{\exp\left(\eta_{t,1} R_{t-1,f_i}\right)}{\exp\left(\eta_{t,1} R_{t-1,h}\right) + \sum_{i'=1}^{T^K} \exp\left(\eta_{t,1} R_{t-1,f_{i'}}\right)}.$$

  end
  Let $b_{t,f} \leftarrow b_{t,h}$ with probability $p_{t,T^K+1} := \frac{\exp(\eta_{t,1} R_{t-1,h})}{\exp(\eta_{t,1} R_{t-1,h}) + \sum_{i'=1}^{T^K} \exp\left(\eta_{t,1} R_{t-1,f_{i'}}\right)}$;
  Sample $b_{t,f} \sim p_t$;
  Set learning rate $\eta_{t,2} \leftarrow \min\left\{\frac{1}{4}, \sqrt{\frac{\log 3}{L_t}}\right\}$;
  for $i \in \{f, g, h\}$ **do**

$$P_{t,i} = \frac{\exp\left(\eta_{t,2} R_{t-1,i}\right)}{\sum_{i' \in \{f,g,h\}} \exp\left(\eta_{t,2} R_{t-1,i'}\right)};$$

  end
  The bidder samples policy $i \sim P_t$ and bids $b_{t,i}$;
  The bidder receives others' highest bid m_t;
  for $i = 1, 2, \cdots, T^K$ **do**

$$R_{t,f_i} \leftarrow R_{t-1,f_i} + r(f_i(v_t); v_t, m_t).$$

  end
  for $i \in \{f, g, h\}$ **do**
    Update $R_{t,i} \leftarrow R_{t-1,i} + r(b_{t,i}; v_t, m_t)$;
  end
end

The algorithm has a tree structure with the nodes in the upper layer representing algorithms instead of specific oracles. In Algorithm 2, the upper nodes are respectively: the algorithm that achieves the regret upper bound in Theorem 5 described in Appendix C.1, "ChEW" algorithm to achieve $\widetilde{O}(\sqrt{T})$ regret bound proposed in [HZF+20], and a single expert which bids $h_t + \sigma_t^{q/(q+1)}$ each time. The probability distribution $P_{t,i}$ runs the multiplicative weights update on the above strategies (see details in Appendix C.2).

B Proof of Main Result in Section 3

B.1 Proof of Regret Upper Bounds in Theorem 1 and Theorem 2

B.1.1 Proof of Upper Bound in Theorem 1.

We prove a slightly stronger result than Theorem 1:

Lemma 1. *If $v_t \equiv 1$ and the bidder observes σ_t at each time t, then the following regret upper bound holds for Algorithm 1:*

$$\sup_{\{m_t, h_t, \sigma_t\}} \mathrm{Reg}(\pi_1) = O\left(\log T + \sqrt{\log T \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}\right),$$

with $\mathrm{Reg}(\pi)$ defined in (2), and the supremum is taken over all m_t sequences and hints that satisfy (3), and the infimum is taken over all possible policies π.

Proof. The following is similar to proof of Theorem 3 in [HZF+20]. As in the standard analysis of multiplicative weights [CBL06], define:

$$\phi_t = \frac{1}{K}\sum_{a=1}^{K} \exp\left(\eta_t \cdot \sum_{s<t} r_{s,a}\right), \quad t = 1, \ldots, T+1.$$

Recall that $K = T+1$ and a^* is the extra expert. We translate every $r_{t,a}$ by $-r_{t,a^*}$ to ensure that $r_{t,a} \in [-1, 1]$ and $r_{t,a^*} = 0$. Then for $t \in [T]$, Jensen's inequality with $\eta_t/\eta_{t+1} \geq 1$ gives

$$\begin{aligned}
(\phi_{t+1})^{\frac{\eta_t}{\eta_{t+1}}} &= \left[\frac{1}{K}\sum_{a=1}^{K} \exp\left(\eta_{t+1} \cdot \sum_{s<t+1} r_{s,a}\right)\right]^{\frac{\eta_t}{\eta_{t+1}}} \\
&\leq \frac{1}{K}\sum_{a=1}^{K} \left[\exp\left(\eta_{t+1} \cdot \sum_{s<t+1} r_{s,a}\right)^{\frac{\eta_t}{\eta_{t+1}}}\right] \\
&= \phi_t \sum_{a=1}^{K} p_{t,a} \cdot \exp\left(\eta_t \cdot r_{t,a}\right) =: \phi_t \mathbb{E}[\exp\left(\eta_t X_t\right)].
\end{aligned}$$

Here X_t is a random variable that takes value $r_{t,a}$ with probability $p_{t,a}$. Now using Bernstein's inequality

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\lambda \mathbb{E}[X] + (e^{\lambda} - \lambda - 1)\mathsf{Var}(X)\right),$$

with $|X - \mathbb{E}[X]| \leq 1$ almost surely, we have

$$\frac{\log \phi_{t+1}}{\eta_{t+1}} - \frac{\log \phi_t}{\eta_t} \leq \mathbb{E}[X_t] + \frac{e^{\eta_t} - \eta_t - 1}{\eta_t}\mathsf{Var}(X_t) \leq \mathbb{E}[X_t] + \eta_t \mathsf{Var}(X_t),$$

where the last inequality is due to $e^x - x - 1 \leq x^2$ for $x \in [0, 1]$. Define $r_t^* := \max_{a \in [K]} r_{t,a}$, we have

$$\mathsf{Var}(X_t) \leq \mathbb{E}[(r_t^* - X_t)^2] \leq 1 \cdot \mathbb{E}[r_t^* - X_t] = r_t^* - \mathbb{E}[X_t].$$

By telescoping and defining $\eta_{T+1} := \eta_T$,

$$\frac{\log \phi_{T+1}}{\eta_T} = \sum_{t=1}^{T}\left[\frac{\log \phi_{t+1}}{\eta_{t+1}} - \frac{\log \phi_t}{\eta_t}\right] \leq \sum_{t=1}^{T} \mathbb{E}[X_t] + \sum_{t=1}^{T} \eta_t \left(r_t^* - \mathbb{E}[X_t]\right). \tag{5}$$

For the left-hand side of (5), we also have

$$\log \phi_{T+1} \geq \eta_T \cdot \max_{a \in [K]} \sum_{s=1}^{T} r_{t,a} - \log K. \tag{6}$$

Combining (5) and (6),

$$\max_{a\in[K]} \sum_{t=1}^{T} r_{t,a} \le \frac{\log K}{\eta_T} + \sum_{t=1}^{T} (1-\eta_t)\cdot \mathbb{E}[X_t] + \sum_{t=1}^{T} \eta_t \cdot r_t^*. \tag{7}$$

Rearranging (7) leads to the following upper bound on the cumulative regret:

$$\max_{a\in[K]} \sum_{t=1}^{T} r_{t,a} - \sum_{t=1}^{T} \mathbb{E}[X_t] \le \frac{\log K}{\eta_T} + \sum_{t=1}^{T} \eta_t r_t^* - \sum_{t=1}^{T} \eta_t \cdot \mathbb{E}[X_t]. \tag{8}$$

Let $V_T := (\log K)/\eta_T + \sum_{t=1}^{T} \eta_t r_t^*$, it remains to upper bound the last term of (8). To do so, note that (7) holds for any intermediate value of $t \in [T]$ as well. Since $\max_{a\in[K]} \sum_{t=1}^{T} r_{t,a} \ge \sum_{t=1}^{T} r_{t,a^*} = 0$, for every $t \in [T]$ we have

$$S_t := \sum_{s=1}^{t} (1-\eta_s)\cdot \mathbb{E}[X_s] \ge -\frac{\log K}{\eta_{t+1}} - \sum_{s=1}^{t} \eta_s \cdot r_s^* = -V_t \ge -V_T,$$

where the last inequality is due to $\eta_{t+1} \ge \eta_T$ and $r_t^* \ge r_{t,a^*} = 0$ for every $t \in [T]$. Consequently,

$$\begin{aligned}
-\sum_{t=1}^{T} \eta_t \cdot \mathbb{E}[X_t] &= -\sum_{t=1}^{T} (S_t - S_{t-1}) \cdot \frac{\eta_t}{1-\eta_t} \\
&= -\sum_{t=1}^{T-1} S_t \cdot \left(\frac{\eta_t}{1-\eta_t} - \frac{\eta_{t+1}}{1-\eta_{t+1}} \right) - S_T \cdot \frac{\eta_T}{1-\eta_T} \\
&\le V_T \sum_{t=1}^{T-1} \left(\frac{\eta_t}{1-\eta_t} - \frac{\eta_{t+1}}{1-\eta_{t+1}} \right) + V_T \cdot \frac{\eta_T}{1-\eta_T} \\
&= \frac{V_T \eta_1}{1-\eta_1} \le V_T,
\end{aligned}$$

where we have used that $1/4 \ge \eta_1 \ge \eta_2 \ge \ldots \ge \eta_T > 0$. Plugging this inequality back into (7) gives

$$\max_{a\in[K]} \sum_{t=1}^{T} r_{t,a} - \sum_{t=1}^{T} \mathbb{E}[X_t] \le 2V_T. \tag{9}$$

Finally it remains to upper bound $\mathbb{E}[V_T]$, where the expectation is taken with respect to the randomness in the hint sequence $\{h_t\}_{t=1}^{T}$. Since the definition of the expert a^* gives that

$$\begin{aligned}
r_t^* &\le (1-m_t) - (1-h_t-\sigma_t^{q/(q+1)})\mathbb{1}(h_t+\sigma_t^{q/(q+1)} \ge m_t) \\
&\le \begin{cases} h_t + \sigma_t^{q/(q+1)} - m_t & \text{if } h_t+\sigma_t^{q/(q+1)} \ge m_t \\ 1 & \text{if } h_t+\sigma_t^{q/(q+1)} < m_t \end{cases},
\end{aligned}$$

we conclude that

$$\begin{aligned}
\mathbb{E}[r_t^*] &\le \mathbb{P}(h_t+\sigma_t^{q/(q+1)} < m_t) + \mathbb{E}[|h_t+\sigma_t^{q/(q+1)} - m_t|] \\
&\le \frac{\mathbb{E}[|h_t-m_t|^q]}{(\sigma_t^{q/(q+1)})^q} + (\mathbb{E}[|h_t-m_t|^q])^{1/q} + \sigma_t^{q/(q+1)} \\
&\le 2\sigma_t^{q/(q+1)} + \sigma_t \le 3\sigma_t^{q/(q+1)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[V_T] &\le \frac{\log K}{\eta_T} + \sum_{t=1}^{T} \eta_t \mathbb{E}[r_t^*] \\
&\le 4\log K + \sqrt{\sum_{t=1}^{T} \sigma_t^{q/(q+1)} \log K} + 3\sum_{t=1}^{T} \sqrt{\frac{\log K}{\sum_{s\le t} \sigma_s^{q/(q+1)}}} \cdot \sigma_t^{q/(q+1)} \\
&\le 4\log K + 7\sqrt{\sum_{t=1}^{T} \sigma_t^{q/(q+1)} \log K},
\end{aligned}$$

where the last inequality follows from

$$\sum_{i=1}^{n} \frac{a_i}{\sqrt{\sum_{j\le i} a_j}} \le \sum_{i=1}^{n} \int_{\sum_{j\le i-1} a_j}^{\sum_{j\le i} a_j} \frac{\mathrm{d}x}{\sqrt{x}} = \int_0^{\sum_{i=1}^n} \frac{\mathrm{d}x}{\sqrt{x}} = 2\sqrt{\sum_{i=1}^{n} a_i}$$

for any non-negative reals $a_1, \cdots, a_n$. Plugging the above upper bound of $\mathbb{E}[V_T]$ into (9) completes the proof of the lemma. □

Theorem 1 follows from Lemma 1 and the following Jensen's inequality:

$$\sqrt{\log T \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}} \le \sqrt{\log T \cdot T \cdot \left(\frac{\sum_{t=1}^{T} \sigma_t}{T}\right)^{\frac{q}{q+1}}} \le \sqrt{\log T \cdot L^{q/(q+1)} \cdot T^{1/(q+1)}}.$$

B.1.2 Proof of Upper Bound in Theorem 2.

To achieve the upper bound of Theorem 2, we construct the same T base experts as Algorithm 1, as well as T additional experts who bid $h_t + i/T, i \in [T]$ at each time t. Then at an additional $O(1)$ cost in the final regret, the additional experts include an expert who bids $h_t + \sqrt{L/T}$ at each time t. Using the same analysis in the proof of Lemma 1, this algorithm achieves a regret upper bound

$$\mathrm{Reg}(\pi_2) \le 2\left(\frac{\log(2T)}{\eta} + \eta \cdot \sum_{t=1}^{T} \mathbb{E}[r_t^*]\right),$$

where $\eta > 0$ is a fixed learning rate, and

$$r_t^* \le \begin{cases} h_t + \sqrt{L/T} - m_t & \text{if } h_t + \sqrt{L/T} \ge m_t \\ 1 & \text{if } h_t + \sqrt{L/T} < m_t \end{cases}.$$

Consequently,

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[r_t^*] &\le \sum_{t=1}^{T} \mathbb{E}[|h_t + \sqrt{L/T} - m_t|] + \sum_{t=1}^{T} \mathbb{P}(h_t + \sqrt{L/T} < m_t) \\
&\le \sqrt{LT} + \mathbb{E}\left[\sum_{t=1}^{T} |h_t - m_t|\right] + \frac{1}{\sqrt{L/T}} \mathbb{E}\left[\sum_{t=1}^{T} |h_t - m_t|\right] \\
&\le 2\sqrt{LT} + L \le 3\sqrt{LT},
\end{aligned}$$

as $1 \le L \le T$. Now choosing $\eta = \min\{1/4, \sqrt{(\log T)/\sqrt{LT}}\}$ leads to the regret upper bound $O\left((\log T)^{\frac{1}{2}} (T \cdot L)^{\frac{1}{4}}\right)$.

B.2 Proof of Regret Lower Bounds in Theorem 1 and Theorem 2

B.2.1 Proof of Lower Bound in Theorem 1.

Proof. We use Le Cam's Two-Point method. Construct hint and minimum bid to win as follows: Let $h_t = \frac{1}{2}, t = 1, \ldots, T$ and σ_t be the same for all t such that $\sigma^{\frac{q}{q+1}} \le \frac{1}{4}$. Consider the following two CDFs for $m_t \in [0, 1]$:

$$G_1(x) = \begin{cases} 0, & \text{if } 0 < x < \frac{1}{2} \\ 2 \cdot (1 - \bar{x} + \delta), & \text{if } \frac{1}{2} < x < \bar{x} \\ 1, & \text{if } \bar{x} < x < 1 \end{cases}, \quad G_2(x) = \begin{cases} 0, & \text{if } 0 < x < \frac{1}{2} \\ 2 \cdot (1 - \bar{x} - \delta), & \text{if } \frac{1}{2} < x < \overline{x} \\ 1, & \text{if } \overline{x} < x < 1 \end{cases},$$

where $\bar{x} := \frac{1}{2} + \frac{1}{2} \cdot \sigma^{\frac{q}{q+1}}$ and let $\delta < \frac{1}{2} \cdot \sigma^{\frac{q}{q+1}}$. Easy to observe the above construction satisfies:

$$\mathbb{E}[|m_t - h_t|^q] \leq 2 \cdot (\frac{1}{2} \cdot \sigma^{\frac{q}{q+1}} + \delta) \cdot \left(\frac{1}{2} \cdot \sigma^{\frac{q}{q+1}}\right)^q \leq \sigma^{\frac{q}{q+1}} \cdot \left(\sigma^{\frac{q}{q+1}}\right)^q = \sigma^q.$$

Let $r_1(v_t, b_t)$ and $r_2(v_t, b_t)$ be the expected instantaneous reward under CDFs G_1 and G_2. Then under the above construction:

$$\begin{aligned}
&\max_{b\in[0,1]} r_1(1,b) = r_1(1, \frac{1}{2}) = \frac{1}{2} \cdot \frac{1-\bar{x}+\delta}{1-\frac{1}{2}} = 1 - \bar{x} + \delta, \\
&\max_{b\in[0,1]} r_2(1,b) = r_2(1,\bar{x}) = 1 - \bar{x}, \\
&\max_{b\in[0,1]} (r_1(1,b) + r_2(1,b)) = r_1(1,\bar{x}) + r_2(1,\bar{x}) = 2 \cdot (1-\bar{x}).
\end{aligned}$$

Therefore, for any $b_t \in [0,1]$,

$$\begin{aligned}
&\left(\max_{b\in[0,1]} r_1(1,b) - r_1(1,b_t)\right) + \left(\max_{b\in[0,1]} r_2(1,b) - r_2(1,b_t)\right) \\
&\geq \left(\max_{b\in[0,1]} r_1(1,b)\right) + \left(\max_{b\in[0,1]} r_2(1,b)\right) - \max_{b\in[0,1]} (r_1(1,b) + r_2(1,b)) \\
&= (1-\bar{x}+\delta) + (1-\bar{x}) - 2\cdot(1-\bar{x}) = \delta.
\end{aligned}$$

Thus we have for any policy π,

$$\begin{aligned}
\sup_G \text{Reg}(\pi) &\geq \frac{1}{2}\mathbb{E}_{G_1}[\text{Reg}(\pi)] + \frac{1}{2}\mathbb{E}_{G_2}[\text{Reg}(\pi)] \\
&= \frac{1}{2}\sum_{t=1}^{T}\left(\mathbb{E}_{P_1^t}\left[\max_{b\in[0,1]} r_1(1,b) - r_1(1,b_t)\right] + \mathbb{E}_{P_2^t}\left[\max_{b\in[0,1]} r_2(1,b) - r_2(1,b_t)\right]\right) \qquad (10) \\
&\geq \frac{1}{2}\sum_{t=1}^{T} \delta \cdot \int \min\{dP_1^t, dP_2^t\} \\
&\geq \frac{1}{2}\sum_{t=1}^{T} \delta \cdot \left(1 - \|P_1^t - P_2^t\|_{\text{TV}}\right) \\
&\geq \frac{1}{2} T\delta \cdot \left(1 - \|P_1^T - P_2^T\|_{\text{TV}}\right),
\end{aligned}$$

where b_t in (10) denotes the bid of the oracle chosen by policy π at time t and P_i^t ($i \in \{1,2\}$) denotes the distribution of all observables $(m_1, \ldots, m_{t-1})$ at the beginning of time t. The KL divergence:

$$\begin{aligned}
D_{\text{KL}}(P_1^T \| P_2^T) &= (T-1) \cdot D_{\text{KL}}(G_1 \| G_2) \\
&= (T-1) \cdot \left(2 \cdot (1-\bar{x}+\delta) \cdot \log \frac{1-\bar{x}+\delta}{1-\bar{x}-\delta} + 2 \cdot \left(\bar{x} - \frac{1}{2} - \delta\right) \cdot \log \frac{\bar{x} - \frac{1}{2} - \delta}{\bar{x} - \frac{1}{2} + \delta}\right) \\
&\leq (T-1) \cdot \left(2 \cdot (1-\bar{x}+\delta) \cdot \left(\frac{1-\bar{x}+\delta}{1-\bar{x}-\delta} - 1\right) + 2 \cdot \left(\bar{x} - \frac{1}{2} - \delta\right) \cdot \left(\frac{\bar{x} - \frac{1}{2} - \delta}{\bar{x} - \frac{1}{2} + \delta} - 1\right)\right) \\
&= 4 \cdot \delta \cdot (T-1) \cdot \left(\frac{1-\bar{x}+\delta}{1-\bar{x}-\delta} - \frac{\bar{x} - \frac{1}{2} - \delta}{\bar{x} - \frac{1}{2} + \delta}\right) \\
&\leq \frac{4T \cdot \delta^2}{(\bar{x} - \frac{1}{2} + \delta)(1-\bar{x}-\delta)} \\
&\leq \frac{16T \cdot \delta^2}{\frac{1}{2} \cdot \sigma^{\frac{q}{q+1}} + \delta}. \\
&\leq \frac{32T \cdot \delta^2}{\sigma^{\frac{q}{q+1}}}.
\end{aligned}$$

Taking the separation parameter $\delta = \min\left\{\frac{1}{2}\cdot\sigma^{\frac{q}{q+1}}, \frac{1}{8}\cdot\sigma^{\frac{q}{2(q+1)}}\cdot T^{-\frac{1}{2}}\right\}$ and substituting into (6) leads to the regret lower bound in Theorem 1:

$$\Omega\left(\sqrt{T\sigma^{\frac{q}{q+1}}}\right) = \Omega\left(\sqrt{L^{\frac{q}{q+1}}\cdot T^{\frac{1}{q+1}}}\right).$$

□

B.2.2 Proof of Lower Bound in Theorem 2.

Proof. At each time t, let $v_t = 1$ and point estimation equals to $\frac{1}{2}$. Define $\varepsilon \in [0, \frac{1}{8}]$ to be some parameter relevant to L. Consider the following two scenarios: (each with probability $\frac{1}{2}$)

- σ_t equals to 0 with probability $p_1 := 1 - 2(\varepsilon - \delta)$, and equals to ε with probability $1 - p_1$, in which case m_t always takes value $h_t + \varepsilon$.
- σ_t equals to 0 with probability $p_2 := 1 - 2(\varepsilon + \delta)$, and equals to ε with probability $1 - p_2$, in which case m_t always takes value $h_t + \varepsilon$.

Easy to observe under this construction the expected value of L:

$$\bar{L} = \sum_{t=1}^{T}\frac{\varepsilon}{2}\cdot(2(\varepsilon+\delta) + 2(\varepsilon-\delta)) = 2\varepsilon^2\cdot T.$$

The above construction also satisfies:

$$\begin{aligned}
&\max_{b\in[0,1]} R_1(1,b) = R_1\left(1,\frac{1}{2}\right) = \frac{1}{2} - \varepsilon + \delta,\\
&\max_{b\in[0,1]} R_2(1,b) = R_2\left(1,\frac{1}{2}+\varepsilon\right) = \frac{1}{2} - \varepsilon,\\
&\max_{b\in[0,1]} (R_1(1,b) + R_2(1,b)) = R_1\left(1,\frac{1}{2}+\varepsilon\right) + R_2\left(1,\frac{1}{2}+\varepsilon\right) = 2\cdot\left(\frac{1}{2}-\varepsilon\right),
\end{aligned}$$

where R_1 and R_2 are expected rewards under the two scenarios. The following steps are similar to previous subsection, for any policy π,

$$\begin{aligned}
\sup_{\{m_t,h_t,\sigma_t\}} \text{Reg}(\pi) &\geq \frac{1}{2}\mathbb{E}_1[\text{Reg}(\pi)] + \frac{1}{2}\mathbb{E}_2[\text{Reg}(\pi)]\\
&= \frac{1}{2}\sum_{t=1}^{T}\left(\mathbb{E}_{P_1^t}\left[\max_{b\in[0,1]} R_1(1,b) - R_1(1,b_t)\right] + \frac{1}{2}\mathbb{E}_{P_2^t}\left[\max_{b\in[0,1]} R_2(1,b) - R_2(1,b_t)\right]\right)\\
&\geq \frac{1}{2}\sum_{t=1}^{T}\delta\cdot\int\min\{dP_1^t, dP_2^t\}\\
&\geq \frac{1}{2}\sum_{t=1}^{T}\delta\cdot\left(1 - \|P_1^t - P_2^t\|_{\text{TV}}\right)\\
&\geq \frac{1}{2}T\delta\cdot\left(1 - \|P_1^T - P_2^T\|_{\text{TV}}\right), \qquad (11)
\end{aligned}$$

with P_1^t and P_2^t defined the same as (10). And the KL divergence

$$\begin{aligned}
D_{\mathrm{KL}}(P_1^T \| P_2^T) &= \sum_{t=1}^{T} \left(2(\varepsilon-\delta) \cdot \log \frac{\varepsilon-\delta}{\varepsilon+\delta} + (1-2(\varepsilon-\delta)) \cdot \log \frac{1-2(\varepsilon-\delta)}{1-2(\varepsilon+\delta)} \right) \\
&\leq \sum_{t=1}^{T} \left(2(\varepsilon-\delta) \cdot \frac{-2\delta}{\varepsilon+\delta} + (1-2(\varepsilon-\delta)) \cdot \frac{4\delta}{1-2(\varepsilon+\delta)} \right) \\
&\leq 4\delta T \cdot \left(-\frac{\varepsilon-\delta}{\varepsilon+\delta} + \frac{1-2\varepsilon+2\delta}{1-2\varepsilon-2\delta} \right) \\
&= 8\delta^2 T \cdot \frac{1}{(\varepsilon+\delta)(1-2\varepsilon-2\delta)} \\
&\leq \frac{16T \cdot \delta^2}{\varepsilon}.
\end{aligned}$$

Taking $\delta = \min\left\{\varepsilon, \frac{1}{4}\sqrt{\frac{\varepsilon}{2T}}\right\}$ and substitute in (11), we have:

$$\sup_{\{m_t, h_t, \sigma_t\}} \mathrm{Reg}(\pi) \geq \frac{1}{4} \min\left\{\varepsilon T, \frac{1}{4\sqrt{2}}\sqrt{T \cdot \varepsilon}\right\},$$

which leads to a lower bound of $\Omega((T \cdot L)^{\frac{1}{4}})$. Note that the construction above requires σ_t to be unknown, otherwise one can achieve 0 regret by bidding hint for $\sigma_t = 0$ and bidding hint + ε for $\sigma = \varepsilon$, which is a technical explanation for the separation in Section 3. □

B.3 Proof of Theorem 3.

Proof. If $L > \left(\sqrt{T}\right)^{\frac{q-1}{q}}$, then $T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} > \sqrt{T}$ and the regret can be lower bounded by $\Omega\left(\sqrt{T}\right)$. So in the following construction, we assume $L \leq \left(\sqrt{T}\right)^{\frac{q-1}{q}}$. First we divide time horizon to $\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor$ equal parts and let σ_t be the same for all t. Construct private values and hints as follows: For $t = i \cdot \left\lfloor \left(\frac{T}{L}\right)^{\frac{q}{q+1}} \right\rfloor + 1, i \cdot \left\lfloor \left(\frac{T}{L}\right)^{\frac{q}{q+1}} \right\rfloor + 2, \ldots, (i+1) \cdot \left\lfloor \left(\frac{T}{L}\right)^{\frac{q}{q+1}} \right\rfloor$,

$$\begin{aligned}
v_t &= \frac{1}{2} + \frac{1}{2} \cdot \frac{i}{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}}, \\
h_t &= \frac{1}{4} + \frac{i}{4} \cdot \sigma^{\frac{q}{q+1}}, \\
m_t &= \begin{cases} \frac{1}{4} + \frac{i}{4} \cdot \sigma^{\frac{q}{q+1}}, & w.p. \quad 1 - \frac{1}{4} \cdot \left(\sigma^{\frac{q}{q+1}} \pm \delta\right) \\ \frac{1}{4} + \frac{i+1}{4} \cdot \sigma^{\frac{q}{q+1}}, & w.p. \quad \frac{1}{4} \cdot \left(\sigma^{\frac{q}{q+1}} \pm \delta\right) \end{cases}
\end{aligned}$$

where $i = 0, 1, 2, \ldots, \left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor - 1$ and δ is the separation parameter similarly defined in the proof of Theorem 1. Since $L \leq \left(\sqrt{T}\right)^{\frac{q-1}{q}}$, we have

$$\frac{1}{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}} \geq \left(\frac{L}{T}\right)^{\frac{q}{q+1}} = \sigma^{\frac{q}{q+1}},$$

which ensures any strategy π that bids in $\left[\frac{1}{4} + \frac{i}{4} \cdot \sigma^{\frac{q}{q+1}}, \frac{1}{4} + \frac{i+1}{4} \cdot \sigma^{\frac{q}{q+1}}\right]$ for the i-th part belongs to 1-Lipschitz and monotone oracle. Therefore, we can now consider the whole time horizon as $\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor$ independent problems, each of which consists of $\left\lfloor \left(\frac{T}{L}\right)^{\frac{q}{q+1}} \right\rfloor$ time steps and has fixed v_t. Substituting $L_i := \left\lfloor \left(\frac{T}{L}\right)^{\frac{q}{q+1}} \right\rfloor \cdot \sigma$, which is L for the i-th subproblem, and applying similar method

to the proof of Theorem 1, we can get:

$$\sup_G \mathrm{Reg}_i(\pi) = \Omega\left(\sqrt{\left(\frac{T}{\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor}\right)^{\frac{1}{q+1}} \cdot \left(\frac{L}{\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor}\right)^{\frac{q}{q+1}}}\right)$$
$$= \Omega\left(\frac{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}}{\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor}\right) = \Omega(1),$$

for each independent problem. Summing over all subproblems leads to the lower bound $\Omega\left(T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}\right)$. □

B.4 Proof of Theorem 4

Proof. We prove that even when L takes expected value $\Theta(1)$, the minimax regret is still lower bounded by $\Omega(\sqrt{T})$. The proof is similar to that of Theorem 3, but by dividing time horizon into $\sqrt{T}$ subproblems. At each time t inside the i-th subproblem, the bidder observes $h_t = \frac{1}{4} + \frac{i \cdot \varepsilon}{4}$ (where $\varepsilon = \frac{1}{\sqrt{T}}$). In the construction of the lower bound in Theorem 2, σ_t equals to 0 with probability $1 - \Theta(\varepsilon)$ and equals to ε with probability $\Theta(\varepsilon)$. Thus,

$$\bar{L} = \mathbb{E}\left[\sum_{t=1}^{T} \sigma_t\right] = T \cdot \varepsilon^2 = \Theta(1).$$

Meanwhile, applying similar method to the proof of Theorem 2, we can get a lower bound of $\Omega\left(\sqrt{\sqrt{T} \cdot \frac{1}{\sqrt{T}}}\right) = \Omega(1)$ for each independent problem, leading to the final lower bound $\Omega\left(\sqrt{T}\right)$. □

C Proof of Main Result in Section 4

C.1 Proof of Theorem 5

C.1.1 Proof of Upper Bounds in Theorem 5

Proof. In the following subsection, we provide a way to achieve $O\left(\sqrt{\log T \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}} \cdot K}\right)$ regret upper bound. *

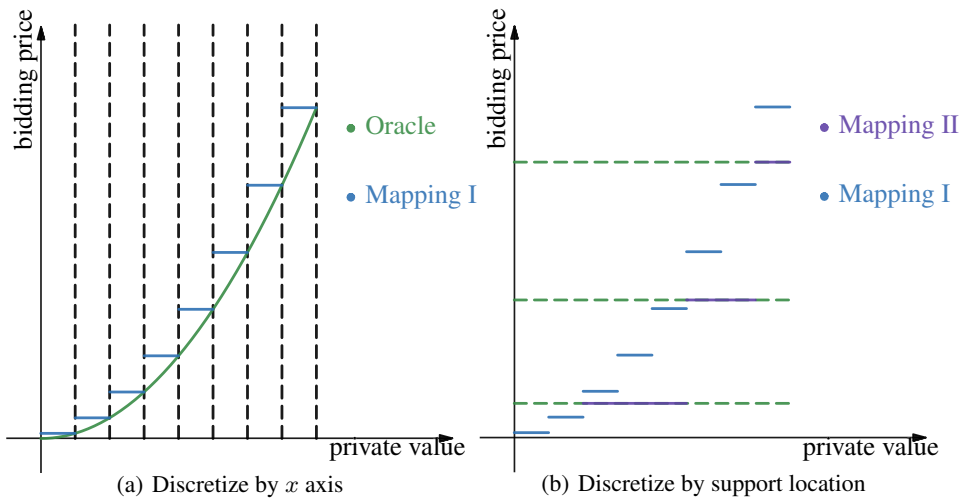
Figure 3 shows any function in oracle can be mapped to a piecewise constant function whose value only takes those in the support set, define this mapped function set to be A. We prove in the appendix that the number of functions in the converted set A is smaller than T^K, then applying the algorithm in Theorem 1's proof directly leads to an upper bound of *

$$O\left(\sqrt{\log\left(|\text{expert set}|\right) \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}\right) = O\left(\sqrt{\log T \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}} \cdot K}\right).$$

To show set A is small enough, let's first imagine walking from $(1, 0)$ to (T, K) with each step either to the positive direction of x-axis or y-axis exactly by 1. There are $T + K - 1$ steps in total and one may choose K of them to go up. Now given any function in A, suppose at $x = 1$ the value equals to the i-th support and at $x = T$ the value equals to the j-th support, which can be considered as points $(1, i)$ and (T, j), $i, j \in \mathbb{Z}$, $0 \le i \le j \le K$. Without loss of monotonicity, we add points $(0, 0)$ and $(T + 1, K)$ to the interval-support pairs of this function, i.e. the function takes value of the

*The other two are described in Appendix A.

*Although in the proof of Theorem 1 we show an upper bound of $O\left(\sqrt{\log T \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}\right)$, the proof schetch can indeed be applied to any finite set of experts.



(a) Discretize by x axis (b) Discretize by support location

Figure 3: Given any 1-Lipschitz and monotone oracle, we first discretize the x-axis into T small intervals, changing the oracle to a piecewise constant function that bids the maximum point for each interval in the oracle; Secondly, we map this piecewise constant function to a piecewise function that only takes support value as bidding price. Easy to verify step 1 leads to $T \cdot O(\frac{1}{T}) = O(1)$ loss, while step 2 leads to a non-negative change to the cumulative reward.

i-th support for the t-th interval, $i \in [K]$, $t \in [T]$, iff we pass point (t, i) in the route from $(0, 0)$ to $(T+1, K)$. The set of routes and set A forms a bijection, both have cardinality:

$$\binom{T+K-1}{K} = \frac{T+K-1}{K} \cdot \frac{T+K-2}{K-1} \cdots \frac{T}{1} \leq T^K.$$

□

C.1.2 Proof of Lower Bounds in Theorem 5

Proof. Consider the three cases separately:

- If $L < \frac{K^{\frac{q+1}{q}}}{T^{\frac{1}{q}}}$, then as in the proof of Theorem 3 we can construct $N = \left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor$ independent problems since $N < K$ in this case. For each independent problem the lower bound is $\Omega(1)$, leading to a total lower bound of $\Omega\left(T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}\right)$.

- If $\frac{K^{\frac{q+1}{q}}}{T^{\frac{1}{q}}} \leq L \leq \frac{T}{K^{\frac{q+1}{q}}}$, we cannot divide into $\left\lfloor T^{\frac{1}{q+1}} L^{\frac{q}{q+1}} \right\rfloor$ subproblems since there are only K values m_t can take. So instead, we divide time horizon into K subproblems:
  For $t = i \cdot \lfloor \frac{T}{K} \rfloor + 1, i \cdot \lfloor \frac{T}{K} \rfloor + 2, \ldots, (i+1) \cdot \lfloor \frac{T}{K} \rfloor$,
  $$v_t = \frac{1}{2} + \frac{1}{2} \cdot \frac{i}{K},$$
  $$h_t = \frac{1}{4} + \frac{i}{4} \cdot \sigma^{\frac{q}{q+1}},$$
  $$m_t = \begin{cases} \frac{1}{4} + \frac{i}{4} \cdot \sigma^{\frac{q}{q+1}}, & w.p. \quad 1 - \frac{1}{4} \cdot \left(\sigma^{\frac{q}{q+1}} \pm \delta\right) \\ \frac{1}{4} + \frac{i+1}{4} \cdot \sigma^{\frac{q}{q+1}}, & w.p. \quad \frac{1}{4} \cdot \left(\sigma^{\frac{q}{q+1}} \pm \delta\right) \end{cases}$$
  where $i = 0, 1, 2, \ldots, K-1$. Observe that the difference between v_t for adjacent subproblem is $\frac{1}{2} \cdot \frac{1}{K}$ and the difference between bid value for adjacent subproblem is at most
  $$2 \cdot \frac{\sigma^{\frac{q}{q+1}}}{4} = \frac{\sigma^{\frac{q}{q+1}}}{2} = \frac{L^{\frac{q}{q+1}}}{2 \cdot T^{\frac{q}{q+1}}} \leq \frac{1}{2} \cdot \frac{1}{K},$$

ensuring the $N = K$ subproblems are indeed independent from each other. Additionally, the separation parameter δ for each subproblem equals to $\sqrt{\frac{\sigma^{\frac{q}{q+1}}}{\frac{T}{K}}} = \sqrt{\frac{K \cdot L^{\frac{q}{q+1}}}{T^{\frac{1}{q+1}}}}$, which is smaller than the separation of m_t: $\sigma^{\frac{q}{q+1}} = \left(\frac{L}{T}\right)^{\frac{q}{q+1}}$. Thus substituting Theorem 1, finally the lower bound is,

$$K \cdot \Omega\left(\left(\frac{T}{K}\right)^{\frac{1}{q+1}} \cdot \left(\frac{L}{K}\right)^{\frac{q}{q+1}}\right) = \Omega\left(\sqrt{K \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}\right).$$

- If $L > \frac{T}{K^{\frac{q+1}{q}}}$, a traditional lower bound gives $\Omega\left(\sqrt{T}\right)$.

□

C.2 Proof of Theorem 6

Proof. Let the learning rate for the upper level $\eta_{t,2} = \min\left\{\frac{1}{4}, \sqrt{\frac{\log 3}{\left\lfloor \sum_{s=1}^{t-1} \sigma_s^{\frac{q}{q+1}} \right\rfloor + 1}}\right\}$ and apply similar analysis as in Appendix B.1.1:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[X_t] &\geq \max_{i \in \{f,g,h\}} \sum_{t=1}^{T} r_{t,i} - 2 \cdot \left(\frac{\log 3}{\eta_{T,2}} + 2 \sum_{t=1}^{T} \eta_{t,2} \cdot 2 \cdot \sigma_t^{\frac{q}{q+1}}\right) \\
&= \max_{i \in \{f,g,h\}} \sum_{t=1}^{T} r_{t,i} - 2 \cdot \left(\sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}} + 4\sqrt{\log 3} \cdot \sum_{t=1}^{T} \frac{\sigma_t^{\frac{q}{q+1}}}{\sqrt{\left\lfloor \sum_{s=1}^{t} \sigma_s^{\frac{q}{q+1}} \right\rfloor + 1}}\right) \\
&\overset{\text{(a)}}{\geq} \max_{i \in \{f,g,h\}} \sum_{t=1}^{T} r_{t,i} - 2 \cdot \left(\sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}} + 8\sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}\right) \\
&= \max_{i \in \{f,g,h\}} \sum_{t=1}^{T} r_{t,i} - 18 \cdot \sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}, \qquad (12)
\end{aligned}$$

where $\sum_{t=1}^{T} \mathbb{E}[X_t]$ is the expected total reward by running Algorithm 1, with expectation taken over both policy randomness and possible m_t sequences. (a) can be considered as taking integral of function $f(x) = \frac{1}{\sqrt{x}}$, but with another piecewise function smaller than it instead. And applying similar method to the lower level of the first node we have:

$$\sum_{t=1}^{T} r_{t,f} \geq \max_{a \in [T^K]} \sum_{t=1}^{T} r_{t,a} - 18 \cdot \sqrt{K \log T \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}. \qquad (13)$$

Combining (12) and (13) and the regret upper bound of ChEW algorithm and choosing hint expert:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[X_t] &\geq \max_{i \in \{f,g,h\}} \left(\max_{a \in [T^K]} \sum_{t=1}^{T} r_{t,a} - \text{Reg}(i)\right) - 18 \cdot \sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}} \\
&= \max_{a \in [T^K]} \sum_{t=1}^{T} r_{t,a} - \min\left\{18 \cdot \sqrt{K \log T \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}, 2 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}, C \cdot \sqrt{T}\right\} - 18 \cdot \sqrt{\log 3 \cdot \sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}},
\end{aligned}$$

where C is a constant number. Therefore, we have:

$$\begin{aligned}\text{Reg}(\pi) &= O\left(\min\left\{\sqrt{\log T\cdot\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}}\cdot K},\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}},\sqrt{T}\right\}\right)+O\left(\sqrt{\log 3\cdot\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}}}\right)\\ &\overset{\text{(b)}}{=} O\left(\min\left\{\sqrt{\log T\cdot\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}}\cdot K},\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}},\sqrt{T}\right\}\right),\end{aligned}$$

while (b) holds since $\sum_{t=1}^{T}\sigma_t^{\frac{q}{q+1}} > L > 1$. □

C.3 Proof of Theorem 7

C.3.1 Proof of Upper Bound in Theorem 7

Proof. Instead of one single hint expert in Algorithm 2, construct T hint experts, with each one bidding a constant gap over h_t, i.e. with the first hint expert bidding $h_t+\frac{1}{T}$ for $t=1,\dots,T$; the second hint expert bidding $h_t+\frac{2}{T}$ for $t=1,\dots,T$; etc. The upper layer then consists of T hint experts and two super nodes, representing ChEW algorithm (g) and modified Algorithm 1 (f). The lower layer of f consists of T^K base experts (constructed as in Appendix C.1) and T hint experts. Let the learning rate for the upper level $\eta_2=\min\left\{\frac{1}{4},\sqrt{\frac{\log(T+2)}{\sqrt{TL}}}\right\}$,

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[X_t] &\geq \max_{i\in\{f,g,h\}}\sum_{t=1}^{T}r_{t,i}-2\cdot\left(\frac{\log(T+2)}{\eta_2}+4\eta_2\sqrt{LT}\right)\\ &= \max_{i\in\{f,g,h\}}\sum_{t=1}^{T}r_{t,i}-10\cdot\sqrt{\log(T+2)\cdot\sqrt{LT}},\end{aligned}\tag{14}$$

And applying similar method to super node f:

$$\sum_{t=1}^{T}r_{t,f}\geq\max_{a\in[T^K]}\sum_{t=1}^{T}r_{t,a}-10\cdot\sqrt{K\log T\cdot\sqrt{TL}}.\tag{15}$$

Combining (14) and (15),

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[X_t] &\geq \max_{i\in\{f,g,h\}}\left(\max_{a\in[T^K]}\sum_{t=1}^{T}r_{t,a}-\text{Reg}(i)\right)-20\cdot\sqrt{\log T\cdot\sqrt{TL}}\\ &= \max_{a\in[T^K]}\sum_{t=1}^{T}r_{t,a}-\min\left\{10\cdot\sqrt{K\log T\cdot\sqrt{TL}},2\cdot\sqrt{TL},C\cdot\sqrt{T}\right\}-20\cdot\sqrt{\log T\cdot\sqrt{TL}},\end{aligned}$$

where C is a constant number. Therefore, we have:

$$\begin{aligned}\text{Reg}(\pi) &= O\left(\min\left\{\sqrt{K\log T\cdot\sqrt{TL}},\sqrt{T}\right\}\right)+O\left(\sqrt{\log T\cdot\sqrt{TL}}\right)\\ &= O\left(\min\left\{\sqrt{K\log T\cdot\sqrt{TL}},\sqrt{T\log T}\right\}\right).\end{aligned}$$

□

C.3.2 Proof of Lower Bound in Theorem 7

The following is similar to proof of lower bound in Theorem 5.

Proof.

- If $L>\frac{T}{K^2}$, as in the proof of Theorem 4 construct $N_0=\left\lfloor\sqrt{\frac{T}{L}}\right\rfloor<K$ independent sub-problems, while for each sub-problem

$$L'=\frac{L}{\sqrt{T/L}}=\sqrt{\frac{L^3}{T}},\quad T'=\frac{T}{\sqrt{T/L}}=\sqrt{TL},$$

and for each sub-problem regret is lower bounded by $\Omega\left(\left(\sqrt{LT}\cdot\sqrt{\frac{L^3}{T}}\right)^{1/4}\right)$, leading to a total lower bound of $\Omega\left(\left(\sqrt{LT}\cdot\sqrt{\frac{L^3}{T}}\right)^{1/4}\cdot\sqrt{\frac{T}{L}}\right)=\Omega(\sqrt{T})$.

- If $L\leq\frac{T}{K^2}$, it is not feasible to construct N_0 independent sub-problems as the optimal bidding value can not take $N_0>K$ values. Instead construct K independent problems, with the separation parameter (see Appendix B.2.2): $\delta=\sqrt{\frac{L}{T}}\cdot\frac{K}{T}<\frac{1}{T}$, leading to a total regret lower bound of $\Omega\left(\sqrt{\sqrt{\frac{T}{K}\cdot\frac{L}{K}}\cdot K}=\Omega\left(\sqrt{K\cdot\sqrt{TL}}\right)\right)$.

□

D Experimental Details

D.1 Description of Experiment 1 in Section 5

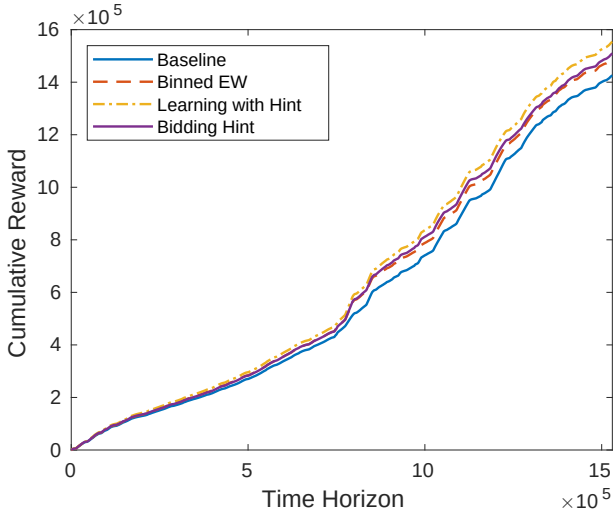
Divide the whole range of private value to D bins, each of which contains v_t's that are close to each other. As long as the bidder observes v_t at time t, we reduce the problem to the bin focusing on the data points with private values close to v_t. Then each bin itself forms a sub-problem described in Section 3. Experiment 1 only serves as an illustration of the effect by hints. The role of hints is threefolds:

- We use hint to help allocating data to different bins. Instead of binning only by private values, we use hint as a side information and conduct binning also based on it. The total number of bins is $M_1\cdot M_2$, while M_1 is the number of discretization for v_t and M_2 is the number of discretization for hints. As for the result on empirical data, we observe $M_2=4$ already leads to rather good performance.
- We use hint to calculate the estimation of instantaneous reward for any given bid b'_t under the assumption that $m_t=b_t$: $r'_{t,a}:=r(b_t;h_t,v_t)$, where b_t is the bid at time t according to oracle a. Then we add this estimated reward to each experts' reward history while sampling among these experts:
$$p_{t,a}=\frac{\exp\left(\eta_t\cdot\left(\sum_{s=1}^{t-1}r_{s,a}+r'_{t,a}\right)\right)}{\sum_{a'\in\mathcal{F}}\exp\left(\eta_t\cdot\left(\sum_{s=1}^{t-1}r_{s,a'}+r'_{t,a'}\right)\right)},t=2,3,\cdots,T.$$
And if σ_t is also observed, we define $r'_{t,a}:=r(b_t;h_t+c_1\cdot\sigma_t,v_t)$ instead, where c_1 is a hyper-parameter to be tuned.
- We include a set of hint experts
$$b_t(a_i):=h_t+\sigma_t^{\Delta_i},\quad i=1,2,\cdots,k,$$
which is close to a combination of algorithms for whether knowing the error, since for real datasets q is often not observed.

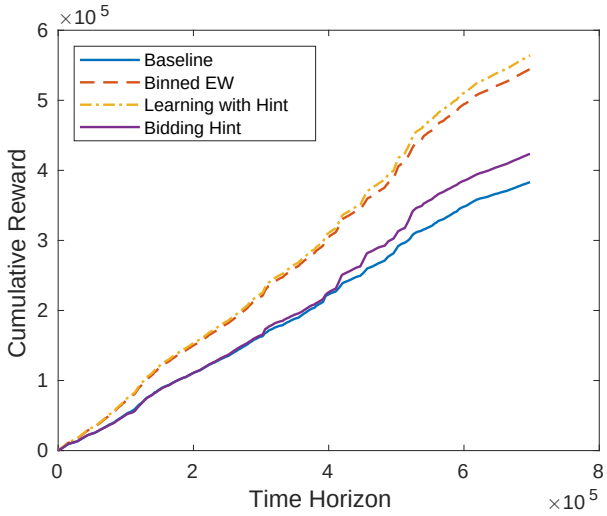
The results in Figure 4 shows the improvement by incorporating hint on other two datasets. The results implies that on datasets whose hint has rather small error, e.g. on dataset 1 bidding hint itself already beats simple online learning algorithm, the improvement by hint is more significant. Namely, 4.38% on dataset 1 with more accurate hint and 3.54% on dataset 2 whose hint is not so good.

D.2 Polynomial Algorithm in Section 5

Consider any 1-Lipschitz & monotone oracle f, since support size is finite, f can be mapped to a discontinuous function f' with $O(1)$ loss, which can be further represented by a series of interval-support pair:
$$\left(0,\frac{1}{D}\right]\leftrightarrow s_{i_1},\quad\left(\frac{1}{D},\frac{2}{D}\right]\leftrightarrow s_{i_2},\quad\ldots,\quad\left(\frac{D-1}{D},1\right]\leftrightarrow s_{i_D},$$



(a) Results on Dataset 1



(b) Results on Dataset 2

Figure 4: Cumulative rewards as a function of time. The dashdot lines stands for incorporating hint into exponential weighting, and the purple solid lines are directly bidding hint. The dotted lines represent binned exponential algorithm.

Algorithm 3: DP algorithm without knowing support locations

Inputs: Time horizon T; support size K;
Initialization: $\text{Reward}_{T,K,T} \leftarrow 0$; $\text{P} \leftarrow 0$;
for $t = 1, 2, \ldots, T$ **do**
  % Calculate Sum_Forward&Sum_Backward Matrix
  $\text{Sum_Forward}_{T,K,T} \leftarrow 1$; $\text{Sum_Backward}_{T,K,T} \leftarrow 1$;
  for $i = 1, 2, \ldots T$ **do**
    for $j = 1, 2, \ldots, T$ **do**
      $\text{Sum_Forward}_{i,1,j} \leftarrow \text{Sum_Forward}_{i-1,1,j} \cdot \exp(\eta_t \cdot \text{Reward}_{i,1,j})$;
      $\text{Sum_Backward}_{i,K,j} \leftarrow \text{Sum_Backward}_{i+1,K,j} \cdot \exp(\eta_t \cdot \text{Reward}_{i,K,j})$;
      for $k = 2, 3, \ldots K-1$ **do**

$$\text{Sum_Forward}_{i,k,j} \leftarrow \sum_{v=1}^{j-1} \Big(\text{Sum_Forward}_{i-1,k-1,v} \cdot \exp(\eta_t \cdot \text{Reward}_{i,k,j})\Big) + \text{Sum_Forward}_{i-1,k,j} \cdot \exp(\eta_t \cdot \text{Reward}_{i,k,j});$$

$$\text{Sum_Backward}_{i,k,j} \leftarrow \sum_{v=j+1}^{T} \Big(\text{Sum_Backward}_{i+1,k+1,v} \cdot \exp(\eta_t \cdot \text{Reward}_{i,k,j})\Big) + \text{Sum_Backward}_{i+1,k,j} \cdot \exp(\eta_t \cdot \text{Reward}_{i,k,j});$$

      end
    end
  end
  % Calculate Probability
  $i \leftarrow \lfloor v_t \cdot T \rfloor$;
  for $j = 1, 2, \ldots T$ **do**

$$\text{P}_j \leftarrow \sum_{k=1,2,\ldots K} \Bigg(\Big(\text{Sum_Forward}_{i-1,k,j} + \sum_{v=1}^{j-1} \text{Sum_Forward}_{i-1,k-1,v}\Big) \cdot \exp(\eta_t \cdot \text{Reward}_{i,k,j}) \cdot \Big(\text{Sum_Backward}_{i+1,k,j} + \sum_{v=j+1}^{T} \text{Sum_Backward}_{i+1,k+1,v}\Big)\Bigg);$$

  end
  for $k = 1, 2, \ldots K$ **do**
    $\text{P}_{\lfloor h_t \cdot T \rfloor} \leftarrow \text{P}_{\lfloor h_t \cdot T \rfloor} + \exp(\eta_t \cdot \text{RH})$;
  end
  Sample $b_t \sim (P / \sum(P))$;
  % Update Reward Matrix
  for $k = 1, 2, \ldots, K$ **do**
    for $j = 1, 2, \ldots T$ **do**
      if $m_t \leq j/T$ **then**
        $\text{Reward}_{i,k,j} \leftarrow \text{Reward}_{i,k,j} + (v_t - j/T)$;
    end
  end
  $\text{RH} \leftarrow \text{RH} + r(h_t; v_t, m_t)$;
end

where $0 \leq s_1 \leq s_2 \leq s_3 \leq \cdots \leq s_K \leq 1$ are the locations of supports in increasing order and $0 \leq i_1 \leq i_2 \leq \cdots \leq i_D \leq K, i_1, i_2, \cdots, i_D \in \mathbb{Z}$. The main idea is to record the cumulative reward for all possible interval-support tuples and use dynamic programming to calculate total reward for some expert sets instead of keeping track of all T^K experts.

Reward[D][K][D]: The first two dimensions represent interval: $[d][k] : (d/D, (d+1)/D] \leftrightarrow s_k$. The third dimension represent the bidding, with steply update

$$\text{Reward}_{i,k,j} \leftarrow \text{Reward}_{i,k,j} + (v_t - j/D)$$

Then we use dynamic programming to calculate the sum of the rewards for several continuous intervals, instead of keeping track of all T^K experts.

Sum_Forward[D][K][D]: Forward DP recording array, representing combined intervals: $[d][K] : (0, d/D] \leftrightarrow \{1, \ldots, s_k\}$ and the third dimension represents bidding for the last interval: $(d/D, (d+1)/D]$. The update calculation is carried out per step before choosing an action.

Sum_Backward[D][K][D]: Backward DP recording array, representing combined intervals: $[d][K] : ((d+1)/D, 1] \leftrightarrow \{s_k + 1, \ldots, K\}$ and the third dimension represents bidding for the first interval: $((d+1)/D, (d+2)/D]$. The update calculation is carried out per step before choosing an action.

Combining the results of Sum_Forward and Sum_Backward, we can calculate reward history for a subset of the T^K experts, which is the only needed quantity for calculating probability in exponential weighting instead of keeping record with an exponential size.