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

**Anonymous Author(s)**
Affiliation
Address
email

## 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: 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 Introduction

With the accelerating proliferation of e-commerce sweeping across the industries [Kha16, KDH17, HL18, FRZP19, WSKS20], 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 (e.g. TVs, radios, newspapers, local billboards) by 20 billion dollars. Since then, this number has been growing while traditional advertising spending will continue to shrink [Wag19].

In this backdrop, the core step that generates the revenue for the digital advertising industry is online ads auctions, which are run and completed automatically each time before an ad is served, with three main entities participating: publishers (sellers), advertisers (bidders) and Ad exchanges. In the past, due to its truthful nature, the second-price auction was a popular auction mechanism and was almost universally adopted for online ads auctions [LR00, Kle04, LRBPR07]. However, very recently there has been an industry-wide shift from second-price auctions to first-price auctions[1] in display ads auctions, which account for 54% of the digital advertising market share (the remaining market

[1] In a first-price auction, the highest bidder wins the auction and pays for the highest price bidded. First-price auctions have been the norm in several more traditional settings, including the mussels auctions [vSK+01]; see also [Esp08] for more discussion.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

share is dominated by search ads, which are still done through second-price auctions [DRS19]), a percentage that has seen continued growth "fueled by the upswing in mobile browsing, social media activities, video ad formats, and the developments in targeting technology" [CMBL20].

Such an industry-wide shift to first-price auctions has occurred for several reasons. First, a larger revenue in first-price auctions for the seller and hence the exchange (which charges a percentage of the winning bid) [Ben18a]. Second, no last-look advantage for exchanges: under second-price auctions, an ad exchange can examine all the submitted bids and *then* raise the floor price (the minimum price the seller is willing to accept) to above the second-highest bid and obtain a larger revenue [Ben18b]. 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. Third, fairness for bidders: a seller would sometimes sell an advertising slot on different exchanges, and take the highest winning bid across the exchanges (this is also known as *header bidding*) [Vro19]. Consequently, under second-price auctions, since each exchange only reports the winning bid (the second-highest bid), it's possible that a winner on one exchange–who bids lower than another winner on a different exchange–ends up winning the final bid, a problem that wouldn't occur under first-price auctions.

Motivated by these considerations, several ad exchanges (AppNexus, Index Exchange and OpenX) started to roll out first-price auctions in 2017 and completed the transition by 2018 [Slu17, App18]. Additionally, under sustained criticism of leveraging last-look advantage in second-price auctions, Google Ad Manager (previously Adx) also completed the move to first-price auctions at the end of 2019 [Dav19] and incorporated additional transparency[2] 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 where several major ad exchanges have just made the transition and many others will be following suit, an important question that arises is: how should a bidder (adaptively) bid in repeated online first-price auctions to maximize his 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), a merit that constitutes the hallmark of second-price auctions.

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 $\tilde{\Theta}(T^{\frac{2}{3}})$; 2) if $m_t$ is adversarial, then one achieves the minimax optimal regret of $\tilde{\Theta}(T^{\frac{3}{4}})$. 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 $\tilde{\Theta}(T^{\frac{1}{2}})$. 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 $\tilde{\Theta}(T^{\frac{1}{2}})$ can be achieved when $m_t$ is adversarial[3]. [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.

However, 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

[2] This is likely an effort to offset the previous negative image, although there was no mention of this in Google's official language.

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

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.

## 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 his past observations of others' bid and $v_t$ for that round, he 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 $\pi$ 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 $\pi$ while the expectation is taken over the randomness inside randomized policy $\pi$. Then we define the regret under policy $\pi$:

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

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 $a$.[4] 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 $(h_t, \sigma_t)$ 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 $\sigma_t$. 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 $\sigma_t$ 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.

## 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.

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

### 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 $t$, then the following regret upper and lower bounds hold:*

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

*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 $\pi$.*

**Theorem 2.** *If $v_t \equiv 1$ and the bidder observes a single hint at each time $t$, then $\forall q \in [1, \infty)$, the following regret upper and lower bounds hold:*

$$\inf_{\pi} \sup_{(m_t, h_t, \sigma_t)} \mathrm{Reg}(\pi) \in O\left((\log T)^{\frac{1}{2}} (T \cdot L)^{\frac{1}{4}}\right) \cap \Omega\left((T \cdot L)^{\frac{1}{4}}\right),$$

*where the supremum is taken over rewards under all possible $m_t$ sequence and hints that satisfy (3), and the infimum is taken over all possible policies $\pi$.*

This main result shows that there is a strict separation in regret bounds compared to the previous subsection when additional information is given. Observe that in the lower bound of Theorem 1, when $q$ becomes larger, the regret lower bound 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 $\tilde{O}(\sqrt{L})$ 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 him to wisely arrange the weight he 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 $\tilde{O}(\sqrt{T})$ 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:

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

**Algorithm 1:** Multiplicative Weights with Hint Intervals

**Input:** Time horizon $T$; Hints accuracy $L$; $q$.
**Output:** A bidding policy $\pi$.
**Initialization:** Construct $K = T + 1$ experts, with the $i$-th ($i = 1, 2, \ldots, T$) base expert bid $\frac{i}{T}$ 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**
Set learning rate $\eta \leftarrow \min\left\{\frac{1}{4}, \sqrt{\frac{\log K}{\sum_{t=1}^{T} \sigma_t^{\frac{q}{q+1}}}}\right\}$;
**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 $\sigma_t$;
  Set $b_{t,a^*} \leftarrow h_t + \sigma_t^{\frac{q}{q+1}}$;
  Let $b_t \leftarrow b_{t,a}$ with probability

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

  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**

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 he 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 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 $\Delta_i$ 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.2)

Notice that in both cases we required knowledge of $L$ to decide learning rate $\eta_t$ 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.

### 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:*

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

*where the supremum is taken over rewards under all possible $(v_t, m_t)$ sequences and hints that satisfy (3), and the infimum is taken over all possible policies $\pi$.*

**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:*

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

*where the supremum is taken over rewards under all possible $(v_t, m_t)$ sequences and hints that satisfy (3), and the infimum is taken over all possible policies $\pi$.*

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. $L \leq 1$. 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 the following key 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. Part of the reason for the sparsity is due to the scenario where the maximum competing bid is the reserve price set by the seller.

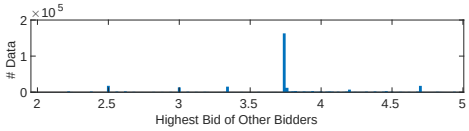



Figure 1: Real data of others' highest bid

Assume that $m_t$ is only supported on $K$ locations. The central question in this section is as follows:

*How does the discontinuity of $K$'s effect on regret bound occur from small support size to large support size, e.g. $K = \sqrt{T}$? 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:*

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

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

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

*where $K \leq \sqrt{T}$.*

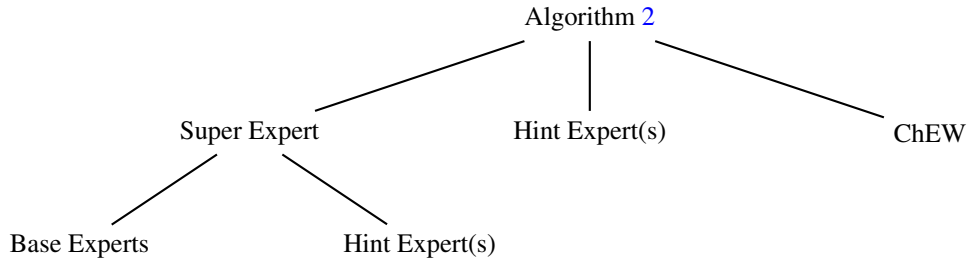
Observe that for $K = \sqrt{T}$, the regret lower bound is the same as that in Theorem 3: $\Omega\left(\min\left\{T^{\frac{1}{q+1}} L^{\frac{q}{q+1}}, \sqrt{T}\right\}\right)$. This implies that for large $K$, hints only help if $L < T^{\frac{q-1}{2q}}$, while for small support size hints are sure to improve over traditional bound, implying a discontinuity on whether hints indeed help on this problem.

The lower bounds in Theorem 5 illustrates how the improvement by hints changes from definite improvement to conditional improvement, as $K$ becomes larger: Consider a worst case example while $L = \Theta(T)$, the regret lower bound under case two is: $\Omega\left(\sqrt{K \cdot T^{\frac{1}{q+1}} \cdot L^{\frac{q}{q+1}}}\right) = \Omega\left(\sqrt{K \cdot T}\right)$, thus implying the second bound has the same magnitude as the first bound only if $K = \Theta(1)$, otherwise for this worst-case $L$ hints cannot improve over traditional lower bound $\sqrt{T}$.

### 4.2 Meta Algorithm Adaptive to Unknown $L$ and $K$

In the proof of Theorem 5 (see appendix D) we prove the three upper bounds can indeed be achieved separately. If a bidder observe the overall quality of hints $L$ at the beginning with $T$, he 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, hints based algorithm with support only beats $O(\sqrt{T})$ [HZF+20] when $L < T/K^{(q+1)/q}$. Additionally, the data-driven method adaptive to $L$ [ACBG02] should pick different learning rate $\eta_t$ for the upper bounds we are aiming for, which leads to trouble when $L$ is not observed in advance.

Algorithm 2 (pseudocode in Appendix A) with meta-experts addresses this problem and ensures to achieve the optimal one among the three upper bound for any $L > 1$. 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 Expert" 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.** *Running Algorithm 2 leads to the optimal one among the upper bounds in Theorem 5, without the total error of hints $L$ given.*

The theorem is also true when $K$ and support locations are unknown together with $L$.[5] At the beginning assume a constant value $K_0$ (e.g. $K_0 = 8$), and run the above algorithm until the number of supports up to time $t$ exceeds this value. If that happens, we double the value of $K$: $K_{i+1} = 2 \cdot K_i$ ($i \in \mathbb{Z}$), and restart the learning algorithm. Since assuming a larger $K$ than reality is no harm for covering all functions in oracle, substituting (4) to this strategy leads to regret upper bound: ($M := \left\lceil \log\left(\frac{K}{8}\right)\right\rceil + 1$)

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

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 expert set as follows. We divide $y$-axis (bidding price) into $T$ small intervals, with the $i$-th interval $[\frac{i}{T}, \frac{i+1}{T}]$ ($i = 0, 1, \cdots, T-1$). 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 would lead to cumulative reward difference $O(1)$ for the whole time horizon, if we let oracle to bid the top of each small interval without loss of generality.) The total number of allocating methods is $\binom{T}{K}$, and for each allocating method $T^K$ experts are enough to cover the functions in oracle (see proof of Theorem 5), thus the size of the complete expert set is:

$$|\text{expert set}| \leq \binom{T}{K} \cdot T^K \leq \left(T^K\right)^2,$$

Then the regret upper bound can be written as: (using result in Appendix B.1)

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

#### 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 is upper bounded by:*

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

*where the supremum is taken over all $m_t$ sequence and hints that satisfies (3), and the infimum is taken over all possible policies $\pi$.*

The upper bound is achieved by replacing each node of hint expert in Algorithm 2 with a set of hint experts that each bids hint at time $t + \Delta_i$, respectively. We relegate the details to the appendix.

## 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 cannot 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 pair of scalars ($v_t$, $m_t$, $h_t$, $\sigma_t$) including

[5] 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.

the private value $v_t$ and the minimum bid $m_t$ needed 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 needed 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 $\sigma_t$ are provided by fitting a lognormal model.

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. The main idea is to use dynamic programming, with space complexity $O(D \cdot K)$ and time complexity $O(D \cdot K \cdot T)$ for $T^K$ experts, where $D$ is parameter for discretization on private values. We implemented Algorithm 3 on dataset with support size 5.

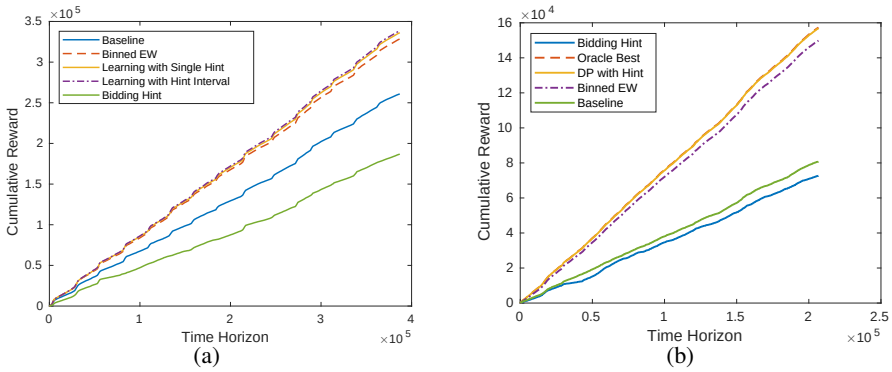



Figure 2: Cumulative reward as a function of time, respectively results for the two subsection. Left hand side uses a binning method to show the improvement by incorporating hints, with the yellow solid line corresponding to learning with single hints and the dashdot line with hint intervals. The dotted line corresponds to binned exponential weighting, and green line is bidding hints itself in each round; Right hand side illustrates the result of implementing polynomial-time version of our algorithm (yellow solid line), and the performance of the best increasing and 1-Lipschitz function in oracles (red dashed line), and purple dashdot line stands for simple binned exponential weighting algorithm.

## 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.

# 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.

[CMBL20] Hana Choi, Carl F Mela, Santiago R Balseiro, and Adam Leary. Online display advertising markets: A literature review and future directions. *Information Systems Research*, 2020.

[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.

[Esp08] Ignacio Esponda. Information feedback in first price auctions. *The RAND Journal of Economics*, 39(2):491–508, 2008.

[FRZP19] Monica Faraoni, Riccardo Rialti, Lamberto Zollo, and Anna Claudia Pellicelli. Exploring e-loyalty antecedents in b2c e-commerce. *British Food Journal*, 121(2):574–589, 2019.

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

[HL18] Heli Hallikainen and Tommi Laukkanen. National culture and consumer trust in e-commerce. *International Journal of Information Management*, 38(1):97–106, 2018.

[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.

[KDH17] Thai Young Kim, Rommert Dekker, and Christiaan Heij. Cross-border electronic commerce: Distance effects and express delivery in european union markets. *International Journal of Electronic Commerce*, 21(2):184–218, 2017.

[Kha16] Abdul Gaffar Khan. Electronic commerce: A study on benefits and challenges in an emerging economy. *Global Journal of Management and Business Research*, 2016.

[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.

[Vro19] Ted Vrountas. Simplifying the header bidding process what marketers should know. https://instapage.com/blog/what-is-header-bidding, 2019. Published: January 29, 2020.

[vSK+01] Frans DJ van Schaik, Jack PC Kleijnen, et al. Sealed-bid auctions: case study. Technical report, 2001.

[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.

[WSKS20] Gerhard Wagner, Hanna Schramm-Klein, and Sascha Steinmann. Online retailing across e-channels and e-channel touchpoints: Empirical studies of consumer behavior in the multichannel e-commerce environment. *Journal of Business Research*, 107:256–270, 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? [N/A]
   (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)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (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)? [N/A]
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]