Regulation of Algorithmic Collusion, Refined: Testing Worst-case Calibrated Regret

Jason D. Hartline **Chang Wang** **Chenhao Zhang**
Northwestern University
hartline@northwestern.edu, {wc, chenhao.zhang.rea}@u.northwestern.edu

Abstract

We study the regulation of algorithmic (non-)collusion amongst sellers in dynamic imperfect price competition by auditing their data as introduced by Hartline et al. [22].

We develop an auditing method that tests whether a seller's worst-case calibrated regret is low. The worst-case calibrated regret is the highest calibrated regret that outcomes compatible with the observed data can generate. This method relaxes the previous requirement that a pricing algorithm must use fully-supported price distributions to be auditable. This method is at least as permissive as any auditing method that has a high probability of failing algorithmic outcomes with non-vanishing calibrated regret. Additionally, we strengthen the justification for using vanishing calibrated regret, versus vanishing best-in-hindsight regret, as the non-collusion definition, by showing that even without side information, the pricing algorithms that only satisfy weaker vanishing best-in-hindsight regret allow an opponent to manipulate them into posting supra-competitive prices.

We motivate and interpret the approach of auditing algorithms from their data as suggesting a per se rule. However, we demonstrate that it is possible for algorithms to pass the audit by pretending to have higher costs than they actually do. For such scenarios the rule of reason can be applied to bound the range of costs to those that are reasonable for the domain.

1 Introduction

The prevailing practice of making pricing decisions with algorithms by sellers in competitive markets has drawn scrutiny from lawmakers and regulators for concerns about price collusion. For example, the US Department of Justice recently filed a lawsuit against RealPage [11], a company providing algorithmic pricing software for landlords in the apartment rental market, for allegedly facilitating price collusion. As Attorney General Merrick Garland stated, "We allege that RealPage's pricing algorithm enables landlords to share confidential, competitively sensitive information and align their rents. Using software as the sharing mechanism does not immunize this scheme from Sherman Act liability...", the legal ground of the regulatory action is based on the argument that the algorithm provided by RealPage enables a covert communication channel for market participants to coordinate and maintain higher than competitive prices. There is one important reason behind this argument: In many jurisdictions such as the US, tacit collusion with the absence of communication is actually not illegal [20, 21].

On the other hand, researchers recently [6, 2, 3] have also discovered a more disturbing fact that popular reinforcement learning algorithms can learn to collude without explicit communication just by engaging in repeated market interactions. These forms of implicit collusion pose new challenges for the regulation of algorithmic collusion.

2nd Workshop on Regulatable ML at NeurIPS 2024.

There are two doctrines of antitrust analysis in the US legal framework: per se rule and rule of reason [23]. The per se rule deems certain business practices, like price-fixing or market division, illegal without requiring further investigation into their actual competitive effects. For example, in response to the algorithmic collusion concern around RealPage, San Francisco recently enacts the per se regulation that precludes any use of algorithms in rental pricing [9]. The rule of reason, on the other hand, involves a more thorough evaluation of business practices, assessing whether they unreasonably restrain trade by examining their purpose, effects, and the overall context of the market.

This paper considers the regulation of algorithmic collusion by auditing from pricing data as proposed by Hartline et al. [22]. In their proposal, they argue that it is feasible to require all sellers deploying pricing algorithms to pass the audit. In other words, sellers do not pass the audit are automatically deemed illegal. Therefore, we reinterpret this proposal as a per se rule.

They then define a per se rule (see the next paragraph) and propose a non-collusion audit implemented by a statistical test on the data collected during the deployment of a seller's algorithm.

Based on the ideas and observations from the theory of online learning, Hartline et al. [22] propose using *calibrated regret* as a quantitative measure for non-collusion for a seller running pricing algorithms. Informally, given a sequence of market conditions and pricing decisions made by a seller, calibrated regret measures how much she can be better off by utilizing the information revealed from her pricing decisions. Low calibrated regret indicates that the seller is close to best responding to the market environment she faces, which implies that she is not colluding since best responding corresponds to competitive behavior. To empirically audit non-collusion of a seller on the collected data, they develop a method for statistically efficiently testing low calibrated regret of the seller. Being unable to pass the test is a violation of their suggested per se rule.

1.1 Our Contributions

We study the framework by Hartline et al. [22] in light of the standard guideline of binary classification: minimizing the number of false negatives and false positives. This is also suggested by Harrington [20] when designing a per se rule specifying a prohibited set of algorithms:

> **false positives** "The more that efficiency-enhancing algorithms are included in the prohibited set, the more harm is created by the associated foregone surplus."
>
> **false negatives** "The more that collusion-promoting algorithms are not included in the prohibited set, the more harm is created because there is collusion that, instead of being prosecuted and shut down, continues unabated."

Our main contribution is three-fold. First, we reduce false positives by providing an improved auditing method that allows algorithms with non-fully supported to be auditable (Section 1.1.1). Second, we give a stronger argument for the need of calibrated regret by showing pricing algorithms minimizing the weaker notion of best-in-hindsight regret can be manipulated into collusion. So the notion of calibrated regret is necessary to reduce false positives (Section 1.1.2). Third, we argue that rule of reason can also be useful by demonstrating that there exists collusive algorithms that could pass the audit with a high inferred cost even if configured with the true cost (Section 1.1.3).

1.1.1 Fewer False Positives

First, we improve in the direction of reducing false positives. We note that there is a significant set of good algorithms that, without modification, is not able to pass the audit of Hartline et al. [22].[1]

Specifically, Hartline et al. [22] assume that the seller's algorithm outputs a distribution of prices in each round. The actual price posted in each round is sampled from the output price distribution. The auditing method computes an estimated regret from a transcript of the pricing algorithm consisting of, in each round: 1) the actual price posted, 2) the observed demand for the posted price, and 3) the distribution of prices, from where the actual posted price is drawn. There is one key requirement: For any pricing algorithm to be auditable, the price distribution in each round must have full support.

[1] Note: Hartline et al. [22] give a procedure for modifying these algorithms so that they can pass the audit.

In other words, in each round, every price level must be posted with at least some probability. We view full-support requirement as restrictive because:

1. The seller might not want to post some prices. First, the seller could possess some side information (that the regulator does not know) that makes him prefer to avoid certain prices. Second, the seller could deliberately avoid some prices due to non-technical reasons (e.g. posting 2.99 instead of 3.00, or avoiding the number 13, etc.).
2. In practice, the exact price distributions of the seller are often unavailable. Asking the seller to submit full price distributions can also be problematic due to privacy issues. To apply the auditing method proposed by Hartline et al. [22], a plausible alternative is to aggregate the prices in a given time window, and use the empirical distribution as the price distribution in that window. If the size of the window is appropriately chosen such that the change in the price distribution is small (for example, when a learning algorithm with small learning rate is used for pricing), then the empirical distribution can be a good approximation to the true price distribution. Of course, this empirical distribution need not be fully-supported.

Remark. In Appendix D we present a formal statement of why aggregating prices to approximate distributions works with small learning rates. An interesting open question is to design a test for small learning rates from data so that the need of price distributions can be removed for sellers with such learning rates.

In this work, we propose a refined auditing method that enables the auditing of algorithms that do not use fully supported price distributions. The refined auditing method continues to use an unbiased estimator for the counterfactual allocations, but it also maintains a *worst-case estimation* for the prices that are not in the support of the price distributions. The new method relaxes the previous requirement that a pricing algorithm must use fully-supported price distributions to be auditable and enables the seller to pass the test by demonstrating her *worst-case calibrated regret* is low. The worst-case regret is the highest regret that outcomes compatible with the observed data can generate.

1.1.2 Fewer False Negatives

Second, we consider false negatives. The calibrated regret that Hartline et al. [22] propose as the non-collusion measure is a strong notion of regret. A common weaker notion is best-in-hindsight regret. Calibrated regret compares the performance of the chosen actions to a counterfactual scenario where the learner may switch among the actions using an arbitrary mapping. Best-in-hindsight regret, on the other hand, compares the performance of the chosen actions to the performance of the best *fixed* action in hindsight. To establish the indispensability of using the stronger notion of calibrated regret for measuring non-collusion, they give a simple example where one seller has side information about buyers' valuation. The seller can utilize her side information to collude with the other seller and have non-positive best-in-hindsight regret. However, she would still have positive calibrated regret in this case.

We give a stronger argument for the need of calibrated regret by showing that a large family of pricing algorithms that minimize the weaker notion of best-in-hindsight regret can be manipulated into collusion. In other words, hindsight-regret algorithms can be susceptible to manipulation, thus, to prevent collusion, a regulator may want to preclude their use.

The argument is inspired by works such as Braverman et al. [5] and Deng et al. [10] showing the vulnerability to manipulation of best-in-hindsight regret minimization algorithms. We construct an instance of the imperfect price competition with two sellers. In our example, the process generating buyers' valuation is stationary and neither of the sellers has any side information about the valuation of the buyer. One seller using a mean-based learning strategy for minimizing best-in-hindsight regret can still be manipulated into maintaining prices above equilibrium level for a significant number of rounds.

1.1.3 Unknown Costs: Per Se Rule v.s. Rule of Reason

Third and finally, we consider the effects of not knowing the seller's cost in the auditing process. In the framework by Hartline et al. [22], the auditor knows the range of the seller's cost but not the exact cost. By their definition, as long as there exists some cost c_* within the range for which the seller's regret is low, the seller is deemed non-collusive. This leads to the following question: Are there

natural algorithms that when configured with the *true* cost c, find outcomes that are considered non-collusive by auditing methods for a higher cost c', while actually being collusive? We provide an affirmative answer to the question by examining a seller using Q-learning that converges to collusive outcomes in simulation environments. The auditing on the seller correctly shows high estimated calibrated regret when a small and precise range of seller's cost is given. However, when the cost range extends beyond seller's true cost by a significant margin, the estimated regret ends up being low.

As a legal implication of this result, solely applying the per se rule of auditing to regulate algorithmic collusion may not be sufficient, particularly when the regulator only has limited information about seller's cost. The knowledge of the cost of the seller can be crucial, and this is when a rule of reason (that investigates the reasonable cost of the seller) can step in.

1.2 Related Work

Calvano et al. [6], Asker et al. [1, 2], Banchio and Skrzypacz [4], Banchio and Mantegazza [3] study various aspect of algorithmic collusion. In the legal domain, Sawyer [27], Gavil [18], Hovenkamp [23] discuss other aspect of antitrust law. Harrington [20], Chassang and Ortner [7] consider the other proposal of regulating algorithmic collusion. Braverman et al. [5], Deng et al. [10] study the manipulation of learning agents. We postpone the detailed discussion of related work to Appendix A.

2 Preliminary

We consider a setting where n sellers repeatedly compete for selling a good in T rounds. Seller i has cost $c_i \in [\underline{c}, \overline{c}]$. In each round t, seller i posts a price $p_i^t \in \mathcal{P}$ where $\mathcal{P}$ is a k-element set of possible price levels.

Let $\overline{p} = \max\{p : p \in \mathcal{P}\}$ be the maximum possible price level. Given all the sellers' prices, the demand (a.k.a. allocation) for seller i is $x_i^t : \mathcal{P}^n \to [0, 1]$. We assume that fixing the prices $\boldsymbol{p}_{-i}$ posted by the sellers other than i, the allocation $x_i^t(p_i, \boldsymbol{p}_{-i})$ is non-increasing in p_i. Seller i's utility at round t posting p is $u_i^t(p) = (p - c_i)x_i^t(p, \boldsymbol{p}_{-i}^t)$. At the end of each round, the seller gets her utility as the feedback.[2] This is known as *bandit feedback* in the literature of online learning. Moreover, the demand can be arbitrary and even adversarial under our framework.

The problem seller i faces is an online-learning problem. Seller i's action in round t can be represented as a price distribution $\pi_i^t \in \Delta(\mathcal{P})$, where $\Delta(\mathcal{P})$ is the set of distributions over $\mathcal{P}$. She posts prices p_i^t according to the distribution π_i^t and obtains the utility resulted from posting p_i^t.

The seller's behavior in a sequence of rounds of competitions can be summarized as a *transcript*. As is the only feedback the regulator can assume the seller gets at the end of each round, the transcript contains the allocation $x_i^t(p_i^t)$ corresponding to the price the seller posted, but not the full demand function $x_i^t(\cdot)$.

Definition 2.1. Call $\mathcal{T}_i^t = \{x_i^s(p^s), p_i^s, \pi_i^s\}_{s=1}^t$ where $p_i^s \sim \pi_i^s$ a *transcript* of length t for seller i. The set of all the length-t transcripts for seller i is denoted as $\mathcal{H}^t$.

As an auditor, given the transcript of the seller, we want to test whether the seller is exhibiting (non-)collusive behavior.

In this work we focus on seller i's behavior, so we will drop the subscript i whenever possible. We denote the sequences $\boldsymbol{x}^T := \{x^t\}_{t=1}^T$, $\boldsymbol{p}^T := \{p^t\}_{t=1}^T$, and $\boldsymbol{\pi}^T := \{\pi^t\}_{t=1}^T$.

Hartline et al. [22] propose that the seller is non-collusive if the transcript satisfies the vanishing calibrated regret property. We define calibrated regret and vanishing calibrated regret as follows.

Definition 2.2. Given the ground-truth $\boldsymbol{x}^T$ and seller's cost c, the *calibrated regret* of the *transcript* for a seller with cost c is

$$R^T(\boldsymbol{x}^T, c) = \max_{\sigma:\mathcal{P}\to\mathcal{P}} \frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{p\sim\pi^t} \left[u(\sigma(p^t), x^t) - u(p, x^t)\right].$$

[2] She might also get other information, but we as auditors cannot directly observe other information.

The seller's calibrated regret is called *vanishing* if $\lim_{T\to\infty} R^T(\boldsymbol{x}^T, c) = 0$.

Unless otherwise noted, we call "calibrated regret" as "regret." Since the auditor does not know the true cost c, as long as there exists a *plausible cost* $c_* \in [\underline{c}, \bar{c}]$ such that $\lim_{T\to\infty} R^T(\boldsymbol{x}^T, c_*) = 0$, the seller is considered *plausibly non-collusive*. The auditing method that Hartline et al. [22] provides is based on estimating the calibrated regret. However, for the auditing method to work properly and provide a meaningful guarantee, it imposes an *auditability requirement* that for all $1 \le t \le T$, the price distribution π^t must be fully-supported. Sellers using algorithms that have vanishing calibrated regret but do not satisfy this requirement are unable to pass the audit without modification.

3 A Framework of Auditing Methods

In this section we present a framework that defines a property called *consistency* which describes that an auditing method *correctly* audits an algorithm (Definition 3.3). Although the auditing method that Hartline et al. [22] propose satisfies a more restrictive consistency property (Definition 3.4), it relies on the *full support requirement*, which means that the pricing algorithm must use every price with non-zero probability. When auditing algorithms that may not randomize over prices with full support, there is missing information because outcomes for prices that are posted with zero probability cannot be estimated. We show that the more restrictive consistency property cannot be satisfied with missing information and this is why we relax it to Definition 3.3. We refer to Definition 3.3 as the *one-sided* consistency requirement and Definition 3.4 as the *two-sided* consistency requirement. Then, under the one-sided consistency requirement, we define the notion of *worst-case [counterfactual] regret* that uses conservative upper bounds on the allocation when there is missing information (Definition 3.8). Finally, we show that a correct auditing method under one-sided consistency must make decisions by considering that the regret of the seller is at least the worst-case regret. This motivates the design of the improved auditing method in Section 4.

From Auditing Methods to Regret Estimators First, we claim that it is without loss of generality to focus on regret estimators when studying auditing methods. This is done via a reduction argument. That is, if we have an approximately correct auditing method, then we also have an approximately correct regret estimator, and vice versa.

To begin we define auditing methods and regret estimators as follows:

Definition 3.1. Given a cost c and regret threshold $r > 0$, an *auditing method* is a mapping $\mathcal{A} : \bigcup_{t\ge 0} \mathcal{H}^t \to \{\mathrm{G}, \mathrm{S}\}$, and the output indicates the regret is greater (G) or smaller (S) than r *assuming* the seller's cost c. A *regret estimator* is a mapping $\mathcal{A} : \bigcup_{t\ge 0} \mathcal{H}^t \to \mathbb{R}$, and the output indicates the estimated regret *assuming* the seller's cost c.

Note that the auditing method proposed by Hartline et al. [22] is based on a regret estimator, so it suffices to reduce regret estimation to auditing.

Proposition 3.2. *Suppose we have an auditing method $\mathcal{A}$ that correctly outputs whether the regret of a transcript $\mathcal{T}^T$ of length T is greater or smaller than r with error probability at most $f(T)$.*[3] *Then there exists an regret estimator of the regret of $\mathcal{T}^T$ up to accuracy ε and error probability at most $\frac{\overline{p} f(T)}{\varepsilon}$.*

The Consistency Requirement From now on we focus on regret estimators. To do a correct hypothesis testing with the regret estimator, we want the regret estimator to be approximately *consistent*, defined below:

Definition 3.3 (Consistency, one-sided)**.** A regret estimator $\mathcal{A}$ is *consistent* if for any sequence $\{\{x^t(p^t)\}_{t=1}^T, \boldsymbol{p}^T, \boldsymbol{\pi}^T\}_{T\ge 1}$ of transcripts, $\varepsilon > 0$, cost c, and sequence of ground-truth sequence of allocations $\{\boldsymbol{x}^T\}_{T\ge 1}$ agreeing with the transcripts,

$$\lim_{T\to\infty} \Pr_{\boldsymbol{p}^T \sim \boldsymbol{\pi}^T} \left[\mathcal{A}(\mathcal{T}^T) < R^T(\boldsymbol{x}^T, c) - \varepsilon\right] = 0.$$

The above definition says that a regret estimator must approximately output at least the true regret of the transcript in the limit.

[3] The error probability typically satisfies $f(T) = o(1)$ so its accuracy increases as T increases.

Before we study the implication of Definition 3.3, we explain why we only require a regret estimator to approximate an upper bound of the true regret instead of approximating the true regret itself—The following definition is tempting:

Definition 3.4 (Consistency, two-sided)**.** A regret estimator $\mathcal{A}$ is *consistent* for a set of transcripts S if, for any sequence $\{\{x^t(p^t)\}_{t=1}^T, \boldsymbol{p}^T, \boldsymbol{\pi}^T\}_{T\geq 1}$ of transcripts in S, $\varepsilon > 0$, cost c, and sequence of ground-truth sequence of allocations $\{\boldsymbol{x}^T\}_{T\geq 1}$ agreeing with the transcripts,

$$\lim_{T\to\infty} \Pr_{\boldsymbol{p}^T\sim\boldsymbol{\pi}^T} \left[|\mathcal{A}(\mathcal{T}^T) - R^T(\boldsymbol{x}^T, c)| \geq \varepsilon\right] = 0.$$

Next we explain why Definition 3.4 is not appropriate in the general case (i.e. for sellers with not fully-supported price distributions). Although the regret estimator in Hartline et al. [22] indeed satisfies the two-sided consistency property (Proposition 3.5), it only works for transcripts with fully-supported price distributions. Unfortunately, the two-sided consistency requirement is too strong for algorithms that accept all the possible transcripts (Proposition 3.6). In other words, there are pricing algorithms that produce transcripts for which the regret cannot be consistently (according to Definition 3.4) estimated.

Proposition 3.5. *Let $\overline{\Delta}(\mathcal{P})$ be the set of fully-supported price distributions over $\mathcal{P}$. The algorithm in Hartline et al. [22] is consistent (two-sided), for the set of transcripts satisfying $\underline{\pi}^T = \min_{t\leq T,p} \pi^t(p) = \omega(T^{(-1/4)})$,*

Proposition 3.6. *No deterministic regret estimator is consistent (two-sided) for the set of all transcripts. In particular, there exists a seller who has vanishing true regret, but her regret cannot be consistently (two-sided) estimated.*

The above proposition shows that it is not possible to get an estimator of the regret satisfying Definition 3.4 in the general case. The auditor could be unable to certify a truly non-collusive seller. This is why we ask for a relaxed property in Definition 3.3 that the regret estimator must output an upper bound of the regret with high probability.

We turn back to the discussion of the one-sided consistency requirement in the general case. Recall that our philosophy is that it is the seller's responsibility to demonstrate enough information that she is non-collusive. The one-sided consistency property ensures that missing information is properly accounted for so that a collusive seller is never deemed as non-collusive because of the regret estimation.

Of course the one-sided consistency requirement does not rule out regret estimators that always output trivial upper bounds of the regret. In the next section we will provide an algorithm that outputs the least possible upper bound.

Finally, Definition 3.3 has an important implication: Whenever there is some missing information, any regret estimator satisfying Definition 3.3 must output at least the *worst-case [counterfactual] regret* of the transcript. We first define worst-case [counterfactual] regret (Definition 3.8) and then show the implication (Proposition 3.9).

Intuitively, the worst-case regret is the highest regret that outcomes compatible with the observed data can generate. We define such compatibility between outcomes and the observed data as follows.

Definition 3.7 (Indistinguishable allocations)**.** Fix the sequence of price distributions $\boldsymbol{\pi}^T$, and let $C^t = \{p \in \mathcal{P} : \pi^t(p) > 0\}$ be the set of price levels that have non-zero probability being posted in round t. Two sequences of allocations $\boldsymbol{x}^T, \boldsymbol{z}^T$ are called *indistinguishable* if $x^t(\cdot)$ and $z^t(\cdot)$ have the same support C^t for every $1 \leq t \leq T$ and

$$x^t(p) = z^t(p) \quad \text{for all} \quad p \in C^t \quad \text{and} \quad 1 \leq t \leq T.$$

The indistinguishable relation is an equivalence relation. If $\boldsymbol{x}^T, \boldsymbol{z}^T$ are indistinguishable, then there is no way to separate them from data.

With the definition of indistinguishable (compatible) allocations, we define the worst-case regret.

Definition 3.8 (Worst-case counterfactual regret)**.** Fix the sequence of price distributions $\boldsymbol{\pi}^T$, and with a given sequence of allocations $\boldsymbol{x}^T = \{x^t\}_{t=1}^T$, the *worst-case [counterfactual] regret* is defined as

$$\overline{R}^T(c, \boldsymbol{x}^T) = \sup\{R^T(c, \boldsymbol{z}^T) : \boldsymbol{z}^T \text{ indistinguishable with } \boldsymbol{x}^T\}.$$

The following proposition implies that, by only looking at the transcript, the auditor cannot rule out the possibility that the true regret is as high as the worst-case regret.

Proposition 3.9. *Any one-sided consistent regret estimator $\mathcal{A}$ must satisfy*

$$\lim_{T\to\infty} \Pr_{\boldsymbol{p}^T\sim\boldsymbol{\pi}^T}\left[\mathcal{A}(\mathcal{T}^T) < \overline{R}^T(\boldsymbol{x}^T, c) - \varepsilon\right] = 0,$$

for any $\varepsilon > 0$, cost c, transcript $\mathcal{T}^T$, and sequence of ground-truth sequence of allocations $\{\boldsymbol{x}^T\}_{T\geq 1}$.

Inspired by Proposition 3.9, in the next section we present an auditing method that enables the seller to pass the test by demonstrating her worst-case regret is low.

4 Testing Worst-case Regret

In this section we present the refined auditing method that estimates the worst-case regret (as defined in the previous section). We then show the following guarantee: With a sufficient amount of data, if the seller's worst-case regret is low, then she passes the audit with high probability, and if the seller's worst-case regret is high at every cost in $[\underline{c}, \bar{c}]$, then she fails the audit with high probability.

First we do a decomposition of calibrated regret so that we can compute it efficiently. Recall that the calibrated regret is defined to be the maximum benefit of deviation by doing a price swap $\sigma : \mathcal{P} \to \mathcal{P}$. A useful decomposition of calibrated regret is to first compute the benefit of changing price p to q, then take the maximum over all possible $q \in \mathcal{P}$, and finally sum the result over all $p \in \mathcal{P}$. Formally, let

$$R^T_{p,q}(c, \boldsymbol{x}^T) = \frac{1}{T}\sum_{t=1}^{T} \pi^t(p)\left[(q-c)x^t(q) - (p-c)x^t(p)\right],$$

then we have

$$R^T(c, \boldsymbol{x}^T) = \sum_p \max_q R^T_{p,q}(c, \boldsymbol{x}^T).$$

The auditing method estimates the worst-case regret based on the above decomposition and the worst-case estimation of allocations from data. The steps are described in a high level as follows.[4]

The input to the general auditing method contains the prices the seller posts in each round $\{p^t\}_{t=1}^T$, the allocations (demands) of the posted prices $\{x^t(p^t)\}_{t=1}^T$, seller's price distributions $\{(\pi^t)\}_{t=1}^T$, and the threshold r. The price distributions need not be fully-supported. The method proceeds in the following steps:

Step 1 We estimate the allocation every round using the transcript. For each round t, let $C^t := \{p \in \mathcal{P} : \pi^t(p_j) > 0\}$ be the support of the price distribution π^t. For every price $p \in C^t$, the propensity score estimator is used to estimate the allocation

$$\hat{x}^t(p) = \begin{cases} x^t(p)/\pi^t(p) & p = p^t, \\ 0 & \text{otherwise.} \end{cases}$$

For the prices that are not in the support, we use the estimator of the allocation at the largest price p' that is smaller than p while being in the support.

$$\hat{h}^t(p) := x^t(p') \quad \text{where} \quad p' = \max\{r \leq p : r \in C^t\}.$$

If no such price exists, then the estimation is capped with 1.

Step 2 We estimate the true regret of the worst-case allocation $\overline{R}^T(c, \boldsymbol{x}^T)$ with the estimator $\widetilde{R}^T(c, \boldsymbol{x}^T)$, built up from the estimator $\widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T)$ for $\overline{R}^T_{p,q}(c, \boldsymbol{x}^T)$. Specifically, we first compute the benefit of substituting price p with q

$$\widetilde{R}^T_{p,q}(c) = \frac{1}{T}\sum_{t=1}^{T} \pi^t(p)\left[(q-c)\hat{h}^t(q) - (p-c)\hat{h}^t(p)\right].$$

[4] A formal pseudocode description can be found at the end of the Appendix (Algorithm 1).

Then the worst-case regret can be estimated by summing the highest benefit of changing each price

$$\widetilde{R}^T(c) = \sum_p \max_q \widetilde{R}^T_{p,q}(c).$$

Step 3 We minimize the worst-case regret over all the possible costs to compute the worst-case plausible regret $\min_{c\in[\underline{c},\bar{c}]} \widetilde{R}^T(c)$. Note that this can be done in polynomial time even with a continuum of costs. In fact, following the observations of Nekipelov et al. [26], for each p, q, $\widetilde{R}^T_{p,q}(c)$ is linear in c. Therefore, $\widetilde{R}^T(c) = \sum_p \max_q \widetilde{R}^T_{p,q}(c)$ is a convex function of c, which can be efficiently minimized over the closed set $[\underline{c}, \bar{c}]$.

Step 4 Finally, we compare the estimated plausible regret plus an additional error margin δ^T with the required threshold. The error margin ensures that the seller cannot pass the audit when the information revealed from the transcript is insufficient to guarantee reliability of the regret estimator. Specifically, if $\widetilde{R}^T(\tilde{c})+\delta^T \le 2r$, then we output PASS, and FAIL otherwise, where

$$\delta^T = \frac{k\overline{p}}{T}\sqrt{2\log\left(\frac{2k^2}{\alpha}\right)\cdot\sum_{s=1}^{T}\left(\frac{1}{\min_{p\in C^s}\pi^s(p)}+1\right)^2}.$$

The following theorem identifies the sample complexity of testing worst-case calibrated regret.

Theorem 4.1 (Sample complexity of testing worst-case regret)**.** *Let c_0 be the seller's true cost and $c_* = \arg\min_{c\in[\underline{c},\bar{c}]} \overline{R}^T(c, \boldsymbol{x}^T)$ be the plausible cost of the seller. Fix confidence level $1-\alpha$, threshold r and let $\underline{\pi} = \min_{p\in C^t, 1\le t\le T}\pi^t(p)$. With our refined auditing method, when the number of rounds*

$$T \ge \log\frac{2k^2}{\alpha}\cdot 2\left(\frac{k\overline{p}}{r}\right)^2\cdot\left(\frac{1}{\underline{\pi}}+1\right)^2,$$

we have

1. *if the seller's true worst-case regret $\overline{R}^T(c_0, \boldsymbol{x}^T) \le r$, she passes w.p. at least $1-\alpha$; and*
2. *if her plausible worst-case regret $\overline{R}^T(c_*, \boldsymbol{x}^T) \ge 2r$, then she fails w.p. at least $1-\alpha$.*

A direct corollary of the above theorem (together with Proposition 3.9) is that the regret estimator in our refined auditing method outputs the least upper bound of the regret of the seller, because Proposition 3.9 asks the regret estimator to output approximately at least the worst-case regret, and the following corollary says it outputs approximately at most the worst-case regret. This implies that our refined auditing method provides the most information about the transcript, given the one-sided consistency requirement. In other words, our method is at least as permissive as any auditing methods that have a high probability of failing algorithmic outcomes with non-vanishing calibrated regret.

Corollary 4.2. *Let the regret estimator in our refined auditing method be $\mathcal{A}$. It satisfies*

$$\lim_{T\to\infty}\Pr_{p^t\sim\pi^t}\left[\mathcal{A}(\mathcal{T}^T) > \overline{R}^T(\boldsymbol{x}^T, c) + \varepsilon\right] = 0,$$

for any $\varepsilon > 0$, cost c, transcript $\mathcal{T}^T$, and sequence of ground-truth sequence of allocations $\{\boldsymbol{x}^T\}_{T\ge1}$.

Proof. A direct corollary from the proof of Theorem 4.1. □

In the following two sections, we study two technical details on the concepts and assumptions used in our auditing, which have practical implications in law. In Section 5 we provide justifications that the more stronger calibrated regret must be used instead of weaker best-in-hindsight regret by arguing that best-in-hindsight regret includes more collusive algorithms and cannot prevent collusion in a unilateral way. In Section 6, we demonstrate that it is possible for algorithms to pass the audit by pretending to have higher costs than they actually do. For such scenarios the rule of reason can be applied to bound the range of costs to those that are reasonable for the domain.

5 Best-in-hindsight Regret is Manipulable

Much online learning literature develops algorithms to satisfy vanishing best-in-hindsight (a.k.a. external) regret. [22] argue that the stronger vanishing calibrated regret is essential for non-collusion by giving an example with side information where a seller colludes while having non-positive best-in-hindsight regret. In this section, we show a stronger argument by demonstrating that even in environments without side information, algorithms can have vanishing best-in-hindsight regret while being susceptible to collusion. This implies

1. If we use the more permissive vanishing best-in-hindsight regret as the definition of non-collusion, then there could be more collusion promoting algorithms passing the audit (recall false negatives).
2. A non-collusion definition using vanishing best-in-hindsight regret would *not* be a unilateral property that an algorithm can satisfy independently of what other algorithms are doing.

Combining the above argument with the fact that calibrated regret minimization leads to approximate correlated equilibria [17], and a manipulator of calibrated-regret-minimization algorithm cannot get more than the payoff of Stackelberg equilibirum [10], it is reasonable to require vanishing calibrated regret in the definition of non-collusion even when there is no side information.

We construct an instance of imperfect price competition without side information and show that one seller using a vanishing best-in-hindsight regret minimization algorithm can be manipulated into posting higher-than-equilibrium prices. In our construction, both sellers (both the manipulator and the manipulated the seller) have no best-in-hindsight regret, while both have non-vanishing calibrated regret. Therefore, there exists a scenario where non-collusion definition of vanishing best-in-hindsight regret fails to identify a collusion.

To begin, consider a setting of dynamic imperfect price competitions with 2 sellers. Let V_1, V_2 be the highest (correlated) equilibrium payoff for seller 1 and seller 2, in which they play a joint distribution of prices π^e. If the equilibrium strategy is pure, then let p_1^e, p_2^e be the prices they play. We define a family of commonly used best-in-hindsight regret minimization algorithms as follows.

Definition 5.1 (γ-mean-based learning, [5])**.** Fix horizon T and $\gamma = o(1)$. Let $\sigma_{p,t} = \sum_{s=1}^{t} u_{p,s}$ be the cumulative utilities for posting price p in the first t rounds. A seller is *γ-mean-based* if the seller posts price p w.p. at most γ as long as there exists another price q such that $\sigma_{q,t} > \sigma_{p,t} + \gamma T$.

Many vanishing best-in-hindsight-regret algorithms, e.g. EXP3, FTPL, are known to be γ-mean-based learning algorithms [5]. In the following theorem, we show that a seller running γ-mean-based learning algorithm is vulnerable to manipulation into collusion. The manipulator can also achieve no best-in-hindsight when doing such manipulation.

Theorem 5.2. *There exists an instance in which the environment is stationary across rounds, both sellers have no side information, and seller 1 can achieve an outcome with the following properties against seller 2 who is using any γ-mean-based learning algorithm:*

1. *(collusion) for $\Omega(T)$ rounds, both play $p_1 > p_1^e, p_2 > p_2^e$ in each round w.h.p.,*[5]
2. *(no loss of payoff) receive expected payoff $V_1'T - o(T), V_2'T - o(T)$ where constants $V_1' > V_1, V_2' > V_2$*[6]*, and*
3. *(no best-in-hindsight regret) both seller 1 and seller 2 have vanishing best-in-hindsight regret.*

In words, Theorem 5.2 says that seller 2 can be manipulated into a significant number (constant fraction) of rounds with supra-competitive prices, while both sellers have no best-in-hindsight regret and get higher-than-equilibrium expected payoffs.

[5] In our construction, the maximum-payoff correlated equilibrium is a pure equilibrium.

[6] In our construction, the payoffs V_1, V_2 are also the Stackelberg equilibrium payoffs.

6 On the Effect of Unknown Costs

In this section, we demonstrate that imprecision in cost information might significantly impact the efficacy of auditing from pricing data and presents challenges to regulating collusion.

Recall that in the auditing problem we assume that the cost of the seller is unknown and the seller passes the audit as long as the *plausible* regret is low (recall that the plausible regret is obtained by minimizing the regret over $[\underline{c}, \overline{c}]$). Therefore, the seller and/or the algorithm can potentially manipulate the cost so that the outcome is actually collusive, but is seen as non-collusive with a higher inferred plausible cost in the auditing method. We now ask, can this really happen? In other words, is it possible for algorithms to pass the audit by pretending to have higher costs than they actually do?

Before providing the answer, we need to clarify the formulation of the question. Note that the following two scenarios are different:

1. A seller deliberately inputs a fake cost to the pricing algorithm, causing the algorithm to post prices higher than what is optimal for her true cost.
2. A seller truthfully reports her cost to the pricing algorithm, but the algorithm finds a collusive outcome that looks competitive with a higher cost.

Since algorithmic collusion refers to collusion facilitated by the algorithm, Item 1 is not *algorithmic* collusion, but Item 2 is. Thus more precisely, our question is

> Are there natural algorithms that when configured with the *true* cost c, find outcomes that are considered non-collusive by auditing methods for a higher cost c', while actually being collusive?

We find an affirmative answer to this question. This has two implications. First, the incomplete knowledge of the cost of the seller could dramatically affect the result of the audit. Even if configured with true costs, collusive algorithms might pass the test with a favorable inferred plausible cost. Second, such behavior is hard to distinguish from genuine competition by only looking at the pricing data, which is a new challenge for auditing *algorithmic* collusion.

A mitigation to this problem is applying rule of reason: A further investigation of the sellers and market contexts is needed to narrow down the cost range.

The details of the simulation experiment leading to our conclusion of this section are described in Appendix B.

7 Conclusion

In this work, we explore several questions around auditing (non-)collusion for pricing algorithms from data based on the framework of Hartline et al. [22]. We motivate and interpret our study under the legal doctrines of antitrust analysis. We develop a refined auditing method that relaxes the previous requirement that a pricing algorithm must use fully-supported price distributions to be auditable by testing the worst-case regret, thus allowing more efficiency-enhancing algorithms to be auditable. We give an example demonstrating that requiring vanishing-calibrated regret as the non-collusion definition being essential to eliminate more collusion-promoting algorithms and prevent collusion unilaterally. Our experiment results show that under the current auditing framework, a regulator with very limited knowledge about a seller's cost may be unable to detect collusive behavior of the seller, which suggest a rule of reason can be useful in antitrust analysis. Open questions include designing a test for small learning rates to remove the need of distributions and improving the sample complexity bound.

References

[1] John Asker, Chaim Fershtman, and Ariel Pakes. Artificial intelligence, algorithm design, and pricing. In *AEA Papers and Proceedings*, volume 112, pages 452–456. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2022.

[2] John Asker, Chaim Fershtman, and Ariel Pakes. The impact of artificial intelligence design on pricing. *Journal of Economics & Management Strategy*, 2023.

[3] Martino Banchio and Giacomo Mantegazza. Adaptive algorithms and collusion via coupling. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC '23, page 208, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701047. doi: 10.1145/3580507.3597726. URL https://doi.org/10.1145/3580507.3597726.

[4] Martino Banchio and Andrzej Skrzypacz. Artificial intelligence and auction design. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 30–31, 2022.

[5] Mark Braverman, Jieming Mao, Jon Schneider, and Matt Weinberg. Selling to a no-regret buyer. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 523–538, 2018.

[6] Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–97, 2020.

[7] Sylvain Chassang and Juan Ortner. Regulating Collusion. *Annual Review of Economics*, 15 (1):177–204, 2023. doi: 10.1146/annurev-economics-051520-021936. URL https://doi.org/10.1146/annurev-economics-051520-021936.

[8] Sylvain Chassang, Kei Kawai, Jun Nakabayashi, and Juan Ortner. Robust screens for noncompetitive bidding in procurement auctions. *Econometrica*, 90(1):315–346, 2022.

[9] City and County of San Francisco. Administrative code - ban on automated rent-setting. https://sfgov.legistar.com/LegislationDetail.aspx?ID=6789588&GUID=89BA28F7-B3B8-44D0-806B-FFDC5FC29015, 2024. Accessed: 2024-09-27.

[10] Yuan Deng, Jon Schneider, and Balasubramanian Sivan. Strategizing against no-regret learners. *Advances in neural information processing systems*, 32, 2019.

[11] Department of Justice. Justice department sues realpage for algorithmic pricing scheme that harms millions of american renters. https://www.justice.gov/opa/pr/justice-department-sues-realpage-algorithmic-pricing-scheme-harms-millions-american-renters, 2024. Accessed: 2024-08-31.

[12] *Addyston Pipe & Steel Co.* v. *United States*, 1899.

[13] *Brooke Group Ltd.* v. *Brown & Williamson Tobacco Corp.*, 1993.

[14] *Chicago Board of Trade* v. *United States*, 1918.

[15] *Standard Oil Co. of New Jersey* v. *United States*, 1911.

[16] Sara Fish, Yannai A Gonczarowski, and Ran I Shorrer. Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806*, 2024.

[17] Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

[18] Andrew I Gavil. Moving beyond caricature and characterization: The modern rule of reason in practice. *S. Cal. L. Rev.*, 85:733, 2011.

[19] Karsten T Hansen, Kanishka Misra, and Mallesh M Pai. Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science*, 40(1):1–12, 2021.

[20] Joseph E Harrington. Developing competition law for collusion by autonomous artificial agents. *Journal of Competition Law & Economics*, 14(3):331–363, 2018.

[21] Joseph E Harrington. The effect of outsourcing pricing algorithms on market competition. *Management Science*, 68(9):6889–6906, 2022.

[22] Jason D Hartline, Sheng Long, and Chenhao Zhang. Regulation of algorithmic collusion. In *Proceedings of the Symposium on Computer Science and Law*, pages 98–108, 2024.

[23] Herbert Hovenkamp. The rule of reason. *Fla. L. Rev.*, 70:81, 2018.

[24] *In re Text Messaging Antitrust Litigation*. *F. 3d*, 782 (No. 14-2301):867, 2015.

[25] Timo Klein. Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, 52(3):538–558, 2021.

[26] Denis Nekipelov, Vasilis Syrgkanis, and Éva Tardos. Econometrics for learning agents. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, page 1–18, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334105. doi: 10.1145/2764468.2764522. URL https://doi.org/10.1145/2764468.2764522.

[27] Laura Phillips Sawyer. *US antitrust law and policy in historical perspective*. Harvard Business School, 2019.

[28] Christopher John Cornish Hellaby Watkins. *Learning from delayed rewards*. PhD thesis, King's College, Cambridge United Kingdom, 1989.

A Detailed Discussion on Related Work

Algorithmic Collusion Most papers studying algorithmic collusion from technical perspectives consider the Q-learning algorithm [28], a common reinforcement learning algorithm. Calvano et al. [6], Klein [25], Asker et al. [1, 2], Banchio and Skrzypacz [4], Banchio and Mantegazza [3] study Q-learning under various settings with simulations and theoretical analysis. They consistently report that Q-learning can find and maintain, without explicit communication, supra-competitive prices (or infra-competitive bids) when in competition with each other. A few other papers have also explored algorithmic collusion beyond Q-learning, such as UCB [19] and large language models [16]. Our simulation in Section 6 follows the setup of Banchio and Skrzypacz [4]. These empirical and theoretical findings and concerns are one of the main motivations of our work.

Legal Landscape of Anti-collusion Analysis US statues regulating price collusion were enacted more than a hundred years ago, long before the era of digital markets and algorithmic pricing. They include the Sherman Act (1890), the Federal Trade Commission Act (1914), and the Clayton Act (1914). The recent court cases such as [24, 13] interpreting these statues for price collusion have affirmed the jurisprudence of requiring express agreement as the prerequisite of establishing liability.

The Sherman Act literally prohibits acts that "in restraint of trade and commerce" without clarifying how it should be applied [27]. In early cases such as [12, 14], the language of the statue is interpreted as applicable to any restraint of trade, which constitutes the per se mode of analysis. The rule of reason doctrine in antitrust first appeared in the US Supreme Court ruling of *Standard Oil Co. of New Jersey* v. *United States* [15]. Led by the then Chief Justice Edward White, the court decided that the Sherman Act should be "construed in the light of reason," hence only applies to *unreasonable* restraints of trade. Over the years, the court has narrowed the domain of per se rules in traditional antitrust cases while incorporating more analysis informed by economic principles to the application of rule of reason. Sawyer [27], Gavil [18] discuss the evolution of the two doctrines. Hovenkamp [23] discuss the scope that rule of reason analysis should be applied in the non-algorithmic antitrust settings. In light of the new challenges posed by algorithmic collusion, Harrington [20] propose adding per se prohibition for certain algorithms to competition laws. We motivate and interpret our work within the legal framework proposed by Harrington [20].

Regulation of Algorithmic Collusion In additional to the auditing approach proposed in Hartline et al. [22], other work Harrington [20] and Chassang and Ortner [7] discuss alternative proposals of regulating algorithmic collusion.

Harrington [20] discuss the approaches of static checking an algorithm's source code and dynamic testing the algorithm with synthetical input to learn its properties. They consider these approaches as means of determining whether the algorithm is prohibited. However, they also suggest that to what extend the prohibition comes from a per se rule or rule of reason depends: Per se rule can be applied for clear collusion-identifying properties checkable with these approaches. Otherwise, rule of reason is more appropriate.

Therefore, given the current development of technology and understanding of algorithmic collusion, applying static checking and dynamic testing on regulating collusion are still more in line of the rule of reason doctrine. Static checking can be used to partially verify certain properties of some algorithms, but usually do not scale well enough to handle complicated properties encoding clear collusive behavior of sophisticated pricing algorithms. On the other hand, the input to pricing algorithms are large in dimension, dynamic, and potentially idiosyncratic across different algorithms. The choice of synthetical input for dynamic checking introduces a significant amount of variability in the process. Finally, these approaches all require access to the algorithms for close inspection, which is a characteristic of rule of reason.

In contrast, the auditing from data approach we consider provides a clear prohibition determination without requiring access to the algorithms themselves, which closely resembles a per se rule (see Table 1 for a possible categorization of proposed methods into dichotomy between per se rule and rule of reason).

Chassang and Ortner [7] propose another regulation based on the relation of regret and collusion observed in Chassang et al. [8]. The regulation approach they propose requires the seller attaching a supervisor algorithm to the pricing algorithm to ensure that the composition satisfies no regret

Regulation	Classification
outright ban on algorithms (e.g. City and County of San Francisco [9])	per se
static checking (check the source code)[20]	leaning towards rule of reason
dynamic testing [20]	leaning towards rule of reason
requiring supervising wrapper [7]	per se
requiring passing auditing ([22] and this paper)	per se

Table 1: Comparison of different proposed regulation of algorithmic collusion

properties. This approach can also be interpreted as a per se rule as it prohibits using algorithms without supervisor attached. Chassang et al. [8] consider the problem of screening non-algorithmic collusion in procurement auctions. Similar to the auditing approach proposed in Hartline et al. [22] that we consider in this work, Chassang and Ortner [7] estimate the demand functions from data and use the demand functions to compute regret-like quantities. However, there are several differences. The framework we consider makes minimal assumptions on the demand functions that a seller faces, namely, the demand is between $[0, 1]$ and monotonically non-increasing in prices. We estimate the demand functions that a single seller faces utilizing the randomization of seller's algorithm without knowledge of other sellers' strategy. In comparison, Chassang et al. [8] consider the estimation problem when the form of the demand functions is known from the auction format and the bids of all bidders are available in the data. To deal with buyer distributions with imperfect competition, their approach would need assumptions on the demand while ours does not. Instead of assuming fixed production cost across rounds, Chassang and Ortner [7] also consider the case when the cost of a seller can be different for each round. Therefore, when computing the regret, the deviation of a seller's strategy has the form of changing the prices across each round proportionally. But in our framework, the deviation can be arbitrary.

Exploitation of a No-regret Learner Braverman et al. [5] consider the problem of repeated selling of an item to an agent using no-regret learning algorithms. They propose the notion of mean-based algorithms and show that mean-based algorithms guaranteeing no best-in-hindsight regret can be manipulated by a seller to extract full surplus. Deng et al. [10] consider the problem of manipulating no-best-in-hindsight-regret learner in general 2-player bimatrix games to get beyond the Stackelberg payoff. Our example in Section 5 is inspired by their work but the construction is tailored in a dynamic, price competition game.

B Details of the Simulation and Results

We show the answer with the simulation below. The configuration resembles that in Banchio and Skrzypacz [4]. We consider two sellers with costs $c_1 = 0.1$ and $c_2 = 0.2$, respectively. The grid of allowable price levels are from 0.05 to 0.95 with step size 0.05. In each round, the buyer's valuations of the two sellers' goods are i.i.d. uniformly distributed over $[0, 1] \times [0, 1]$. The sellers post prices and the reward of the seller is the expected payment from the buyer, net her own cost. At the end of each round, each seller records her posted price, the demand, and her price distribution. The transcript of seller $i\,(i = 1, 2)$ also contains the price posted by seller $-i$ for evaluating the true regret (but the auditing method will not use this information). The competition lasts for $T = 10^6$ rounds and the experiment with the same setup is repeated 100 times. Sellers compete with each other using the Q-learning algorithm. We use the same hyper-parameters as those in Banchio and Skrzypacz [4]. That is, an ε-greedy strategy with optimistic initialization and exploration probability $\varepsilon = 0.001$. The Q-table is updated according to the standard rule

$$Q^{t+1}(p) = (1 - \alpha)Q^{t+1}(p) + \alpha(u^t(p) + \gamma \max_{q \in \mathcal{P}} Q^t(q)) \quad \forall p \in \mathcal{P},$$

where the learning rate $\alpha = 0.05$ and discount factor $\gamma = 0.99$.

In Figure 1 we confirm that Q-learning exhibits collusive behavior. Q-learning converges to strategies that both sellers post prices greater or equal to 0.9 in most cases. The Nash equilibrium of the game is that the sellers post 0.65 and 0.7, respectively.

After the transcripts are generated, we audit the transcripts. As a baseline, we use the prices posted by the opponent to compute the true expected regret of the seller. We then audit the transcripts with our auditing method. The auditing result is shown in Figure 2.

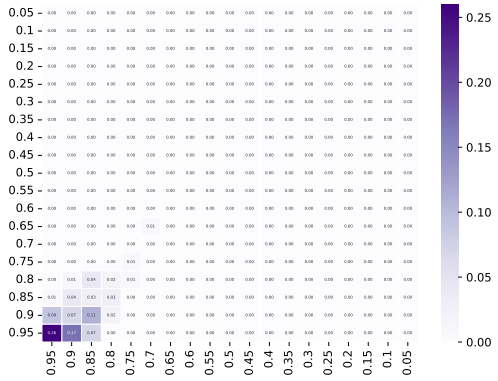
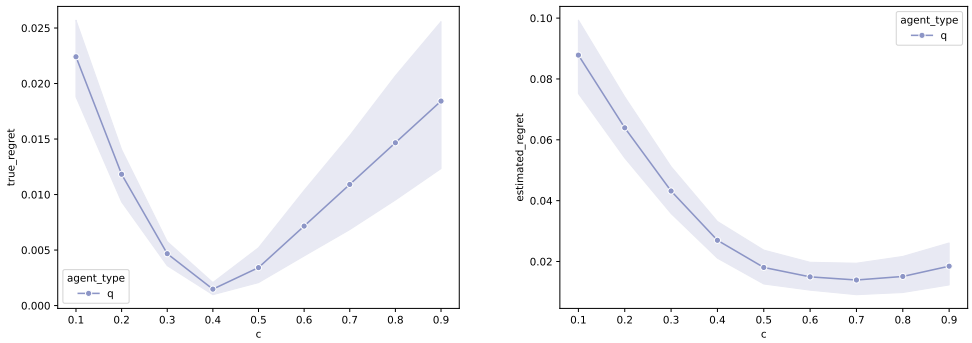


Figure 1: The frequencies of each pair of strategies in the last 10 rounds of the competition are shown in the heatmap. Both the x- and y-axis denote the possible price levels.



(a) (b)

Figure 2: The true regret and estimated regret, plotted against different assumed costs of seller 1.[7]The seller maintains exploration.

In Figure 2 we note that although the true regret eventually increases as the assumed cost of the seller increases (Panel (a)), the auditing method is unable to discover collusion as the estimated regret under high costs is low (Panel (b)). The auditing on the seller correctly shows high estimated calibrated regret when a small and precise range of seller's cost is given (around 0.1). However, when the cost range extends beyond seller's true cost by a significant margin, the estimated regret ends up being low (around 0.8). This means that, although we know Q-learning algorithm converges to collusive prices, if we assume that the cost of the seller is greater or equal to approximately 0.8, then Q-learning algorithm turns out to have a low estimated calibrated regret. In other words, when configured with a low cost (such as $c = 0.1, 0.2$), Q-learning algorithm finds outcomes that look competitive for a higher cost c' (such as $c' = 0.7, 0.8$). This answers the question we raised at the beginning of this section.

A possible explanation of the phenomenon is to consider the extreme case. Let the highest possible cost $\overline{c} = 1$ and consider a seller posting 1. Then if we assume her true cost is 1, then she always has no regret, since any deviation results in a non-positive payoff. However, note that the phenomenon in Figure 2 is not as trivial as the extreme case, because the seller's lowest plausible regret is achieved at prices strictly lower than the price she is posting.

[7]That of seller 2 is similar.

C Omitted Proofs

C.1 Proof of Proposition 3.2

Proof. Note that the regret of any transcript is bounded by $\overline{p}$. We discretize the interval with step size ε and do $\ell = \overline{p}/\varepsilon$ audits with thresholds $r_1 = \varepsilon, r_2 = 2\varepsilon, \ldots, r_\ell = \overline{p}$ simultaneously using $\mathcal{A}$. Each threshold auditing returns either $[0, r_i]$ or $[r_i, \overline{p}]$, indicating which interval the true regret is in. Call this interval J_i. Let $J = \bigcap_{i=1}^{k} J_i$.

Consider the "good event" that all the auditing are correct, then J's length is at most ε, and in this case (when $|J| \le \varepsilon$) we output the midpoint of J as the estimator. Otherwise, we draw a guess uniformly random from $[0, \overline{p}]$. By union bound, with probability at least $1 - \frac{\overline{p} f(T)}{\varepsilon}$ the good event happens, and the estimator is of accuracy at least ε. □

C.2 Proof of Proposition 3.5

Proof. From the proof of Lemma A.1 in [22], we have for any $\varepsilon > 0$

$$P_{\ge\varepsilon}^{T} = \Pr_{\boldsymbol{p}^T \sim \boldsymbol{\pi}^T}\left[|\mathcal{A}(\mathcal{T}^T) - R^T(\boldsymbol{x}^T, c)| \ge \varepsilon\right] \le 2k^2 \exp\left(-\frac{\varepsilon^2}{2k^2 \sum_{t=1}^{T} d^2}\right)$$

where $k = |\mathcal{P}|$ and

$$d = \frac{1}{T}\left(\frac{1}{\underline{\pi}^T} + 1\right)\overline{p}.$$

By assumption on the transcript, we have

$$P_{\ge\varepsilon}^{T} \le 2k^2 \exp\left(-\frac{\varepsilon^2 T}{2k^2 \overline{p}^2 (\underline{\pi}^T + 1)^2}\right) = o(T).$$

Therefore, we have $\lim_{T\to\infty} P_{\ge\varepsilon}^{T} = 0$. □

C.3 Proof of Proposition 3.6

Proof. Fix any regret estimator $\mathcal{A}$. Assume for contradiction that two-sided consistency holds. Sort the prices in $\mathcal{P}$ as $p_1 < p_2 < \cdots < p_k$.

Consider the following example. For all $1 \le t \le T$ we have $\pi^t(p_k) = 0$ and $\pi^t(p_{k-1}) \ne 0$. Pick an arbitrary positive constant $a \le 1$ and let $\boldsymbol{x}^T$ be such that for all $1 \le t \le T$, $x^t(p_i) = a$ and $x^t(p_k) = 0$ for $1 \le i < k$. Let $\boldsymbol{z}^T$ be another sequence of allocations such that

$$z^t(p) = x^t(p) \quad \text{for} \quad p = p_1, \ldots, p_{k-1} \quad \text{and} \quad z^t(p_k) = x^t(p_{k-1}) = a$$

for all $1 \le t \le T$.

By the assumption that two-sided consistency always holds, we have

$$\mathcal{A}(\{x^t(p^t), p^t, \pi^t\}_{t=1}^{T}) \xrightarrow{P} R^T(c, \boldsymbol{x}^T), \quad \mathcal{A}(\{z^t(p^t), p^t, \pi^t\}_{t=1}^{T}) \xrightarrow{P} R^T(c, \boldsymbol{z}^T).$$

We claim that by construction the random variables $\mathcal{A}(\{x^t(p^t), p^t, \pi^t\}_{t=1}^{T})$ and $\mathcal{A}(\{z^t(p^t), p^t, \pi^t\}_{t=1}^{T})$ are equal w.p. 1. Therefore $R^T(c, \boldsymbol{x}^T) = R^T(c, \boldsymbol{z}^T)$ holds. In fact,

$$\begin{aligned}\Pr_{p_t \sim \pi_t}\left[\{x^t(p^t), p^t, \pi^t\}_{t=1}^{T} \ne \{z^t(p^t), p^t, \pi^t\}_{t=1}^{T}\right] &\le \sum_{t=1}^{T} \Pr_{p^t \sim \pi^t}[z^t(p^t) \ne x^t(p^t)] \\ &= \sum_{t=1}^{T} \Pr[p^t = p_k] = \sum_{t=1}^{T} \pi^t(p_k) = 0.\end{aligned}$$

This means the transcripts are the same w.p. 1, so any deterministic algorithm's outputs are the same w.p. 1.

Next we aim at showing that $R^T(c, \boldsymbol{z}^T) > R^T(c, \boldsymbol{x}^T)$, so we get a contradiction. Since $\pi^t(p_k) = 0$ and $x^t(p) = z^t(p)$ for all $p \neq p_k$ and $1 \leq t \leq T$ we have

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)(p-c)x^t(p) = \frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)(p-c)z^t(p).$$

So it suffices to show that

$$\max_{\sigma}\frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)(\sigma(p)-c)z^t(\sigma(p)) > \max_{\sigma}\frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)(\sigma(p)-c)x^t(\sigma(p)). \quad (1)$$

Note that the optimizer of the LHS is $\tau(p) = p_k$ for all $p \in \mathcal{P}$ and the optimizer of the RHS is $\rho(p) = p_{k-1}$. Equation (1) follows since

$$\sum_{t=1}^{T}\sum_{p}\pi^t(p)(p_k-c)z^t(p_k) - \sum_{t=1}^{T}\sum_{p}\pi^t(p)(p_{k-1}-c)x^t(p_{k-1}) = \sum_{t=1}^{T}\sum_{p}a\pi^t(p)(p_k-p_{k-1}) > 0.$$

This completes the proof. □

Remark. We note that this example further implies that there exists an algorithm with vanishing regret, but from its transcript, there is no two-sided consistent estimator for its regret. In fact, consider the following simple scenario: Let the opponent always play price p_k and the ground truth allocation be $\boldsymbol{x}^T$. Assume the seller being audited is best responding to her opponent. We cannot consistently (one-sided) estimate her regret according to the proof.

C.4 Proof of Proposition 3.9

We first observe the following lemma.

Lemma C.1. *Fix the sequence of price distributions $\boldsymbol{\pi}^T$, for any $\boldsymbol{x}^T \sim_{\boldsymbol{\pi}^T} \boldsymbol{z}^T$ and deterministic regret estimator $\mathcal{A}$*

$$\Pr_{p^t\sim\pi^t}\left[\mathcal{A}(\{x^t(p^t), p^t, \pi^t\}_{t=1}^T) = \mathcal{A}(\{z^t(p^t), p^t, \pi^t\}_{t=1}^T)\right] = 1.$$

Proof. Similar to the proof of Proposition 3.6

$$\begin{aligned}\Pr_{\{p^t,\pi^t\}_{t=1}^T\sim\mathcal{M}}&[\{x^t(p^t), p^t, \pi^t\}_{t=1}^T \neq \{z^t(p^t), p^t, \pi^t\}_{t=1}^T] \\ &\leq \sum_{t=1}^{T}\Pr_{p^t\sim\pi^t}[z^t(p^t) \neq x^t(p^t)] \\ &= \sum_{t=1}^{T}\Pr[p^t \notin C^t] = 0. \qquad \text{(because } \boldsymbol{x}^T \sim_{\boldsymbol{\pi}^T} \boldsymbol{z}^T\text{)}\end{aligned}$$

So the transcripts are the same conditioned on $\boldsymbol{\pi}^T$ and the result follows. □

Proof of Proposition 3.9. Let $\boldsymbol{x}_*^T$ be a sequence of allocations that achieves $\overline{R}^T(c, \boldsymbol{x}^T)$. By the one-sided consistency requirement

$$\lim_{T\to\infty}\Pr_{p^t\sim\pi^t}[\mathcal{A}(\{x_*^t(p^t), p^t, \pi^t\}_{t=1}^T) < \overline{R}^T(c, \boldsymbol{x}^T) - \varepsilon] = 0.$$

But Lemma C.1 implies that $\mathcal{A}(\{x^t(p^t), p^t, \pi^t\}_{t=1}^T) = A(\{x_*^t(p^t), p^t, \pi^t\}_{t=1}^T)$ w.p. 1, and the proposition follows. □

C.5 Proof of Theorem 4.1

To prove the theorem, we present a few useful lemmas. We first characterize the location of the worst-case allocations. We then show that the algorithm can consistently estimate the worst-case regret.

Lemma C.2. *Fix any $\boldsymbol{\pi}^T$. Let $[\boldsymbol{x}^T]$ be an equivalence class under the relation $\sim_{\boldsymbol{\pi}^T}$. Consider the following construction: Pick any $\boldsymbol{x}^T \in [\boldsymbol{x}^T]$ and set*

$$z_*^t(p^t) = \begin{cases} x^t(p^t) & (\textit{if } p^t \in C^t), \\ x^t(p) \textit{ where } p = \min\{q \le p^t : q \in C^t\} & (\textit{otherwise}). \end{cases}$$

Then $\boldsymbol{z}_*^T$ is well-defined, $\boldsymbol{z}_*^T \in [\boldsymbol{x}^T]$, and $R^T(c, \boldsymbol{z}_*^T) = \sup_{\boldsymbol{z}^T \sim_{\boldsymbol{\pi}^T} \boldsymbol{x}^T} R^T(c, \boldsymbol{z}^T)$.

Proof. Since any $\boldsymbol{x}^T$ in the equivalence class agrees on the prices that are in C^t for all $1 \le t \le T$, and we also set $\boldsymbol{z}_*^T$'s allocation there the same, we have that $\boldsymbol{z}_*^T$ is well-defined and $\boldsymbol{z}_*^T \in [\boldsymbol{x}^T]$. To see that $\boldsymbol{z}_*^T$ achieves the supremum, note that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)\left[(\sigma(p)-c)z_*^t(\sigma(p)) - (p-c)z_*^t(p)\right] \ge \frac{1}{T}\sum_{t=1}^{T}\sum_{p}\pi^t(p)\left[(\sigma(p)-c)x^t(\sigma(p)) - (p-c)x^t(p)\right] \tag{2}$$

for any $\boldsymbol{x}^T \in [\boldsymbol{x}^T]$ and any mapping σ. In fact, since $\boldsymbol{z}_*^T \sim_{\boldsymbol{\pi}^T} \boldsymbol{x}^T$ we have

$$\sum_{t=1}^{T}\sum_{p}\pi^t(p)(p-c)z_*^t(p)] = \sum_{t=1}^{T}\sum_{p}\pi^t(p)(p-c)x^t(p),$$

and

$$\sum_{t=1}^{T}\sum_{p}\pi^t(p)(\sigma(p)-c)z_*^t(\sigma(p)) \ge \sum_{t=1}^{T}\sum_{p}\pi^t(p)(\sigma(p)-c)x^t(\sigma(p))$$

because $z_*^t(p) \ge x^t(p)$ for every $p \in \mathcal{P}$ and $1 \le t \le T$, by construction of $\boldsymbol{z}^T$.

The lemma now follows from Equation (2) and the fact that if $f(\sigma) \ge g(\sigma)$ everywhere, then $\max_\sigma f(\sigma) \ge \max_\sigma g(\sigma)$. □

Lemma C.3. *Given a sequence of allocations $\boldsymbol{x}^T$. Let $k = |\mathcal{P}|$ be the number of price levels and $\overline{p}$ be the highest price. Given cost c, conditional on observing the sequence of price distributions $\boldsymbol{\pi}^T$, for any fixed sequence of allocations $\boldsymbol{\pi}^T$,*

$$\Pr[|\widetilde{R}_{p,q}^T(c, \boldsymbol{x}^T) - \overline{R}_{p,q}^T(c, \boldsymbol{x}^T)| \ge \varepsilon] \le 2\exp\left(-\frac{\varepsilon^2}{2k^2\sum_{t=1}^T d_t^2}\right)$$

where

$$d^t = \frac{1}{T}\left(\frac{1}{\min_{p' \in C^t}\pi^t(p')} + 1\right)\overline{p}.$$

Now we state the proof of Theorem 4.1.

Proof. 1. Starting from Lemma C.3, we claim that for any fixed c

$$\Pr[\widetilde{R}^T(c, \boldsymbol{x}^T) - \overline{R}^T(c, \boldsymbol{x}^T) \ge \varepsilon] \le 2k^2\exp\left(-\frac{\varepsilon^2}{2k^2\sum_{t=1}^T d_t^2}\right).$$

To bound the probability of

$$\Pr\left[\widetilde{R}^T(c, \boldsymbol{x}^T) - \overline{R}^T(c, \boldsymbol{x}^T) \ge \varepsilon\right] = \Pr\left[\sum_p \max_q R_{p,q}^T(c, \boldsymbol{x}^T) - \sum_p \max_q \overline{R}_{p,q}^T(c, \boldsymbol{x}^T) \ge \varepsilon\right],$$

note that

$$\begin{aligned}
&\Pr\left[\sum_p \max_q \widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) - \sum_p \max_q \overline{R}^T_{p,q}(c, \boldsymbol{x}^T) \geq \varepsilon\right] \\
\leq &\Pr\left[\exists p \in \mathcal{P}, \exists p' \in \mathcal{P}, \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T) \geq \frac{\varepsilon}{|\mathcal{P}|}\right] \\
\leq &k^2 \exp\left(-\frac{\varepsilon^2}{2k^2 \sum_{t=1}^T d_t^2}\right) \qquad \text{(union bound).}
\end{aligned}$$

Plug in the lower bound of T, $\varepsilon = \delta^T$ and $c = c_0$, and δ^T, we have $\Pr[\widetilde{R}^T(c_0, \boldsymbol{x}^T) - \overline{R}^T(c_0, \boldsymbol{x}^T) \geq \delta^T] \leq \alpha$. By definition of $\tilde{c}$ and c_0, we have $\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) \geq \widetilde{R}^T(c_0, \boldsymbol{x}^T)$. When the seller satisfies $\min_{p \in C^t, 1 \leq t \leq T} \pi^t(p) \geq \underline{\pi}$, we have $\delta^T \leq r/2$. Therefore, when $\overline{R}^T(c_0, \boldsymbol{x}^T) \leq r$, we have

$$\begin{aligned}
&\Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) + \delta^T \geq 2r] \\
\leq &\Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) + \delta^T \geq r + 2\delta^T] \\
\leq &\Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) + \delta^T \geq \overline{R}^T(c_0, \boldsymbol{x}^T) + 2\delta^T] \\
\leq &\Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(c_0, \boldsymbol{x}^T) \geq \delta^T] \\
\leq &\alpha
\end{aligned}$$

and the seller passes with probability at least $1 - \alpha$.

2. Note that since $\tilde{c}$ is a random variable, we can not use the same argument for fixed c_0 to bound the probability $\Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(c_*, \boldsymbol{x}^T) \leq -r]$ by plugging $c = \tilde{c}$. Instead, observe that

$$\begin{aligned}
&\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(\tilde{c}, \boldsymbol{x}^T) \leq -r\right] \\
\leq &\Pr\left[\exists c, \widetilde{R}^T(c, \boldsymbol{x}^T) - \overline{R}^T(c, \boldsymbol{x}^T) \leq -r\right] \\
= &\Pr\left[\exists c, \sum_p \max_q \widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) - \sum_p \max_q \overline{R}^T_{p,q}(c, \boldsymbol{x}^T) \leq -r\right] \\
= &\Pr\left[\exists c, \sum_p \max_q \overline{R}^T_{p,q}(c, \boldsymbol{x}^T) - \sum_p \max_q \widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) \geq r\right] \\
\leq &\Pr\left[\exists c, \exists p \in \mathcal{P}, \exists p' \in \mathcal{P}, \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) \geq \frac{r}{|\mathcal{P}|}\right] \\
= &\Pr\left[\exists c, \exists p \in \mathcal{P}, \exists p' \in \mathcal{P}, \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right].
\end{aligned}$$

Taking union bound over $p \in \mathcal{P}$ and $q' \in \mathcal{P}$, we have

$$\begin{aligned}
&\Pr\left[\exists c, \exists p \in \mathcal{P}, \exists q' \in \mathcal{P}, \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right] \\
\leq &\sum_{p \in \mathcal{P}} \sum_{p' \in \mathcal{P}} \Pr\left[\exists c, \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,q'}(c, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right].
\end{aligned}$$

Observe that $\widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T)$ is linear in c hence, when $c \in [\underline{c}, \bar{c}]$,

$$\Pr\left[\exists c, \widetilde{R}^T_{p,p'}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right]$$

$$\leq \Pr\left[\widetilde{R}^T_{p,q'}(\underline{c}, \boldsymbol{x}^T) - \overline{R}^T_{p,q'}(\underline{c}, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|} \cup \widetilde{R}^T_{p,q'}(\bar{c}, \boldsymbol{x}^T) - \overline{R}^T_{p,q'}(\bar{c}, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right]$$
$$\leq \Pr\left[\widetilde{R}^T_{p,q'}(\underline{c}, \boldsymbol{x}^T) - \overline{R}^T_{p,q'}(\underline{c}, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right] + \Pr\left[\widetilde{R}^T_{p,q'}(\bar{c}, \boldsymbol{x}^T) - \overline{R}^T_{p,q'}(\bar{c}, \boldsymbol{x}^T) \leq -\frac{r}{|\mathcal{P}|}\right].$$

Combining Lemma C.3, we get

$$\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(\tilde{c}, \boldsymbol{x}^T) \leq -r\right] \leq 2k^2 \exp\left(-\frac{\varepsilon^2}{2k^2 \sum_{t=1}^T d_t^2}\right).$$

Plug in the bound of T and $\varepsilon = \delta^T$, we have $\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(\tilde{c}, \boldsymbol{x}^T) \leq -\delta^T\right] \leq \alpha$. Therefore when $\overline{R}^T(c_*, \boldsymbol{x}^T) \geq 2r$, by definition of c_* and $\tilde{c}$, we have

$$\begin{aligned}
&\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) + \delta^T \leq 2r\right] \\
\leq &\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - 2r \leq -\delta^T\right] \\
\leq &\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(c_*, \boldsymbol{x}^T) \leq -\delta^T\right] \\
\leq &\alpha.
\end{aligned}$$

Hence the seller fails with probability at least $1 - \alpha$. □

C.5.1 Proof of Lemma C.3

Proof. Let

$$\widetilde{r}^t_{p,q} = \frac{1}{T}\left(\pi^t(p)\left[(q-c)\hat{h}^t(q) - (p-c)\hat{h}^t(p)\right]\right), \overline{r}^t_{p,q} = \frac{1}{T}\left(\pi^t(p)\left[(q-c)x^t_*(q) - (p-c)x^t_*(p)\right]\right),$$

we claim that $\mathbb{E}_{p^t \sim \pi^t}\left[\widetilde{r}^t_{p,q}\right] = \overline{r}^t_{p,q}$. In fact, by definition of $\hat{h}^t$, we have

$$\hat{h}^t(p) = \hat{x}^t(p') \text{ where } p' = \max\{r \leq p : r \in C^t\}. \tag{3}$$

We have $\hat{h}^t(p') = \hat{x}^t(p')$, which implies that $\mathbb{E}_{p^t \sim \pi^t}\left[\hat{h}^t(p')\right] = x^t(p')$. By definition of x^t_*, we have $x^t(p') = x^t_*(p)$. Apply the same reasoning we also have $\mathbb{E}_{p^t \sim \pi^t}\left[\hat{h}^t(q)\right] = x^t_*(q)$. Hence, by linearity of expectation we get $\mathbb{E}_{p^t \sim \pi^t}\left[\widetilde{r}^t_{p,q}\right] = \overline{r}^t_{p,q}$. Since $p \leq \overline{p}, q \leq p\max$, and $\hat{x}^t(p') \leq 1/\pi^t(p')$ for all $p' \in C^t$, we also have that

$$|\widetilde{r}^t_{p,q} - r^t_{p,q}| \leq \frac{1}{T}\left(\frac{1}{\min_{p' \in C^t} \pi^t(p')} + 1\right)\overline{p}.$$

Note that

$$\widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,q}(c) = \sum_{t=1}^T (\widetilde{r}^t_{p,q} - \overline{r}^t_{p,q}).$$

Applying Azuma's inequality, we get

$$\Pr[\widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,q}(c, \boldsymbol{x}^T) \geq \varepsilon] \leq k^2 \exp\left(-\frac{\varepsilon^2}{2k^2 \sum_{t=1}^T d_t^2}\right).$$

By a similar argument on the other side we also have

$$\Pr[\widetilde{R}^T_{p,q}(c, \boldsymbol{x}^T) - \overline{R}^T_{p,q}(c, \boldsymbol{x}^T) \leq -\varepsilon] \leq k^2 \exp\left(-\frac{\varepsilon^2}{2k^2 \sum_{t=1}^T d_t^2}\right).$$

We get the desired result by combining two sides of the inequality. □

v_1/v_2	0	1	2	3
0	0	0	0	$\frac{67}{600}+\frac{1}{3}\varepsilon$
1	0	0	0	$\frac{1}{30}-\frac{4}{3}\varepsilon$
2	0	0	0	$\frac{1}{100}+\varepsilon$
3	$\frac{1}{40}$	$\frac{9}{25}$	0	$\frac{23}{50}$

Table 2: No-best-in-hindsight-regret playing is not enough: value distribution

C.6 Proof of Theorem 5.2

Proof. Fix $\gamma = o(1)$. Let $\varepsilon = \sqrt{\gamma}$. To prove the theorem we first provide the construction.

Example C.4. There are $(1 + 1.1)T$ rounds of interaction. The buyer's valuation (v_1^t, v_2^t) is supported on $\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$ and the two sellers can post any price $p_i^t \in \mathcal{P} = \{0, 1, 2, 3\}$. Both sellers have cost $c_1 = c_2 = 0$. The joint distribution of (v_1^t, v_2^t) is shown in Table 2 and is i.i.d. across rounds. We also assume that the buyer break ties randomly and he chooses to buy if buying gets utility 0.

We first note that with such a valuation, the buyer never chooses to buy nothing because either $v_1^t = 3$ or $v_2^t = 3$ with probability 1. It follows that

Claim C.5. *Given prices (p_1, p_2), the buyer buys good 1 if and only if $v_1 - v_2 > p_1 - p_2$, buys good 2 if and only if $v_1 - v_2 < p_1 - p_2$, and chooses randomly between seller 1 and 2 if $v_1 - v_2 = p_1 - p_2$.*

Proof. If the buyer buys good 1 then $v_1 - v_2 > p_1 - p_2$ is necessary. If $v_1 - v_2 > p_1 - p_2$ but he does not buy good 1, this means $0 > v_1 - p_1 > v_2 - p_2$ so he buys nothing. But by construction the buyer never does this. □

The claim enables us to write the demand function only using the distribution of $v_1 - v_2$:

$$x_1(p_1, p_2) = \Pr[v_1 - v_2 > p_1 - p_2] + \frac{1}{2}\Pr[v_1 = v_2], \quad x_2(p_1, p_2) = 1 - x_1(p_1, p_2). \tag{4}$$

Using Equation (4) we construct the ex-ante payoff matrix in Table 3 (note that playing price 0 is a dominated action so we omit it here). The highest payoff correlated equilibrium of this game is a

p_1/p_2	1	2	3
1	$0.615, 0.385$	$0.85+\frac{\varepsilon}{2}, 0.3-\varepsilon$	$\frac{523}{600}+\frac{\varepsilon}{3}, 0.385-\varepsilon$
2	$0.77, 0.615$	$1.23, 0.77$	$1.7+\varepsilon, 0.45-\frac{3\varepsilon}{2}$
3	$0.615, 0.795$	$1.155, 1.23$	$1.845, 1.155$

Table 3: No-best-in-hindsight-regret playing is not enough: payoff matrix

pure NE where they play $(p_1, p_2) = (2, 2)$. The equilibrium payoff is $(1 + 1.1)T \cdot (1.23, 0.77) = (2.583T, 1.617T)$.

We claim there exists a manipulation such that the sellers play $(p_1, p_2) = (1, 1)$ in each round with high probability for $T - o(T)$ rounds, and then switch to collude by playing $(p_1, p_2) = (3, 3)$ in each round with high probability for $1.1T - o(T)$ rounds .

We first assume the claim is true. Then:

1. Since $1.845 > 1.23, 1.155 > 0.77$, for $\Omega(T)$ rounds, both play $p_1 > p_1^e, p_2 > p_2^e$ in each round with high probability. This shows the first point of the theorem.
2. By linearity of expectation, the total expected payoff is now $(T - o(T)) \cdot (0.615 + 1.1 \times 1.845, 0.385 + 1.1 \times 1.155) = (2.6445T - o(T), 1.6555T - o(T)) > (2.583T, 1.617T)$, which is higher than the equilibrium payoff by $\Omega(T)$. This shows the second point of the theorem.

3. Consider seller 1's best fixed action in hindsight:

   (a) If she plays price 1, the expected payoff is $(T-o(T))\cdot 0.615+(1.1T-o(T))\cdot(\frac{523}{600}+\frac{\varepsilon}{3}) < 1.574T-o(T)$.
   (b) If she plays price 2, the expected payoff is $(T-o(T))\cdot 0.77+(1.1T-o(T))\cdot(1.7+\varepsilon) = 2.64T-o(T)$.
   (c) If she plays price 3, the expected payoff is $(T-o(T))\cdot 0.615+(1.1T-o(T))\cdot 1.845 = 2.6445T-o(T)$.

   The best-in-hindsight price is 3 with expected payoff $2.6445T-o(T)$. But we just showed seller 1's total expected payoff in the manipulation is also $2.6445T-o(T)$, thus she has vanishing regret $o(T)$.

   Consider seller 2's best fixed action in hindsight:

   (a) If she plays price 1, the expected payoff is $(T-o(T))\cdot 0.385+(1.1T-o(T))\cdot 0.795 < 1.26T-o(T)$.
   (b) If she plays price 2, the expected payoff is $(T-o(T))\cdot(0.3-\varepsilon)+(1.1T-o(T))\cdot 1.23 = 1.653T-o(T)$.
   (c) If she plays price 3, the expected payoff is $(T-o(T))\cdot(0.385-\varepsilon)+(1.1T-o(T))\cdot 1.155 = 1.6555T-o(T)$.

   The best-in-hindsight price is 3 with expected payoff $1.6555T-o(T)$. But we just showed seller 2's total expected payoff in the manipulation is also $1.6555T-o(T)$, thus she has vanishing regret $o(T)$.

Note that seller 1 has non-vanishing calibrated regret because the best response to price 1 is price 2, and seller 2 has non-vanishing calibrated regret because the best response to price 3 is price 2.

Next we show the claim: How to manipulate a γ-mean-based seller 2 into a collusion under this setting.

Seller 1 manipulates as follows: He first plays $p_1 = 1$ for T rounds, and then switch to playing $p_1 = 3$ for the remaining $1.1T$ rounds.

The following claim follows the definition of γ-mean-based strategy.

Claim C.6. *With a γ-mean-based algorithm:*

1. *For each $T \geq t \geq O(\varepsilon T)$, seller 2 posts $p_2 = 1$ in round t w.p. at least $1-\gamma$.*
2. *For each $T + O(\varepsilon T) \leq t \leq T + 1.1T$, seller 2 posts $p_2 = 3$ in round t w.p. at least $1-\gamma$.*

Proof. We write the cumulative reward of playing prices 1, 2, and 3 as follows:

$$r_1(t) = \begin{cases} 0.385t & t \leq T \\ 0.385T + 0.795(t-T) & t \geq T \end{cases},$$

$$r_2(t) = \begin{cases} (0.3-\varepsilon)t & t \leq T \\ (0.3-\varepsilon)T + 1.23(t-T) & t \geq T \end{cases},$$

$$r_3(t) = \begin{cases} (0.385-\varepsilon)t & t \leq T \\ (0.385-\varepsilon)T + 1.155(t-T) & t \geq T \end{cases}.$$

It follows that between $t = \varepsilon(1+1.1)T$ and $t = T$ we have $r_1(t) \geq \max(r_2(t), r_3(t)) + \gamma T$. Just after $t = T$ price 1 has an advantage of εT but price 3 quickly comes as the best choice after $T+\varepsilon T/(1.155-0.795) \leq T+3\varepsilon T$. Price 3 still dominates until $t = 0.085T/(1.23-1.155) < 1.1T$ (this is when price 2 becomes the best). Therefore, for each $T \geq t \geq O(\sqrt{\gamma}T)$ seller 2 posts $p_2 = 1$ w.p. at least $1-\gamma$, and for each $T + O(\varepsilon T) \leq t \leq 1.1T + T$ seller 2 posts $p_2 = 3$ w.p. at least $1-\gamma$. □

Remark. Although the above argument suffices for our purposes, we remark that the claim still holds even if the sellers only gets a realization of the buyer's decision each round (instead of getting

the expected reward of her strategy) by using a concentration argument and the length of the time window where the sellers are collusive only suffer an $o(T)$ loss.

This ends the proof. □

Remark. In the example provided above, a calibrated no-regret algorithm will not be manipulated into collusion.

To see this, first note the Stackelberg outcome of the game is the same as the highest-payoff correlated equilibrium (CE). By [10] the manipulator cannot get more than the Stackelberg payoff, therefore she cannot get more than the CE payoff in our example. But the outcome that both sellers post prices higher than the CE prices have a higher payoff than the CE for the manipulator. Therefore this cannot happen.

D Aggregating Prices to Approximate Distributions

Recall from Section 1.1.1 the motivations of designing the refined auditing method that we describe in the introduction. One of the motivations is to allow testing aggregated empirical distributions when price distribution data is not available. We provide a formal statement of why this is feasible for our auditing method.

Proposition D.1. *Consider T rounds and k price levels bounded in $[0,1]$. In round i the seller posts price $p_i \sim \pi_i$. Suppose $\|\pi_i - \pi_{i+1}\|_\infty \le \varepsilon$ for all $1 \le i \le T-1$.*[8] *Then there exists an algorithm that uses only price samples $p_1, \dots, p_T$ and outputs estimated price distributions $\tilde{\pi}_1, \dots, \tilde{\pi}_T$ such that, with probability at least $1-\delta$*

$$\|\tilde{\pi}_i - \pi_i\|_\infty \le \sqrt[3]{4\varepsilon \log \frac{2Tk}{\delta}},$$

for all $1 \le i \le T$.

Hence when the price distributions are not available and the rate of change in price distributions are low, it is possible to use the aggregated price distributions as the input to the auditing method and the following steps are the same.

Proof. Let the price set be $\mathcal{P}$ where $|\mathcal{P}| = k$. We use the aggregation method as stated in the introduction. Let the window length be L. For each π_i, we use the empirical distribution of prices in the window centered at round i to approximate π_i. Let the price distributions in the window be $F_1, \dots, F_L$ and $p_1 \sim F_1, \dots, p_L \sim F_L$. Then for any fixed $p \in \mathcal{P}$ and j we have $\mathbb{E}[\mathbf{1}_{\{p_j \le p\}}] = F_j(p)$. Azuma–Hoeffding now gives

$$\Pr\left[\left|\frac{1}{L}\sum_{j=1}^{L}(\mathbf{1}_{\{p_j \le p\}} - F_j(p))\right| \ge t\right] \le 2\exp(-2Lt^2).$$

With a union bound over all price levels, with probability at least $1-2k\exp(-2Lt^2)$ the aggregated empirical distribution $\tilde{F}$ satisfies $\|\tilde{F} - \frac{1}{L}\sum_{j=1}^{L} F_j\|_\infty \le t$. By the assumption of the lemma and the triangle inequality we have $\|F_i - \frac{1}{L}\sum_{j=1}^{L} F_j\|_\infty \le L\varepsilon$ for all i. This implies $\|\tilde{F} - F_i\| \le t + L\varepsilon$ w.p. at least $1-2k\exp(-2Lt^2)$. With a union bound on all T rounds, w.p. at least $1-2kT\exp(-2Lt^2)$, the aggregated estimator is $(t+L\varepsilon)$-close to the true distribution in the ℓ_∞ distance in every round. Setting $\delta = 2kT\exp(-2Lt^2)$ gives the error bound

$$\frac{\log \frac{2Tk}{\delta}}{2t^2}\epsilon + t.$$

The lemma follows by setting $\frac{\log \frac{2Tk}{\delta}}{2t^2}\epsilon = t$ and solving for the optimal t. □

[8] For example, the multiplicative weights update (MWU) algorithm with learning rate ε satisfies the condition. See the ending remark of the proof of this lemma for details. Designing a test for such a condition is left as an open question.

Remark. We provide an example application of the lemma. We show that the multiplicative weights update (MWU) algorithm with reward in $[0,1]$ satisfies the condition. Let $V_1,\dots,V_k$ be the current cumulative reward for actions $1,\dots,k$. Now, the maximum change in the probability of playing action 1 occurs when the reward of action 1 is 1 and those for the other actions are 0. We bound the change as follows:

$$\begin{aligned}
&\frac{(1+\varepsilon)^{V_1+1}}{(1+\varepsilon)^{V_1+1}+(1+\varepsilon)^{V_2}+\cdots+(1+\varepsilon)^{V_k}} - \frac{(1+\varepsilon)^{V_1}}{(1+\varepsilon)^{V_1}+(1+\varepsilon)^{V_2}+\cdots+(1+\varepsilon)^{V_k}} \\
\le &\frac{(1+\varepsilon)^{V_1+1}}{(1+\varepsilon)^{V_1+1}+(1+\varepsilon)^{V_2}+\cdots+(1+\varepsilon)^{V_k}} - \frac{(1+\varepsilon)^{V_1}}{(1+\varepsilon)^{V_1+1}+(1+\varepsilon)^{V_2}+\cdots+(1+\varepsilon)^{V_k}} \\
= &\frac{(1+\varepsilon)^{V_1+1}-(1+\varepsilon)^{V_1}}{(1+\varepsilon)^{V_1+1}+(1+\varepsilon)^{V_2}+\cdots+(1+\varepsilon)^{V_k}} \le \frac{(1+\varepsilon)^{V_1+1}-(1+\varepsilon)^{V_1}}{(1+\varepsilon)^{V_1+1}+k-1} \\
= &\frac{\varepsilon}{1+\varepsilon+\frac{k-1}{(1+\varepsilon)^{V_1}}} \le \frac{\varepsilon}{\varepsilon+1} \le \varepsilon.
\end{aligned}$$

Therefore, the lemma can be used with MWU sellers (this includes a lot of MWU-style algorithms, such as EXP3).

Algorithm 1: Auditing (non)-collusion via testing the worst-case regret

Data: $\{p^t\}_{t=1}^T, \{x^t(p^t)\}_{t=1}^T, \{(\pi^t)\}_{t=1}^T$

for $t \in [T]$ **do**
  $C^t \leftarrow \{p \in \mathcal{P} : \pi^t(p_j) > 0\}$;
  for $p \in C^t$ **do**
    $\hat{x}^t(p) \leftarrow \begin{cases} x^t(p)/\pi^t(p), & p = p^t, \\ 0, & \text{otherwise} \end{cases}$;
  end
  for $p \in \mathcal{P}$ **do**
    $\hat{h}^t(p) \leftarrow \hat{x}^t(p')$ where $p' = \max\{r \le p : r \in C^t\}$;
  end
end

Solve the following programming and let the solution be $\tilde{c}$, defined as *estimated plausible cost*:

$$\min_{c \in [\underline{c}, \bar{c}]} \quad \widetilde{R}^T(c)$$

where

$$\widetilde{R}^T(c) = \sum_p \max_q \widetilde{R}^T_{p,q}(c),$$

$$\widetilde{R}^T_{p,q}(c) = \frac{1}{T}\sum_{t=1}^T \pi^t(p)\left[(q-c)\hat{h}^t(q) - (p-c)\hat{h}^t(p)\right].$$

Let

$$\delta^T = \frac{k\overline{p}}{T}\sqrt{2\log\left(\frac{2k^2}{\alpha}\right)\cdot\sum_{s=1}^T\left(\frac{1}{\min_{p\in C^s}\pi^s(p)}+1\right)^2}.$$

if $\widetilde{R}^T(\tilde{c}) + \delta^T \le 2r$ **then**
  Output PASS;
else
  Output FAIL;
end