Regulation of Algorithmic Collusion, Refined: Testing Pessimistic Calibrated Regret

JASON D. HARTLINE, CHANG WANG, and CHENHAO ZHANG, Northwestern University, USA

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. [23].

We develop an auditing method that tests whether a seller's pessimistic calibrated regret is low. The pessimistic calibrated regret is the highest calibrated regret of outcomes compatible with the observed data. 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 any 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. This manipulation cannot be excluded with a non-collusion definition of vanishing best-in-hindsight regret.

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.

CCS Concepts: • **Applied computing** → **Law**; • **Theory of computation** → **Online learning algorithms**; **Algorithmic game theory and mechanism design**.

Additional Key Words and Phrases: Algorithmic collusion, Antitrust, Regulation of algorithms, Algorithmic pricing

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 [12], 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 (a.k.a. supra-competitive prices). There is one important reason behind the strong emphasis of communication in this argument: In many jurisdictions such as the US, overt communication is a prerequisite to establishing legal liability of collusion. Collusion with the absence of communication is actually not illegal according to the current legal theory [21, 22].

On the other hand, researchers recently [3, 4, 7] 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.

There are two doctrines of antitrust analysis in the US legal framework: per se rule and rule of reason [24]. The per se rule deems certain business practices, like price-fixing or market division, illegal without requiring further

Authors' address: Jason D. Hartline, hartline@northwestern.edu; Chang Wang, wc@u.northwestern.edu; Chenhao Zhang, chenhao.zhang.rea@u.northwestern.edu, Northwestern University, Evanston, IL, USA.

investigation into their actual competitive effects. For example, in response to the algorithmic collusion concern around RealPage, San Francisco has recently enacted the per se regulation that precludes any use of algorithms in rental pricing [10]. 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. See Section 1.2 for a more comprehensive categorization of currently proposed regulations of algorithmic collusion.

This paper considers the regulation of algorithmic collusion by auditing pricing data as proposed by Hartline et al. [23]. In their proposal, they argue that it is feasible to require all sellers to deploy pricing algorithms to pass the audit. This means that it would not harm the market to deem those who do not pass as illegal. The regulation to require algorithms to pass such an audit falls within the per se doctrine.

They then provide details of the 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.

The non-collusion audit they propose is based on ideas and observations from the theory of online learning. At its core, they propose using *calibrated regret* as a quantitative measure for how far from non-collusion a seller running pricing algorithms is. 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 further utilizing the information revealed from her current pricing decisions. Low calibrated regret indicates that the seller is already 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 from the collected data, they develop a statistically efficient test of 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. [23] in light of the standard guideline of binary classification: minimizing the number of false positives and false negatives. This classification is also suggested by Harrington [21, VI. A] when designing a per se rule specifying a prohibited set of algorithms.

false positives A competitive algorithm that is unable to pass the audit is a false positive. We want to allow a range of competitive algorithms as broad as possible to pass the audit.

false negatives An algorithm that enables supra-competitive prices and passes the audit is a false negative. We want as many algorithms that enable supra-competitive prices to fail the audit as possible.

In the light of the above guideline, our main contribution is three-fold. First, we reduce false positives by providing an improved auditing method that drops the prerequisite that an algorithm must use fully-supported distribution to pass the audit, thus allowing a broader range of competitive algorithms to pass (Section 1.1.1). Second, Hartline et al. [23] show the need of vanishing calibrated regret as the (non-)collusion definition by providing an example under which testing best-in-hindsight regret passes supra-competitive algorithms when one seller can have side information related to the demand. The construction relies strongly on the usefulness of the side information. We show the stronger insufficiency of testing vanishing best-in-hindsight regret to exclude supra-competitive prices even when neither seller has side information. Thus, using vanishing calibrated regret as the non-collusion notion is important to reduce false negatives even without side information (Section 1.1.2).

Third, we argue that rule of reason can also be useful by demonstrating that there exist supra-competitive 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 1.1. 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. [23].[1]

Specifically, Hartline et al. [23] 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 which 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. In other words, in each round, every price level must be posted with at least some probability. We view this 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 her prefer to avoid certain prices. Second, the seller could deliberately avoid some prices for 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. [23], 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 distributions is small in that window (for example, when a learning algorithm with a 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 1.1. Lemma 4.3 and Corollary 4.4 present a formal statement of why auditing by 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 reliably 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 *pessimistic 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 that her *pessimistic calibrated regret* is low. The pessimistic regret is the highest regret of counterfactual outcomes compatible with the observed data.

In addition, we also clarify the interpretation of the non-collusion definition. Informally, the calibrated regret estimated both in Hartline et al. [23] and our work is only the regret of a particular run of the seller's algorithm. The estimated regret is not of the algorithm itself. Therefore, the non-collusion definition is consequentialist and not deontological.[2] We argue that this philosophy is appropriate not only from a legal perspective but also from a technical perspective. See Remark 2.4.

1.1.2 Fewer False Negatives. Second, we consider 1.1. The calibrated regret that Hartline et al. [23] propose as the measure of distance to non-collusion is a strong notion of regret. A common weaker notion is best-in-hindsight (a.k.a. external) 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

[1] Note: Hartline et al. [23] give a procedure for modifying these algorithms so that they can pass the audit.
[2] Consequentialism is the perspective of judging an act only by its outcome, while deontology is the perspective of judging an act by its inherent property.

hand, compares the performance of the chosen actions to the performance of the best *fixed* action in hindsight. They argue that the stronger notion of vanishing calibrated regret is indispensable to exclude certain supra-competitive prices by giving a simple example where one seller has side information about buyers' valuations. 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 non-vanishing calibrated regret in this case. The previous construction relies strongly on the usefulness of side information. Intuitively, regulation using calibrated regret is necessary in this case because the seller must best respond to such useful information. In this paper, we give a stronger argument for the need of calibrated regret for reducing false negatives. We show that even *without any side information*, a large family of pricing algorithms that minimize the weaker notion of best-in-hindsight regret can be *manipulated into playing supra-competitive prices*, and this is undetectable by a regulatory definition of non-collusion that uses best-in-hindsight regret. In other words, best-in-hindsight-regret algorithms can be susceptible to manipulation, thus, to prevent collusion and reduce false negatives, a regulator may want to discourage the use of algorithms that only satisfy this weaker property regardless of what side information can exist.

Before we describe the argument, we must be clear on what kind of manipulation we are concerned about. In the context of collusion auditing, we say a learning algorithm is *manipulated into playing supra-competitive prices*, if: a) the manipulator uses a strategy that leads the learning algorithm into an outcome where both of them have prices and payoffs exceeding correlated equilibria[3], *and* b) the manipulation cannot be excluded with the regulatory definition of non-collusion (e.g., no best-in-hindsight regret).

Note that we are only concerned about manipulative behavior that the regulatory non-collusion definition cannot exclude. The presence of such unexcluded manipulative behavior implies that certain algorithms can enable supra-competitive prices and evade the regulation, resulting in false negatives. On the other hand, manipulative behavior that have already been excluded by the definition do not result in false negatives.

Our technique is inspired by works such as Braverman et al. [6] and Deng et al. [11] showing the susceptibility to manipulation of best-in-hindsight regret minimization algorithms. The manipulator in the example of Deng et al. [11] has non-vanishing best-in-highlight regret, but in our construction, the manipulator has vanishing best-in-highlight regret. Our construction is a strengthening with respect to their construction in this sense.

More specifically, we construct an instance of the imperfect price competition with two sellers. In our example, the process that generates buyers' valuations is stationary and neither seller has any side information about the valuations of the buyer (or any other side information). 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. Our construction has two important features. First, *both* sellers play a constant fraction of rounds with supra-competitive prices in expectation, and *both* extract revenue higher than correlated equilibrium, meaning the process is non-negligible and benefits both parties. Second and more importantly, *both sellers have vanishing best-in-hindsight regret*, meaning such supra-competitive outcome would not be prevented by a per se regulation requiring vanishing best-in-hindsight regret.

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. [23], 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

[3]Recall that a *correlated equilibrium* is a joint distribution over players' actions, such that for every player i and every action a_i, playing a_i is a best response for i conditioned on seeing a_i, given that everyone else will play according to the distribution.

seller's regret is low, the seller is deemed competitive. This leads to the following question: Are there natural algorithms that when configured with the *true* cost c, find outcomes that are considered competitive by auditing methods for a higher cost c', while actually being supra-competitive? We provide an affirmative answer to the question by examining sellers using the Q-learning algorithm [5] that converge to supra-competitive 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 goes far beyond the seller's actual cost, the estimated regret becomes low.

A legal implication of this result is that it may not be sufficient to solely apply the per se rule of auditing to regulate algorithmic collusion, particularly when the regulator only has limited information about the seller's cost upfront. In such settings, the regulator should further investigates the reasonable cost of the seller in the market context, which is an application of rule of reason.

1.2 A Categorization of Proposed Regulations of Algorithmic Collusion

As we have mentioned previously, we interpret proposed regulations of algorithmic collusion and motivate ours under the legal framework of antitrust analysis: per se rule v.s. rule of reason. In this section, we provide a more comprehensive analysis of those proposed methods and categorize them into the two doctrines. The conclusion of this subsection is summarized in Table 1. More comparisons of technical aspects and other related work are included in Section 1.3.

Examples of Regulation	Classification
outright ban on algorithms (e.g., City and County of San Francisco [10])	per se
setting up formal non-collusive specifications and requiring proof of verification	per se
inspecting source code to judge the degree of (non-)competitiveness within context	rule of reason
requiring passing dynamic unit tests with specific protocols	per se
use context-specific input to dynamically test and analyze pricing behavior	rule of reason
requiring monitor wrapper [8]	per se
requiring passing data audit (Hartline et al. [23] and this paper)	per se
analyzing cost under context and audit (Hartline et al. [23] and this paper)	rule of reason

Table 1. Comparison of different proposed regulations of algorithmic collusion

The San Francisco ban on automated rent-setting [10] is a per se rule that considers all algorithmic pricing on the rental market illegal. Clearly, such regulation stifles innovation by excluding rent-setting algorithms that promote competition and improve market efficiency.

Harrington [21] discuss the approaches of *static checking* an algorithm's source code[4] without running it and *dynamic testing* the algorithm[5] with synthetical inputs to learn its properties related to price collusion and conclude whether it should be prohibited. However, they suggest that to what extent the prohibition comes from a per se rule or rule of reason depends on details: Prohibitions from clear collusion-identifying properties checkable with these approaches can be classified as per se rules. Otherwise, classification into rule of reason is more appropriate.

We list a few proposals using static checking or dynamic testing and classify them.

Static checking approaches range from having human experts manually review the source code to automatically generating machine-checkable proofs or counterexamples. One possible per se regulation of algorithmic collusion with

[4]Here "source code" refers to any formal representation of an algorithm, which also includes, for instance, the architecture and weights of neural network models.
[5]Here "algorithm" refers to the executable form of the algorithm that a seller is going to deploy.

static checking is to use the technique of static verification with formal logic: The regulator sets up formal *specifications* that exclude collusion-identifying properties and requires the sellers to submit the proofs that their algorithms are verified against the specifications. Algorithms without a proof of verification are prohibited. On the other hand, if the regulator inspects the source code of pricing algorithms to judge its (non)-competitiveness within a specific context, such an approach is in line with the rule of reason.

Dynamic testing an algorithm involves running the algorithms with certain inputs. One possible per se rule the regulator might use is to prepare a set of test cases and announce a uniform protocol for conducting tests of all algorithms of sellers. The protocol will specify how the test is conducted (e.g., by having the seller send a representative taking the algorithm to a brick-and-mortar site of the regulator agency or uploading the algorithm to a cloud service) and pass-fail criteria. The regulator prohibits algorithms that fail the test following the protocol. Alternatively, the regulator can analyze the pricing behavior of each seller's algorithm by feeding it with context-specific inputs. Such regulation would be aligned with the rule of reason.

Chassang and Ortner [8] propose another regulation based on the relation of regret and collusion observed in Chassang et al. [9]. The regulation approach they propose requires the seller to attach a monitor algorithm, known as a "pricing wrapper," to the pricing algorithm to ensure that the composition satisfies certain no-regret properties. This approach can also be interpreted as a per se rule because it prohibits any algorithm that cannot be obtained by the composition of an existing pricing algorithm and the pricing wrapper.

Finally, our proposal of auditing from data provides a clear prohibition determination without requiring access to the algorithms themselves while making minimal assumptions on the input space of the algorithms. While it is a per se rule, it can be augmented with the rule of reason analysis to narrow the range of what costs are reasonable.

1.3 Related Work

Algorithmic Collusion. Most papers studying algorithmic collusion from technical perspectives consider the Q-learning algorithm [29], a common reinforcement learning algorithm. Asker et al. [2, 3], Banchio and Mantegazza [4], Banchio and Skrzypacz [5], Calvano et al. [7], Klein [26] 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 [20] and large language models [17]. Our simulation in Section 6 follows the setup of Banchio and Skrzypacz [5]. These empirical and theoretical findings and concerns are among the main motivations of our work.

Legal Landscape of Anti-collusion Analysis. US statutes 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 *Brooke Group Ltd.* v. *Brown & Williamson Tobacco Corp.* [14], *In re Text Messaging Antitrust Litigation* [25] interpreting these statutes for price collusion have affirmed the jurisprudence of requiring express agreement as the prerequisite of establishing liability.

The Sherman Act literally prohibits acts that are "in restraint of trade and commerce" without clarifying how it should be applied [28]. In early cases such as *Addyston Pipe & Steel Co.* v. *United States* [13] and *Chicago Board of Trade* v. *United States* [15], the language of the statute 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* [16]. 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 [28] and Gavil [19] discuss the evolution of the two doctrines. Hovenkamp [24] discuss the scope where rule of reason analysis should be applied in the non-algorithmic antitrust settings. In light of the new challenges posed by algorithmic collusion, Harrington [21] propose adding per se prohibition for certain algorithms to competition laws. We motivate and interpret our work within the legal framework proposed by Harrington [21].

Regulation of Algorithmic Collusion. In addition to the auditing approach proposed in Hartline et al. [23], Harrington [21] and Chassang and Ortner [8] discuss alternative proposals of regulating algorithmic collusion. See Section 1.2 for a legal perspective for on this works.

The regulation of requiring non-regret monitor wrapper proposed by Chassang and Ortner [8] is based on the observations in Chassang et al. [9], which consider the problem of screening non-algorithmic collusion in procurement auctions. Similar to the auditing approach proposed in Hartline et al. [23], Chassang et al. [9] 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. [9] 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 et al. [9] 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. [6] 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. [11] consider the problem of manipulating no-best-in-hindsight-regret learner in general 2-player bimatrix games to get beyond the Stackelberg payoff[6]. Our example in Section 5 is inspired by their work but the construction is tailored for a dynamic price competition game.

Concurrently and independently, Arunachaleswaran et al. [1] provide an example (referred to as "their result" thereafter) where two sellers could be supra-competitive compared to correlated equilibria when *one* seller is using a vanishing calibrated regret algorithm. Their result is different from our result in Section 5 in the following two ways.

First, we take a different perspective on collusion and the results convey different messages. In Arunachaleswaran et al. [1], their concern is about the *cause* of algorithmic collusion, and they claim that their result shows algorithmic collusion can arise even if there is no "threat" between the sellers. On the other hand, our paper's question is about the *exclusion* of algorithmic collusion. We show that using no-best-in-hindsight regret as the non-collusion definition is not appropriate as it does not exclude some collusion even when side information is guaranteed not to exist. Although as is shown in Arunachaleswaran et al. [1], a seller running vanishing-calibrated-regret algorithm can still be manipulated

[6]Stackelberg payoff is the payoff in a *Stackelberg equilibrium*. Recall that in a Stackelberg equilibrium, a player (called leader) *commits* to a strategy such that when the other player (called follower) best responds to the strategy, the payoff of the leader is maximized.

to supra-competitive prices, this scenario will be prevented by our audit as the manipulating seller will not have non-vanishing calibrated regret and would be caught by the regulator.

Second, even in a pure technical sense, our results are different from theirs. In their construction, only one seller is guaranteed to have no regret (calibrated regret or best-in-hindsight regret), while in our example, both sellers have vanishing best-in-hindsight regret. In fact, it is impossible to guarantee that both sellers have no calibrated regret while they post supra-competitive prices. Our results reaffirm the importance for *both* sellers to have vanishing calibrated regret for competitive prices in a different way from them.

1.4 Limitations

Finally, we discuss a few caveats of our framework in this subsection.

First, if the seller uses pricing algorithms that do not optimize, she could also have high regret. Therefore, she would not pass our audit despite not being competitive. However, we argue that this is not a significant issue from the perspective of a per se rule. Sellers employ algorithmic pricing for the purpose of maximizing their profit and there are a variety of optimizing algorithms readily available. Therefore, using an algorithm that does not optimize is unlikely to serve any reasonable business purpose for a profit-seeking seller.

Second, as is discussed in Section 1.1.3, when the cost information of the seller possessed by the regulator is limited, the range of plausible costs can be large. In this situation, there exist algorithms with correctly configured true costs that exhibit supra-competitive behavior while still being able to pass the audit due to the existence of a plausible cost with low regret. Therefore, the imprecision of cost information might significantly impact the auditing result and presents challenges to regulating collusion from pricing data. 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.[7]

Third, the current theoretical analysis gives a large sample complexity bound on our auditing method. Therefore, it is more applicable for regulating algorithmic collusion in online market environments with high-frequency transactions and a large amount of data. An open question is to improve the sample complexity of our auditing method.

2 PRELIMINARIES

We consider a setting where n sellers repeatedly compete for selling a good in T rounds. Seller i has cost $c_i \in [\underline{c}, \bar{c}]$. 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 allocation as the feedback.[8] 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 strategy 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 resulting 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 $x_i^t(\cdot)$.

[7] Note: Third-party algorithms could be required to log the costs that they are configured with.
[8] She might also get other information, but we as auditors cannot directly observe other information.

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.

We will consider the non-collusion definition from Hartline et al. [23], which is a unilateral property. Therefore, from now on we focus on a single seller's behavior, and 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. [23] 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

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

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

Unless noted otherwise, we use "regret" to refer to "calibrated 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. [23] provide 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.

We conclude this section with two remarks on the settings.

The first remark says that the k-element price set can be chosen by the seller without affecting our results. And the second remark clarifies the randomness under the definition of calibrated regret in our framework.

Remark 2.3. The k-element price set can be *endogenous* to our model, meaning it can be chosen by the seller instead of being prescribed in advance. For example, the regulator first announces a continuum of possible prices $[0, h]$ and parameters of the audit, and the seller decides to discretize the set into k price levels according to her (learning) algorithm. Then the regulator can conduct an audit assuming all the prices in $[0, h]$ can be posted, and there will be an additional "discretization loss" in the regret estimation. See Footnote 10 for details of handling this in our auditing method.

Remark 2.4. We note a technical subtlety in Definition 2.2. Definition 2.2 is *not* the calibrated regret of the *algorithm* that generates the transcript. Consider the following example: A seller flips a coin to decide whether to run a collusive algorithm or a non-collusive algorithm. Then the quantity in Definition 2.2 is either the calibrated regret of the collusive algorithm, or that of the non-collusive algorithm, depending on the outcome of the coin flip. The practical implication of this subtlety is contained in the following question: Suppose the auditor has evidence that the seller's algorithm is doing such a coin flip, but the outcome is that she happens to run the non-collusive branch, does he flag the seller? Definition 2.2 implies that our answer is no, hence our definition of non-collusion is *consequentialist* and not *deontological*. In fact, we can show that in the realm of algorithmic collusion, we are only *able* to use a consequentialist definition, because a deontological collusion detection using only observable data is technically impossible. See Appendix A for the technical details.

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. [23] 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 *pessimistic (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 pessimistic 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 a probably approximately correct auditing method, then we also have a probably approximately correct regret estimator, and vice versa.

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

Definition 3.1. For a fixed cost c and regret threshold $r > 0$, an *auditing method* is a mapping $\mathcal{A} : \bigcup_{t\geq 0} \mathcal{H}^t \to \{\mathrm{G}, \mathrm{S}\}$, 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 $\mathcal{A} : \bigcup_{t\geq 0} \mathcal{H}^t \to \mathbb{R}$, and the output indicates the estimated regret *assuming* the seller's cost c.

Note that the auditing method proposed by Hartline et al. [23] is based on a regret estimator, so it suffices to reduce regret estimation to auditing.

Lemma 3.2. *Suppose we have an auditing method $\mathcal{A}$ that satisfies the following property:*

- *If the regret of a transcript $\mathcal{T}^T$ of length T is smaller than or equal to r, then it outputs S (smaller), and*
- *if the regret of the transcript $\mathcal{T}^T$ is larger than or equal to $r + \varepsilon$, then it outputs G (greater),*

with probability at least $1 - f(\varepsilon, T)$. Then there exists an estimator of the regret of $\mathcal{T}^T$ up to accuracy ε and error probability at most $\frac{\overline{p} f(\varepsilon,T)}{\varepsilon}$.

All proofs, unless otherwise specified, are deferred to Appendix B.

The Consistency Requirement. From now on we focus on regret estimators. To do 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 $\{\{x^t(p^t)\}_{t=1}^T, \boldsymbol{p}^T, \boldsymbol{\pi}^T\}_{T\geq 1}$ of transcripts, $\varepsilon > 0$, cost c, and sequence of ground-truth sequence of allocations $\{\boldsymbol{x}^T\}_{T\geq 1}$ agreeing with the transcripts,

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

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

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. [23] 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. *The algorithm in Hartline et al. [23] 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 competitive 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.

Returning 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 competitive. The one-sided consistency property ensures that missing information is properly accounted for so that a supra-competitive seller is never deemed as competitive 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 *pessimistic (counterfactual) regret* of the transcript. We first define pessimistic (counterfactual) regret (Definition 3.8) and then state the implication (Proposition 3.9).

Intuitively, the pessimistic 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 indistinguishability 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 allocations, we define the pessimistic regret.

Definition 3.8 (Pessimistic 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 *pessimistic (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 pessimistic 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 pessimistic regret is low.

4 TESTING PESSIMISTIC REGRET

In this section, we present the refined auditing method that estimates the pessimistic regret (as defined in the previous section). We then show the following guarantee: With a sufficient amount of data, if the seller's pessimistic regret is low, then she passes the audit with high probability, and if the seller's pessimistic 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 pessimistic regret based on the above decomposition and the pessimistic estimation of allocations from data. The steps are described at a high level as follows.[9]

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:

[9] A formal pseudocode description can be found at the end of Appendix B.

Step 1 We estimate the allocation for 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.

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

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

Step 2 We estimate the true regret of the pessimistic 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

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

Then the pessimistic 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 pessimistic regret over all the possible costs to compute the pessimistic plausible regret:

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

Note that this can be done in polynomial time even with a continuum of costs. In fact, following the observations of Nekipelov et al. [27], 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 (piece-wise linear) 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 the reliability of the regret estimator. Specifically, if $\widetilde{R}^T(\tilde{c}) + \delta^T \le 2r$,[10] 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 pessimistic calibrated regret.

Theorem 4.1 (Sample complexity of testing pessimistic regret)**.** *Consider the environment with maximum possible price level $\overline{p}$ and $k = |\mathcal{P}|$ price levels. 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 $\underline{\pi} = \min_{p \in C^t, 1 \le t \le T} \pi^t(p)$ be the minimum probability in the support of price distributions. With our refined auditing method, when the number of rounds*

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

[10] If the choice of the price set $\mathcal{P}$ is endogenous (see Remark 2.3), then we add a discretization loss to the estimated regret. Specifically, let the price levels be $q_1 < \cdots < q_k$ and $d = \max_{0 \le i \le k+1}(q_i - q_{i-1})$, where $q_0 = 0, q_{k+1} = h$. Output result by comparing $\widetilde{R}^T(\tilde{c}) + \delta^T + d$ with $2r$ similarly as Step 4.

we have

(1) if the seller's true pessimistic regret $\overline{R}^T(c_0, \boldsymbol{x}^T) \le r$*, she passes with probability at least* $1-\alpha$*; and*
(2) if her plausible pessimistic regret $\overline{R}^T(c_*, \boldsymbol{x}^T) \ge 2r$*, then she fails with probability 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 pessimistic regret, and the following corollary says it outputs approximately at most the pessimistic regret. This corollary 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 method that has 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* $\{\boldsymbol{x}^T\}_{T\ge 1}$*.*

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

An extension: Auditing with aggregated price distributions. Recall that one of the motivations for designing the refined auditing method is to allow testing aggregated empirical distributions when price distribution data is not available. We end this section by showing that as long as the rate of change in price distributions is low, it is possible to run the same auditing method with the aggregated price distributions as a plug-in estimator for the true distributions. We do this by first showing a lemma that the aggregated price distributions are good approximators to the true price distributions, and then providing a sample complexity guarantee.

Lemma 4.3. *Consider* T *rounds and* k *price levels bounded in* $[0,1]$*. In round* t *the seller posts price* $p^t \sim \pi^t$*. Suppose* $\|\pi^t - \pi^{t+1}\|_\infty \le \varepsilon$ *for all* $1 \le t \le T-1$*.*[11] *Then there exists an algorithm that uses only price samples* $\{p^t\}_{t=1}^T$ *and outputs estimated price distributions* $\{\hat{\pi}^t\}_{t=1}^T$ *such that, with probability at least* $1-\delta$

$$\|\hat{\pi}^t - \pi^t\|_\infty \le \left(4\varepsilon \log\frac{2Tk}{\delta}\right)^{1/3},$$

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

Corollary 4.4. *Consider the environment with maximum possible price level* $\overline{p}$ *and* $k = |\mathcal{P}|$ *price levels. 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* $1-\delta$*, threshold* r*. Suppose the seller's pricing algorithm has the property that, for all* T'*, the price distributions* $\{\pi^t\}_{t=1}^T$ *generated by running the algorithm for* T' *rounds satisfy*

(1) $\|\pi^t - \pi^{t+1}\|_\infty \le (T')^{-\gamma}$ *for all* $1 \le t \le T'-1$ *and some* $\gamma > 0$*, and*
(2) for some $\underline{\pi}$*, for all* $p \in C^t$ *and* $1 \le t \le T$*,* $\pi^t(p) \ge \underline{\pi}$*, i.e., the minimum probability in the support of price distributions is at least* $\underline{\pi}$*,*

[11] 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.

then there exists an auditing method without access to $\{\pi^t\}_{t=1}^T$*, when the number of rounds*

$$T \geq \max\left\{t_0, \left(\frac{4(8\overline{p}k + r\underline{\pi})^3}{r^3\underline{\pi}^6}\right)^{2/\gamma}, \frac{16k^2}{r^2}\log\frac{8k^2}{\delta}\cdot\left(\frac{1}{\underline{\pi}}+1\right)^2\cdot\overline{p}^2, \left(\frac{4}{\underline{\pi}^3}\right)^{2/\gamma}\right\} + 1,$$

where $t_0 = \sup\left\{t \in \mathbb{R} : t^{\frac{\gamma}{2}} \leq \log\frac{8tk^3}{\delta}\right\}$*, for which*

(1) if the seller's true pessimistic regret $\overline{R}^T(c_0, \boldsymbol{x}^T) \leq r$*, she passes with probability at least* $1-\delta$*; and*
(2) if her plausible pessimistic regret $\overline{R}^T(c_*, \boldsymbol{x}^T) \geq 2r$*, then she fails with probability at least* $1-\delta$*.*

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 stronger calibrated regret must be used instead of weaker best-in-hindsight regret by arguing that best-in-hindsight regret fails to exclude supra-competitive algorithms even when useful side information is guaranteed not to exist. 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. But Hartline et al. [23] argue that the stronger vanishing calibrated regret is essential for the regulation of (non-)collusion. They do this with an example *with side information* where a seller posts supra-competitive prices while having non-positive best-in-hindsight regret.

Intuitively, regulation using calibrated regret is necessary with side information because we need to test whether the seller best responds to such useful information. A natural question is: If side information is guaranteed not to exist, can we use the weaker notion of (vanishing) best-in-hindsight regret? We give a negative answer 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 *manipulable* (to be defined shortly) to supra-competitive prices. Therefore, if we use the more permissive vanishing best-in-hindsight regret as the definition of non-collusion, then there could be more anti-competitive algorithms passing the audit, even if side information is guaranteed not to exist.

Combining the above argument with the fact that (simultaneous) calibrated regret minimization leads to approximate correlated equilibria [18], we conclude that it is important to require vanishing calibrated regret in the definition of non-collusion even when there is no side information.

We first define what we mean by manipulation, and then construct the manipulation example.

Definition 5.1 (Manipulation)**.** Fix an instance of imperfect price competition with two sellers and a property $\mathcal{A}$ of pricing algorithms. Suppose seller 2's algorithm satisfies property $\mathcal{A}$. A *manipulation of* $\mathcal{A}$*-algorithm* is an algorithm of seller 1 such that, when played against seller 2's algorithm, yields an outcome with the following properties:

(1) (supra-competitive) *Both* sellers have payoffs *strictly higher* than any correlated equilibrium.
(2) (undetectable) *Both* seller 1 and seller 2's algorithms satisfy property $\mathcal{A}$.

Next, we show a manipulation of no-best-in-hindsight-regret algorithms (i.e. $\mathcal{A}$ = no best-in-hindsight regret). The manipulation implies that there exists a scenario where the non-collusion definition of vanishing best-in-hindsight regret fails to identify a supra-competitive scenario.

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 consider a family of commonly used best-in-hindsight regret minimization algorithms defined as follows.

Definition 5.2 (γ-mean-based learning, Braverman et al. [6])**.** Fix horizon T and $\gamma = o(1)$. Let $\sigma_{p,t} = \sum_{s=1}^{t} u^s(p)$ 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., exponential-weight algorithm for exploration and exploitation, follow the perturbed leader) are known to be γ-mean-based learning algorithms [6]. In the following theorem, we show that a seller running a γ-mean-based learning algorithm is suspectible to manipulation into supra-competitive outcomes. The manipulator can also achieve no best-in-hindsight when doing such a manipulation.

Theorem 5.3. *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) (supra-competitive) for $\Omega(T)$ rounds, both play $p_1 > p_1^e, p_2 > p_2^e$ in each round with high probability,[12]
(2) (no loss of payoff) receive expected payoff $V_1'T - o(T), V_2'T - o(T)$ where constants $V_1' > V_1, V_2' > V_2$, and
(3) (no best-in-hindsight regret) both seller 1 and seller 2 have vanishing best-in-hindsight regret.

In words, Theorem 5.3 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.

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 present challenges to regulating collusion.

Recall that in the auditing problem, we assume that the cost of the seller is unknown to the regulator and the seller passes the audit as long as the *plausible* regret is low (recall that the plausible regret is the minimum regret that a seller with cost in $[\underline{c}, \bar{c}]$ can have). Therefore, it is conceivable that algorithms could find a supra-competitive outcome for the true costs but pass the audit for different costs. We ask whether such a phenomenon can actually occur. More precisely, our question is

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

We find an affirmative answer to this question with two implications. First, the imprecise knowledge of the cost of the seller could dramatically affect the result of the audit. Even when configured with true costs, supra-competitive

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

algorithms might still pass the test because the outcome is consistent with low regret for some other plausible costs. 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 range of reasonable costs.

Remark 6.1. 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 supra-competitive outcome that looks competitive with a higher cost.

Since algorithmic collusion refers to collusion facilitated by the algorithm, we consider only scenario 2 as algorithmic collusion.

Details of the simulation and results. We demonstrate by simulation the possibility that algorithms can find a supra-competitive outcome that is seen as competitive for a high inferred plausible cost.[13] The instance of dynamic imperfect price competition is the same as that in Hartline et al. [23] and the configuration of the algorithm resembles that in Banchio and Skrzypacz [5]. 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 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 same simple stateless Q-learning algorithm as that in Banchio and Skrzypacz [5], which only keeps track of the estimated continuation payoff for each price level $p \in \mathcal{P}$ in the Q-table. The hyper-parameters are chosen as follows. The Q-learning uses an ε-greedy strategy with optimistic initialization and exploration probability $\varepsilon = 0.01$. 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 $\gamma = 0.99$. Note that this stateless Q-learning in competition is a mis-specified model. Thus the Q-table payoffs are statistically inaccurate and this inaccuracy is the channel by which the algorithms achieve supra-competitive prices.

In Figure 1 we confirm that Q-learning exhibits supra-competitive behavior. Q-learning converges to strategies that both sellers post prices $(0.6, 0.65)$ in most cases (with payoffs 0.169 and 0.122). The $(0.6, 0.65)$ outcome is supra-competitive as its prices exceed the highest payoff correlated equilibrium of the game. The highest payoff equilibrium of the game is that the sellers post 0.5 and 0.55 (with payoffs 0.159 and 0.114), respectively. Note from the values of the payoffs that in this example, the payoff difference between the sellers' supra-competitive outcome and the equilibrium is small, thus the regret of the sellers is expected to be at a relatively small scale $10^{-2} \sim 10^{-3}$.

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 using our auditing method. The auditing result is shown in Figure 2.

[13] Code for replicating the simulation can be found at https://github.com/wangchang327/collusion-cslaw25.
[13] That of seller 2 is similar.

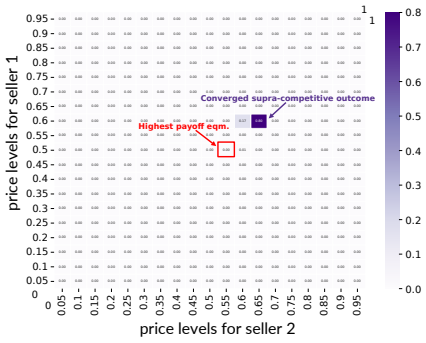
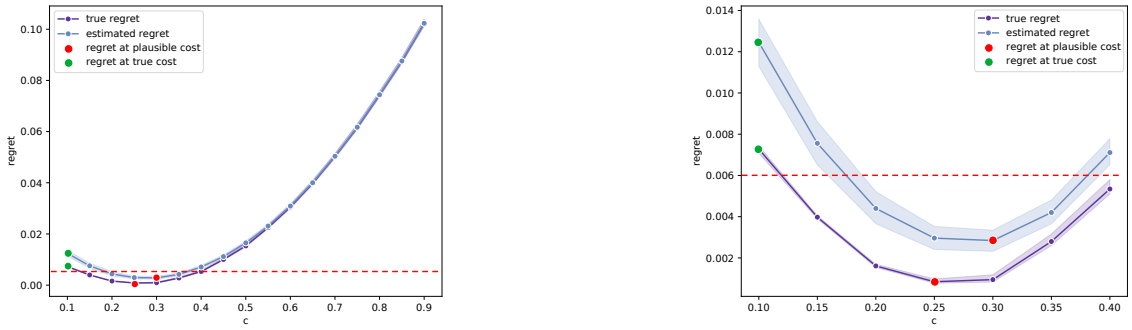


Fig. 1. The frequencies of each pair of strategies in the last 10 rounds of the competition are shown in the heatmap. The highest payoff equilibrium is marked in red.



(a) Range of costs $[0.1, 0.9]$

(b) Range of costs $[0.1, 0.4]$

Fig. 2. The true regret and estimated regret plotted against different assumed costs of seller 1, with a zoomed-view on the right. The red dashed line shows the example threshold 6×10^{-2} for estimated regret. The seller maintains exploration. The true regret is the regret estimated with the knowledge of the ground truth demand, while the estimated regret is the regret estimated using our auditing method.

In Figure 2 we observe that the plausible estimated regret of the seller is around 3×10^{-3} at $c' = 0.3$.[14] The true cost of the seller is in fact $c = 0.1$, for which the estimated regret is around 12×10^{-3}. If we set the auditing threshold to be, say, 6×10^{-3}, then the algorithm should be flagged as supra-competitive as the regret at $c = 0.1$ is more than 6×10^{-3}. But since the plausible regret is approximately $3 \times 10^{-3} < 6 \times 10^{-3}$, the seller will not be flagged. A similar pattern also applies to the true regret curve (i.e., even when the regret estimation is exact) at $c' = 0.25$. In Figure 3, we also plot the true regret at the true cost $c = 0.1$ and at the plausible cost $c' = 0.25$ with different time horizons. We observe that both the regret at the true cost and the plausible cost converge at $T = 10^6$. The regret at the true cost exceeds the regret at the plausible cost by a significant margin. Excluding the loss due to exploration, the regret at the true cost is not vanishing and the regret at the plausible cost is vanishing.

From the observations above, we conclude that although we know the Q-learning algorithm converges to supra-competitive prices, if we assume that the viable costs of the seller include $c' = 0.3$, then the Q-learning algorithm turns

[14]Recall that in this example, the payoff difference between the sellers' supra-competitive outcome and the equilibrium is small, thus the regret of the sellers is at a relatively small scale $10^{-2} \sim 10^{-3}$.

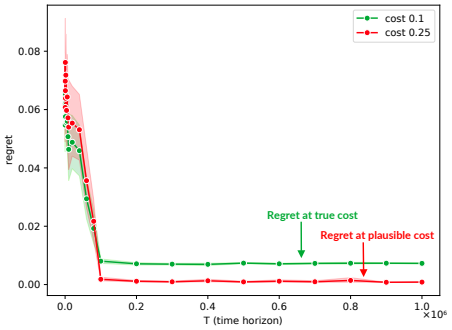


Fig. 3. The true regret at the true cost $c = 0.1$ and at the plausible cost $c' = 0.25$ plotted with different time horizons.

out to have a low plausible calibrated regret. In other words, when configured with a low cost (such as $c = 0.1$), the Q-learning algorithm finds outcomes that look competitive for a higher cost c' (such as $c' = 0.3$). The results of this experiment confirm that it is possible for algorithms to find a supra-competitive outcome that is seen as competitive at a high inferred plausible cost.

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. [23]. 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 pessimistic 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 is essential to eliminate more collusion-promoting algorithms even without side information. 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 supra-competitive behavior of the seller, which suggest a rule of reason can be useful in antitrust analysis. Open questions include: (a) designing a test for small learning rates to remove the need to record price distributions and (b) improving the sample complexity bound.

ACKNOWLEDGMENTS

We thank the reviewers of ACM CS&Law'25 for helpful feedback and revision recommendations. We also thank the attendees of the 2nd Workshop on Regulatable ML at NeurIPS'24 and the 2024/2025 NSF TRIPODS workshop for the discussions and feedback. The authors are partially supported by the National Science Foundation (CCF-2229162 and ECCS-2216970).

REFERENCES

Eshwar Ram Arunachaleswaran, Natalie Collina, Sampath Kannan, Aaron Roth, and Juba Ziani. Algorithmic collusion without threats. *arXiv preprint arXiv:2409.03956*, 2024.

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.

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

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.

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.

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.

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

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.

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

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.

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

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.

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

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

Chicago Board of Trade v. *United States*, 1918.

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

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

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

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

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.

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

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

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.

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

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

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

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.

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

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

APPENDICES

A REGRET OVER FULL RANDOMNESS OF THE ALGORITHM CANNOT BE ESTIMATED CONSISTENTLY

As we have discussed in Definition 2.2 and Remark 2.4, the regret we are estimating is not the true regret of the algorithm itself. It might be tempting to consider the latter which is a stronger notion and is the usual regret definition in the analysis of an algorithm. In this appendix, we show that it is not possible to estimate consistently even with fully-supported price distributions. Details in this appendix require the knowledge of definitions from Section 3.

Observe that for online learning algorithms, the distributions $\boldsymbol{\pi}^T$ are potentially random since the sampling distribution at $t+1$, π^{t+1}, depends on the realization of the price p^t sampled at t. This motivates the question of estimating the regret with respect to the full randomness of the algorithm. We first define an online posted-pricing algorithm as follows:

Definition A.1 (Online posted-pricing algorithm)**.** Call $\mathcal{T}^t = \{x^s(p^s), p^s, \pi^s\}_{s=1}^t$ a *transcript* of length t. The set of all the length-t transcripts is denoted as $\mathcal{H}^t$. Let $\mathcal{H}^0 = \emptyset$. An *online posted-pricing algorithm* is a mapping $\mathcal{M} : \bigcup_{t\geq 0} \mathcal{H}^t \to \Delta(\mathcal{P})$. Given the ground-truth $\boldsymbol{x}^T$, a T-round realization of algorithm $\mathcal{M}$ is a transcript of length T where

$$\pi^{t+1} = \mathcal{M}(\{x^s(p^s), p^s, \pi^s\}_{s=1}^t), \quad p^{t+1} \sim \pi^{t+1} \quad (0 \leq t \leq T).$$

Note that given the ground-truth $\boldsymbol{x}^T$ and T, a posted-pricing algorithm can be identified as a joint distribution (that depends on the ground-truth allocation) over $\{p^t, \pi^t\}_{t=1}^T$. The distribution of p^t conditioned on π^t is π^t.

Using this identification, we can write the expected regret of algorithm $\mathcal{M}$ given the ground-truth allocation $\boldsymbol{x}^T$ for a seller with cost c as

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

Unfortunately, it is impossible to estimate this regret consistently, even if all the price distributions are fully-supported.

Proposition A.2. *There does not exist a regret estimator such that*

$$\lim_{T\to\infty} \Pr_{\{p^t,\pi^t\}_{t=1}^T \sim \mathcal{M}} \left[|\mathcal{A}(\mathcal{T}^T) - R^T(\mathcal{M}, \boldsymbol{x}^T, c)| \geq \varepsilon \right] = 0,$$

even if we restrict attention to algorithms with fully-supported price distributions.

The intuition of the above proposition is that by looking at the transcript, the regret estimator has limited information about how the price distributions are generated, so it cannot consistently make inferences about the regret over the randomness of the price distributions.

Proof. Suppose not. Pick two algorithms $\mathcal{M}, \mathcal{M}'$, cost c and $\boldsymbol{x}^T$ such that $R^T(\mathcal{M}, \boldsymbol{x}^T, c) = r_1, R^T(\mathcal{M}', \boldsymbol{x}^T, c) = r_2$, and $r_1 - r_2 = \delta > 0$. Create a new algorithm $\overline{\mathcal{M}}$ as follows: Before the first round, the algorithm secretly flips a fair coin and secretly decides to run $\mathcal{M}$ or $\mathcal{M}'$ for infinite time according to the result of the flip. Note that the regret estimator sees an infinite sequence of transcript, generated by either $\mathcal{M}$ or $\mathcal{M}'$, w.p. $\frac{1}{2}$ each. Therefore,

$$\lim_{T\to\infty} \Pr_{\{p^t,\pi^t\}_{t=1}^T \sim \overline{\mathcal{M}}} \left[|\mathcal{A}(\mathcal{T}^T) - r_1| \leq \varepsilon \right] = \lim_{T\to\infty} \Pr_{\{p^t,\pi^t\}_{t=1}^T \sim \overline{\mathcal{M}}} \left[|\mathcal{A}(\mathcal{T}^T) - r_2| \leq \varepsilon \right] = \frac{1}{2}.$$

But note that $r_2 < R^T(\overline{\mathcal{M}}, \boldsymbol{x}^T, c) < r_1$. Set $\varepsilon_0 < \frac{1}{4}\min\{R^T(\overline{\mathcal{M}}) - r_1, r_2 - R^T(\overline{\mathcal{M}})\}$ we then have

$$\lim_{T\to\infty} \Pr_{\{p^t,\pi^t\}_{t=1}^T \sim \overline{\mathcal{M}}} \left[|\mathcal{A}(\mathcal{T}^T) - R^T(\overline{\mathcal{M}}, \boldsymbol{x}^T, c)| \geq \varepsilon_0\right] = 0,$$

a contradiction. □

B OMITTED PROOFS

B.1 Proof of Lemma 3.2

Proof. The estimation algorithm: 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 $[r_i + \varepsilon, \overline{p}]$, indicating which interval the true regret is in. Call this interval J_i. Let $J = \bigcap_{i=1}^k J_i$. If J contains only a single point or $|J| \leq \varepsilon$, then we output the single point or the midpoint of J. Otherwise, we draw a guess uniformly random from $[0, \overline{p}]$.

Analysis: Consider the "good event" that all the $\overline{p}/\varepsilon$ audits satisfy the property in the statement of Lemma 3.2. Suppose the true regret is in $[j\varepsilon, (j+1)\varepsilon]$. Then conditioned on the good event, for all audits with threshold $r \geq (j+1)\varepsilon$ they output S, and for all audits with $r \leq (j-1)\varepsilon$ they output G, and for the audit with threshold $r = j\varepsilon$, its output can be arbitrary. Observe that in this case the intersection of the outputs of audits except that with threshold $j\varepsilon$ is exactly $[j\varepsilon, (j+1)\varepsilon]$. The final intersection is either $\{j\varepsilon\}$, $\{(j+1)\varepsilon\}$, or $[j\varepsilon, (j+1)\varepsilon]$ depending on the output of the audit with threshold $j\varepsilon$ (which we do not have guarantee with). But all three possible outputs are within ε error of the true regret. We conclude the proof by noting that union bound gives that the good event happens with probability at least $1 - \frac{\overline{p}f(\varepsilon,T)}{\varepsilon}$. □

B.2 Proof of Proposition 3.5

Proof. From the proof of Lemma A.1 in Hartline et al. [2024], we have for any $\varepsilon > 0$

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

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

$$P^T_{\geq\varepsilon} \leq 2k^2 \exp\left(-\frac{\varepsilon^2 T}{2k^2\overline{p}^2(\underline{\pi}^T + 1)^2}\right) = o(T).$$

Therefore, we have $\lim_{T\to\infty} P^T_{\geq\varepsilon} = 0$. □

B.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 $1 \leq t \leq T$ we have $\pi^t(p_k) = 0$ and $\pi^t(p_{k-1}) \neq 0$. Pick an arbitrary positive constant $a \leq 1$ and let $\boldsymbol{x}^T$ be such that for all $1 \leq t \leq T$, $x^t(p_i) = a$ and $x^t(p_k) = 0$ for $1 \leq i < k$. 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,

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

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

Next, we aim to show 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 $1 \le t \le T$ 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

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

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 (B.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 B.1. 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.

B.4 Proof of Proposition 3.9

We first observe the following lemma.

Lemma B.2. *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

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

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 B.2 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. □

B.5 Proof of Theorem 4.1

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

Lemma B.3. *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*

$$z_*^t(p) = \begin{cases} x^t(p) & (\textit{if } p\in C^t), \\ x^t(p') & (\textit{if } p' = \max\{r\le p : r\in C^t\} \textit{exists}), \\ 1 & (\textit{otherwise}). \end{cases}$$

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

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T\sum_p \pi^t(p)\left[(\sigma(p)-c)z_*^t(\sigma(p)) - (p-c)z_*^t(p)\right] \\
&\qquad\qquad \geq \frac{1}{T}\sum_{t=1}^T\sum_p \pi^t(p)\left[(\sigma(p)-c)x^t(\sigma(p)) - (p-c)x^t(p)\right] \qquad \text{(B.2)}
\end{aligned}$$

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

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

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

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

Lemma B.4. *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, conditioned on observing the sequence of price distributions $\boldsymbol{\pi}^T$, for any fixed sequence of allocations $\boldsymbol{\pi}^T$,*

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

where

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

We defer the proof of this lemma after presenting the proof of Theorem 4.1.

Now we state the proof of Theorem 4.1.

Proof. (1) Starting from Lemma B.4, we claim that for any fixed c

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

To bound the probability of

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

note that

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

Plugging 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 $\widetilde{R}^T(\tilde{c},\boldsymbol{x}^T) \leq \widetilde{R}^T(c_0,\boldsymbol{x}^T)$. 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

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

$$\begin{aligned}
&\le \Pr[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) - \overline{R}^T(c_0, \boldsymbol{x}^T) \ge \delta^T] \\
&\le \Pr[\widetilde{R}^T(c_0, \boldsymbol{x}^T) - \overline{R}^T(c_0, \boldsymbol{x}^T) \ge \delta^T] && (\text{since } \widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) \le \widetilde{R}^T(c_0, \boldsymbol{x}^T)) \\
&\le \alpha
\end{aligned}$$

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

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

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

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

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

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 $c \in [\underline{c}, \bar{c}]$,

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

Combining Lemma B.4, we get

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

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

$$\Pr\left[\widetilde{R}^T(\tilde{c}, \boldsymbol{x}^T) + \delta^T \le 2r\right]$$

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

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

B.5.1 Proof of Lemma B.4.

Proof. Let

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

we claim that $\mathbb{E}_{p^t\sim\pi^t}\left[\widetilde{r}^t_{p,q}\right] = \overline{r}^t_{p,q}$. In fact, first note that when $p \notin C^t$, we have $\pi^t(p) = 0$ and hence $\widetilde{r}^t_{p,q} = \overline{r}^t_{p,q} = 0$ and the equality holds. Therefore, we assume from now that $\pi^t(p) \neq 0$. By definition of $\hat{x}^t$, when $p' = \max\{r \leq p : r \in C^t\}$ exits, we have

$$
\hat{x}^t(p) = \begin{cases} x^t(p')/\pi^t(p') & p' = p^t, \\ 0 & \text{otherwise.} \end{cases}, \tag{B.3}
$$

which implies that $\mathbb{E}_{p^t\sim\pi^t}\left[\hat{x}^t(p')\right] = x^t(p')$. By definition of x^t_*, we have $x^t(p') = x^t_*(p)$. When p' does not exist we have $\hat{x}^t(p) = x^t_*(p) = 1$. Applying the same reasoning we also have $\mathbb{E}_{p^t\sim\pi^t}\left[\hat{x}^t(q)\right] = x^t_*(q)$. 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 $p \leq \overline{p}, q \leq \overline{p}$, 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

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

By a similar argument on the other side, we also have

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

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

B.6 Proof of Lemma 4.3

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 $F_1, \ldots, F_L$ and $p_1 \sim F_1, \ldots, p_L \sim F_L$. 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's inequality 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. □

Remark B.5. 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 $V_1, \ldots, V_k$ be the current cumulative reward for actions $1, \ldots, k$. 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).

B.7 Proof of Corollary 4.4

The new algorithm is the same as the algorithm mentioned in Section 4 with the estimated distribution using *Lemma* 4.3, except that at the end the threshold δ^T to add is

$$k\frac{\overline{p}\rho'}{\underline{\pi}}\left(\frac{1}{\underline{\pi} - \rho'} + 1\right) + \sqrt{\log\frac{8k^2}{\delta} \cdot 2\left(\frac{1}{\underline{\pi}} + 1\right)^2 \overline{p}^2 \cdot \frac{1}{T}},$$

where $\rho' = \sqrt[3]{T^{-\gamma}\log\frac{8Tk^3}{\delta}}$.

Step 1. To prove Corollary 4.4, we first prove a variant of Lemma B.4.

Let $\check{x}^t(p)$ defined to be $\hat{x}^t(p)$ with π^t is substituted with $\hat{\pi}^t$ estimated according Lemma 4.3. Let

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

and $\rho = \sqrt[3]{4T^{-\gamma}\log\frac{2Tk}{\delta}}$ be the bound on the ℓ_∞ distance between the estimated $\hat{\pi}^t$ and π^t. With probability $1-\delta$, we have

$$\begin{aligned}|\check{r}^t_{p,q} - \widetilde{r}^t_{p,q}| &= \frac{1}{T}\left|(q-c)[\hat{\pi}^t(p)\check{x}^t(q) - \pi^t(p)\hat{x}^t(q)] - (p-c)[\hat{\pi}^t(p)\check{x}^t(p) - \pi^t(p)\hat{x}^t(p)]\right| \\ &\le \frac{1}{T}\left(\left|(q-c)[\hat{\pi}^t(p)\check{x}^t(q) - \pi^t(p)\hat{x}^t(q)]\right| + \left|(p-c)[\hat{\pi}^t(p)\check{x}^t(p) - \pi^t(p)\hat{x}^t(p)]\right|\right) \\ &\le \frac{1}{T}\overline{p}\left(\left|\hat{\pi}^t(p)\check{x}^t(q) - \pi^t(p)\hat{x}^t(q)\right| + \left|\hat{\pi}^t(p)\check{x}^t(p) - \pi^t(p)\hat{x}^t(p)\right|\right).\end{aligned}$$

For the first term in the parenthesis we have

$$\begin{aligned}|\hat{\pi}^t(p)\check{x}^t(q) - \pi^t(p)\hat{x}^t(q)| \le& |\hat{\pi}^t(p)\check{x}^t(q) - \hat{\pi}^t(p)\hat{x}^t(q) + \hat{\pi}^t(p)\hat{x}^t(q) - \pi^t(p)\hat{x}^t(q)| \\ \le& |\hat{\pi}^t(p)\check{x}^t(q) - \hat{\pi}^t(p)\hat{x}^t(q)| + |\hat{\pi}^t(p)\hat{x}^t(q) - \pi^t(p)\hat{x}^t(q)| \\ \le& |\hat{\pi}^t(p) - \pi^t(p)||\hat{x}^t(q) + \hat{\pi}^t(p)||\check{x}^t(q) - \hat{x}^t(q)|.\end{aligned}$$

Since we assume that the estimation in Lemma 4.3 succeeded, we have $|\hat{\pi}^t(p) - \pi^t(p)| \le \rho$. By assumption that $\pi^t(q) \ge \underline{\pi}$, we have $\hat{x}^t(q) \le x^t(q)/\pi^t(q) \le 1/\underline{\pi}$. Therefore

$$|\hat{\pi}^t(p) - \pi^t(p)|\hat{x}^t(q) \le \frac{\rho}{\underline{\pi}}.$$

When T is large enough so that $\underline{\pi} > \rho$ holds, π^t and $\hat{\pi}^t$ will have the same support with high probability. From now let us assume it is the case (and add this requirement on T in the end as well). By definition of $\check{x}^t$ and $\hat{x}^t$, for $q' = \max\{r \le q : \hat{\pi}^t(r) > 0\}, q'' = \max\{r \le q : \pi^t(r) > 0\}$, we have

$$\check{x}^t(q) - \hat{x}^t(q) = \frac{x^t(q')}{\hat{\pi}^t(q')} - \frac{x^t(q'')}{\pi^t(q'')}.$$

As $\hat{\pi}^t$ and π^t have the same support, we know that $q' = q''$. Therefore

$$\check{x}^t(q) - \hat{x}^t(q) \le x^t(q')\left|\frac{\pi^t(q') - \hat{\pi}^t(q')}{\hat{\pi}^t(q')\pi^t(q')}\right| \le 1 \cdot \frac{\rho}{(\underline{\pi}-\rho)\underline{\pi}}$$

as $x^t(q') \le 1, |\pi^t(q') - \hat{\pi}^t(q')| \le \rho, \pi^t(q') \ge \underline{\pi}, \hat{\pi}^t(q') \ge \pi^t(q') - \rho$, and by the construction of T that we have $\underline{\pi} > \rho$.

For the second term in the parenthesis, by definition of $\check{x}^t, \hat{x}^t$ as well as the assumption that $\hat{\pi}^t$ and π^t have the same support we have

$$\left|\hat{\pi}^t(p)\check{x}^t(p) - \pi^t(p)\hat{x}^t(p)\right| = 0.$$

Therefore, combining the two terms, with $1-\delta$ probability

$$|\check{r}^t_{p,q} - \widetilde{r}^t_{p,q}| \le \frac{1}{T}\frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho} + 1\right).$$

It follows that with probability at $1-\delta$

$$\check{R}^T(c, \boldsymbol{x}^T) - \widetilde{R}^T(c, \boldsymbol{x}^T) \le \frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho} + 1\right).$$

Then by the triangle inequality and Lemma B.4, we have, with probability at least $1-2\delta$,

$$\check{R}^T(c, \boldsymbol{x}^T) - \overline{R}^T(c, \boldsymbol{x}^T) \le \frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho} + 1\right) + \sqrt{\log\frac{2}{\delta} \cdot 2\left(\frac{1}{\underline{\pi}} + 1\right)^2 \overline{p}^2 \cdot \frac{1}{T}}.$$

For the sake of brevity from now on we let

$$f(T,\delta) = \frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho}+1\right) + \sqrt{\log\frac{2}{\delta}\cdot 2\left(\frac{1}{\underline{\pi}}+1\right)^2\overline{p}^2\cdot\frac{1}{T}}.$$

Observe that $f(T,\delta)$ is weakly decreasing with δ.

Step 2. Next, we prove the first point of Corollary 4.4.

We claim that for any $c \in [\underline{c},\bar{c}]$

$$\Pr\left[\check{R}^T(c,\boldsymbol{x}^T) - \overline{R}^T(c,\boldsymbol{x}^T) \geq kf(T,\delta)\right] \leq 2k^2\delta,$$

which implies

$$\Pr\left[\check{R}^T(c,\boldsymbol{x}^T) - \overline{R}^T(c,\boldsymbol{x}^T) \geq kf\left(T,\frac{\delta}{2k^2}\right)\right] \leq \delta.$$

Monotonicity of $f(\delta,T)$ in δ further implies

$$\Pr\left[\check{R}^T(c,\boldsymbol{x}^T) - \overline{R}^T(c,\boldsymbol{x}^T) \geq kf\left(T,\frac{\delta}{4k^2}\right)\right] \leq \delta.$$

In fact,

$$\Pr\left[\check{R}^T(c,\boldsymbol{x}^T) - \overline{R}^T(c,\boldsymbol{x}^T) \geq kf(T,\delta)\right] = \Pr\left[\sum_p \max_q \check{R}^T_{p,q}(c,\boldsymbol{x}^T) - \sum_p \max_q \overline{R}^T_{p,q}(c,\boldsymbol{x}^T) \geq kf(T,\delta)\right],$$

and

$$\begin{aligned}
&\Pr\left[\sum_p \max_q \check{R}^T_{p,q}(c,\boldsymbol{x}^T) - \sum_p \max_q \overline{R}^T_{p,q}(c,\boldsymbol{x}^T) \geq kf(T,\delta)\right]\\
\leq&\Pr\left[\exists p\in\mathcal{P},\exists p'\in\mathcal{P},\check{R}^T_{p,p'}(c,\boldsymbol{x}^T) - \overline{R}^T_{p,p'}(c,\boldsymbol{x}^T) \geq \frac{kf(T,\delta)}{|\mathcal{P}|}\right]\\
\leq&2k^2\delta && \text{(union bound).}
\end{aligned}$$

Recall the choice of parameters: $\delta^T = kf\left(T,\frac{\delta}{4k^2}\right)$ and T satisfies that $kf\left(T,\frac{\delta}{4k^2}\right) \leq \frac{r}{2}$. Suppose the seller satisfies $R^T(c_0,\boldsymbol{x}^T) \leq r$, the probability that she fails the test

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

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

Step 3. Finally, we prove the first point of Corollary 4.4. We claim for $\tilde{c} = \operatorname{argmin}_{c\in[\underline{c},\bar{c}]}\check{R}^T(c,\boldsymbol{x}^T)$ that

$$\Pr\left[\check{R}^T(c_*,\boldsymbol{x}^T) - \overline{R}^T(c_*,\boldsymbol{x}^T) \leq -kf\left(T,\frac{\delta}{4k^2}\right)\right] \leq \delta.$$

In fact, note that

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

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

$$\begin{aligned}
&\Pr\left[\exists c,\exists p\in\mathcal{P},\exists q'\in\mathcal{P},\check{R}^T_{p,p'}(c,\boldsymbol{x}^T)-\overline{R}^T_{p,p'}(c,\boldsymbol{x}^T)\le -\frac{kf(T,\delta)}{|\mathcal{P}|}\right]\\
\le&k^2\Pr\left[\exists c,\check{R}^T_{p,p'}(c,\boldsymbol{x}^T)-\overline{R}^T_{p,q'}(c,\boldsymbol{x}^T)\le -\frac{kf(T,\delta)}{|\mathcal{P}|}\right].
\end{aligned}$$

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

$$\begin{aligned}
&\Pr\left[\exists c,\check{R}^T_{p,p'}(c,\boldsymbol{x}^T)-\overline{R}^T_{p,p'}(c,\boldsymbol{x}^T)\le -\frac{kf(T,\delta)}{|\mathcal{P}|}\right]\\
\le&\Pr\left[\check{R}^T_{p,q'}(\underline{c},\boldsymbol{x}^T)-\overline{R}^T_{p,q'}(\underline{c},\boldsymbol{x}^T)\le -\frac{kf(T,\delta)}{|\mathcal{P}|}\cup\check{R}^T_{p,q'}(\bar{c},\boldsymbol{x}^T)-\overline{R}^T_{p,q'}(\bar{c},\boldsymbol{x}^T)\le -\frac{kf(T,\delta)}{|\mathcal{P}|}\right]\\
\le&\Pr\left[\check{R}^T_{p,q'}(\underline{c},\boldsymbol{x}^T)-\overline{R}^T_{p,q'}(\underline{c},\boldsymbol{x}^T)\le -f(T,\delta)\right]+\Pr\left[\check{R}^T_{p,q'}(\bar{c},\boldsymbol{x}^T)-\overline{R}^T_{p,q'}(\bar{c},\boldsymbol{x}^T)\le -f(T,\delta)\right]\\
\le&4\delta k^2.
\end{aligned}$$

Recall the choice of parameters: $\delta^T=kf\left(T,\frac{\delta}{4k^2}\right)$ and T satisfies that $kf\left(T,\frac{\delta}{4k^2}\right)\le\frac{r}{2}$. Plugging in the bound of T and δ^T, we have $\Pr\left[\check{R}^T(\tilde{c},\boldsymbol{x}^T)-\overline{R}^T(\tilde{c},\boldsymbol{x}^T)\le-\delta^T\right]\le\delta$.

Therefore, when the seller's regret $\overline{R}^T(c_*,\boldsymbol{x}^T)\ge 2r$, by definition of c_* and $\tilde{c}$, we have

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

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

Step 4. Finally, we do the algebra to verify our bound on T indeed satisfies that $kf\left(T,\frac{\delta}{4k^2}\right)\le\frac{r}{2}$ and $\underline{\pi}>\rho$. Recall that

$$T\ge\max\left\{t_0,\left(\frac{4(8\overline{p}k+r\underline{\pi})^3}{r^3\underline{\pi}^6}\right)^{2/\gamma},\frac{16k^2}{r^2}\log\frac{8k^2}{\delta}\cdot\left(\frac{1}{\underline{\pi}}+1\right)^2\cdot\overline{p}^2,\left(\frac{4}{\underline{\pi}^3}\right)^{2/\gamma}\right\}+1,$$

where $t_0 = \sup\left\{t \in \mathbb{R} : t^{\frac{\gamma}{2}} \le \log \frac{8tk^3}{\delta}\right\}$, and

$$f\left(T, \frac{\delta}{4k^2}\right) = \frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho}+1\right) + \sqrt{\log\frac{8k^2}{\delta}\cdot 2\left(\frac{1}{\underline{\pi}}+1\right)^2\overline{p}^2\cdot\frac{1}{T}}.$$

We show that in this case

$$k\frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho}+1\right) \le \frac{r}{4}, \quad \text{and} \quad k\sqrt{\log\frac{8k^2}{\delta}\cdot 2\left(\frac{1}{\underline{\pi}}+1\right)^2\overline{p}^2\cdot\frac{1}{T}} \le \frac{r}{4}.$$

In fact, since $T \ge t_0 + 1$, we have $T^{\frac{\gamma}{2}} > \log\frac{8Tk^3}{\delta}$, hence

$$\rho \le \sqrt[3]{\frac{4}{T^{\gamma/2}}} \le \frac{r\underline{\pi}^2}{8\overline{p}k + r\underline{\pi}}.$$

It follows that

$$k\frac{\overline{p}\rho}{\underline{\pi}}\left(\frac{1}{\underline{\pi}-\rho}+1\right) \le k\frac{\overline{p}\rho}{\underline{\pi}}\cdot\frac{2}{\underline{\pi}-\rho} \le k\frac{\overline{p}\frac{r\underline{\pi}^2}{8\overline{p}k+r\underline{\pi}}}{\underline{\pi}}\cdot\frac{2}{\underline{\pi}-\frac{r\underline{\pi}^2}{8\overline{p}k+r\underline{\pi}}} = k\frac{\overline{p}\frac{r\underline{\pi}^2}{8\overline{p}k+r\underline{\pi}}}{\underline{\pi}}\cdot\frac{2}{\frac{8\overline{p}k\underline{\pi}}{8\overline{p}k+r\underline{\pi}}} = \frac{r}{4}.$$

The verification of the second $\le \frac{r}{4}$ follows immediately using $T \ge \frac{16k^2}{r^2}\log\frac{8k^2}{\delta}\cdot\left(\frac{1}{\underline{\pi}}+1\right)^2\cdot\overline{p}^2$. The verification that $\underline{\pi} > \rho$ follows similarly from $T \ge \left(\frac{4}{\underline{\pi}^3}\right)^{2/\gamma} + 1$ and $T \ge t_0 + 1$.

This concludes the proof.

B.8 Proof of Theorem 5.3

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

Example B.6. 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 breaks ties randomly

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

and she 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 B.7. *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 she does not buy good 1, then $0 > v_1 - p_1 > v_2 - p_2$ so she buys nothing. But by construction, the buyer never does this. □

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

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

Using Equation (B.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 pure NE where they play $(p_1, p_2) = (2, 2)$.

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

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 an expected payoff of $2.6445T - o(T)$. 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 an expected payoff of $1.6555T - o(T)$. 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: She first plays $p_1 = 1$ for T rounds, and then switches to playing $p_1 = 3$ for the remaining $1.1T$ rounds.

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

Claim B.8. *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 B.9. Although the above argument suffices for our purposes, we remark that the claim still holds even if the sellers only get 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 supra-competitive will only suffer an $o(T)$ loss.

This ends the proof. □

Algorithm 1: Auditing (non)-collusion via testing the pessimistic 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

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

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