Detecting Algorithmic Collusion by Algorithms

Yonghong An Yongzhi Xu Yu Zhu*

November 16, 2025

Preliminary draft, comments welcome

Abstract

Algorithms can facilitate collusive behaviors among competing firms. It is challenging for the antitrust authority to monitor and detect algorithmic collusion due to complicated price patterns and frequent price changes. In this paper, we study two important issues assuming that the antitrust authority employs algorithms: how firms respond to an algorithmic antitrust authority and what price patterns the algorithmic antitrust authority would detect. In a framework of quantity competition between two firms, we develop a static theoretical model describing the interaction between the algorithmic authority, and then simulate the behavior of the authority and firms using a Q-learning algorithm. We find that the simulation results are partially consistent with the static model, and that the antitrust authority's algorithms can effectively boost firms' quantities and reduce the possibility of algorithmic collusion. The results are robust to alternative algorithms, different levels of auditing strictness, asymmetry of learning rate between the authority and firms, and replacing quantity-based auditing with price-based auditing. The effectiveness of auditing mainly relies on firms' cost structure and the authority's incentive to audit.

Keywords: algorithmic pricing, collusion, antitrust

JEL Classifications: L1, L13, L31

*Yonghong An: Department of Economics, Texas A&M University; email: yonghongan@tamu.edu. Yongzhi Xu: Department of Economics, Texas A&M University; email: yz_xu@tamu.edu. Yu Zhu: School of Economics, Renmin University; email: zhuyuzlf57@gmail.com.

1 Introduction

As the explosion of Artificial intelligence (AI), algorithms have been increasingly used to price goods and services in competitive markets. With the help of AI, algorithms are capable of learning and discovering profit-enhancing collusive pricing rules. Recent empirical and simulation studies have shown that algorithms may find and maintain supracompetitive prices without communication between competing firms. On the simulation side, Calvano et al. (2020b) and Klein (2021) demonstrate that algorithmic pricing adopted by competing, profit maximizing firms can lead to substantially higher prices than their competitive counterparts. On the empirical side, Assad et al. (2024) provides empirical evidence of algorithmic collusion in Germany's retail gasoline markets. Adopting algorithms increases gas stations' markup by 20 to 30%. The findings are consistent with the results of simulation studies.[1]

The recent development in algorithmic pricing raises concerns regarding the possibility of algorithmic collusion for government authorities. Both the U.S. Federal Trade Commission (FTC) and the European Commission have already considered the possible impacts of using algorithms, artificial intelligence, and predictive analytics in business decisions and their conduct on consumer welfare.[2] However, it is still an open question how antitrust agencies identify collusive pricing rules when firms use algorithms. Recently, Calvano et al. (2020a) suggests that the authorities identify the collusive pricing rules by checking properties of prices derived from economic theory and studies of human collusion. However, it is unclear whether price patterns of algorithmic collusion are consistent with those predicted by economic theory.

Instead of detecting algorithmic collusion based on predetermined pricing rules implied by economic theory and human collusion, we consider a different and novel approach to detect algorithmic collusion in this paper. The main idea is that the authorities also employ algorithms to audit firms' prices and find the collusive price patterns. We address several important questions on detecting algorithmic collusion. First, whether detecting collusion using algorithms can be effective. Second, what pricing patterns an algorithmic authority identifies to be collusive. Third, how competitive firms using algorithms respond to the algorithmic authority.

We first develop a theoretical model describing the auditing game, in which two firms produce a homogeneous product and compete in quantity, and they can collude with a contact cost. An antitrust authority tries to identify and penalize the supracompetitive behavior of

[1] The empirical evidence is indirect because the time of adoption of the pricing algorithms is not observed but inferred.

[2] Please see Commission et al. (2018) and OECD (2017).

the firms. We solve the game under two cost structures: firms' costs are the same and fixed, and each firm's cost is randomly drawn from a binary distribution. We prove the existence of equilibria under both cost structures. Under fixed cost, firms either produce at the Cournot level or the monopoly level. The authority employs a pure strategy to audit: it audits when both firms' quantities are lower than the Cournot level and does not audit otherwise. When marginal costs are random, firms collude with positive probability and the authority audits using a mixed strategy provided that the benefit-cost ratio for the authority is sufficiently large. The static game provides a benchmark for the comparison of the simulation results.

We simulate the model by using a Q-learning algorithm, one of the most widely used algorithms in reinforcement learning. The algorithm mimics a rational agent's behavior in a finite Markov decision process and learns the value as a function of action and state variables to maximize its present value of discounted payoff. In the simulation setup, we consider a Cournot model where two competing firms with private marginal costs set quantities for identical products by algorithms. The antitrust authority employs Q-learning to monitor and detect collusion by the following decision rule: Whenever both firms' quantities are lower than the Cournot quantities at a predetermined threshold, the firms are identified to be collusive. The antitrust authority is rewarded by successfully detecting collusion and firms incur loss due to detected collusive behavior. The state variables for firms are both firms' quantities and the authority's auditing decision in the previous period. In addition to these variable, the authority also observes firms' quantities in the current period as its state variable.

In the simulation experiments, we consider two setups: with and without the antitrust authority. For both setups, we consider two cost structures: both firms have the same constant marginal cost, and both firms' costs are randomly drawn from a binary distribution with the support of high and low costs. In all the combined scenarios above, we simulate the learning process of firms' quantities and payoff, market price, the authority's auditing probability and payoff, consumer surplus, and total surplus. We further check how those simulated objectives change with the authority's incentive, measured by its benefit-cost ratio.

The simulation results demonstrate that without the authority, firms learn to collude and produce 18% and 5% below the static Cournot equilibrium when costs are fixed and random, respectively. Also, it takes longer for firms to collude when the costs are random. When the antitrust authority joins the game, the convergence process of the firms changes substantially. The authority's auditing is effective in improving the quantities to the Cournot level or above. When marginal cost is fixed, the authority's auditing is successful regardless of the incentive of the authority. A generic pattern is that the firms first produce below the Cournot quantity and the authority audits with a high probability around 50%. As a

response, firms quickly adjust their quantities above the Cournot quantity, then gradually learn to produce at the Cournot level.

When marginal costs are random, the auditing is effective only when the authority's incentive is sufficiently strong (the benefit-cost ratio is greater than 2). When the incentive is weak, the auditing is not successful: even though firms increase quantities as a consequence of the relatively higher auditing probability at the beginning, the converged quantity is still lower than the Cournot level. As the benefit-cost ratio increases to 2 and 4, however, the authority's highest auditing probability increases from 50significantly to 80% and 90%, respectively. The firms respond to the auditing by increasing quantities significantly, which converge to 0.350 and 0.349, respectively, both of which are higher than the Cournot quantity. As the quantity increases, the authority learns to audit with smaller probabilities.

The authority's auditing probabilities conditional on the firms' quantities in the current period show that the auditing behavior is consistent with the static game under fixed marginal cost: the authority audits with probability one whenever both firms' quantities are lower than the Cournot level, otherwise, the authority does not audit at all. Under random marginal costs, the authority's auditing behavior is largely consistent with the prediction of the static game. It can learn to identify the four equilibria quantity combinations where it does not audit. The authority audits when the quantities deviate from three of the four equilibria above. However, the static equilibrium predicts that auditing occurs only at one of the four equilibria.

To check the robustness of our findings, we change (1) the penalty parameter of the authority such that the auditing is less strict and (2) the learning rate of the authority such that it learns slower than the firms. We find that our findings are robust to those changes.

We also check the simulation results by changing the model setup. First, we employ the Actor-Critic algorithm to replace Q-learning. Different from Q-learning that learns value function, Actor–Critic is a hybrid policy/value method — it explicitly learns a policy (actor) guided by a value estimate (critic). Q-learning uses discrete action selection via argmax, while Actor–Critic can handle continuous actions more easily and tends to learn smoother policies. The simulation results using Actor-Critic provide the similar main message as Q-learning: the auditing is effective in restricting firms from learning to collude and the effectiveness depends on cost structure. The major difference between the two sets of results is that Actor-Critic never audits with probability 1, and it audits with probability 0 only at the four equilibria discussed above. The comparison of the two algorithm demonstrates that our findings are robust to the algorithms, too. The second model change is that the state variable is changed from quantity to price. This is motivated by the fact that in some markets, only price rather than quantity is observed. In the modified model, firms' state

variables are the market price and the authority's auditing decision in the previous period, and the authority's state variables include the state variables of firms and the current period market price. We find from the simulation results that the authority audits with higher probabilities at the early stage, therefore the game converges faster than in the baseline case. An interpretation of the discrepancy is that the information in market price is "more aggregated" than two quantities. For all the quantities combinations that correspond to the same market price, the authority may only audits some of the combinations. However, when price is the state variable, the authority likely audits this price with probability 1.

The main contribution of our paper to the literature on algorithmic pricing is to investigate the possibility of detecting collusion by algorithms. As we discussed above, the existing literature on detecting collusion due to algorithmic pricing still focuses on the price patterns derived from economic theory or human collusion (e.g., see Calvano et al. (2020a)). We show that it is possible to detect collusion successfully using algorithms. These results are the first in the literature on detecting algorithmic collusion by algorithms. Our results shed light on antitrust practice in the presence of algorithmic collusion due to algorithmic pricing by competing firms. A related work to ours is Johnson et al. (2023). This paper discusses how a platform can design policies to promote competition and prohibit algorithmic collusion on the platform. The main idea is that the platform's policy can affect the demand of a firm by promoting those products with low prices such that colluding to increase prices is not profitable. Nevertheless, the platform policies do not apply to market competition because the antitrust authority is not supposed to intervene the market.

Another contribution of our paper is to show the impacts of cost uncertainty on algorithmic pricing and algorithmic collusion. The growing literature on algorithmic pricing mainly focuses on the effects of algorithms on competition, leaving information fixed. For example, Asker et al. (2023) finds that algorithms may not lead to collusion if learning is synchronous. Wang et al. (2023) analyzes the competition between learning algorithms and rule-based algorithms. Dou et al. (2023) studies the impacts of AI in capital markets. None of them explores how information structures of costs affect algorithmic pricing. We find that cost randomness plays a crucial role in algorithmic collusion and auditing. The auditing is more effective under random costs because firms find it more difficult to collude.

The outline of the remaining paper is as follows. Section 2 presents a static model to describe the competition between firms and the monitoring and auditing behavior of an antitrust authority. Section 3 summarizes the Q-learning and simulation setup. Section 4 summarizes simulation results and robustness checks. Section 5 extends our simulation studies to alternative models, and Section 6 concludes. We provide simple analyses for our model in the static case in the Appendix.

2 Theoretical Models

To set a basis for simulation analysis, we introduce Cournot games in two scenarios: Firms compete non-cooperatively in quantities and firms compete cooperatively in quantities with communication costs. In both games, we study their static equilibrium with and without monitoring of an antitrust authority.

When the game is non-cooperative, firms produce at the Cournot-Nash equilibrium. It is easy to compute that a firm with cost c_i produces $q_i^c = (2b + \bar{c} - 3c_i)/(6a), \quad i = 1, 2$ at the equilibrium, where $\bar{c}$ is the average cost of the two firms. The expected profit of this firm is $\mathbb{E}[\pi_i|c_i] = (2b + \bar{c} - 3c_i)^2/(36a), \quad i = 1, 2$. Since firms are not colluding, the strategy for the antitrust authority at the equilibrium is not to audit at all.

In the remainder of this section, we study a cooperative Cournot game in two cost structures: fixed marginal costs and random marginal costs.

2.1 Model setup

In this section, we setup the collusion game without and with an antitrust authority.

Two firms compete in a three stage game. In stage 0, nature randomly selects a firm, which will be able to contact the other firm for colluding in stage 1. The firm selected can pay a cost ζ, which is drawn from a distribution $F_\zeta(\cdot)$, to contact the other firm for collusion. In stage 1, each firm draws a marginal cost of production from either a degenerate distribution (c) or a two-point distribution that takes value c_l and c_h with equal probability, where $c_l < c_h$ and denote $\bar{c} = (c_l + c_h)/2 = c$. The costs are private information and independent across firms. In stage 2, firms face an inverse demand function $P = b - aQ$, where Q is the aggregate supply from the two firms, and produce with constant marginal costs. The firms engage in Cournot competition if the selected firm does not contact the other firm in stage 2. Otherwise, firms reveal their marginal costs to each other and bargain over the pair of production with the threat point being the Cournot equilibrium with known costs. We use Kalai bargain with firms having equal bargaining power and assume there is no side payment.

Now we introduce an antitrust authority. It observes the quantities produced by both firms in stage 2 and then decide whether to investigate at a cost ξ. If it investigates, the two firms need to submit data on their marginal costs. Then the antitrust authority calculates the quantities produced in Cournot equilibrium with incomplete information. If any of the observed quantities from the firms falls below $\theta \leq 1$ fraction of the Cournot quantities, it decides there is a collusion and a colluding firm needs to pay a penalty κ. The authority obtains a payoff v if it successfully detects a collusion. We consider the case that θ is close

to 1.

This completes the description of the game. We next solve the model using backward induction.

2.2 Fixed marginal costs

We first consider the game where both firms have the same marginal cost c. One can show that if firms collude, the total quantity is $q^m = (b-c)/(2a)$ and profit of a firm is $\Pi^m(c) = q^m(b - aq^m - c)/2 = (b-c)^2/(4a)$. If firms play Cournot equilibrium, then the total quantity is $q^c = 2(b-c)/(3a)$ and profit of a firm is $\Pi^c(c) = (b-c)^2/(9a)$. When there is no antitrust authority, firms collude if and only if the benefit of collusion is greater than the contact cost, i.e.,

$$\Pi^m(c) - \Pi^c(c) = \frac{(b-c)^2}{4a} - \frac{(b-c)^2}{9a} = \frac{5(b-c)^2}{36a} \geq \zeta.$$

This implies that the collusion probability without an antitrust authority is

$$\mathbb{P}(collude) = F_\zeta(\Pi^m(c) - \Pi^c(c)) = F_\zeta\left(\frac{5(b-c)^2}{36a}\right),$$

Once we introduce the antitrust authority, firms' strategy will change. Because c is degenerate and known to all players, the authority audit if and only if $q < \theta q^c$. Therefore, if firms decide to collude, they either choose $q = \theta q^c$ or q^m. One can show that the profit for a firm from producing $q = \theta q^c$ is $(3-2\theta)\theta(b-c)^2/(9a)$. The quantity q satisfies

$$q = \begin{cases} \theta q^c & \text{if } \frac{(3-2\theta)\theta(b-c)^2}{9a} > \frac{(b-c)^2}{4a} - \kappa \\ q^m & \text{if } \frac{(3-2\theta)\theta(b-c)^2}{9a} < \frac{(b-c)^2}{4a} - \kappa \end{cases}.$$

Note that $\frac{(3-2\theta)\theta(b-c)^2}{9a} < \frac{(b-c)^2}{4a}$ for all $\theta \in (0,1)$. Therefore, if κ is too small, firms choose to collude and produce q^m even with monitoring. In such a case, the benefit from colluding declines from $5(b-c)^2/(36a)$ discussed above to $5(b-c)^2/(36a) - \kappa$, and the colluding probability reduces to

$$\mathbb{P}_1 = F_\zeta\left(\frac{5(b-c)^2}{36a} - \kappa\right).$$

If κ is sufficiently large, firms choose θq^c over q^m and the benefit from colluding is

$$\frac{(3-2\theta)\,\theta\,(b-c)^2}{9a} - \frac{(b-c)^2}{9a} = \frac{(3\theta-2\theta^2-1)\,(b-c)^2}{9a}.$$

The benefit strictly increases in the tolerance parameter θ when $0 < \theta \leq 3/4$ and then decreases. Then, the colluding probability is

$$\mathbb{P}_2 = F_\zeta\left(\frac{(3\theta-2\theta^2-1)\,(b-c)^2}{9a}\right).$$

Based on the behavior of the firms, the equilibrium strategy of the antitrust authority is to audit whenever the quantity is no less than θq^c and not to audit whenever the quantity is greater than θq^c. We summarize the equilibrium in the following proposition.

Proposition 1 *When both firms have marginal cost c, there is an equilibrium where firms collude with probability* $\mathbb{P}_1$ *and each produce* θq^c *for small* κ *and with probability* $\mathbb{P}_2$ *for large* κ *and each produce* q^m*. The equilibrium auditing strategy is*

$$m\,(q) = \begin{cases} 0 & \text{if } q \geq \theta q^c \\ 1 & \text{otherwise} \end{cases}.$$

2.3 Model with random marginal costs

In this section, we consider a collusion game where firms have random marginal costs, i.e., $c_i \in \{c_L, c_H\}$. We first analyze the case without the antitrust authority using backward induction.

Stage 2: There are two possibilities depending on whether the firm selected by nature in stage 1 contacts the other firm in stage 2. If the selected firm does not contact the other firm in stage 2, the two firms engage in Cournot competition without knowing the marginal costs of their opponents. After some algebra, one can show a firm with cost c_i produces $q^c(c_i) = \frac{2b+\bar{c}-3c_i}{6a}, \quad i = l, h$. The expected profit is $\mathbb{E}[\pi_i|c_i] = \frac{(2b+\bar{c}-3c_i)^2}{36a}, \quad i = l, h$. If the selected firm contacts the other firm in stage 2, the two firms jointly determine their quantities (q_1, q_2). They maximize firm 1's surplus subject to firm 2 getting the same surplus:

$$\begin{aligned} \boldsymbol{\pi}(c_1, c_2) &= \max_{q_1, q_2}(b - aq_1 - aq_2 - c_1)q_1 - M(c_1, c_2) \\ \text{st. } (b - aq_1 - aq_2 - c_1)q_1 - M(c_1, c_2) &= (b - aq_1 - aq_2 - c_2)q_2 - M(c_2, c_1), \end{aligned} \tag{1}$$

where $M(c_1, c_2)$ is the profit of a firm in a Cournot game when the firm has a marginal cost

c_1 and its opponent has a marginal cost c_2. It can be shown that

$$M(c_1, c_2) = \frac{(2b + c_2 - 3c_1)^2}{36a}. \tag{2}$$

One can show that the bargaining problem above leads to a unique solution. Let $\boldsymbol{q}^n(c_1, c_2)$ be quantity solved when two firms' costs are c_1 and c_2, respectively, where and $\boldsymbol{\pi}(c_1, c_2)$ be the corresponding profit. Define $\boldsymbol{\pi}(c_1, c_2)$ as the corresponding profit. Notice that if $c_1 = c_2 = c$, then bargaining achieves the monopoly outcome and the quantity produced by a firm is $\boldsymbol{q}^n(c, c) = (b - c)/(4a)$, and a firm's profit is $\boldsymbol{\pi}^n(c, c) = (b - c)^2/(4a)$.

Stage 1: Only the firm selected in the first stage needs to make a decision. The expected profit from colluding is

$$\Pi^n(c) = \frac{\boldsymbol{\pi}^n(c, c_l) + \boldsymbol{\pi}^n(c, c_h)}{2}. \tag{3}$$

Let $\Pi^c(c)$ be the profit under Cournot competition

$$\Pi^c(c) = \frac{(2b + \bar{c} - 3c)^2}{36a}. \tag{4}$$

The firm contacts the other firm if $\zeta < \Pi^c(c) - \Pi^n(c)$, which implies a contact probability of $F_\zeta(\Pi^c(c) - \Pi^n(c))$.

We summarize the findings above in the following proposition.

Proposition 2 *When firms' marginal costs are randomly drawn from $\{c_L, c_H\}$ with equal probability and there is no antitrust authority. There exists a unique equilibrium such that if costs are (c_1, c_2), then*

1. *Firms produce $(q^c(c_1), q^c(c_2))$ if they do not collude. They produce $(\boldsymbol{q}^n(c_1, c_2), \boldsymbol{q}^n(c_2, c_1))$ if they collude.*

2. *Firms collude with probability*

$$\mathbf{P}^n(c_1, c_2) = \frac{F_\zeta(\Pi^c(c_1) - \Pi^n(c_1)) + F_\zeta(\Pi^c(c_2) - \Pi^n(c_2))}{2}. \tag{5}$$

It is obvious that if there is no contact cost, i.e., $\zeta = 0$, two firms would collude for sure and produce $(\boldsymbol{q}^n(c_1, c_2), \boldsymbol{q}^n(c_2, c_1))$.

2.3.1 Model with an Antitrust Authority

Now we introduce the antitrust authority and analyze the model equilibrium. Depending on the costs of two firms, we summarize all the scenarios below.

- If cost are (c_l, c_l), firms produce (q_l^c, q_l^c) if none of the firms contacts the other firm for collusion. If one of the firms contacts the other, they produce $(\theta q_l^c, \theta q_l^c)$, (q_h^c, q_h^c) and $(\theta q_h^c, \theta q_h^c)$, which yield the same profit.[3]

- If costs are (c_h, c_l), firms produce (q_h^c, q_l^c) if collusion is not successful. If collusion is successful, they produce $(\theta q_h^c, \theta q_l^c)$. If costs are (c_l, c_h), the results are similar.

- If costs are (c_h, c_h), they produce (q_h^c, q_h^c) when collusion fails and produce $(\theta q_h^c, \theta q_h^c)$.

Next, we derive the probability of collusion in all the cases above.

If a firm has cost c_l, the expected profit from Cournot competition is

$$\Pi^c(c_l) = \frac{(2b + \bar{c} - 3c_l)^2}{36a}. \tag{6}$$

If the firm contacts the other firm, the profit is as follows.

$$\Pi^n(c_l) = \frac{\theta(2b + \bar{c} - 3c_l)\left[b - 2\theta(b - \bar{c})/3 - c_l - \theta(\bar{c} - c_l)/2\right]}{6a}. \tag{7}$$

Notice that $\Pi^n(c_l) > \Pi^c(c_l)$ for all θ close to 1 if $b > \bar{c}$. Then the firm contacts the other firm if and only if $\zeta < \Pi^n(c_l) - \Pi^c(c_l)$, which implies that the probability of contacting is $\mathbf{P}^r(c_l) = F_\zeta\left(\Pi^n(c_l) - \Pi^c(c_l)\right)$.

Similarly, if the firm's cost is c_h, its profit under Cournot competition is

$$\Pi^c(c_h) = \frac{(2b + \bar{c} - 3c_h)^2}{36a}. \tag{8}$$

If it contacts the other firm, the profit is

$$\Pi^n(c_h) = \frac{\theta(2b + \bar{c} - 3c_h)\left[b - 2\theta(b - \bar{c})/3 - c_h - \theta(\bar{c} - c_h)/2\right]}{6a}. \tag{9}$$

Then the probability of contacting is $\mathbf{P}^r(c_h) = F_\zeta\left(\Pi^n(c_h) - \Pi^c(c_h)\right)$.

Based on the probability of contact, we can obtain the probability of collusion, denoted by $\mathbf{P}^n(\cdot, \cdot)$, for each pair of costs. After some derivation, one can show $\mathbf{P}^n(c_l, c_l) = \mathbf{P}^r(c_l)$, $\mathbf{P}^n(c_l, c_h) = \mathbf{P}^n(c_h, c_l) = (\mathbf{P}^r(c_l) + \mathbf{P}^r(c_h))/2$, and $\mathbf{P}^n(c_h, c_h) = \mathbf{P}^r(c_h)$.

[3]We focus on the case where firms do not collude on (q_h^c, q_l^c), (q_l^c, q_h^c), $(\theta q_h^c, \theta q_l^c)$ and $(\theta q_l^c, \theta q_h^c)$. This occurs if θ is sufficiently close to 1 and q_h^c is not significantly higher than q_l^c.

Lastly, we derive the distribution of the quantities produced if the costs are (c_l, c_l). Let $\gamma(x, y)$ be the probability that firm 1 produces x and firm 2 produces y. Then $\gamma(x, y) > 0$ only if $(x, y) = (\theta q_l^c, \theta q_l^c)$, (q_h^c, q_h^c) and $(\theta q_h^c, \theta q_h^c)$. To make the authority indifferent between monitoring or not if the quantities produced are (q_h^c, q_h^c),

$$\frac{\mathbf{P}^n(c_l, c_l)\gamma(q_h^c, q_h^c)}{\mathbf{P}^n(c_l, c_l)\gamma(q_h^c, q_h^c) + 1 - \mathbf{P}^n(c_h, c_h)} v - \xi = 0, \tag{10}$$

which implies

$$\gamma(q_h^c, q_h^c) = \frac{(1 - \mathbf{P}^n(c_h, c_h))\xi}{\mathbf{P}^n(c_l, c_l)(v - \xi)}. \tag{11}$$

Similarly, to make the authority indifferent between monitoring or not if the quantities produced are $(\theta q_h^c, \theta q_h^c)$,

$$\frac{\mathbf{P}^n(c_l, c_l)\gamma(\theta q_h^c, \theta q_h^c)}{\mathbf{P}^n(c_l, c_l)\gamma(\theta q_h^c, \theta q_h^c) + \mathbf{P}^n(c_h, c_h)} v - \xi = 0, \tag{12}$$

which implies

$$\gamma(\theta q_h^c, \theta q_h^c) = \frac{\mathbf{P}^n(c_h, c_h)\xi}{\mathbf{P}^n(c_l, c_l)(v - \xi)}. \tag{13}$$

To ensure $\gamma(\theta q_h^c, \theta q_h^c) + \gamma(q_h^c, q_h^c) < 1$, we need v/ξ to be sufficiently large (need simulations with $v/\xi > 2$)

Next, we derive the monitoring probability of the authority. We consider the following

$$m(q_1, q_2) = \begin{cases} 0 & \text{if } q_1 \geq \theta q_l^c \text{ and } q_2 \geq \theta q_l^c \\ \in (0, 1) & \text{if } (q_1, q_2) = (q_h^c, q_h^c) \text{ or } (\theta q_h^c, \theta q_h^c) \\ 1 & \text{otherwise} \end{cases}. \tag{14}$$

If the costs are (c_l, c_l), the firms are indifferent between $(\theta q_l^c, \theta q_l^c)$, (q_h^c, q_h^c) and $(\theta q_h^c, \theta q_h^c)$. This implies

$$\begin{aligned} (b - 2a\theta q_h^c - c_l)\, \theta q_h^c - m\,(\theta q_h^c, \theta q_h^c)\, \kappa &= (b - 2a\theta q_l^c - c_l)\, \theta q_l^c, \\ (b - 2a q_h^c - c_l)\, q_h^c - m\,(q_h^c, q_h^c)\, \kappa &= (b - 2a\theta q_l^c - c_l)\, \theta q_l^c. \end{aligned}$$

The monitoring probabilities are

$$m\left(\theta q_h^c, \theta q_h^c\right) = \frac{\left(b - 2a\theta q_h^c - c_l\right)\theta q_h^c - \left(b - 2a\theta q_l^c - c_l\right)\theta q_l^c}{\kappa}, \tag{15}$$

$$m\left(q_h^c, q_h^c\right) = \frac{\left(b - 2aq_h^c - c_l\right) q_h^c - \left(b - 2a\theta q_l^c - c_l\right)\theta q_l^c}{\kappa}. \tag{16}$$

Proposition 3 *Suppose that (1) θ is sufficiently close to 1, (2) v/ξ is sufficiently large, and (3) $b > \bar{c}$. There exists an equilibrium that satisfies the following. Firms produce Cournot quantity if they do not collude. They collude with probability probability $\mathbf{P}^n(c_i, c_j)$ if the costs are c_i and c_j, $i = l, h$ and $j = l, h$. If they collude, the following holds.*

1. *If the costs are (c_l, c_l), they randomize between $(\theta q_l^c, \theta q_l^c)$, (q_h^c, q_h^c) and $(\theta q_h^c, \theta q_h^c)$ with probability $1 - \gamma(q_h^c, q_h^c) - \gamma(\theta q_h^c, \theta q_h^c)$, $\gamma(q_h^c, q_h^c)$ and $\gamma(\theta q_h^c, \theta q_h^c)$, respectively.*

2. *If costs are (c_i, c_j), where $i = l, h$, $j = l, h$ and $i \neq j$, then firms produce $(\theta q_i^c, \theta q_j^c)$.*

The authority audits with probability 0 *if the quantities are $(\theta q_h^c, \theta q_l^c)$ or $(\theta q_l^c, \theta q_h^c)$, or $(\theta q_l^c, \theta q_l^c)$. It audits with probabilities $m(\theta q_h^c, \theta q_h^c)$ and $m(q_h^c, q_h^c)$ if the quantities are $(\theta q_h^c, \theta q_h^c)$ and (q_h^c, q_h^c), respectively. It audits with probability 1 for other quantities.*

If ζ is degenerate with a value 0, the two firms collude with probability 1 regardless of their costs. Then firms do not produce q_h^c, q_h^c. To see this, note that if their costs are (c_h, c_h), producing $(\theta q_h^c, \theta q_h^c)$ always leads to a higher profit than q_h^c, q_h^c. If the costs are (c_h, c_l), they produce $(\theta q_h^c, \theta q_l^c)$ or $(\theta q_l^c, \theta q_h^c)$. If costs are (c_l, c_l), they randomize between $(\theta q_l^c, \theta q_l^c)$ and $(\theta q_h^c, \theta q_h^c)$. To make the authority indifferent between investigating or not, $\gamma(\theta q_h^c, \theta q_h^c) = \xi/(v - \xi)$.

3 Q-learning

In this section, we present a multi-agent reinforcement learning (MARL) approach of the auditing game discussed in Section 2. Reinforcement learning is one of the three basic machine learning paradigms, alongside supervised learning and unsupervised learning. It is concerned with how an gent takes actions to maximize the total reward in a Markov Decision Process (MDP) by learning the environment from her own past experiences. When multiple agents interact and employ reinforcement learning, it is referred to as multi-agent reinforcement learning.

The specific reinforcement learning algorithm we use is Q-learning, which is motivated by the dynamic programming problem in a MDP (Watkins (1989) and Sutton et al. (1998)). Q-learning allows an agent to learn the optimal policy with little knowledge of the underlying

environment, and its convergence is guaranteed if all actions are repeatedly sampled in all states and the action-values are represented discretely (Watkins and Dayan (1992)).

3.1 Single Agent Problems

Consider an unknown stationary Markov Decision process faced by a single agent. In each period $t = 0, 1, 2, \ldots$, the agent observes a state $s_t \in \mathcal{S}$ and then chooses an action $a_t \in \mathcal{A}$. Both the state space $\mathcal{S}$ and action space $\mathcal{A}$ are finite and time-invariant, and $\mathcal{A}$ is state-independent. The payoff received by the agent is $\pi_t = \pi(s_t, a_t)$, which could be random, then the system moves from state s_t to the next state $s_{t+1} \in \mathcal{S}$.

Let $a^*(s)$ represent an optimal policy, which is a mapping from the state space $\mathcal{S}$ to the action space $\mathcal{A}$ that maximizes the expected present value of discounted payoff

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t \pi_t\right] = \mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t \pi(s_t, a^*(s_t))\right], \tag{17}$$

where $\delta < 1$ represents the discount factor. Let $V(s)$ be the value in state s

$$V(s) = \max_{a \in A} \Big\{ \mathbb{E}[\pi|s, a] + \delta\, \mathbb{E}[V(s')|s, a] \Big\}, \tag{18}$$

which is the maximum discounted payoff in state s, and $Q(s, a)$ be the choice-specific value function

$$Q(s, a) = \mathbb{E}[\pi|s, a] + \delta\, \mathbb{E}\left[\max_{a' \in A} Q(s', a')|s, a\right], \tag{19}$$

which represent the expected discounted payoff of taking action a at state s and choose optimal policy function $a^*(s)$ in the future. Notice that Q-function is related to the value function by $V(s) = \max_{a \in A} Q(s, a)$.

Because both $\mathcal{A}$ and $\mathcal{S}$ are finite, Q-function is simply a matrix. If the Q-matrix were known *ex ante*, the optimization problem could be solved by searching the maximizer of the specific row of Q-matrix corresponding to state s, or

$$a^*(s) = \arg\max_{a \in A} Q(s, a). \tag{20}$$

Therefore, as long as the Q-matrix is known, without knowing any underlying model, the agent is able to solve the optimization problem.

However, the Q-matrix is unknown. The idea of Q-learning is to estimate the Q-matrix using the following iterative procedure. Starting from an arbitrary initial matrix Q_0, the

algorithm chooses an action a_t in state s_t for each time period. After observing the payoff π_t, the algorithm updates one cell of Q-matrix according to the following learning rule while keeping other cells $s_t \neq s$ and $a_t \neq a$ unchanged:

$$Q_{t+1}(s,a) = (1-\alpha)Q_t(s,a) + \alpha\left[\pi_t + \delta \max_{a\in A} Q_t(s',a)\right], \tag{21}$$

where $Q_t(s,a)$ is the "un-updated" element of the Q-matrix, the learning rate parameter $\alpha \in (0,1)$ captures the weight the agent puts on "new information" $[\pi_t + \delta \max_{a\in A} Q_t(s',a)]$. Since α is a constant, "new information" is weighted equally in each period but the weight on any given piece of "old information" decays across time.

The agent may stuck to a suboptimal policy by applying the algorithm above. For example, assume that a^* is the unique maxima in state s, such that for any action $a' \neq a^*$, we have $Q(s,a') < Q(s,a^*)$. If the initial Q-matrix Q_0 is chosen such at $Q_0(s,a^*) < Q_0(s,a')$, then it is possible that the algorithm only updates $Q_t(s,a')$ rather than $Q_t(s,a^*)$. In such a scenario, the algorithm will stuck at a' and never learn that a^* is the optimal action in state s. To avoid such an issue and to estimate a^* and Q-matrix starting from an arbitrary initial matrix Q_0, the algorithm is allowed to "make mistakes", or to explore non-optimal actions. The method we use in our analysis is the ε-greedy model. The idea is that the algorithm exploits (chooses the currently optimal action) with probability $1-\varepsilon_t$ and to explore (randomize uniformly across all actions) with probability ε_t in period t. The probability ε_t decays with time and is assumed to be $\varepsilon = e^{-\beta t}$, with $\beta > 0$. The algorithm is characterized by the couple (α, β).

3.2 MARL: Quantity competition with auditing

We now present our MARL approach to describing the quantity competition and auditing modeled in Section 2.

In a market with $n = 2$ firms producing a homogeneous product, firm i sets the optimal quantity q_i to maximize its profit given its marginal cost c_i and the demand $D(p)$. The optimization problem of firm i is solved by using the Q-learning algorithm discussed above.

An antitrust authority j monitors the market quantities (prices) to detect colluding behavior with it objective being maximizing consumer welfare. When all the quantities (or those quantities corresponding to a substantial portion of demand) in the market reach a pre-set threshold, the authority conducts an investigation into the possible collusion. Specifically, the antitrust authority occurs a cost $\xi_a > 0$ for auditing. Firms submit their marginal costs to the authority upon auditing and the authority learns the equilibrium (Cournot) quantities

(q_1^e, q_2^e). The authority claims that firm i is colluding if $q_i \leq \theta q_i^e, i = 1, 2$ and the payoff of the authority is $v_a > \xi_a$. If $q_i > \theta q_i^e$ for some i, then firm i is innocent and the authority's penalty is ζ_a. The payoff of the authority is 0 without auditing. The antitrust authority observes $q_{i,t-1}$ and $q_{i,t}, \forall i$.[4]

If a firm is audited, a compliance cost ξ_f occurs. If it is found colluding, there will be a cost (including penalty and other losses, e.g., reputation cost) κ_f which can be extended to be a function of q_i^e/q_i for firm i. If there is no auditing, the payoff function is the same as in the case without the antitrust authority. In sum, the firm's cost is $\xi_f + \kappa_f \cdot \mathbb{1}(q_i \leq \theta q_i^e)$.

The timing and information structure of the game are as follows. At the beginning of period t, each firm observes both firms' quantities and the authority's auditing decision in period $t-1$, and its own marginal cost in period t, then chooses its quantity in period t. The authority observes the quantities in period t and $t-1$, as well as its auditing decision in period $t-1$ to decide whether to audit at time period t. All the benefit, costs, payoffs, and penalty are realized at the end of t.

Firms and the authority all use a Q-learning algorithm. The Q-function of the antitrust authority is

$$\begin{aligned} Q_a(s_a, d) &= d \cdot \big(-\xi_a + \mathbb{E}[\mathbb{1}\left\{\min(q^i/q_i^e) \leq \theta\right\} \cdot v_a - \mathbb{1}\left\{\min(q^i/q_i^e) \geq \theta\right\} \cdot \zeta_a | q_{t-1}, q_t]\big) \\ &+ \delta \mathbb{E}[\max_{d'} Q_a(d', s') | d, s_a] \\ &= d \cdot \left\{ \mathbb{E}[\mathbb{1}\left\{\min(q_i/q_i^e) \leq \theta\right\} \cdot v_a - \left(\mathbb{1}\left\{\min(q_i/q_i^e) \geq \theta\right\} \cdot \zeta_a + \xi_a\right) | q_{t-1}, q_t] \right\} \\ &+ \delta \mathbb{E}[\max_{d'} Q_a(d', s') | d, s_a], \end{aligned} \tag{22}$$

where $d \in \{0, 1\}$ is the authority's binary auditing decision, δ is the discount factor, $s_a \equiv (q_{1,t-1}, q_{2,t-1}, q_{1t}, q_{2t}, o_{t-1})$ is the state variable, and o_{t-1} is the authority's auditing decision defined as follows.

$$o_{t-1} = \begin{cases} 0, \text{ if } d_{t-1} = 0 \\ 1, \text{ if } d_{t-1} = 1, \ \min_i(q_i/q_i^e) \geq \theta \\ 2, \text{ if } d_{t-1} = 1, \ q_i/q_i^e \leq \theta, \ q_j/q_j^e \geq \theta \\ 3, \text{ if } d_{t-1} = 1, \ q_i/q_i^e \geq \theta, \ q_j/q_j^e \leq \theta \\ 4, \text{ if } d_{t-1} = 1, \ \max_i(q_i/q_i^e) \leq \theta. \end{cases} \tag{23}$$

[4]Alternatively, we can assume only q_{t-1} and q_t are observed by the authority. In such case, firms are required to submit their quantities in addition to their marginal costs if an audit occurs.

Firm i's Q-function is

$$\begin{aligned} Q_i(s, q_i) &= \mathbb{E}[(p(q_i, q_j) - c_i) \cdot q_i - d \cdot (\xi_f + \kappa_f \cdot \mathbb{1}(q_i/q_i^e \leq \theta))|q_{t-1}, q_i, c_i, o_{t-1}] \\ &+ \delta \mathbb{E}[\max_{q_i'} Q_i(q_i', s')|q_i, c_i, s_f], \end{aligned} \tag{24}$$

where the state variable of the firm is $s_f \equiv (q_{1,t-1}, q_{2,t-1}, o_{t-1})$.

It worth noting that in the learning process above, a firm has no knowledge about its rival's cost, even in the case where both firms have the same and constant marginal cost. The antitrust authority learns firms' marginal costs from a subpoena whenever it makes an auditing decision. However, the authority needs to continue to learn firms cost because it has no knowledge of firms' cost structure at all.

As is well known, there is no theoretical guarantee of convergence for a MARL approach. This is because in MARL, each agent's learning makes the "environment" non-stationary for other agents, so the usual single-agent convergence arguments fail. Nevertheless, in our simulation studies we almost always experience convergence, as in some other studies, e.g., Calvano et al. (2020b) and Johnson et al. (2023).

3.3 Simulation setup

In our simulation experiments, we specify the demand as a linear function $p = p_0 - \sum_{i=1}^{n} q_i$ with $p_0 = 2$. The average marginal cost for both firms is $\bar{c} = 1$. We consider two cases of costs: (1) marginal cost is $c = 1$ for both firms, and (2) both firms' marginal costs are either $c_l = 0.75$ or $c_h = 1.25$ with equal probability. In setting (1), firms produce $q = 1/3$ in the static Cournot equilibrium. In setting (2), firms produce $q_l = 11/24$ and $q_h = 5/24$ corresponding to c_l and c_h, respectively. We choose the action (quantity) set to be $n_a = 15$ equally spaced points on $\mathcal{A} = [2/15, 29/60]$ that contains the quantities in the static Cournot equilibrium as interior points.

The two firms and the authority have discount factor $\delta = 0.95$, learning parameter $\alpha = 0.05$, the greedy index is $\beta = 5 \times 10^{-6}$, and the greedy parameter, i.e., the probability of exploration is $\epsilon_t = e^{-\beta t}$. To simplify our analysis by keeping those more essential parameters, we set both the compliance cost of firms ξ_f and the penalty to the authority ζ_a to be zero. The authority's auditing cost is $\xi_a = 0.05$, its payoff from successfully auditing a collusion is $v_a \in \{0.05, 0.1, 0.2\}$ and firm's loss $\kappa_f = 0.05$ if they are audited to be collusive. The auditing threshold is $\theta = 1.0$.

In the simulation experiments, we consider two setups: without antitrust authority and with antitrust authority. In each of the setups, we simulate the game under two cost structures. In each of the four scenarios above, we vary the authority's incentive relative to firm's

colluding cost, which is measure by the ratio $v_a/\xi_a \in \{1, 2, 4\}$. In all the 12 cases above, we simulate firms' quantities and payoffs, market price, the authority's probability of auditing and its payoff, consumer surplus, and social welfare. For each time period, we simulate 100 times and present the average of the results across 100 simulations. Convergence is achieved if an optimal choice does not change over 10^5 periods. There is no convergence if an optimal choice keeps changing up to 10^8 periods. For quantities and auditing probabilities, we present their convergence process. In addition, we also illustrate auditing probabilities at convergence for different cost combinations using heatmaps. For firms' prices, payoffs, the authority's payoff, consumer surplus, and social welfare, we present the results at convergence. Specifically, the payoff, consumer surplus, and social welfare are their present values.

4 Simulation Results

In this section, we present our baseline results of simulation. The results without an antitrust authority are in the first subsection and the ones with an antitrust authority are in the second subsection.

We simulate 1000 times for all the settings and present the average of the results across 1000 times. When we present the convergence process of simulations, each data point is the average of the first 100 periods for every 10^5 periods. For results at convergence, we take the average of the last 100 periods after convergence.

4.1 Without an antitrust authority

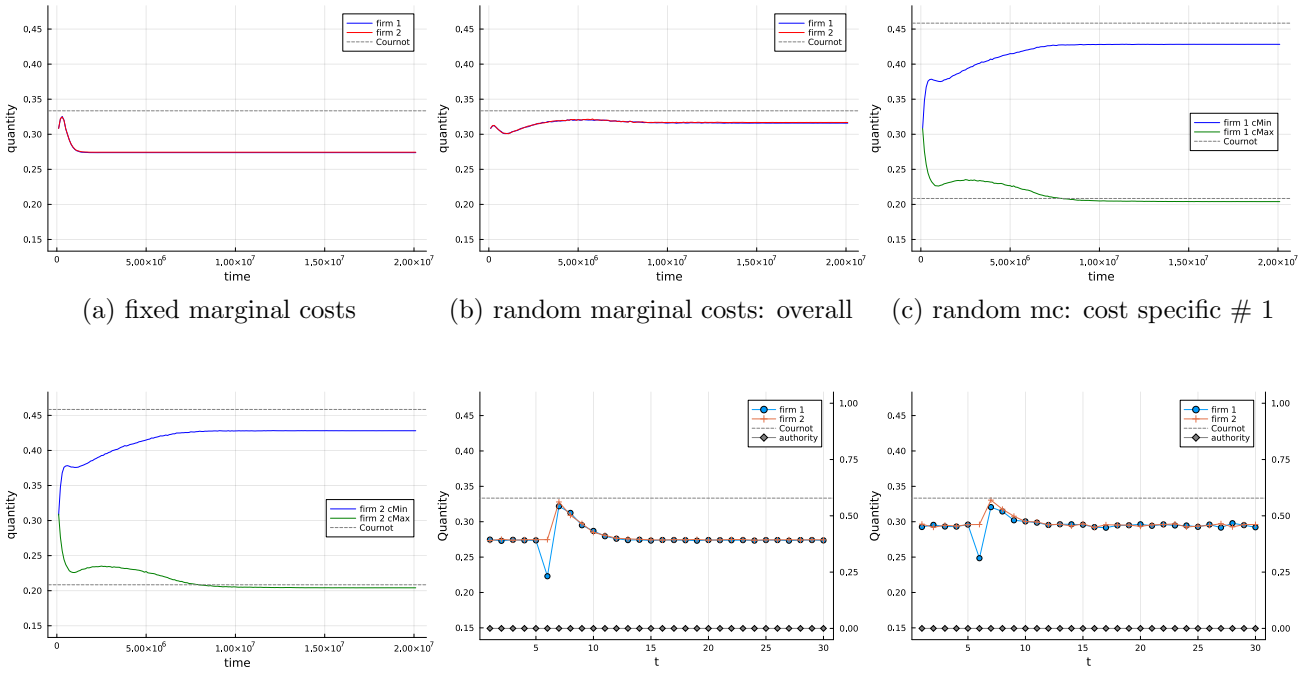
In this section, we present the simulation results when there is no auditing. In these simulations, we investigate firms behavior and its dependence on cost structure.

We first present the convergence process of quantities in Figure 1 . Subplot (a) is for the scenario where both firms have the constant marginal cost $c_i = 1$. The subplot illustrates that both firms never produce at or above Cournot equilibrium. They produce relatively high at the beginning but quickly collaborate to produce 0.274, which is 18% lower than the quantity at the Cournot equilibrium, but still higher than the monopoly quantity 0.25. When firms' costs are random, i.e., taking value $c_l = 0.75$ or $c_h = 1.25$ with equal probability, as shown in subplot (b), the two firms also learn to collude produce lower than Cournot equilibrium. However, it takes longer for them to collude, and more importantly, the quantity at the collusion level is 0.316, which is only 5% lower than the Cournot equilibrium. A comparison of subplots (a) with (b) shows that randomness of marginal costs is crucial in determining

the outcome of algorithmic collusion.

To further explore the dependence of firms' learning process on the cost structures, we provide subplots (c)- (f) in Figure 1. In subplot (c), we collect all the periods where firm 1's cost is $c_h = 1.25$ (cMax) or $c_l = 0.75$ (cMin) and plot firm i's corresponding quantities. The higher and lower dash lines are Cournot quantities $q_l^c = 11/24$ and $q_h^c = 5/24$ for cost c_h and c_l, respectively. Subplot (d) presents the corresponding results for firm 2. Due to the symmetry of the two firms, subplots (c) and (d) are almost the same. The two subplots illustrate how firms learn asymmetrically when they receive a draw of higher or lower marginal cost. The two learning curves converge to 0.430 and 0.204, which are 6% and 2% lower than the corresponding Cournot equilibrium.

Figure 1: Firm's quantities without auditing



(a) fixed marginal costs (b) random marginal costs: overall (c) random mc: cost specific # 1

(d) random mc: cost specific # 2 (e) fixed mc: deviation (f) random mc: deviation

In subplots (e)-(f), we illustrate how a firm respond to its opponent's deviation from the learning process in the two cost structures. Specifically, focusing on the quantities after convergence, we let firm 1's quantity drop two grid points from its current quantity, and simulate the quantities for the two firms. From subplot (e), we find that when costs are fixed, firm 2 will respond to firm 1's cut of quantity by increasing its own quantity. Firm 1 responds back immediately in the next period by following firm 2 to produce a similar quantity. After that, both firms basically produce the same quantity during the process of

Table 1: Price, payoff, and welfare at convergence

			Payoff			Payoff at Cournot		Welfare	
cost	v_a/ξ_a	price	firm 1	firm 2	authority	firm 1	firm 2	CS	TS
Panel A: without auditing									
fixed	–	1.452	2.464	2.469	–	2.222	2.222	3.016	7.948
random	–	1.368	2.592	2.607	–	2.522	2.537	4.264	9.463
Panel B: with auditing									
fixed	1	1.319	2.159	2.158	0.000	2.222	2.222	4.643	8.961
fixed	2	1.332	2.219	2.219	0.000	2.222	2.222	4.457	8.894
fixed	4	1.332	2.217	2.217	0.000	2.222	2.222	4.464	8.898
random	1	1.338	2.526	2.522	0.000	2.547	2.544	4.597	9.646
random	2	1.301	2.394	2.381	0.029	2.540	2.528	5.174	9.978
random	4	1.301	2.337	2.355	0.141	2.524	2.539	5.176	10.008

Notes: All the results are average of last 100 periods after convergence.

learning to the quantity at convergence. By comparison, subplot (f) tells a different story. Firm 2's quantity rises as a response to the sudden drop of firm 1's quantity. However, firm 1 does not produce the same quantity as firm 2 even though its quantity increases until five periods later. It takes more periods for the two firms return to the convergence process when costs are random. This pattern is due to the fact that it is easier for firms to learn to cooperate when there is no cost randomness. When costs are random, it is difficult for a firm to identify its rival's deviation is due to a shock to cost or non-cooperating deviation. Therefore, a punishment is less likely to be implemented effectively.

Panel A of Table 1 presents price, present value of payoff for the two firms, of consumer surplus and social welfare at convergence under the two cost structures (fixed cost and random costs). The price under fixed cost is 6% higher than that under the random costs because the converged quantity in the former case is smaller as illustrated in subplots (a)-(b) in Figure 1. Interestingly, the expected payoff is larger when cost is random. This is because when cost is c_L the benefit of algorithmic collusion is very large as the quality is much lower than Cournot $q_l^c = 11/24$, as showing in subplots (c)-(d). On average, firms are better off in the case of random costs. Not surprisingly, both the consumer surplus and total surplus are also higher in the case of random costs.

4.2 With an antitrust authority

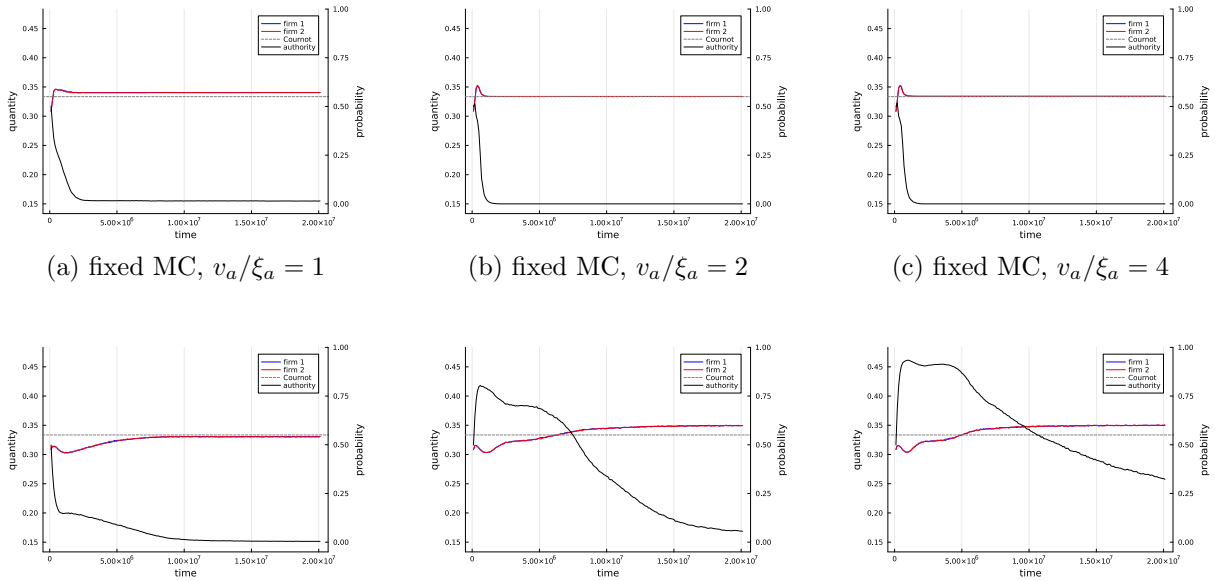
In this subsection, we present the simulation results when an antitrust authority joins the game. We first report the convergence process of firms' quantities and the authority's auditing probability, then discuss the auditing probability at the convergence. For all the simulations in this section, we consider all the five values of o_{t-1} in the state variable $s_a \equiv (q_{1,t-1}, q_{2,t-1}, q_{1t}, q_{2t}, o_{t-1})$. The results are similar for all the five values, we only present the results of $o_{t-1} = 0$ for simplicity.

4.2.1 The convergence process of quantity and auditing probability

We first present the convergence process of firms' quantities and the authority's auditing probability in Figure 2, where subplots (a)-(c) are for fixed marginal costs with the authority's benefit-cost ratio being 1, 2, and 4, respectively. Subplots (d)-(f) are corresponding results for random marginal costs.

In both Figure 2 with Figure 1, even though the initial values are exactly the same, it is obvious that the authority's auditing effectively improves the quantity to the Cournot level or above. The only exception is when marginal costs are random and the authority's benefit-cost ratio is 1. An explanation of this exception will be provided later. When

Figure 2: Quantities and auditing probabilities: convergence process



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$

(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

marginal cost is fixed, the authority's auditing is successful regardless the incentive of the authority. The firms first produce below the Cournot quantity and the authority audits with a beginning probability around 50%. As a response, firms quickly adjust their quantities above the Cournot quantity, then gradually learn to produce at the Cournot level. Note that the quantities in subplot (a) does not converge to the Cournot level. A possible cause is that when the incentive of the authority is low, the starting auditing probability is relatively lower. Firms have no enough data to learn about the authority's auditing rule, leading to higher quantity than the Cournot level. This is evidenced by the fact that it takes longer for the quantity to converge in subplot (a) than in (b) and (c).

By contrast, when marginal costs are random, the pattern is different. In subplot (d), the authority's incentive is low, the auditing is not successful: even though firms increase quantities as a consequence of the relative higher auditing probability at the beginning, the converged quantity is 0.331, still lower than the Cournot level 1/3. As the benefit-cost ratio increases to 2 and 4, however, the authority's highest auditing probability increases from 50% significantly to 80% and 90%, respectively. The firms respond to the auditing by increasing quantities significantly, which converge to 0.350 and 0.349, respectively, both are higher than the Cournot quantity. As the quantity increases, the authority learns to audit with smaller probabilities.

An interesting pattern is that firms fail to learn to produce the Cournot quantity when the authority's benefit-cost ratio is 2 or 4. Instead, the quantity at convergence is higher than Cournot. To further investigate the causes, we plot in Figure 3 the convergence process of firms' quantities and the authority's auditing probabilities when firms' marginal costs are c_l and c_h. Because the two firms are symmetric, the patterns are the same, and we only present the results for firm 1. It is evidenced from the figure that the authority audits the low quantity with a higher probability and the higher average quantity at convergence is mainly due to the larger quantity when firm's marginal cost is c_h.

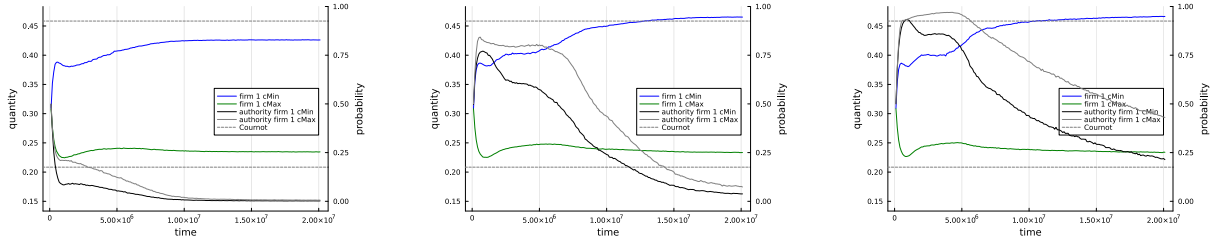
This is consistent with the results in Proposition 3: when firm 1 produces the lower quantity q_h^c, then the two firms may produce (q_h^c, q_h^c) or (q_h^c, q_l^c) where the quantities (q_h^c, q_h^c) will be audited with a positive probability. On the other hand, if firm 1 produce the higher quantity q_l^c, then two firms may produce (q_l^c, q_h^c) or (q_l^c, q_l^c) and there will be no auditing at the equilibrium. Overall, the low quantity is audited and the high quantity is not at the equilibrium. The authority learns to that direction, even though the auditing probability for c_l is still greater than zero after the convergence. As a result of the authority's auditing behavior, firms produce at much higher level than q_h^c when the cost is c_h, while producing slightly higher than q_l^c when the cost is c_l. The figure also reveals that as the larger incentive of the authority leads to a larger difference of auditing probabilities for low and high quantities.

In summary, the results in the Figure demonstrate that the behavior of firms is mainly consistent with the equilibrium at which firms communicate to cooperate.

4.2.2 Auditing behavior at convergence

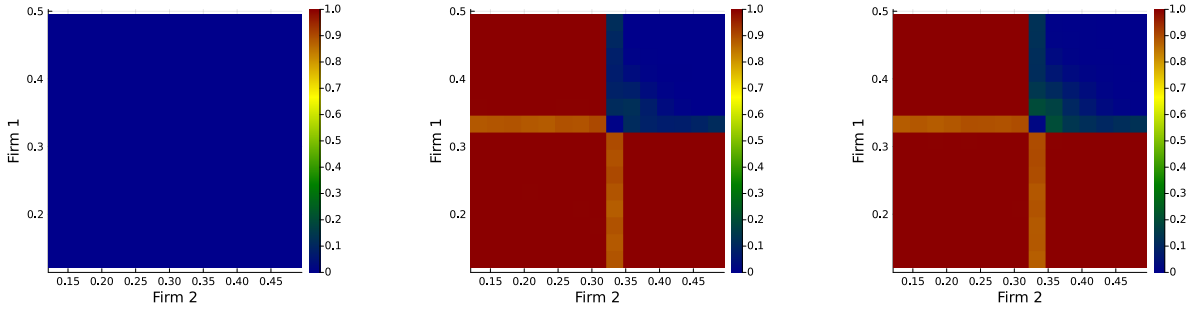
We present the antitrust authority's auditing behavior at convergence using heatmaps in Figure 4. The heatmaps are employed to illustrate the antitrust authority's policy function at convergence, i.e., the dependence of the auditing decision on the state variable $s_a \equiv (q_{1t}, q_{2t}, q_{1,t-1}, q_{2,t-1}, o_{t-1})$. Let the antitrust authority's auditing decision conditional on the

Figure 3: Quantities and auditing probabilities: cost specific

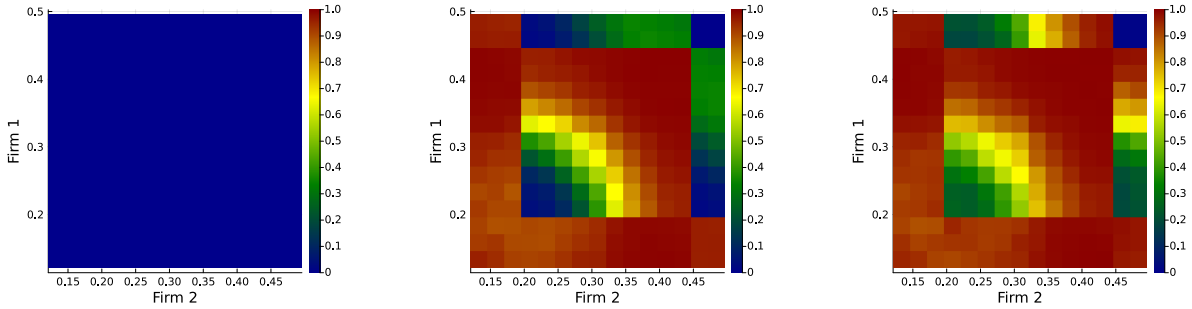


(a) firm 1, $v_a/\xi_a = 1$ (b) firm 1, $v_a/\xi_a = 2$ (c) firm 1, $v_a/\xi_a = 4$

Figure 4: Auditing probabilities at convergence



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$



(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

state variable at convergence is $d(i,j,k,l) \in \{0,1\}$, where $i,j,k,l \in \{1,2,\cdots,15\}$ are the i-th,j-th, k-th, and l-th grid points on the quantity support $\mathcal{A}$, the state variable $o_{t-1} = 0$ is dropped for simplicity. $d(i,j,k,l)$ can be computed using the learned Q-matrix at convergence. To get the auditing decision conditional on the quantities at the current period, i.e., $d(i,j)$, we use a weighted average over all the possible values of (k,l), i.e., the quantities of the previous period, as follows.

$$w_{ijkl} = \frac{\sum_t \mathbb{1}\{q_{1t} = i, q_{2t} = j, q_{1,t-1} = k, q_{2,t-1} = l, o_{t-1} = 0\}}{\sum_t \mathbb{1}\{q_{1t} = i, q_{2t} = j, o_{t-1} = 0\}}, \tag{25}$$

where the summation is over all the history of convergence process. The auditing decision conditional on firms' quantity (i,j) is

$$d(i,j) = \sum_{k,l} w_{ijkl} d(i,j,k,l). \tag{26}$$

Note that the auditing decision $d(i,j,k,l) \in \{0,1\}$, then $d(i,j) \in [0,1]$ measures the probability of auditing by the authority when the quantities are the (i,j)-th grid point.

In Figure 4, we plot $d(i,j)$ for $i,j \in \{1,2,\cdots,15\}$, where we transfer (i,j)-th grid point to quantity values on $\mathcal{A}$. The first row shows heatmaps for the fixed-cost setting at three benefit–cost ratios, $v_a/\xi_a \in \{1,2,4\}$. The second row reports results for the random cost setting. The first observation from the figure is that the auditing behavior depends crucially on the authority's incentives under both cost structures. Under both cost structures, auditing probability is zero when the authority's benefit equals its cost ($v_a/\xi_a = 1$). Increasing the benefit–cost ratio from one to two leads to a significant change in auditing behavior; however, a further increase of the ratio from two to four produces little additional change.

Under the fixed marginal costs, when $v_a/\xi_a > 1$ the auditing probability is close to zero when both firms' quantities are higher than the Cournot quantity, which is $q^c = 1/3$ (the boundary cell of cold and warm colors is $(1/3, 1/3)$), and the auditing probability is one whenever one of the firm produces more than $1/3$. Such an auditing pattern is consistent with the results in Proposition 1—the authority does not audit when both firms produces at least q^c. Otherwise, the authority audits with probability one. It worth noting that when one of the firms produce at the Cournot quantity, the auditing probabilities slightly deviate from the pattern above. For instance, when firm 1 produces at the Cournot quantity, the auditing probabilities are slightly smaller than 1 and larger than 0, respectively, if firm 2 produces less than and more than the Cournot quantity. The pattern becomes more evident as the authority's incentive increases from 2 to 4. This reveals that at the convergence, the authority only partially implements the auditing strategy described in Proposition 1 when

one of the firms produce just at the Cournot quantity, indiciting that the learning process of the authority is difficult in presence of a quantity at the boundary.

Next, we discuss the auditing pattern under the random cost structure. In subplot (e), there are four larger cells with blue color, meaning no auditing. In each of those four larger cells, the most left-bottom small cell represents an equilibrium point discussed in Proposition 3. The two at the top from left to right are (q_h^c, q_l^c) and (q_l^c, q_l^c). The other two at the bottom from left the right are (q_h^c, q_h^c) and (q_l^c, q_h^c). This subplot illustrates that the authority does not audit at those equilibrium quantities, and also when both firms produce slightly more than the equilibrium. In subplot (f), however, the authority audits in three of the above four larger cells, except the one (q_l^c, q_l^c). Recall Proposition 3 states that at the equilibrium with contact between two firms, the authority will never audit at the quantity (q_l^c, q_l^c), and audit with a positive probability at the quantity (q_h^c, q_h^c). These predictions are consistent with subplot (f).

At the other two larger cells presented by (q_h^c, q_l^c) and (q_l^c, q_h^c), Proposition 3 predicts no auditing at the equilibrium. However, subplot (f) shows that when the authority's incentive is strong enough, the auditing probability at those quantities can be positive, even though the probability is small than at (q_h^c, q_h^c).

The figure also illustrates that on the 45° line from the cell (q_h^c, q_h^c) (the left-bottom cell) to (q_l^c, q_l^c) (the right-upper cell), the auditing probability increases from about 10% to 100%. Especially, for all those right-upper cells of the Cournot equilibrium under the fixed marginal cost $(1/3, 1/3)$, the auditing probability reaches 1 very quickly. The rationale behind such an auditing pattern is that those quantities along the 45° are likely produced by firms with cost (c_l, c_l) and the likelihood is higher as the quantities move to the right-upper direction. Thus, the authority learns to audit with higher probability in that direction. Similarly, when we move from (q_h^c, q_h^c) to the right to (q_h^c, q_l^c) and to above to (q_l^c, q_h^c), the cells on the path can be collusive outcomes for the cost (c_h, c_l) and (c_l, c_h), respectively, and the authority audits with an increasing probabilities along those two directions.

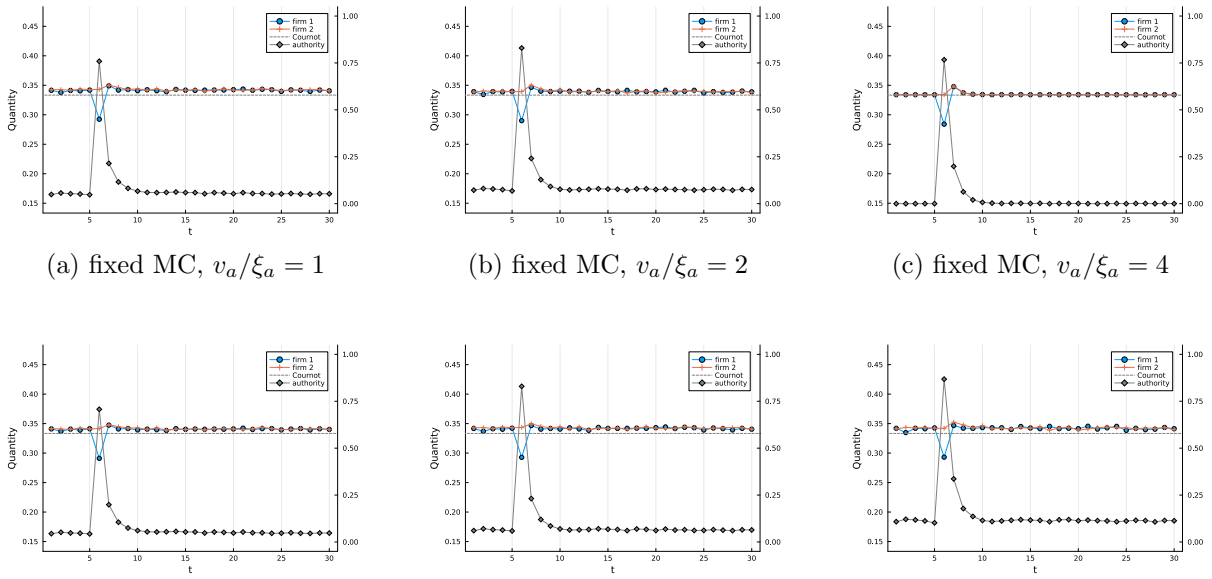
To summarize, the auditing behavior at the convergence is largely consistent with the theoretical model with firms' communication. With higher incentives, the authority was able to audit effectively when firms produce off-equilibrium quantities with positive probability.

4.2.3 Deviation

To test effectiveness of the authority's auditing, we arbitrarily impose perturbations to firm 1's quantity after convergence and simulate the response of the authority. The results are summarized in Figure 5. As in previous figures, the three subplots in the upper row are for fixed marginal cost and the ones in the bottom row are for random marginal costs.

In each subplot, at the beginning, the quantities of the two firms and the auditing probabilities are at convergence. The auditing probability is stable at the minimum. We perturb firm 1's quantity by letting it drop two grid points in period six.[5] From the subplots, we observe that the authority responds promptly by increasing the auditing probability from the minimum to above 75%. As a result, firm 1 immediately goes back and produces more than the quantity before the perturbation, then quickly adjusts the quantity to the level before the perturbation in the eighth period. The authority, at the same time, decreases its auditing probability significantly to the level before the perturbation in the tenth period.

Figure 5: Quantities and auditing probabilities: Deviation



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$

(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

The subplots also illustrate that the response of the authority to firm's sudden quantity drop does not rely on the authority's incentive much. When the authority's incentive ratio v_a/ξ_a is 2 and 4, the highest auditing probability in period 6 is slight higher than that when the ratio is 1, in both cost structures.

Note that our deviation analysis is different from those in Calvano et al. (2020b), where firms are perturbed to the direction non-collusive outcomes, i.e., higher prices, and simulations are conducted to test whether firms can go back to the collusive behavior. By contrast, our perturbation is toward the more collusive direction and the purpose is to test whether the authority can respond to the change by increasing the auditing probability. Our simulation results in Figure 5 demonstrate that the authority successfully captures the sudden

[5]The period here is relative after convergence.

change and push the quantities back to the level at convergence by immediately increasing the auditing probability.

4.3 Robustness

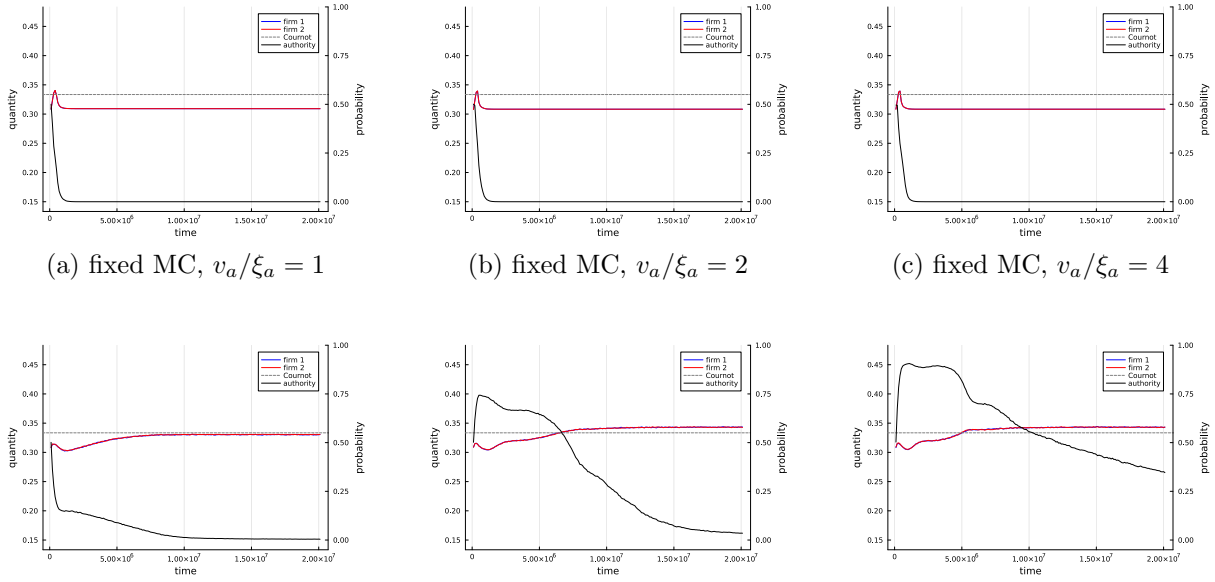
In this section, we check robustness of our results to the baseline simulation settings. We make two changes to our baseline setting. First, the authority's auditing threshold θ is lowered from $\theta = 1$ to $\theta = 10/11$, i.e., the authority audits when a firm's quantity is below 10/11 of the Cournot quantity—a less strict auditing rule. Second, we allow asymmetric learing between firms and the authority by setting the authority's learning parameters (α, β) to be half of that of the two firms. We find that our simulations results are robust to these variations.

In Figures 6-7, we present the simulation results for a lower auditing threshold ($\theta = 10/11$). The convergence process of quantities and auditing probabilities in Figure 6 display a similar pattern to our baseline results. The only difference is that when marginal cost is fixed, the converged quantities are about 0.30, lower than the threshold θ. This is consistent with Proposition 1 where $\theta q^c = (10/11) \times (1/3) = 0.3$ is an equilibrium quantity. Similarly, the auditing probabilities at convergence illustrated in the upper panel of Figure 7 display the same pattern as in our baseline results, with the only difference being the boundary cell is $(0.3, 0.3)$. The authority audits with probability 1 whenever any firm's quantity is greater than 0.3. The results show that even though the authority's auditing is effective, firms learn to produce less when the auditing is less strict.

For the random marginal costs, the results in Figures 6 are also similar to the baseline simulations, with the quantity at convergence slightly lower due to the less strict auditing rule: in subplots (e) and (f), the quantities at convergence are 0.343 and 0.342, respectively, while the numbers are 0.350 and 0.349 in the baseline results. In Figure 7, the four cells indicating static equilibrium quantities are (q_h^c, q_h^c), $(\theta q_l^c, q_h^c)$, $(q_h^c, \theta q_l^c)$, and $(\theta q_l^c, \theta q_l^c)$, i.e., the authority lowers the auditing quantity to θq_l^c for the cost c_l while the quantity under c_h does not change.

In the asymmetric learning setting, the authority's learning parameters are set to be $\alpha = 0.025, \beta = 2.5 \times 10^{-6}$, while the firms' parameters are $\alpha = 0.05, \beta = 5 \times 10^{-6}$ as in our baseline setting. Figures 8-9 summarize the simulation results for the asymmetric learning setting. The results are almost identical to the baseline ones. The only difference is that when the marginal costs are random, it takes slightly more periods for the authority. This is demonstrated by the observation that the auditing probability in the asymmetric case is smaller than in the baseline case for a given time period. Nevertheless, the asymmetry of

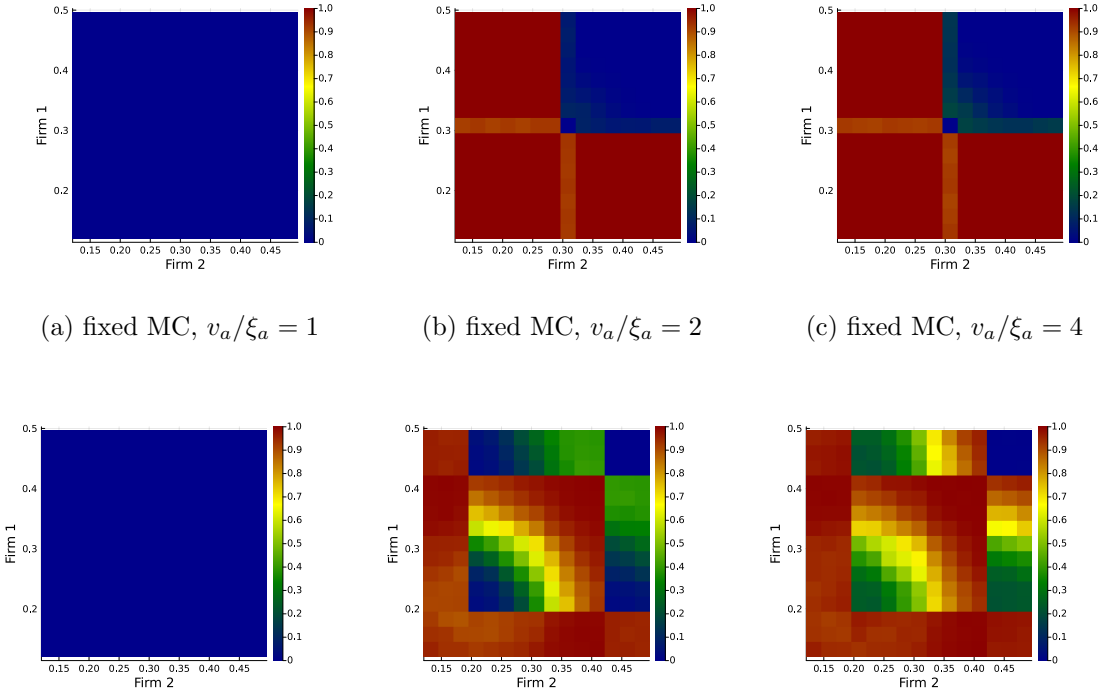
Figure 6: Lower threshold: Convergence process of quantities and auditing probabilities



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$

(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

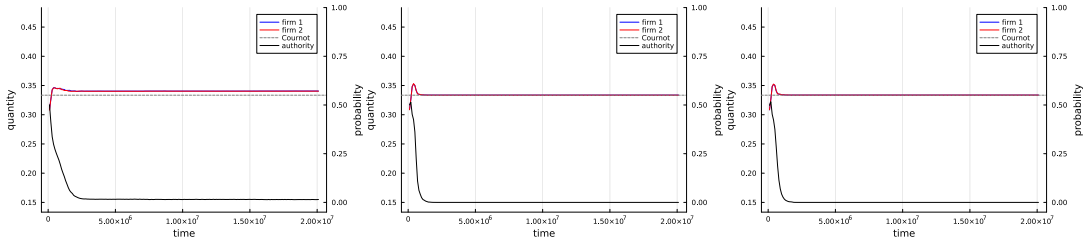
Figure 7: Lower threshold: Auditing probabilities at convergence



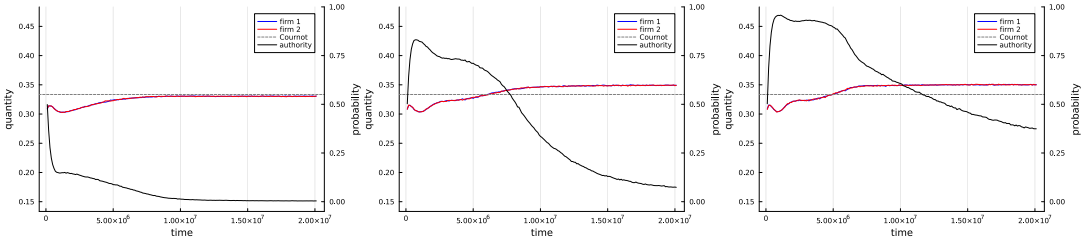
(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$

(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

Figure 8: Asymmetric learning: Convergence process of quantities and auditing probabilities

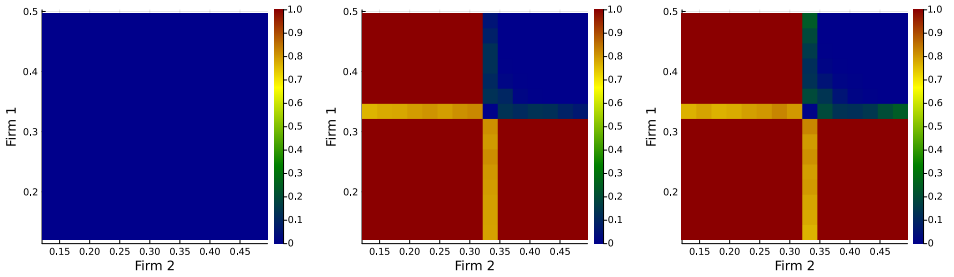


(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$

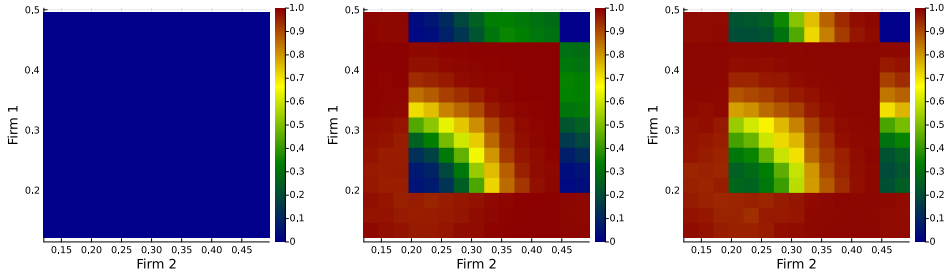


(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

Figure 9: Asymmetric learning: Auditing probabilities at convergence



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$



(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

learning between the authority and firms does not affect the outcome: the authority audits effectively as in the baseline simulations.

5 Alternative models

In this section, we change the model setup for our simulation to investigate whether the authority's auditing is still effective. Specifically, we first change the Q-learning algorithm to the Actor-Critic algorithm to check the dependence of our baseline results on algorithms. Next, considering that in some markets only price but not quantity can be audited, we modify the model such that the authority audits price rather than quantity.

5.1 Actor-Critic

In this section, we discuss Actor-Critic algorithm and the simulation results by using this algorithm. Actor–Critic is another popular algorithm of reinforcement learning. The main idea of Actor-Critic is that it combines two components: the actor, which learns a policy (how to act, i.e., what action to take in each state), and the critic, which evaluates the actor's actions by estimating the value function (how good a state or action is). The critic provides feedback to the actor to improve its policy.

5.1.1 The algorithm

As we discussed before, Q-learning is a value-based method — it learns the action-value function $Q(s, a)$ directly and chooses the action with the highest Q-value. By contrast, Actor–Critic is a hybrid policy/value method — it explicitly learns a policy (actor) guided by a value estimate (critic). Q-learning uses discrete action selection via argmax, while Actor–Critic can handle continuous actions more easily and tends to learn smoother policies. In our simulations, the authority's decision given a state variable s_a in a binary decision in $\{0, 1\}$ if Q-learning is employed, and the decision is a probability in $[0, 1]$ if Actor-Critic is employed. In other words, the players using Actor-Critic can use mixed strategy to audit while only pure strategy is possible by using Q-learning. We check whether the Actor-Critic algorithm employed by the firms and the authority change our baseline results.

We first present the details of Actor-Critic for a single agent. In each period t, a firm selects an action $a_t = a \in \mathcal{A}$ in state $s_t = s \in \mathcal{S}$ according to a stochastic policy $\pi(a|s, \boldsymbol{\lambda}) \equiv \Pr(a_t = a|s_t = s, \boldsymbol{\lambda})$ where $\boldsymbol{\lambda} \in \mathbb{R}^d, d = |\mathcal{A}| \times |\mathcal{S}|$ is the policy's parameter vector. As in Q-learning, we maintain that both the state variable s and the action a are discrete. Therefore, we parametrize numerical preferences $h(s, a, \boldsymbol{\lambda}) \in \mathbb{R}$ for each state-action pair.

The actions with the highest preferences in each state are given the highest probabilities of being selected. We choose a commonly used exponential soft-max distribution of $\pi(a|s, \boldsymbol{\lambda})$

$$\pi(a|s, \boldsymbol{\lambda}) = \frac{\exp\left(h(s, a, \boldsymbol{\lambda})\right)}{\sum_{a' \in \mathcal{A}} \exp\left(h(s, a', \boldsymbol{\lambda})\right)}. \tag{27}$$

We choose $h(s, a, \boldsymbol{\lambda})$ to be linear in λ

$$h(s, a, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \cdot \boldsymbol{x}(s, a), \tag{28}$$

where $\boldsymbol{x}(s, a)$ is vector of length $|\mathcal{A}| \times |\mathcal{S}|$. Its elements are 1 for the state-action pair (s, a) and 0 for the state-action pair $(s', a'), s \neq s', a \neq a'$. For example, $s \in \{s_1, s_2\}, a \in \{a_1, a_2\}$, then $\boldsymbol{x} = [x(s_1, a_1)\ \ x(s_1, a_2)\ \ x(s_2, a_1)\ \ x(s_2, a_2)]'$. Let $\boldsymbol{\lambda} = [\lambda_1\ \ \lambda_2\ \ \lambda_3\ \ \lambda_4]'$. Then

$$h(s_1, a_1, \boldsymbol{\lambda}) = [\lambda_1\ \ \lambda_2\ \ \lambda_3\ \ \lambda_4] \cdot [1\ \ 0\ \ 0\ \ 0]' = \lambda_1. \tag{29}$$

The Actor-Critic algorithm learns policy $\pi(a|s, \boldsymbol{\lambda})$ by learning $\boldsymbol{\lambda}$ through a critic and an actor. The critic evaluates the actions through temporal-difference (TD) update rule similar with Q-learning through

$$Q(s_{t+1}, a_{t+1}) = (1 - \alpha) Q(s_t, a_t) + \alpha \left(r_t + \sum_{a \in \mathcal{A}} \pi_t(a|s_{t+1}, \boldsymbol{\lambda}_t) Q(s_{t+1}, a) \right), \tag{30}$$

where Q-function is defined as in Q-learning, r_t is the reward in t, and α is the learning rate as defined in Q-learning. Note that in Q-learning, the second term on the right-hand-side is $\max_{a \in \mathcal{A}} Q(s_{t+1}, a)$, yielding a binary action, i.e., a pure strategy. Here, however, the softmax policy induces a probability of action on $[0, 1]$, i.e., a mixed strategy. Consequently, no separate ε-greedy exploration is required.

Next, the actor updates the policy parameters $\boldsymbol{\lambda}$ in the direction of the policy gradient, using the advantage defined as the difference between the realized and expected value of an action:

$$A(s_t, a_t) = Q(s_t, a_t) - \sum\nolimits_{a \in \mathcal{A}} \pi_t(a; s_t, \boldsymbol{\lambda}) Q(s_t, a). \tag{31}$$

The policy parameters $\boldsymbol{\lambda}$ are then updated according to the updating rule

$$\boldsymbol{\lambda}_{t+1} = \boldsymbol{\lambda}_t + \alpha A(s_t, a_t) \nabla \log \pi_t(a_t|s_t, \boldsymbol{\lambda}_t). \tag{32}$$

To be specific, for the softmax policy function, the gradients are defined as

$$\nabla \log \pi_t(a; s_t, \boldsymbol{\lambda}) = \begin{cases} 1 - \pi_t(a|s_t, \boldsymbol{\lambda}), & \text{if } a = a_t, \\ -\pi_t(a|s_t, \boldsymbol{\lambda}), & \text{otherwise.} \end{cases} \tag{33}$$

To setup the simulations, both the two firms and the antitrust authority adopt Actor-Critic algorithm. The MARL approach adopted in Q-learning is also employed for Actor-Critic. All the details of the auditing game are the same as presented in Section 3.2. All the parameters are set the same as in our baseline simulations.

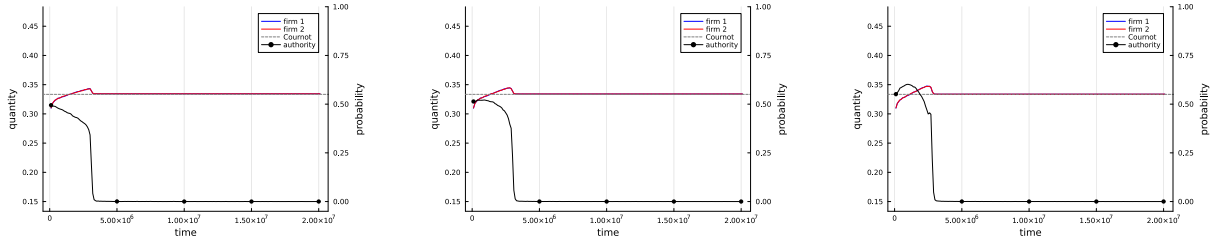
5.1.2 Simulation results

We present the convergence process of firms' quantities and the authority's auditing probability in Figure 10. Under both cost structures, the pattern of quantity and probability is similar to the baseline results: auditing is effective in lowering firms' quantities and the effectiveness increases in the authority's incentive.

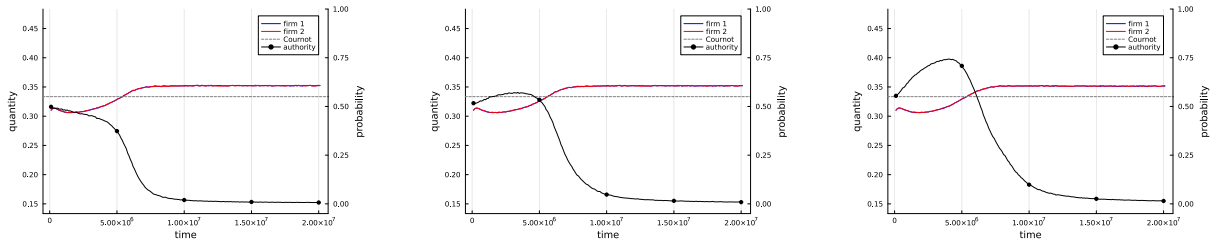
There are several differences between the results in Figure 10 and the baseline results. First, the auditing using actor-critic is more effective than using Q-learning. The highest auditing probability in the former case is much smaller than that in the latter case. For example, in subplot (f) of Figure 10, the highest auditing probability is about 75% while the corresponding probability is 90% in Figure 2. Moreover, when marginal cost is fixed and the benefit-cost ratio v_a/ξ_a is 1, the quantities converge to the Cournot quantity using actor-critic, while the converged quantity is larger than the Cournot quantity using Q-learning. Similarly, when the costs are random, the converged quantities are above the Cournot quantity using actor-critic while they are below the Cournot quantity using Q-learning. The reason why the two algorithms lead to quantitatively different results is that the actor-critic algorithm can audit with any probability between 0 and 1, depending on the quantities. The Q-learning algorithm, however, can only audit with probability 1 or 0, i.e., there always exists over-auditing or under-auditing.

This is evidenced by the heatmaps of auditing probabilities at convergence in Figure 11. It is evident that when the marginal cost is fixed, the authority effectively learned not to audit when the quantity is at the Cournot equilibrium (q^c, q^c) (the blue cell in the center). Different from Q-learning, the authority using Actor-Critic audits with positive probabilities when both firms' quantities are higher than q^c and audits with probabilities less than 1 if any firm's quantity is lower than q^c. The difference implies that Actor-Critic, relative to Q-learning, over-audits and under-audits when firms' quantities are higher and lower, respectively than the equilibrium. As the authority's benefit-cost ratio v_a/ξ_a increases, the

Figure 10: Actor-Critic: Convergence process of quantities and auditing probabilities



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$



(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

auditing probabilities move toward the one using Q-learning, i.e, the auditing probabilities are smaller when both firms' quantities are higher than the Cournot, while they are larger when any of the firm's quantity lower than the Cournot.

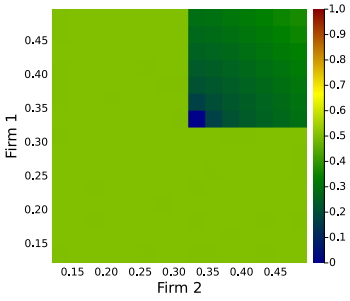
When the marginal costs are random, the authority is able to identify the equilibrium cells (q_h^c, q_l^c), (q_l^c, q_l^c), (q_h^c, q_h^c), and (q_l^c, q_h^c) and audits with a similar pattern to the baseline case. By comparing Figure 11 with Figure 4, we find that (1) the auditing is more effective by using Actor-Critic than Q-learning when the authority's incentive is not strong ($v_a/\xi_a = 1$). Recall that the authority does not audit at all in the baseline case. When the incentive is stronger, the authority audits with probabilities less than 1 in those cells with potential collusion by using Actor-Critic, while the probabilities are 1 in the baseline case. Nevertheless, as the benefit-cost ratio v_a/ξ_a increases from 2 to 4, the auditing pattern moves toward the the baseline pattern.

5.2 Price auditing

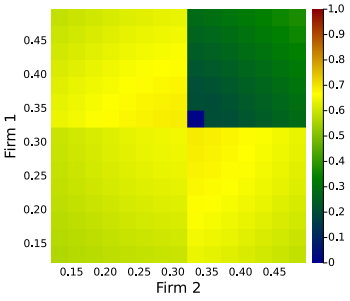
In some markets, the antitrust authority may access to prices rather than quantities. We modify our model to accommodate price auditing of the antitrust auditing.

In this modified model, firms still compete for quantity. However, a firm can only observe the market price, but not quantities, and the authority's auditing decision in the previous

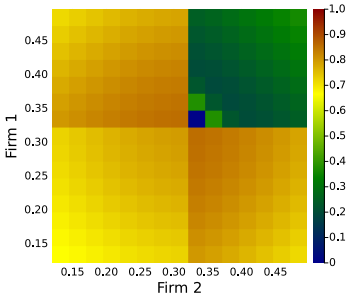
Figure 11: Actor-Critic: Auditing probabilities at convergence



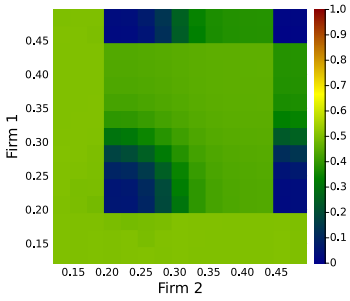
(a) fixed MC, $v_a/\xi_a = 1$



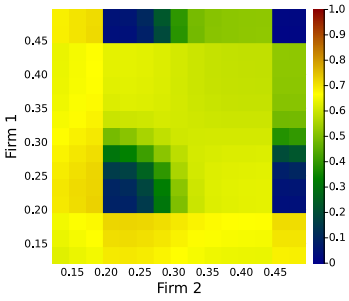
(b) fixed MC, $v_a/\xi_a = 2$



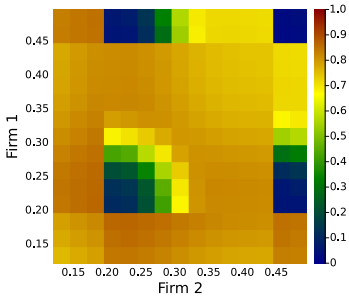
(c) fixed MC, $v_a/\xi_a = 4$



(d) random MC, $v_a/\xi_a = 1$



(e) random MC, $v_a/\xi_a = 2$

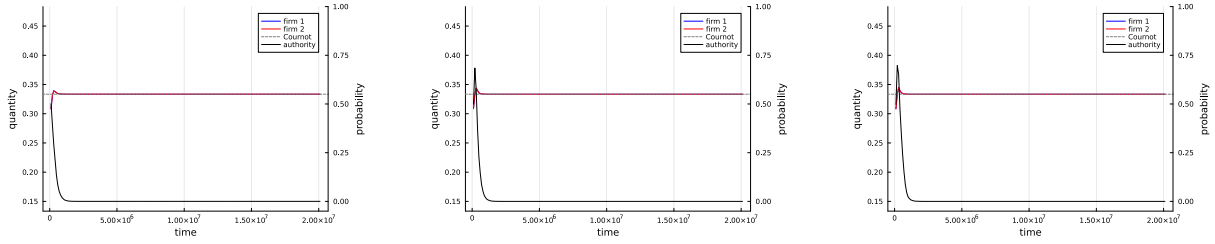


(f) random MC, $v_a/\xi_a = 4$

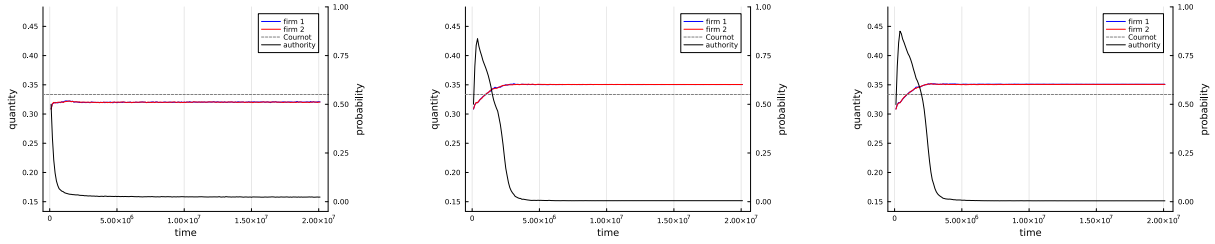
period. The authority can observe the market prices in the previous and the current period, as well as its own auditing decision in the previous period. Therefore, the state variable for firms and the authority are $s_f = (p_{t-1}, o_{t-1})$ and $s_a = (p_{t-1}, p_t, o_{t-1})$, where p_{t-1} and p_t are market prices in period $t-1$ and t, respectively, and $p_t = b - a(q_{1t} + q_{2t})$. After redefine the state variables, the remaining part of the MARL approach follows exactly Section 3.2.

We present the convergence process of firms' quantities and the authority's auditing probability in Figure 12. An evident difference of these results from the baseline ones is that the authority audits with higher probabilities at early stage, therefore the game converges faster than in the baseline case. An interpretation of the discrepancy is that the information in market price is "more aggregated" than two quantities. For all the quantities combinations (q_1, q_2) that are corresponding to the same market price p, the authority may only audits some of the combinations with probability 1 and does not audit the remaining ones. However, when price is the state variable, the authority likely audits this price with probability 1. The higher auditing probability pushes the firms to produce the quantities at convergence quickly.

Figure 12: Price: Convergence process of quantities and auditing probabilities



(a) fixed MC, $v_a/\xi_a = 1$ (b) fixed MC, $v_a/\xi_a = 2$ (c) fixed MC, $v_a/\xi_a = 4$



(d) random MC, $v_a/\xi_a = 1$ (e) random MC, $v_a/\xi_a = 2$ (f) random MC, $v_a/\xi_a = 4$

6 Concluding Remarks

This research sheds light on antitrust practice in presence of algorithmic collusion due to algorithmic pricing of competing firms. Assuming an antitrust authority also employs algo-

rithms to monitor firms and detect their possible collusive behavior, we show by simulation experiments that the effectiveness of the algorithmic authority in detecting collisions relies on information structure of firms as well as the authority's incentive and firms' costs of collusion. No matter whether firms' marginal costs are fixed or random, even such information is private, the authority can successfully prohibit firm's collusion and greatly improve consumer surplus. The outcome under the random cost structure is better than the fixed cost. The simulation results are robust to learning algorithm, alternative state variable, auditing threshold, and symmetric learning parameters between the authority and firms.

Our study provides the first piece of evidence on detecting collusion using algorithms. It identifies those important factors that affects the effectiveness of algorithmic detection. A natural direction for future research is to translate the conceptual detection framework into empirical tools that can be deployed in real markets. One promising avenue is to apply algorithmic-detection methods to high-frequency retail pricing data, where the prevalence of automated pricing makes markets particularly susceptible to algorithmic coordination. Researchers could also test reinforcement-learning–based detection procedures in online environments such as travel platforms, e-commerce retailers, or food-delivery marketplaces, where rich data and rapid price adjustments create ideal settings for algorithmic experimentation.

References

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

Stephanie Assad, Robert Clark, Daniel Ershov, and Lei Xu. Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market. *Journal of Political Economy*, 0(0):000–000, 2024. doi: 10.1086/726906.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, Joseph E Harrington Jr, and Sergio Pastorello. Protecting consumers from collusive prices due to ai. *Science*, 370(6520): 1040–1042, 2020a.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020b.

Federal Trade Commission et al. Ftc hearing# 7: The competition and consumer protection issues of algorithms, artificial intelligence, and predictive analytics, 2018.

Winston Wei Dou, Itay Goldstein, and Yan Ji. Ai-powered trading, algorithmic collusion, and price efficiency. *Available at SSRN 4452704*, 2023.

Justin P Johnson, Andrew Rhodes, and Matthijs Wildenbeest. Platform design when sellers use pricing algorithms. *Econometrica*, 91(5):1841–1879, 2023.

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

Policy Roundtable-Algorithms OECD. Collusion–note from the european union. Technical report, DAF/COMP/WD, 2017.

Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Qiaochu Wang, Yan Huang, Param Vir Singh, and Kannan Srinivasan. Algorithms, artificial intelligence and simple rule based pricing. *Available at SSRN 4144905*, 2023.

Christopher J. C. H. Watkins. Learning from delayed rewards. *Ph.D. dissertation,*, 1989.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3):279–292, 1992.

Appendix

In this appendix, we provide some details of the proof for Proposition 3.

We assume that firms collude only if both firms get positive profit from colluding. Assume c_i takes two values c_h and c_l, $c_h > c_l$.

1. When two firms' costs are (c_l, c_l). All the possible quantities are (1) Cournot $q_l^c = \frac{2b+c-3c_l}{6a}$; (2) Collusion #1: $q_i = \theta q_l^c$; (3) Cournot (c_l, c_h): $q = (q_l^c, q_h^c) \equiv (\frac{2b+c-3c_l}{6a}, \frac{2b+c-3c_h}{6a})$; (4) Cournot (c_l, c_h): $q = (q_h^c, q_l^c)$; (5) $q = (\theta q_l^c, \theta q_h^c)$; (6) $q(\theta q_h^c, \theta q_l^c)$ (7) Cournot $q = (q_h^c, q_h^c)$; (8) $q = (\theta q_h^c, \theta q_h^c)$.

If θ is sufficiently close to 1, the only possible outcomes are (1),(2), (7), and (8). If they choose to collude, then choice (2) is better than (1). We compare (2) with (7) and (8) to obtain the monitoring probability. In this case, two firms are symmetric.

$$\begin{aligned}(b - 2a\theta q_h^c - c_l)\,\theta q_h^c - m\,(\theta q_h^c, \theta q_h^c)\,\kappa &= (b - 2a\theta q_l^c - c_l)\,\theta q_l^c,\\ (b - 2aq_h^c - c_l)\,q_h^c - m\,(q_h^c, q_h^c)\,\kappa &= (b - 2a\theta q_l^c - c_l)\,\theta q_l^c.\end{aligned}$$

The monitoring probabilities are

$$m\,(\theta q_h^c, \theta q_h^c) = \frac{(b - 2a\theta q_h^c - c_l)\,\theta q_h^c - (b - 2a\theta q_l^c - c_l)\,\theta q_l^c}{\kappa}, \tag{.34}$$

$$m\,(q_h^c, q_h^c) = \frac{(b - 2aq_h^c - c_l)\,q_h^c - (b - 2a\theta q_l^c - c_l)\,\theta q_l^c}{\kappa}. \tag{.35}$$

When costs are (c_l, c_h), all possible quantities are (1) (q_l^c, q_h^c); (2) $(\theta q_l^c, \theta q_h^c)$; (3) (q_h^c, q_h^c); (4) $(\theta q_h^c, \theta q_h^c)$.

$$\text{firm with } c_l : (b - 2aq_h^c - c_l)\,q_h^c - m\,(q_h^c, q_h^c)\,\kappa = (b - a\theta q_l^c - a\theta q_h^c - c_l)\,\theta q_l^c, \tag{.36}$$

$$\text{firm with } c_h : (b - 2aq_h^c - c_h)\,q_h^c - m\,(q_h^c, q_h^c)\,\kappa = (b - a\theta q_l^c - a\theta q_h^c - c_h)\,\theta q_h^c. \tag{.37}$$

Take difference between two equations in Eq.(.36)

$$-a(q_h^c)^2\theta^2 + a(q_l^c)^2\theta^2 + bq_h^c\theta - bq_l^c\theta - c_hq_h^c\theta + c_hq_h^c - c_lq_h^c + c_lq_l^c\theta$$

When $\theta = 1$, the difference is

$$\begin{aligned}& -a(q_h^c)^2 + a(q_l^c)^2 + bq_h^c - bq_l^c - c_lq_h^c + c_lq_l^c\\ =\ & -a\big[(q_h^c - q_l^c)(q_h^c + q_l^c)\big] + (b - c_l)(q_h^c - q_l^c)\\ =\ & (q_h^c - q_l^c)\big[b - c_l - a(q_h^c + q_l^c)\big] > 0,\end{aligned} \tag{.38}$$

which implies that the two equations above cannot hold simultaneously. When θ is close to one, no deviation. Only (1) and (2) are possible outcomes. When costs are (c_h, c_l), the results are the same.

When costs are (c_h, c_h), the possible quantities are (3) and (4) above.

To sum up,

- If cost are (c_l, c_l), firms produce (q_l^c, q_l^c) if none of the firms contacts the other firm for collusion. If one of the firms contacts the other, they produce $(\theta q_l^c, \theta q_l^c)$, (q_h^c, q_h^c) and $(\theta q_h^c, \theta q_h^c)$, which yield the same profit.

- If costs are (c_h, c_l), firms produce (q_h^c, q_l^c) if collusion is not successful. If collusion is successful, they produce $(\theta q_h^c, \theta q_l^c)$. If costs are (c_l, c_h), the results are similar.

- If costs are (c_h, c_h), they produce (q_h^c, q_h^c) when collusion fails and produce $(\theta q_h^c, \theta q_h^c)$.