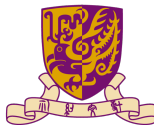


# Algorithmic Collusion of Competitive Dynamic Pricing: Past, Present and Future

Chen Tang <sup>1</sup>

<sup>1</sup>School of Data Science, Chinese University of Hong Kong, Shenzhen

July 2024



香港中文大學 (深圳)  
The Chinese University of Hong Kong, Shenzhen

The paper for this presentation is available [HERE](#).  
More information about the author can be found [HERE](#).



# Introduction to Dynamic Pricing

Dynamic pricing studies how to adapt prices in response to demand fluctuations.



It's widely been used in industries:

- Electricity Market;
- Online Retailing;
- Airline Companies.

I recommend [den Boer, 2015] as a general review for dynamic pricing.

- ① 2 to 5 percent growth in sales;
- ② 5 to 10 percent growth in profit margins.



- It can lift consumer satisfaction during the purchasing.

# Dynamic Pricing under Competition

Most previous research has focused on three primary aspects:

- ① Given a competitive pricing environment, what is the theoretical equilibrium of the game? Is it unique or not?
  - One-period / Multi-stage;
  - Capacity / Inventory Constraint;
  - Homogeneous / Heterogeneous Prodducts.
- ② The validity of the *market response hypothesis*:
  - If both firms use a monopoly model to learn the environment, can the game converge to the equilibrium?
- ③ Strategy of dynamic pricing in a competitive environment.

The study of algorithmic collusion is a new subfield **only after 2020**.

# A taste on algorithmic collusion

Using the prisoners dilemma as the toy model:

**Table 1:** The payoff matrix of the toy model

		Firm 2	
		<i>H</i>	<i>L</i>
Firm 1	<i>H</i>	(2, 2)	(0.5, 3)
	<i>L</i>	(3, 0.5)	(1, 1)

In the static game (one-period), the unique Nash equilibrium is both setting low prices.

What if both firms use the RL algorithm (such as Q-learning) to determine their prices, and both firms set high prices?

[Calvano et al., 2019] first observed this phenomenon in 2019 and called this as **algorithmic collusion** (unrigorous definition).

The greatness of the competition.

# One-shot Deviation Principle

The same payoff matrix, but multi-stage:

**Table 2:** The payoff matrix of the toy model

		Firm 2	
		H	L
Firm 1	H	(2, 2)	(0.5, 3)
	L	(3, 0.5)	(1, 1)

In the dynamic game, the SPNE is not unique.

Collusion can be sustained if both firms use the one-shot deviation principle.

Assume the discount rates are  $\delta_1$  and  $\delta_2$  respectively, a reliable threat can exist if:

$$2 \cdot \sum_{t=0}^{\infty} \delta_1^t \geq 3 + 1 \cdot \sum_{t=1}^{\infty} \delta_1^t$$

$$2 \cdot \sum_{t=0}^{\infty} \delta_2^t \geq 3 + 1 \cdot \sum_{t=1}^{\infty} \delta_2^t$$

$$\delta_1 \geq \frac{2}{3}, \delta_2 \geq \frac{2}{3}$$

*Folk Theorem:* collusion can occur as long as players are patient enough.

# Spurious Collusion & Genuine Collusion

A reliable collusion should exhibit a reward-punishment mechanism for the opponent's deviations.

[Calvano et al., 2023] categorizes algorithmic collusion into two types:

- *Genuine Collusion*: High prices are the result of firms not recognizing the potential for greater short-term profits through deviation, rather than due to an explicit reward-punishment scheme.
- *Spurious Collusion*: High prices are maintained through a reward-punishment mechanism where the AI algorithm exhibits behavior akin to the one-shot deviation principle.

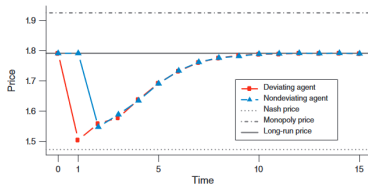


Figure 1: This figure is from figure 4 in [Calvano et al., 2020].

# Authentic Collusion

There are many critics on these findings, the most representative voice is:

- Even though genuine collusion demonstrates a reward-punishment scheme to deviations, it can be outperformed (hacked) by other algorithms. Thus, this is not really ‘spurious’.

Arnoud V. den Boer gives a rigorous definition of collusion, and I summarize it as the authentic collusion:

- *Authentic Collusion*: collusion sustained by a theoretically reliable threat, and the collusive algorithm performs well enough against the opponent’s alternative algorithms.

These definitions are not just tautology!



# The directions of research on algorithmic collusion

The research on algorithmic collusion can be broadly categorized into two streams based on different directions:

- ① This stream of research focuses on **convergence**, investigating whether specific AI algorithms could converge to spurious collusion, genuine collusion, or merely competitive outcomes.
- ② This stream of research focuses on **equilibrium**, studying whether a specific algorithmic environment could formulate a collusive equilibrium.

The *convergence* and *equilibrium* are not the same concepts. In many numerical experiments, the game can converge to a scenario that is not any equilibrium.

## Why Does Algorithmic Collusion Matter?

It's about the fairness of competition, a fair market can bring better welfare for the society.

Collusion is usually illegal (anti-monopoly law) in most regions, but using reinforcement learning may collude in a legal way.

Another realistic concern is that more and more companies are trying to use AI to price their products ([Spann et al., 2024]).

- ① Introduction
- ② Algorithmic Collusion under RL
- ③ Authentic Algorithmic Collusion
- ④ Regulate Algorithmic Collusion
- ⑤ Conclusions
- ⑥ Reference

# Introduction on Reinforcement Learning: Q-learning 1

Five key elements of reinforcement learning (in the context of dynamic pricing):

- ① States ( $\mathcal{S}$ ): a finite set representing all possible states of the environment (the firms' historical prices);
- ② Actions ( $\mathcal{A}(\mathcal{S})$ ): a finite set of actions available to the decision-maker, which may depend on the current state (prices to be determined);
- ③ Transition Probabilities ( $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ ): The probability of transitioning to a new state given the current state and the action taken (updated historical prices);
- ④ Rewards Function ( $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ ): A function that assigns a reward for each state-action pair (revenue);
- ⑤ Discount Factor ( $\delta \in [0, 1]$ ): A factor that discounts future rewards based on their distance from the present.

A very famous algorithm is called **Q-learning**, which maintains a Q-function  $Q(s, a)$  for every state-action pair:

- At each time  $t$  given the current state  $s_t$ , the algorithm has two options:
  - ① With prob  $\epsilon$ : randomly pick an action from  $\mathcal{A}$ ;
  - ② With prob  $1 - \epsilon$ : select  $\arg \max_a Q(s_t, a)$ ;
- Note that  $\epsilon$  is usually a decreasing function on  $t$ ;
- Then the agent receives reward  $r_{it}$  from the environment.

# Introduction on Reinforcement Learning: Q-learning 2

---

## Algorithm 1 Q-Learning with $\epsilon$ -Greedy Exploration

---

```

1: Initialize action-value function  $Q(s,a)$  arbitrarily for all  $s$  in  $\mathcal{S}$  and  $a$  in  $\mathcal{A}$ 
2: Initialize  $Q'(s,a) \leftarrow 0$ 
3: Initialize a learning rate  $\alpha$ , discount factor  $\delta$ , exploration rate  $\epsilon$ , and threshold  $\delta$ 
4:  $s \leftarrow$  initial state
5: while  $|Q - Q'|_{\infty} > \delta$  do
6:    $Q \leftarrow Q'$ 
7:    $a \leftarrow \operatorname{argmax} Q(s,a)$  with probability  $1 - \epsilon$                                 ▶ Exploitation
8:    $a \leftarrow$  random action from  $\mathcal{A}$  with probability  $\epsilon$                             ▶ Exploration
9:   Take action  $a$ , observe reward  $r$  and next state  $s'$ 
10:   $Q'(s,a) \leftarrow Q(s,a) + \alpha(r + \delta \cdot \max_{a'} Q(s',a') - Q(s,a))$ 
11:   $s \leftarrow s'$                                                                 ▶ Transition to the next state
12: end while
13: return  $\pi^*(s) = \operatorname{argmax}_a Q(s,a)$ 
    
```

---

Most of the RL algorithms use a value function approximation scheme:

- They maintain the Q-matrix for all state-action pairs;
- Use different approaches to estimate the Q-matrix (deep learning, sampling, etc)

# Introduction on Reinforcement Learning: Bandit Learning

There is a type of *state-free* algorithm called the bandit algorithm.

UCB(upper confidence bound) algorithm:

- Only maintaining the value function for the action space  $Q(a)$  rather than  $Q(s, a)$ .

$$v_{a,t} = \frac{1}{N_{a,t-1}} \sum_{t' \leq t-1} r_{t'} 1\{a_{t'} = a\} + \sqrt{\frac{2 \log(1/\delta)}{N_{a,t-1}}}$$

Besides the value-based algorithms, policy gradient algorithms are proposed without relying on the estimation of the value function.

# Algorithmic Collusion: The initial paper 1

The first paper on algorithmic collusion is [Calvano et al., 2020]. Assume there are  $n$  different agents, the demand for firm  $i$  at time  $t$  is:

$$d_{i,t} = \frac{e^{\frac{a_i - p_{i,t}}{\mu}}}{\sum_{j=1}^n e^{\frac{a_j - p_{j,t}}{\mu}} + e^{\frac{a_0}{\mu}}} \quad (1)$$

They discretize the firms' action spaces: for each firm, we can compute its one-shot Bertrand-Nash equilibrium (**competitive price**)  $p^{com}$  and the monopoly (**collusive**) price  $p^{col}$ . Then, taking the interval  $[p^{com} - \xi(p^{col} - p^{com}), p^{col} + \xi(p^{col} - p^{com})]$ ,  $\xi > 0$ , and splitting the interval into  $m$  discrete prices, we can get the action space for each agent. The state space is defined as the prices in the last  $k$  epochs. Each agent utilizes the Q-learning algorithm to make decisions.

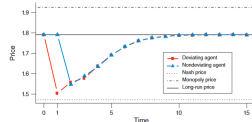
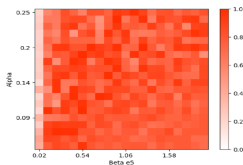
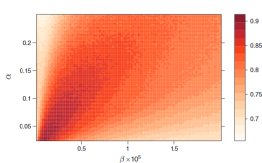
In the baseline experiment,  $k = 1$ ,  $m = 3$ , and only two firms are competing.

# Algorithmic Collusion: The initial paper 2

The authors define the **average profit gain**  $\Delta$ :

$$\Delta \equiv \frac{\bar{\pi} - \pi^{com}}{\pi^{col} - \pi^{com}}$$

- When  $\Delta$  nears 1, the game is in collusive scenario; when it nears 0, the game is in competitive scenario;
- The left figure shows the Q-learning would converge to the collusive scenario, and the middle figure is my replication;
- The right figure shows that the agent can respond strategically to the opponent's deviation, meaning it's a genuine collusion.



The results are robust in other parameter settings.



# Q-learning can Collude Even more Smartly 1

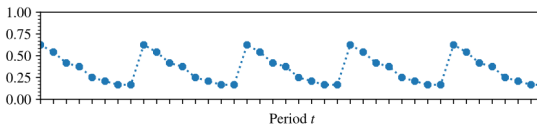
[Klein, 2021] tests the Q-learning in another environment. Consider two agents

$i \in \{1, 2\}$  price in infinitely discrete time indexed by  $t \in \{0, 1, 2, \dots\}$ . Price adjustments occur sequentially: agent 1 can only adjust its price in odd-numbered periods, agent 2 in even-numbered periods. Each agent selects its price from a discrete set  $P = \{0, \frac{1}{k}, \frac{2}{k}, \dots, 1\}$ . Firm  $i$  profit at time  $t$  is derived as  $\pi_i(p_{it}, p_{jt}) = p_{it} D_i(p_{it}, p_{jt})$ , its objective function can be expressed as  $\sum_{s=0}^{\infty} \delta^s \pi_i(p_{i,t+s}, p_{j,t+s})$ . The demand is linear:

$$D_i(p_{it}, p_{jt}) = \begin{cases} 1 - p_{it} & \text{if } p_{it} < p_{jt} \\ 0.5(1 - p_{it}) & \text{if } p_{it} = p_{jt} \\ 0 & \text{if } p_{it} > p_{jt} \end{cases}.$$

## Q-learning can Collude Even more Smartly 2

The **Edgeworth price cycles** is an MPE for this game: periodic price jumps reset a gradual price decline.



However, unlike theoretical analysis in past works, these price cycles are deterministic: It is always the same firm that undertakes the costly action of “resetting” the price cycle by jumping up in price rather than undercutting, with the other firm able to free-ride on this.

# State-free Q-learning algorithm can collude

[Asker et al., 2022], [Asker et al., 2023] study the state-free Q-learning in the Bertrand model:

- ① The state space for the Q-learning algorithm is reduced to a singleton, indicating that the algorithm does not rely on historical prices;
- ②  $\delta = 0$ , implying that the algorithm considers only the one-period payoff;
- ③ There is no  $\epsilon$  exploration; instead, at each epoch, the agent persistently uses the greedy action;

This experiment is very **important**:  $\delta = 0$  implies there is no equilibrium that can collude, but the game can still converge to the supra-competitive prices.

# Bandit Algorithm can Collude

[Hansen et al., 2021] assumes a linear demand:

$$d_j^*(p_j, p_{-j}) = \alpha - \beta p_j + \gamma p_{-j}.$$

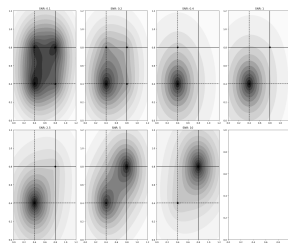
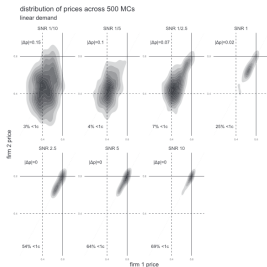
$$\pi_t(p_j, p_{-j}) = p_j d_j^*(p_j, p_{-j}) + \varepsilon_{j,t},$$

$$\varepsilon_{j,t} \sim U\left[-\frac{1}{\delta}, \frac{1}{\delta}\right]$$

$\delta$  is the SNR. Both firms use a UCB-tuned algorithm to make decisions:

$$V_{k,t} = \overline{\pi_{k,t}^2} - \bar{\pi}_{k,t}^2 + \sqrt{\frac{2 \log t}{n_{k,t}}},$$

$$\text{UCB-tuned}_{k,t} = \bar{\pi}_{k,t} + \sqrt{\frac{\log t}{n_{k,t}}} \min\left(\frac{1}{4}, V_{k,t}\right),$$



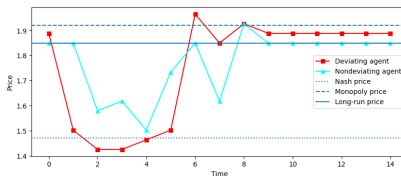
# Other Algorithmic Collusion

Algorithm	Environment	Collusion	Type of Collusion
Q-learning Calvano, Calzolari, Denicolo, and Pastorello (2020)	Calvano	Yes	Genuine
DQN Hettich (2021)	Calvano	Yes	Genuine
Actor-Critic Martello (2022)	Calvano	Yes	Genuine
LLM Fish et al. (2024)	Calvano	Yes	Unkown
Perfect Synchronous State-Free Q-learning Asker et al. (2022)	Bertrand	No	/
Asynchronous State-Free Q-learning Asker et al. (2023)	Bertrand	Yes	Spurious
Sequential Q-learning Klein (2021)	Sequential Bertrand	Yes	Unkown
UCB-tuned Hansen et al. (2021)	Linear Duopoly	Yes	Unkown

In summary, between 2019 - 2022, many **numerical research** has shown that some RL algorithms can converge to collusion in the pricing game.

## Criticism 1: What looks like collusion need not be collusion

- The decision made by the RL algorithm is not that 'smart':
  - For example, in the research by [Klein, 2021], the players do not exhibit the patterns exactly the same as MPE.
  - [Epivent and Lambin, 2024]: the algorithm will start a price war to any price modification.
- The algorithms can sometimes behave irrationally.



Many scholars think these collusions are caused by the insufficient exploration inherent in the Q-learning algorithm.

## Criticism 2 & 3

- ① *Number effect*: as the number of firms participating in the market increases, the tacit collusion induced by Q-learning may be mitigated. This phenomenon has been verified by several numerical studies including [Hettich, 2021], [Asker et al., 2022], [Abada and Lambin, 2023];
- ② It's not likely for firms to adopt the Q-learning algorithm to price when it can be outperformed by other simpler algorithms, such as Exp3 ([den Boer et al., 2022]).

## Criticism 4: Conditions Are Harsh

[den Boer et al., 2022] refined the definition of equilibrium to accommodate the dynamics of dynamic pricing using  $\epsilon$ -greedy Q-learning in a duopoly market.

$(\delta - \epsilon)$ -best-response: A strategy  $\sigma^{(i)} = \{\sigma^{(i)}(s) : s \in \mathcal{A}_i \times \mathcal{A}_{-i}\}$  of player  $i$  is called  $(\delta - \epsilon)$ -best-response to strategy  $\sigma^{(-i)} = \{\sigma^{(-i)}(s) : s \in \mathcal{A}_{-i} \times \mathcal{A}_i\}$  if the following equation holds for all states  $s$  and  $s_{-i}$ :

$$\begin{aligned} \sigma^{(i)}(s) \in \arg \max_{a \in \mathcal{A}_i} \frac{\epsilon}{|\mathcal{A}_{-i}|} \sum_{a_{-i} \in \mathcal{A}_{-i}} \{r_i(a, a_{-i}) + \delta V_{\sigma^{(-i)}}^{(i)}(a, a_{-i})\} \\ + (1 - \epsilon) \{r_i(a, \sigma^{(-i)}(s_{-i}, s_i)) + \delta V_{\sigma^{(-i)}}^{(i)}(a, \sigma^{(-i)}(s_{-i}, s_i))\}, \end{aligned}$$

where the value function  $V_{\sigma^{(-i)}}^{(i)}(s)$  is defined by:

$$\begin{aligned} V_{\sigma^{(-i)}}^{(i)}(s) := \max_{p \in \mathcal{A}_i} \frac{\epsilon}{|\mathcal{A}_{-i}|} \sum_{a_{-i} \in \mathcal{A}_{-i}} \{r_i(p, a_{-i}) + \delta V_{\sigma^{(-i)}}^{(i)}(p, a_{-i})\} \\ + (1 - \epsilon) \{r_i(p, \sigma^{(-i)}(s_{-i}, s_i)) + \delta V_{\sigma^{(-i)}}^{(i)}(p, \sigma^{(-i)}(s_{-i}, s_i))\}, \end{aligned}$$

A strategy pair  $(\sigma^{(i)}, \sigma^{(-i)})$  is called a  $(\delta - \epsilon)$ -strategy-equilibrium if they are mutually  $(\delta - \epsilon)$ -best-response to each other.

This definition of best response aligns with **the exploratory nature of Q-learning**.

For any  $\delta$ , there exists an  $\epsilon^*(\delta) \in (0, 1)$  such that for all  $\epsilon > \epsilon^*(\delta)$ , the only  $(\delta - \epsilon)$ -strategy-equilibrium is both firms use the competitive price.



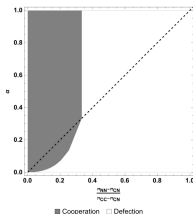
# Theoretical Evidence: 1

- [den Boer et al., 2022]'s  $(\delta - \epsilon)$ -*best-response* is great and can explain the slow convergence in most numerical experiments;
- However, this framework takes the perspective of equilibrium rather than convergence, sometimes the game can converge to the collusive outcome even though the collusive one is not an equilibrium ([Asker et al., 2023]).

To date, there is **no universal** conclusion on the cause of algorithmic collusion by the reinforcement learning algorithm.

## Theoretical Evidence: 2

- [Banchio and Skrzypacz, 2022] proposes that algorithmic collusion relies on *spontaneous coupling*, which results in overestimating the Q-value of the high prices.
- Spontaneous coupling arises from the correlation of Q-values due to the updation being dependent on both firms' actions.
- This research presents sufficient conditions when spontaneous coupling does not occur (can explain some observations in [Asker et al., 2022].)
- [Dolgoplov, 2024] employs evolutionary game theory to analyze the convergence of **state-free** Q-learning agents in a prisoner's dilemma game.
- When both agents use logit (Boltzmann) exploration, the convergence is contingent upon a closed-form relationship between the learning rate and the payoffs.



- [Brown and MacKay, 2023] examine the asymmetric setting where firms can price based on an opponent's most recent price;
- It found that the price level will be supra-competitive under the MPE.

In conclusion, under specific circumstances and parameter settings, the use of RL algorithms can **converge to** a collusive scenario with a theoretical guarantee.

# Theoretical Evidence: 3

## Theoretical evidence beyond reinforcement learning:

- [Cho and Williams, 2024] proved that firms using a simple model averaging and least squares estimation to price converge to a collusive scenario in a linear Bertrand duopoly setting;
- [Lamba and Zhuk, 2022] studied a pricing game where firms price their goods solely based on the opponent's price during the previous epochs, it is proved that under the Markov Perfect Equilibrium, the average price set by both firms is supra-competitive;
- Other insights can be drawn from the theoretical studies on the collusion of Cournot games ([Shi and Zhang, 2020], [Possnig, 2023]).

# Empirical Evidence: 1

The most compelling evidence comes from [Clark et al., 2023]

$$y_{mt} = \alpha_m + \alpha_t + \beta_1 T_{mt}^1 + \beta_2 T_{mt}^2 + \epsilon_{mt},$$

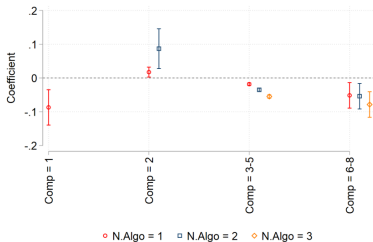
- $y_{mt}$ : represents the price margin in market  $m$  at time  $t$ ;
- $T_{mt}^1$ : dummy variable that indicate whether market  $m$  at time  $t$  has **only** one gas stations using algorithms to price;
- $T_{mt}^2$ : dummy variable that indicate whether market  $m$  at time  $t$  has **exactly** two gas stations using algorithms to price.

Using instrumental variables, the authors identified that in duopoly markets, the adoption of algorithmic pricing by a single firm does not significantly affect the price margin, whereas the adoption by both firms can increase station-level margins by 28%.

Algorithmic pricing can significantly increase the market prices.

## Empirical Evidence: 2

[Wieting and Sapi, 2021] provide an intuitive evidence:



In a duopoly market, the adaptation of AI pricing algorithms by both firms can increase the market price.

- Another empirical evidence comes from [Musolf, 2022];
- However, all empirical research only reveals that algorithmic pricing can increase the market price, not indeed to be RL algorithms.

## Summary on Previous Findings

- The reinforcement learning algorithm can result in collusive situations in many numerical experiments;
- The simulation can not generate a stable and 'smart' enough result, some scholars question the validity of the collusion;
- Theoretical analysis suggests that in some settings, the Q-learning can converge to the collusion with guarantee;
- Empirical papers find that AI algorithms in competitive cases can increase product prices.

## My Own Remarks

- Critics may argue that the collusion, including the reward-punishment scheme, is not a well-defined equilibrium and that simple reinforcement learning;
- Addressing the incomplete nature of algorithmic collusion under RL is a meaningful discussion, but this does not negate the significance of algorithmic collusion under RL algorithms.

**The convergence of a pricing game under RL algorithm does not need to be a theoretically well-defined equilibrium.**

- Algorithmic collusion is a problem that deserves consideration.

- 1 Introduction
- 2 Algorithmic Collusion under RL
- 3 Authentic Algorithmic Collusion**
- 4 Regulate Algorithmic Collusion
- 5 Conclusions
- 6 Reference



# Definition on Authentic Collusion: 1

[den Boer, 2023] offers a fantastic framework for the duopoly game.

- Consider a repeated game between two firms, denoted by  $i = 1$  and  $i = -1$ , selling a homogeneous product;
- At time  $t \in \mathbb{N}$ , each firm selects a price  $P_{it}$  from  $\mathcal{P}_i$ ;
- Demand  $D_{it}, D_{-it}$  obtained by firms at time  $t$  is drawn from the **unknown** demand function  $d(\mathbf{P}_t)$ ;
- Each firm receives a revenue  $R_{it} = P_{it} \cdot D_{it}$ ;
- The algorithm (pricing strategy)  $\pi_i \in \Pi_i$  is defined as the mapping from the *state*  $H_{it} := (P_{is}, P_{-is}, D_{is} \text{ for } 1 \leq s \leq t)$  to the probability distribution over the action space  $\mathbb{P}(\mathcal{P}_i)$ .

We can obtain the expected cumulative reward for each firm:

$$\phi_i(\pi_i, \pi_{-i}, d)$$

*Regret* of an algorithm is defined as the difference from the reward under the **optimal fixed** action:

$$\text{Regret}_i(\pi_i, \pi_{-i}, d) := \sup_{P_i^* \in \mathcal{P}_i} \phi_i(P_i^*, \pi_{-i}, d) - \phi_i(\pi_i, \pi_{-i}, d)$$

## Definition on Authentic Collusion: 2

Requirements on authentic collusion:

- 1 The definition of algorithm already implicitly assumes that there is no illegal communication between the two firms;
- 2 Algorithmic collusion needs to achieve supra-competitive payoffs:

$$\phi_i(\tilde{\pi}_i, \tilde{\pi}_{-i}, d) \geq \phi_i^{comp}(d),$$

where  $\phi_i^{comp}(d)$  is the expected payoff during a fully competitive setting.

- 3 A 'smart' collusion must have a good enough performance against alternative algorithms:

$$Regret_i(\tilde{\pi}_i, \pi_{-i}, d) \leq \epsilon$$

for some self-determined  $\epsilon > 0$ .

- 4 Finally, an algorithm can achieve authentic collusion if it is the best response when the opponent uses the same algorithm:

$$\phi_i(\tilde{\pi}_i, \tilde{\pi}_{-i}, d) \geq \phi_i(\pi_i, \tilde{\pi}_{-i}, d)$$

## Definition on Authentic Collusion: Summary

- ① *Spurious collusion*: No illegal communication but can achieve supra-competitive outcomes;
- ② *Genuine collusion*: Based on the preceding requirements, the collusive algorithm should exhibit a reward-punishment scheme toward deviation (a subset of alternative algorithms);
- ③ *Authentic collusion*: Based on the preceding requirements, the collusive algorithm should have low regret against all alternative algorithms and is the best response to the collusive algorithm.

One big limitation in this paper:

- This framework does not intuitively distinguish between genuine collusion and authentic collusion;
- Because it's challenging to define deviation in genuine collusion using rigorous notations.

Currently, there have been two authentic algorithms.

# Authentic Collusion 1: Composite SPSA

[Meylahn and V. den Boer, 2022] provides the first authentic collusive algorithm.  
Important Assumption in this paper:

- The demands are public information for both firms;

Then both firms can learn the collusive and competitive prices using Simultaneous Perturbation Kiefer–Wolfowitz recursion:

- Each firm maintains an estimation  $\hat{p}_i^{col}(n)$ , where  $n$  is a counter
- In each two consecutive times, the firm charges  $\hat{p}_i^{col}(n) + c_n \omega_i^{col}(n)$  and  $\hat{p}_i^{col}(n) - c_n \omega_i^{col}(n)$ ,  $c_n$  is a tuning sequence and  $\omega_i^{col}(n)$  is a random variable equals 1 and  $-1$ , both with probability  $\frac{1}{2}$ ;
- After observing the joint revenue  $\tilde{r}^+$  and  $\tilde{r}^-$  in two consecutive times, both firm take gradient descent to update  $\hat{p}_i^{col}(n)$ :

$$\hat{\nabla} r(n) := \frac{\tilde{r}^+ - \tilde{r}^-}{2c_n \omega_i^{col}(n)}$$

- Similarly. both firms can learn  $\hat{p}_i^{com}(n)$ .

# Authentic Collusion 1: Composite SPSA

- The full process includes five stages to learn the revenue under the collusive and competitive cases, then both firms decide to collude or compete based on the comparison of the revenue under the two cases;
- The algorithm would change to the 'competitive' module if it detects the opponent is not using the algorithm to price;
- It is proved to be an authentic collusive algorithm.

Two drawbacks of Composite SPSA:

- ① It requires knowledge of the mutual demand function.
- ② The SPSA algorithm requires synchronization. i.e. if the two firms do not initiate the same algorithm simultaneously, they may not end up charging the collusive prices.

## Authentic Collusion 2: Collude-or-compete

[Loots and den Boer, 2023]'s *Collude-or-compete* solved the above two problems.

$$\lambda_j(\mathbf{p}; \theta) := \frac{\nu_j(p_j; \theta)}{1 + \nu_1(p_1; \theta) + \nu_2(p_2; \theta)}$$

- To learn the collusive price with convergence guarantee, the algorithm needs to estimate  $\Theta$  consistently. To estimate  $\Theta$ , both player's demand information is needed;
- Their algorithm can reveal the private demand information by pricing:
  - ① If the demand is 1, the opponents demand is 0;
  - ② If the demand is 0, how should we know the opponents demand, 0 or 1?

How to extract the demand information from the prices?

- Maintain three estimations of the parameters:  $\hat{\theta}_{t,j}^{(0,0)}, \hat{\theta}_{t,j}^{(0,1)}, \hat{\theta}_{t,j}^{(1,0)}$ ;
- Calculate the collusive price  $(p_j^{(0,0)}(t+1), p_{-j}^{(0,0)}(t+1)), (p_j^{(0,1)}(t+1), p_{-j}^{(0,1)}(t+1)), (p_j^{(1,0)}(t+1), p_{-j}^{(1,0)}(t+1))$ ;
- When  $d_{t,j} = 0$ , assume that  $d_{t,-j} = 0$ , and set price at  $p_j^{(0,0)}(t+1)$ ;
- At time  $t+1$ , when she observes the opponent's price is  $p_{-j}^{(0,1)}(t+1)$ , she would know that the opponent's demand at time  $t$  is 1.

The authors proved the convergence of this algorithm.

## Authentic Collusion 2: Collude-or-compete

More about the *Collude-or-compete* algorithm.

What if the opponents choose other algorithms (not to collude)?

- The collude-or-compete includes two modules:
- The collude module (in the last slide) is trying to learn the parameters of the environment;
- The complete module can deter the opponents' incentive to use other pricing strategies.

The algorithm has a mechanism to decide when to switch between the two modules, and the algorithm also has a mechanism to synchronize both firm's timing.

This algorithm has been proven to be an authentic collusive algorithm.

- 1 Introduction
- 2 Algorithmic Collusion under RL
- 3 Authentic Algorithmic Collusion
- 4 Regulate Algorithmic Collusion**
- 5 Conclusions
- 6 Reference



# Mutual Impact between Algorithmic Collusion and the Market

## The Market on Algorithmic Collusion

- [Miklós-Thal and Tucker, 2019] found that more accurate demand forecasting algorithms can reduce potential opportunities for collusion;
- [Johnson et al., 2023] discovers that by restricting the size of the product assortment, the market price set by the firms using AI algorithm tends to decrease;
- [Xu et al., 2023] verified that the occurrence of algorithmic collusion via Q-learning depends on the platform's recommendation strategy.

## Algorithmic Collusion on the Market

[Yang et al., 2023] shows that the presence of potential algorithmic collusion can prohibit the regulator from deterring price discrimination in some settings.

# Regulatory Measures: 1

Currently, there are three approaches to sustain algorithmic collusion:

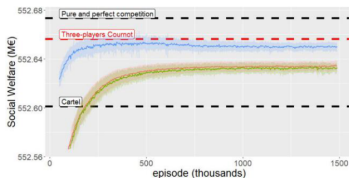
- ① Direct inspection of firms' pricing algorithms;
- ② Intervention in the market, where authorities act as a player with incumbent firms;
- ③ Design market structures that inherently block algorithmic collusion.

The first direction is proposed by [Harrington, 2018] and has been more practically studied by [Hartline et al., 2024].

## Regulatory Measures: 2

[Abada and Lambin, 2023] proposes that the regulator can participate in the market itself.

- Specifically, the regulator can implement a simple reward-punishment strategy: bidding aggressively when the market appears collusive and bidding conservatively when the market appears competitive (this paper studies the circumstance of the electricity trading market).
- Numerical results indicate that this strategy can substantially enhance market welfare. The blue line represents the regulator using the reward-punishment strategy and the green line represents the three-firm competing scenario.

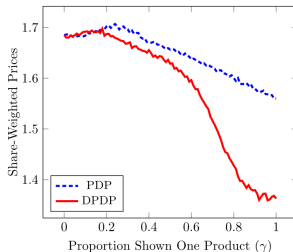


- The authors also test alternative strategies: the regulator can employ the same Q-learning algorithm, but instead of using the payoff as the reward, the regulator updates Q-values using the welfare. However, this approach has little impact on alleviating algorithmic collusion.

## Regulatory Measures: 3

[Johnson et al., 2023] centers on strategies to deter algorithmic collusion for platform owners.

- Consider a platform where multiple firms employ Q-learning to set their prices, while the platform determines the assortment of products recommended to consumers;
- The authors propose two recommendation principles that can deter algorithmic collusion:
  - Limit the size of the assortment.
  - Recommend the lowest-priced product first, and if no other firm offers products with a significantly lower price, then never recommend the products of other firms.



- 1 Introduction
- 2 Algorithmic Collusion under RL
- 3 Authentic Algorithmic Collusion
- 4 Regulate Algorithmic Collusion
- 5 Conclusions**
- 6 Reference

# Tiny & Conclusions

- ① Many numerical studies have demonstrated that Q-learning and some other reinforcement learning algorithms can lead to supra-competitive prices in oligopoly pricing games.
- ② Some theoretical work proves that, under appropriate parameters, Q-learning (as well as some other algorithms) can converge to collusive prices in certain simple settings. Some of the convergences achieved are well-defined equilibriums, whereas the majority are not.
- ③ Empirical evidence has identified that algorithmic pricing (not necessarily RL algorithms) can increase prices under competitive conditions;
- ④ Although the current evidence is incomplete, and the seeming collusion is widely believed to be due to imperfect exploration, it still poses a great threat to social welfare because algorithmic collusion does not need to be a theoretically well-defined equilibrium; it just happens there.
- ⑤ Collusions observed under RL algorithms can be categorized as spurious or genuine based on whether they exhibit a reward-punishment scheme for deviation. However, these types of algorithms can be outperformed by alternative algorithms, making them not likely to be practiced.
- ⑥ Authentic collusion is a robust definition of algorithmic collusion, which must perform well against all alternative algorithms and be the best response toward itself (This has been beyond the scope of reinforcement learning).
- ⑦ Authentic collusive algorithms have been developed, but current versions require to know (or infer) the opponent's demand information.
- ⑧ Algorithmic collusion can be mitigated by the number effect. Regulatory intervention through active participation in the market, employing a reward-punishment strategy, can largely eliminate collusion.

# Further Research: 1

## Numerical Research

- Incorporating constraints such as inventory cost and strategic consumer behaviors can make the simulations more representative of reality;
- Current numerical research only uses 'static strategies', using molecular strategies ([Kimbrough et al., 2005]) can make the results more explainable;
- Can AI algorithm lead to price discrimination and unfair pricing in other environments ([Spann et al., 2024])?
- Algorithmic collusion in other fields ([Cont and Xiong, 2024], [Gu, 2023]).

## Emirical Research

- Identify the number effect;
- The mutual impact between the market structure and algorithmic collusion;
- Current empirical research has identified instances of algorithmic collusion, but the algorithms in these studies are only AI algorithms, not necessarily RL algorithms.

## Theretical Research

- A more rigorous framework to differentiate genuine collusion and authentic collusion;
- Future work could aim to provide more robust proofs of the convergence of Reinforcement Learning algorithms to collusive outcomes;
- Current authentic collusive algorithms need to infer the opponent's demand information; future research could design collusive algorithms without relying on this channel.

## Further Research: 2

### Other research directions:

- Research on the regulating aspect is lacking. Although in settings where a platform acts as an intermediary to determine which products are accessible to consumers, there are relatively practical solutions to deter algorithmic collusion ([Johnson et al., 2023]); However, in settings where products are directly accessible to consumers without platform filtration, regulatory measures have a huge potential to be investigated ([Harrington, 2018], [Abada and Lambin, 2023], [Hartline et al., 2024]);
- The recommender systems described in [Johnson et al., 2023], [Xu et al., 2023] may be more accurately modeled by assortment optimization problems in the OM literature. Under the current assortment optimization model, what will the market dynamics converge to when sellers use RL algorithms to price?
- Future studies could devise the optimal coupon distribution strategy or explore alternative market structure modifications to improve platform payoffs and social welfare.



## Q&amp; A

Any Questions?

Thanks!

- 1 Introduction
- 2 Algorithmic Collusion under RL
- 3 Authentic Algorithmic Collusion
- 4 Regulate Algorithmic Collusion
- 5 Conclusions
- 6 Reference

- [Abada and Lambin, 2023] Abada, I. and Lambin, X. (2023).  
Artificial intelligence: Can seemingly collusive outcomes be avoided?  
*Management Science*.
- [Asker et al., 2022] Asker, J., Fershtman, C., and Pakes, A. (2022).  
Artificial intelligence, algorithm design, and pricing.  
In *AEA Papers and Proceedings*, volume 112, pages 452–56.
- [Asker et al., 2023] Asker, J., Fershtman, C., and Pakes, A. (2023).  
The impact of artificial intelligence design on pricing.  
*Journal of Economics & Management Strategy*.
- [Banchio and Skrzypacz, 2022] Banchio, M. and Skrzypacz, A. (2022).  
Artificial intelligence and auction design.  
In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 30–31.
- [Brown and MacKay, 2023] Brown, Z. Y. and MacKay, A. (2023).  
Competition in pricing algorithms.  
*American Economic Journal: Microeconomics*, 15(2):109–156.
- [Calvano et al., 2019] Calvano, E., Calzolari, G., Denicolò, V., and Pastorello, S. (2019).  
Algorithmic pricing what implications for competition policy?  
*Review of industrial organization*, 55:155–171.

- [Calvano et al., 2020] Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2020).  
Artificial intelligence, algorithmic pricing, and collusion.  
*American Economic Review*, 110(10):3267–3297.
- [Calvano et al., 2023] Calvano, E., Calzolari, G., Denicoló, V., and Pastorello, S. (2023).  
Algorithmic collusion: Genuine or spurious?  
*International Journal of Industrial Organization*, 90:102973.
- [Cho and Williams, 2024] Cho, I.-K. and Williams, N. (2024).  
Collusive outcomes without collusion: Algorithmic pricing in a duopoly model.  
*Available at SSRN 4753617*.
- [Clark et al., 2023] Clark, R., Assad, S., Ershov, D., and Xu, L. (2023).  
Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market.  
*Journal of Political Economy*.
- [Cont and Xiong, 2024] Cont, R. and Xiong, W. (2024).  
Dynamics of market making algorithms in dealer markets: Learning and tacit collusion.  
*Mathematical Finance*, 34(2):467–521.

[den Boer, 2015] den Boer, A. V. (2015).

Dynamic pricing and learning: Historical origins, current research, and new directions.

*Surveys in operations research and management science*, 20(1):1–18.

[den Boer, 2023] den Boer, A. V. (2023).

Algorithmic collusion: A mathematical definition and research agenda for the OR/MS community.

*Available at SSRN 4636488*.

[den Boer et al., 2022] den Boer, A. V., Meylahn, J. M., and Schinkel, M. P. (2022).

Artificial collusion: Examining supracompetitive pricing by Q-learning algorithms. *Amsterdam Law School Research Paper*, (2022-25).

[Dolgoplov, 2024] Dolgoplov, A. (2024).

Reinforcement learning in a prisoner's dilemma.

*Games and Economic Behavior*, 144:84–103.

[Epivent and Lambin, 2024] Epivent, A. and Lambin, X. (2024).

On algorithmic collusion and reward-punishment schemes.

*Economics Letters*, page 111661.

[Gu, 2023] Gu, C. (2023).

Can dynamic pricing algorithm facilitate tacit collusion? an experimental study using deep reinforcement learning in airline revenue management.

- [Hansen et al., 2021] Hansen, K. T., Misra, K., and Pai, M. M. (2021).  
Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms.  
*Marketing Science*, 40(1):1–12.
- [Harrington, 2018] Harrington, J. E. (2018).  
Developing competition law for collusion by autonomous artificial agents.  
*Journal of Competition Law & Economics*, 14(3):331–363.
- [Hartline et al., 2024] Hartline, J. D., Long, S., and Zhang, C. (2024).  
Regulation of algorithmic collusion.  
In *Proceedings of the Symposium on Computer Science and Law*, pages 98–108.
- [Hettich, 2021] Hettich, M. (2021).  
Algorithmic collusion: Insights from deep learning.  
*Available at SSRN 3785966*.
- [Johnson et al., 2023] Johnson, J. P., Rhodes, A., and Wildenbeest, M. (2023).  
Platform design when sellers use pricing algorithms.  
*Econometrica*, 91(5):1841–1879.
- [Kimbrough et al., 2005] Kimbrough, S. O., Lu, M., and Murphy, F. (2005).  
Learning and tacit collusion by artificial agents in cournot duopoly games.  
*Formal modelling in electronic commerce*, pages 477–492.

- [Klein, 2021] Klein, T. (2021).  
Autonomous algorithmic collusion: Q-learning under sequential pricing.  
*The RAND Journal of Economics*, 52(3):538–558.
- [Lamba and Zhuk, 2022] Lamba, R. and Zhuk, S. (2022).  
Pricing with algorithms.  
*arXiv preprint arXiv:2205.04661*.
- [Loots and den Boer, 2023] Loots, T. and den Boer, A. V. (2023).  
Data-driven collusion and competition in a pricing duopoly with multinomial logit demand.  
*Production and Operations Management*, 32(4):1169–1186.
- [Meylahn and V. den Boer, 2022] Meylahn, J. M. and V. den Boer, A. (2022).  
Learning to collude in a pricing duopoly.  
*Manufacturing & Service Operations Management*, 24(5):2577–2594.
- [Miklós-Thal and Tucker, 2019] Miklós-Thal, J. and Tucker, C. (2019).  
Collusion by algorithm: Does better demand prediction facilitate coordination between sellers?  
*Management Science*, 65(4):1552–1561.



[Musolff, 2022] Musolff, L. (2022).  
 Algorithmic pricing facilitates tacit collusion: Evidence from e-commerce.  
*In Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 32–33.

[Possnig, 2023] Possnig, C. (2023).  
*Reinforcement Learning and Collusion*.  
 Department of Economics, University of Waterloo.

[Shi and Zhang, 2020] Shi, Y. and Zhang, B. (2020).  
 Multi-agent reinforcement learning in cournot games.  
*In 2020 59th IEEE Conference on Decision and Control (CDC)*, pages 3561–3566.  
 IEEE.

[Spann et al., 2024] Spann, M., Bertini, M., Koenigsberg, O., Zeithammer, R., Aparicio, D., Chen, Y., Fantini, F., Jin, G. Z., Morwitz, V., Leszczyc, P. P., et al. (2024).  
 Algorithmic pricing: Implications for consumers, managers, and regulators.  
 Technical report, National Bureau of Economic Research.

[Wieting and Sapi, 2021] Wieting, M. and Sapi, G. (2021).  
 Algorithms in the marketplace: An empirical analysis of automated pricing in e-commerce.  
*Available at SSRN 3945137*.

- [Xu et al., 2023] Xu, X., Lee, S., and Tan, Y. (2023).  
 Algorithmic collusion or competition: the role of platforms' recommender systems.  
*arXiv preprint arXiv:2309.14548.*
- [Yang et al., 2023] Yang, Z., Lei, X., and Gao, P. (2023).  
 Regulating discriminatory pricing in the presence of tacit collusion.  
*Available at SSRN.*