

ALGORITHMIC COLLUSION BY REINFORCEMENT-LEARNING PRICING AGENTS: SIMULATION EVIDENCE AND IMPLICATIONS FOR FINANCIAL MARKETS AND COMPETITION LAW

Dolgov V.L., Moscow Technological Institute (MTI), Moscow
vladikd4@gmail.com

Abstract. Pricing algorithms built on reinforcement learning can reach supra-competitive outcomes with no agreement and no communication between the firms that deploy them. I report an independent simulation of Q-learning agents in a Bertrand oligopoly with logit demand and connect the results to competition policy for financial markets. Across 32 duopoly sessions the mean collusion index is 0.778, with a 95 per cent confidence interval of [0.735, 0.821] and per-firm profits 40 per cent above the Bertrand-Nash benchmark. Not one session ended at the competitive outcome. In a separate experiment on a common price grid, the index falls from 0.807 with two agents to 0.577 with three and 0.471 with four, while the relative profit gain over the competitive benchmark rises from 40 to 69 per cent. A forced deviation to the competitive price triggers a price war in the next period and a return to the previous level within five to seven periods. Removing the rival's past price from the state space eliminates the effect: the index drops to -0.018, with a confidence interval that contains zero, and prices settle at the static equilibrium. Coordination here follows from what the algorithm observes. Neither Section 1 of the Sherman Act nor Article 101 of the Treaty on the Functioning of the European Union reaches conduct of this kind, and the RealPage proposed final judgment of November 2025 governs data inputs and tool design, leaving autonomous learning untouched.

Keywords: algorithmic collusion; reinforcement learning; Q-learning; tacit collusion; competition policy; financial markets; market making; antitrust.

JEL classification: L41; K21; D43; G14; C63.

1. Introduction

Competition law in the United States and the European Union rests on the concept of an agreement. Section 1 of the Sherman Act prohibits contracts, combinations and conspiracies in restraint of trade [1]. Article 101 of the Treaty on the Functioning of the European Union prohibits agreements and concerted practices between undertakings [2]. Parallel conduct without such an agreement escapes both prohibitions, and the drafters had in mind a world where coordination required people to talk to each other.

Six years of research have undermined that assumption. Calvano and co-authors placed Q-learning agents in a repeated pricing game and watched them settle above the competitive price and punish deviations, with no instruction to coordinate [3]. Klein confirmed the result under sequential moves [4]. Large language models acting as pricing agents reach supra-competitive prices within a few rounds [5], algorithmic market makers in securities markets charge markups over the competitive spread [6], and reinforcement-learning speculators sustain collusive profits in asset markets [7]. The one large-scale empirical study, on German retail gasoline, found that stations adopting algorithmic pricing raised margins by roughly 9 per cent on average, and that market-level margins in duopolies rose by about 38 per cent once both stations had adopted [8].

Financial markets combine the conditions that make this behaviour likely. Quotes and transaction prices are public within milliseconds, the same participants meet again within the hour, and a single instrument may have only a handful of significant liquidity providers. Gauci argues that the resulting conduct is economically indistinguishable from cartel behaviour yet falls outside Article 101 because the doctrine demands concertation, and he treats the problem as doctrinal rather than evidentiary [9]. Supervisors approach the same phenomenon from a different direction, treating correlated algorithmic decisions as a source of systemic vulnerability [10, 11].

This article adds two things to that literature. I ran an independent simulation instead of relying on published results, which lets me measure the effect of market structure and of the agent's information set under one specification and report confidence intervals for both. I then added a treatment that no previous study reports: the rival's past price is removed from the agent's state, leaving everything else unchanged. That treatment isolates the mechanism, and it points to a remedy a regulator could write down and verify. The legal discussion in Section 6 works through what current doctrine reaches, using the RealPage proposed final judgment of November 2025 [12] as the outer edge of enforcement practice.

2. Literature review

2.1. Simulation evidence

Calvano, Calzolari, Denicolo and Pastorello built the reference specification: Q-learning agents, a repeated Bertrand game with logit demand, and a normalised index of collusion [3]. Their agents converged above the static equilibrium and adopted strategies with a punishment phase. The paper is six years old, which in this field is not recent, but the environment and the

index it introduced remain the standard against which later work is measured, and I use both here. Klein extended the setting to sequential pricing and found that supra-competitive outcomes survive alternating moves [4]. Brown and MacKay showed that differences in pricing frequency raise equilibrium prices with no learning dynamics at all, so algorithmic pricing raises competitive concerns through more than one channel [13].

One recent result complicates the standard account. Arunachaleswaran and co-authors prove that near-monopoly prices can arise as a Nash equilibrium between a no-regret learner and a best-responding competitor, with no threat or punishment strategy anywhere in the construction [14]. Their conclusion is that the regulatory definition of algorithmic collusion should not be tied to threat-based behaviour. I return to this in Section 4.4, because my own mechanism result runs in the opposite direction within my setting. Bichler, Durmann and Oberlechner survey the computer science and economics strands and note that theory still lags well behind the volume of experimental evidence [15]. Hartline, Long and Zhang come at the problem from mechanism design and ask what a regulator could require of an algorithm to rule out collusive outcomes in the first place [16, 17].

2.2. Large language models as pricing agents

Fish, Gonczarowski and Shorrer replaced the tabular learner with large language models told to maximise long-run profit in a repeated Bertrand oligopoly [5]. Under one prompt, prices settled around 1.89 against a Bertrand-Nash level near 1.47 and a monopoly level near 1.92, with profits around 33 against a competitive benchmark near 22. A second prompt, differing only in a sentence about undercutting, produced an average near 1.69. Two results carry over to policy. The models reached supra-competitive prices within a few rounds, since the capability comes from pretraining rather than from training on the market, and the gap between the two prompts reached 12 to 15 per cent. A firm that deploys such a model has little basis to predict which outcome it will get.

2.3. Empirical evidence

Assad, Clark, Ershov and Xu ran the first empirical test of the relationship between algorithmic pricing and competition, identifying adopting stations in the German retail gasoline market through structural break tests [8]. Station-level margins rose by about 9 per cent on average after adoption. The market-level picture is sharper and more instructive: in duopolies where one station adopted, margins did not move, and where both adopted, market-level

margins rose by 3.2 cents per litre, close to 38 per cent. Adoption by a single firm changes little. The effect belongs to the market.

2.4. Financial market evidence

Colliard, Foucault and Lovo study Q-learning market makers facing adverse selection [6]. Their algorithms handle adverse selection as theory predicts and still charge a markup over the competitive price, and the markup falls off slowly: competitive pricing arrives at roughly ten algorithmic market makers, against the two that theory requires. Dou, Goldstein and Ji model informed speculators trading on reinforcement learning and separate two mechanisms [7]. The first is a price-trigger strategy that emerges when noise trading risk is low. The second comes from a learning bias: adverse outcomes cut the estimated value of aggressive strategies more sharply than favourable outcomes raise it, so the algorithm prunes aggressive trading from its policy across the whole parameter range. Price informativeness and liquidity fall under both.

2.5. Legal literature

The OECD background note of 2023 set out the categories of concern [18], and the 2025 report on G7 jurisdictions supplies the comparative picture [19]. Authorities work with four scenarios: algorithms implementing an explicit agreement, hub-and-spoke arrangements built on a shared third-party algorithm, vertical arrangements, and tacit collusion produced by autonomous learning. Enforcement has reached the first three. On the fourth, the OECD records that the absence of case practice makes it hard to judge how relevant the scenario is, and that accountability for autonomous algorithmic behaviour remains unresolved without precedent [19]. Gauci reaches the same conclusion for EU financial services and maps the instruments that might close the gap, among them MiFID II, the Market Abuse Regulation and the AI Act [9]. The European debate is collected in a 2026 dossier of the Concurrences review [20]. Table 1 summarises the studies discussed above.

Table 1. Selected studies of algorithmic collusion

Study	Method	Setting	Main result
Calvano et al. (2020) [3]	Q-learning simulation	Bertrand duopoly, logit demand	Supra-competitive prices sustained by punishment strategies
Klein (2021) [4]	Q-learning simulation	Sequential pricing	Supra-competitive outcomes survive alternating moves
Brown & MacKay (2023) [13]	Theory and empirics	Online retail	Differences in pricing frequency raise equilibrium prices
Assad et al. (2024) [8]	Empirical, structural breaks	German retail gasoline	Market-level margins up about 38% once both duopolists adopt

Study	Method	Setting	Main result
Arunachaleswaran et al. (2025) [14]	Theory, no-regret learning	Sequential Bertrand pricing	Near-monopoly prices without threats or punishment
Fish et al. (2026) [5]	LLM agents experiment	Repeated Bertrand oligopoly	Prices near monopoly level; outcome sensitive to prompt wording
Dou et al. (2025) [7]	Theory and simulation	Informed speculation in asset markets	Two mechanisms; price informativeness and liquidity decline
Colliard et al. (2026) [6]	Q-learning simulation	Securities market making	Markups persist; competition needs about ten market makers
Gauci (2026) [9]	Doctrinal analysis	EU financial services law	Parallel algorithmic conduct falls outside Article 101

Source: author's compilation from [3, 4, 5, 6, 7, 8, 9, 13, 14].

3. Methodology

3.1. Economic environment

The environment follows Calvano and co-authors [3]. Each of n firms posts a price for a differentiated product in every period. Demand takes the logit form: the quantity sold by firm i depends on the exponential of the difference between the product quality index and its own price, scaled by a horizontal differentiation parameter, relative to the same expression summed over all products and the outside option. Quality equals 2 for all products, the outside option is 0, the differentiation parameter is 0.25, and marginal cost is 1 for every firm.

I computed both benchmarks numerically instead of copying them from the source, iterating best responses for the static equilibrium and maximising symmetric joint profit for the collusive one. For the duopoly the Bertrand-Nash price is 1.473 and the fully collusive price is 1.925; per-firm profits are 0.2229 and 0.3375. With three firms the two prices are 1.370 and 2.000, and with four firms 1.331 and 2.054.

The action set is a grid of k prices that spans the interval between the benchmarks and extends past each by ten per cent of the distance between them. The baseline uses $k = 15$, as in the reference specification [3]. One consequence matters for reading the results: the lowest grid point sits at 1.428, below the static equilibrium of 1.473, so an agent pricing at the competitive level lands on a grid point slightly below it and the collusion index defined in Section 3.3 can take small negative values.

3.2. Learning agents

Each firm is an independent Q-learning agent in the formulation of Watkins and Dayan [21], updating the estimated value of a state-action pair towards the realised profit plus the discounted maximum estimated value of the successor state; the treatment follows Sutton and Barto [22]. I cite both as sources for the method, not as recent literature. The learning rate is

0.15 and the discount factor 0.95 in the baseline. Action selection is epsilon-greedy, with the exploration probability decaying exponentially in the period index at rate β , set to 0.00001. Values start at the discounted average profit over all action profiles, an optimistic initialisation that keeps the agent exploring early on.

The state is the vector of prices posted by all firms in the previous period. An agent can therefore condition its action on what its rival did, and a punishment strategy becomes representable. Section 4.4 reports the alternative: the state holds only the agent's own previous price, and the agent has no way to detect a deviation.

Agents learn nothing about the payoff function, the number of competitors or the structure of demand. They do not observe rival profits. No communication channel exists between them. The only feedback is the profit realised in the period.

3.3. Measure of collusion

The outcome measure is the collusion index from the reference literature [3]: average per-firm profit after learning, minus the static Bertrand-Nash profit, divided by the difference between the fully collusive and the Bertrand-Nash profit. Zero marks the competitive outcome and one marks full collusion. Profits come from 50,000 periods of purely greedy play after training, with exploration switched off, so the index reflects the learned policy and not the exploration path.

3.4. Experimental design

I ran five experiments. The baseline consists of 32 independent duopoly sessions. The market structure experiment compares two, three and four agents on a common grid of 11 prices, which keeps the comparison internally consistent as the state space grows with the number of firms¹. Two robustness experiments vary the learning rate and the exploration decay rate. The state space experiment removes the rival's price from the state. Table 2 lists the parameters. Sessions within an experiment run in parallel from independent random seeds; the code is Python with numpy vectorisation across sessions².

Table 2. Parameters of the simulation experiments

Experiment	Agents	Price grid	Sessions	Training periods	Varied parameter
Baseline	2	15	32	1 000 000	None

¹ The state space grows as the number of price points raised to the power of the number of firms, so a 15-point grid with four agents implies just over 50,000 states per agent. The 11-point grid keeps memory requirements and convergence times comparable across the three market structures.

² Session-level results, the random seeds, the exact package versions and the scripts that produce every figure in this article are in the repository named in the data availability statement.

Experiment	Agents	Price grid	Sessions	Training periods	Varied parameter
Market structure	2, 3, 4	11	16 each	1 000 000	Number of agents
Exploration decay	2	15	16 each	1 500 000	Decay rate beta
Learning rate	2	15	16 each	1 000 000	Learning rate alpha
State space	2	15	16	1 000 000	Content of the state

Source: author's compilation.

4. Results

4.1. Baseline duopoly

Figure 1 traces the average posted price through learning. Prices drop below the starting level while the agents explore, then climb as the exploration probability decays, and settle well above the static equilibrium. The terminal average of 1.764 sits at 64 per cent of the distance between the Bertrand-Nash price and the fully collusive price.

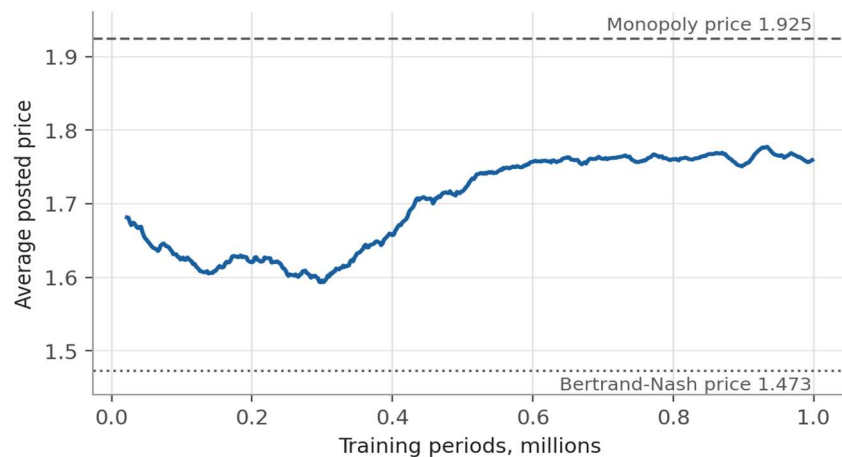


Figure 1. Average posted price through learning, mean of 32 duopoly sessions

Source: own simulation results.

Figure 2 shows how the collusion index is distributed across the 32 sessions. The mean is 0.778 with a standard deviation of 0.123, which gives a 95 per cent confidence interval of [0.735, 0.821]. The minimum across sessions is 0.549 and the maximum 0.996. No session converged to the competitive outcome. In profit terms the mean corresponds to 0.312 per firm against the competitive benchmark of 0.223, a gain of 40 per cent. Table 3 collects the descriptive statistics.

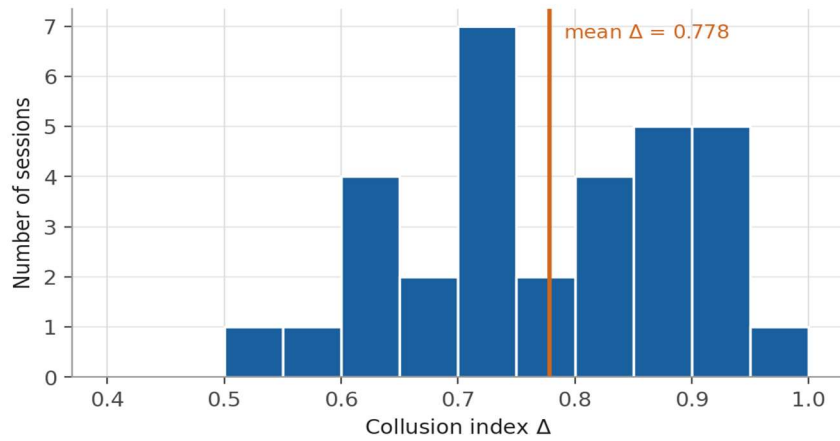


Figure 2. Distribution of the collusion index across 32 baseline sessions

Source: own simulation results.

Table 3. Baseline duopoly results, 32 sessions

Indicator	Value
Bertrand-Nash price / profit	1.473 / 0.2229
Fully collusive price / profit	1.925 / 0.3375
Mean posted price after learning	1.764
Mean collusion index	0.778
Standard deviation of the index	0.123
95% confidence interval of the mean	[0.735, 0.821]
Minimum / maximum index	0.549 / 0.996
Share of sessions with index above 0.5	100%
Implied profit gain over the competitive benchmark	40.0%

Source: own simulation results.

4.2. Market structure

Figure 3 reports the collusion index for two, three and four agents. The index falls from 0.807 in the duopoly, with a confidence interval of [0.751, 0.863], to 0.577 with three agents, interval [0.507, 0.647], and 0.471 with four, interval [0.442, 0.500]. The intervals do not overlap, and a Welch test on the duopoly against the four-agent case gives $t = 10.5$ with 22.5 degrees of freedom. The share of sessions above half the monopoly rent falls from 100 per cent to 81 and then to 25. The direction reproduces the finding in the reference literature that extra participants make coordination harder to hold [3].

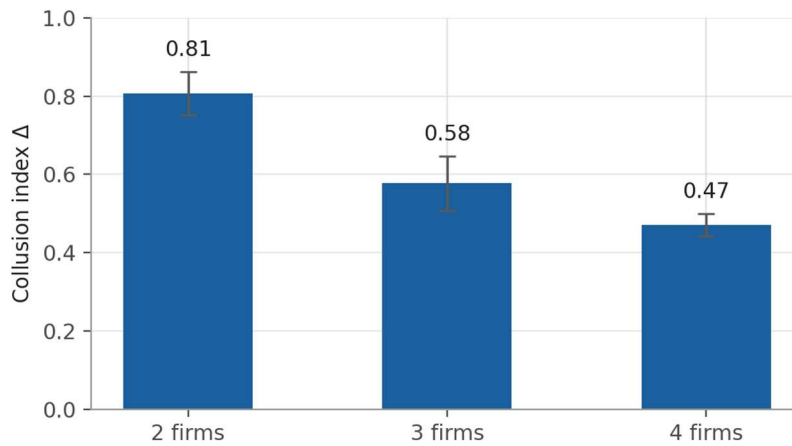


Figure 3. Collusion index by number of competing agents, 16 sessions per configuration, whiskers show 95 per cent confidence intervals

Source: own simulation results.

Read on its own, the normalised index understates what happens in less concentrated markets. Competitive profit falls as firms are added, so the same index translates into a larger relative gain: profits exceed the competitive benchmark by 62 per cent with three agents and 69 per cent with four, against 40 per cent in the duopoly. Concentration governs how close the outcome comes to the monopoly rent. It does not govern how much the participants gain relative to competition, and on that second measure the less concentrated markets look worse.

4.3. Robustness

Table 4 reports the sensitivity of the baseline to the learning rate and to the exploration schedule, together with the state space treatment discussed in the next subsection. The first row is a 16-session run under baseline parameters, which serves as the comparison base for every treatment in the table; its mean of 0.776 differs from the 0.778 of Section 4.1 because that figure comes from the larger 32-session run under a different seed. Across the five parameter configurations the index stays between 0.653 and 0.776, and in each of them at least 88 per cent of sessions clear half the monopoly rent. A faster decay of exploration shortens the learning phase and lowers the index to 0.673. The slowest decay widens dispersion, since some sessions had not fully converged within the training horizon. None of this depends on a lucky choice of learning parameters.

Table 4. Sensitivity of the collusion index to learning parameters and to the state space

Configuration	Mean index	95% CI	Std. dev.	Mean price	Share above 0.5
Baseline, 16 sessions (alpha 0.15, beta 0.00001)	0.776	[0.721, 0.831]	0.112	1.760	100%
Learning rate 0.05	0.653	[0.613, 0.694]	0.083	1.682	94%

Configuration	Mean index	95% CI	Std. dev.	Mean price	Share above 0.5
Learning rate 0.30	0.705	[0.626, 0.784]	0.162	1.735	88%
Exploration decay 0.00005, 1.5 m periods	0.673	[0.610, 0.736]	0.129	1.704	88%
Exploration decay 0.00001, 1.5 m periods	0.768	[0.705, 0.831]	0.129	1.750	100%
Exploration decay 0.000004, 1.5 m periods	0.724	[0.630, 0.817]	0.191	1.730	88%
Own past price only in the state	-0.018	[-0.062, 0.026]	0.090	1.469	0%

Source: own simulation results. The three exploration rows share a 1.5 million period horizon and are compared with each other; the remaining rows use 1 million periods.

4.4. The mechanism: punishment and the state space

Two experiments get at what sustains the outcome. The first traces the response to a deviation. After training, one agent is forced for a single period to post the grid price closest to the static equilibrium, and both agents then follow their learned policies. Figure 4 shows the average path across sessions. The rival cuts its own price from 1.779 to 1.545 in the following period, both agents stay below the pre-deviation level for several periods, and prices recover to roughly 1.75 within five to seven periods. A deviation draws a temporary price war, after which the previous level returns. Neither agent was instructed to punish, and neither can send a message to the other.

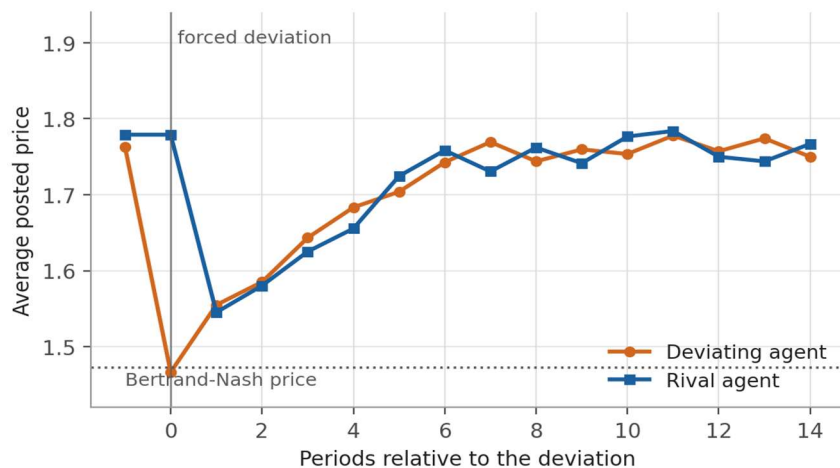


Figure 4. Average price paths around a forced deviation, 32 sessions

Source: own simulation results.

I did not test this interpretation formally against the alternative that the paths simply reflect movement through the learned policy from an off-equilibrium state, and the work of Arunachaleswaran and co-authors is a reminder that supra-competitive prices do not require a

threat structure at all [14]. What Figure 4 establishes is narrower: in this environment, deviation is followed by a targeted price cut and then by recovery.

The second experiment removes the rival's past price from the state, leaving each agent with only its own previous price. Every other parameter is unchanged and the seed matches the corresponding baseline run. The collusion index falls from 0.776 to -0.018, with a confidence interval of $[-0.062, 0.026]$ that contains zero, and a Welch test on the two groups gives $t = 22.0$. The mean price falls from 1.760 to 1.469 against a static equilibrium of 1.473, the small gap reflecting the grid point at 1.466 that lies just below the equilibrium. None of the 16 sessions ended above half the monopoly rent. Figure 5 places the two treatments side by side.

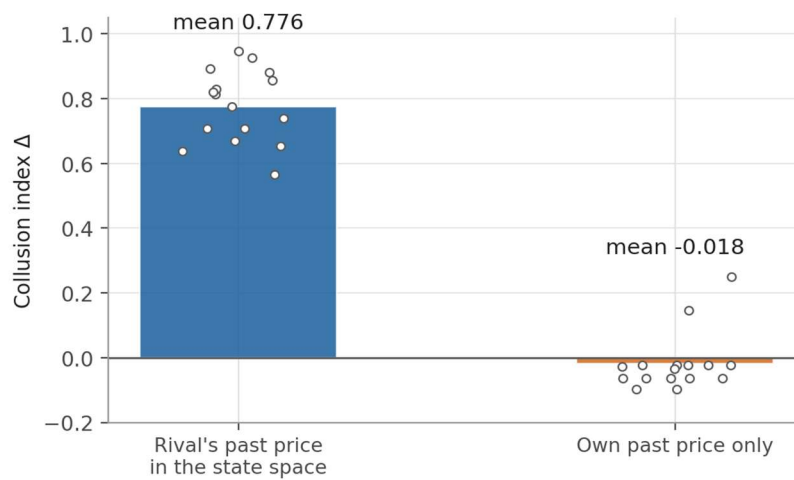


Figure 5. Collusion index with and without the rival's past price in the state space, 16 sessions each

Source: own simulation results.

Supra-competitive pricing is therefore not a general property of reinforcement learning applied to a pricing problem. In this environment it depends on the agent observing rival conduct, which is what makes a punishment strategy representable. An agent that cannot detect a deviation cannot threaten one, and its learned policy converges to the static equilibrium. Section 6 takes up what follows for regulation.

5. Implications for financial markets

Financial markets meet the structural conditions under which the effect documented above is most likely to appear. Quotes and transaction prices are published within milliseconds, so the observation of rival conduct that Section 4.4 identifies as the operative mechanism comes free. The same participants interact many times a day, so the number of periods needed for convergence maps onto a short calendar interval. Reliance on similar model architectures and

on a concentrated set of technology vendors, documented by the Financial Stability Board [10, 11], raises the chance that participants run structurally similar algorithms.

One link in that chain deserves more caution than the others. My market structure experiment covers two to four agents, and whether the number of economically significant liquidity providers in a given instrument falls in that range is an empirical question I do not settle here. The relevant benchmark comes from Colliard, Foucault and Lovo, whose algorithmic market makers reach competitive pricing only at around ten participants [6]. That threshold, rather than the textbook duopoly, is what a supervisor would need to check instrument by instrument.

The evidence that does exist for financial markets points the same way. Colliard and co-authors find markups that decline slowly in the number of algorithmic market makers [6]. Dou, Goldstein and Ji find supra-competitive profits among reinforcement-learning speculators, and one of their two mechanisms operates across the whole parameter range, which limits what an intervention aimed at a single channel can achieve [7].

Three areas of financial activity look exposed. Market making and liquidity provision come first, where the price variable is the spread and the participant count in a given instrument is often small. Retail lending and insurance follow, where algorithmic underwriting sets the rate offered to comparable applicants and tacit coordination shows up as a higher price of credit instead of a wider spread. Fee setting on retail investment platforms belongs in the same group, with public prices that adjust often.

The consequences run past the transfer from customers to intermediaries. Collusion among informed traders reduces price informativeness and liquidity [7], and supervisors already treat correlated algorithmic decisions as a systemic vulnerability [10, 11, 23]. The International Monetary Fund has recommended recalibrating circuit breakers and margin requirements and widening the monitoring of large traders, given the speed at which algorithmic strategies now operate [24]. The European Central Bank points to the concentration and interconnectedness that come with reliance on a small group of technology providers [25]. Coordination without agreement therefore sits between two regulatory domains: a competition problem in its distributive effect, a market quality problem in its effect on price formation. Neither competition authorities nor financial supervisors have claimed it.

6. Implications for competition law

The conduct examined here escapes the central prohibitions of both systems. Section 1 of the Sherman Act requires a contract, combination or conspiracy [1], and Article 101 requires

an agreement, a decision of an association or a concerted practice [2]. In the simulation there is no communication, no exchange of information, no common instruction. Each firm runs an algorithm that maximises its own profit from its own realised payoffs. The result is supra-competitive pricing held in place by a punishment strategy, which is the economic substance of a cartel, produced by conduct that satisfies no element of either prohibition.

Case law makes the gap concrete. In *Twombly* the Supreme Court held that allegations of parallel conduct do not state a claim under Section 1 without factual matter suggesting an agreement, which leaves a plaintiff facing independently deployed algorithms with nothing to plead [26]. European doctrine reaches a similar place from the other direction. In *Wood Pulp II* the Court of Justice held that parallel behaviour cannot be treated as proof of concertation unless concertation is the only plausible explanation [27], and in *T-Mobile Netherlands* the Court confirmed that a concerted practice requires contact between undertakings, however brief [28]. Independent learning from public prices supplies an obvious alternative explanation and involves no contact at all. Gauci draws the same conclusion for EU financial services and argues that the difficulty is doctrinal, not evidentiary [9].

The 2025 OECD review of G7 jurisdictions confirms that no enforcement practice has formed on the fourth scenario [19]. Authorities have proceeded under traditional theories: algorithms used to implement an explicit agreement, as in the posters and frames case decided by the United Kingdom authority in 2016 [29]; hub-and-spoke arrangements built on a shared third-party algorithm; and vertical arrangements. The review records that accountability for autonomous algorithmic behaviour remains unresolved in the absence of precedent [19].

The proposed final judgment filed by the United States Department of Justice against RealPage on 24 November 2025 marks the current frontier of enforcement and shows where it stops [12]³. The theory of harm was the unlawful sharing of competitively sensitive information among competitors and the alignment of pricing that removes independent centres of decision-making. The remedy bars the use of non-public competitor data in the real-time operation of the recommendation engine, bars automatic implementation of recommendations without an override, and bars product features that systematically favour price increases over decreases; competitor data at least twelve months old may still be used for model training, subject to carve-outs. The judgment governs the data that enters the algorithm and the design of the tool. An

³ Docket entries for the RealPage defendant show a stipulation and proposed final judgment filed on 24 November 2025. A final judgment against a co-defendant, Greystar Management Services, was entered on 2 March 2026; as of September 2026 the docket records no entry of final judgment against RealPage itself.

algorithm that learns from its own payoffs, which is the configuration examined here, lies outside it.

Table 5 sets the two regimes side by side. The European Union has adopted a horizontal regulation on artificial intelligence [30], with obligations for stand-alone high-risk systems postponed to 2 December 2027 and for systems embedded in products to 2 August 2028 under the simplification package agreed in May 2026 [31]. Pricing algorithms are not classified as high-risk on the basis of their competitive effects, and the regulation addresses safety, transparency and fundamental rights, not market conduct.

Table 5. Treatment of algorithmic pricing under the two principal competition regimes

Dimension	United States	European Union
Legal basis	Section 1 of the Sherman Act [1]	Article 101 TFEU [2]
Required element	Contract, combination or conspiracy	Agreement or concerted practice
Leading authority on parallel conduct	Twombly [26]	Wood Pulp II [27]; T-Mobile Netherlands [28]
Algorithm implementing an agreement	Covered	Covered
Hub-and-spoke via common vendor	Covered; RealPage proposed judgment [12]	Covered in principle; limited practice [19]
Autonomous tacit collusion	Not covered	Not covered [9]
Regulation of data inputs	Established for pricing tools by settlement terms [12]	Information exchange doctrine applies; nothing specific to pricing algorithms
Sector-specific instrument	None specific to pricing algorithms	AI Act, high-risk obligations from December 2027, not framed around competition [30, 31]

Source: author's compilation from [1, 2, 9, 12, 19, 26, 27, 28, 30, 31].

Three responses follow from Section 4. Regulation of inputs is the one already in use, along the lines of the RealPage judgment, and it needs no legislative change. Its limit shows up in my own results: the agents here never received competitor data as an input and coordinated anyway, because posted prices are observable in the market itself.

Regulation of the state space follows from the experiment in Figure 5. If the collusive outcome depends on the agent conditioning its action on rival conduct, a requirement that pricing algorithms in designated markets exclude rival price histories from their decision state removes the mechanism instead of the symptom. An inspector can verify compliance from the model specification, which is not true of an assessment of intent, and the requirement can be tested before deployment. Related proposals for constraining algorithmic behaviour by design have been worked out in the mechanism design literature [16, 17].

This remedy has a weakness that deserves stating. An algorithm barred from observing rival prices still observes its own sales, and realised demand carries information about what

rivals did. A sufficiently rich learner might reconstruct the rival's behaviour indirectly and rebuild the punishment channel. My agents received exactly that feedback, in the form of realised profit, and did not learn to coordinate without the rival's price in the state, which suggests the indirect channel is weak for tabular learners in this environment. Whether it stays weak for richer function approximators is untested. The rule also costs responsiveness to genuine market movements, which argues for narrow markets instead of a general prohibition.

Detection based on conduct is the third option and the cheapest to start. Financial supervisors already receive granular quote and transaction data, and that data carries the signature in Figure 4: a deviation, a punishment phase, a return to the previous level within a small number of periods. An analyst can see the pattern without access to the algorithm, and it separates coordinated behaviour from parallel responses to a common shock, since a common shock does not produce a response aimed at one participant's deviation. Screening of this kind sits closer to existing market abuse surveillance than to a conventional antitrust investigation, which puts financial supervisors in a better position than competition authorities to notice it first.

7. Limitations

The results come from a stylised model. Demand is stationary, costs are symmetric, products are symmetrically differentiated, and the action space is a coarse grid of fifteen prices. Real pricing problems involve shifting demand, asymmetric costs and far larger action spaces. The agents remember one period; longer memory would enlarge the strategy space, and I do not examine its effect.

I assess convergence through the stability of the greedy policy over the evaluation window instead of a formal criterion on the value function. The configuration with the slowest exploration decay may not have converged fully within its training horizon, which shows up as the wider dispersion in Table 4. With 16 to 32 sessions per configuration the confidence intervals are wide enough that small differences between parameter settings should not be read as effects.

Section 5 argues from structural analogy, not from evidence. I show that a particular class of algorithms coordinates in a particular environment and identify the mechanism that sustains it in that environment; I do not show that such coordination is present in any real market, financial or otherwise. The empirical record on that question still rests on the retail markets studied by Assad and co-authors [8]. The mechanism result is also specific to this

setting, and the theoretical work on collusion without threats suggests that removing the punishment channel need not remove supra-competitive pricing in every environment [14].

8. Conclusion

Independent Q-learning agents pricing in a Bertrand oligopoly reach outcomes that are collusive in substance and lawful in form. Across 32 duopoly sessions the mean collusion index was 0.778, with a confidence interval of [0.735, 0.821] and profits 40 per cent above the competitive benchmark, and no session produced a competitive outcome. Adding participants lowers the index, to 0.577 with three agents and 0.471 with four, while the relative profit gain rises. Concentration shapes how much of the monopoly rent the agents capture, not whether they beat competition.

The mechanism was identified rather than assumed. A forced deviation draws a price war and a return to the previous level within five to seven periods, and removing the rival's past price from the state eliminates the effect: the index falls to -0.018 and prices settle at the static equilibrium. In this environment, coordination follows from what the algorithm observes.

Financial markets supply the conditions the effect needs. Prices are public, interaction is frequent, reliance on similar models and common vendors is documented by supervisors [10, 11], and the existing work on algorithmic market makers and reinforcement-learning speculators points the same way [6, 7]. Competition law does not reach the conduct in either principal jurisdiction, and the RealPage judgment, the current frontier of enforcement, governs data inputs and tool design [12].

Two responses look workable. A constraint on what a pricing algorithm may observe attacks the mechanism and can be verified from the model specification. Supervisory screening for the deviation and punishment signature uses data that supervisors already collect. Further work should test longer memory specifications and heterogeneous algorithms competing against one another, and calibrate the environment to a specific market, starting with dealer quotes in over-the-counter instruments where participants are few and quotations are observable.

Data and code availability

The simulation code, the session-level results behind every table and figure, the random seeds and the scripts that generate the figures are available at <https://github.com/Vldslv4/algorithmic-collusion-ql> under the MIT licence. Running the reproduction script regenerates all reported numbers from scratch.

References

- [1] Sherman Antitrust Act, 15 U.S.C. Section 1 (1890).
- [2] Consolidated Version of the Treaty on the Functioning of the European Union, Article 101. Official Journal of the European Union, C 326, 26 October 2012.
- [3] Calvano, E., Calzolari, G., Denicolo, V., & Pastorello, S. (2020). Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review*, 110(10), 3267-3297. <https://doi.org/10.1257/aer.20190623>
- [4] Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *The RAND Journal of Economics*, 52(3), 538-558. <https://doi.org/10.1111/1756-2171.12383>
- [5] Fish, S., Gonczarowski, Y. A., & Shorrer, R. I. (2026). Algorithmic Collusion by Large Language Models. *arXiv:2404.00806v5*, 5 March 2026. <https://doi.org/10.48550/arXiv.2404.00806>
- [6] Colliard, J.-E., Foucault, T., & Lovo, S. (2026). Algorithmic Pricing and Liquidity in Securities Markets. *The Review of Financial Studies*, advance access, published online 17 February 2026. <https://doi.org/10.1093/rfs/hhag010>
- [7] Dou, W. W., Goldstein, I., & Ji, Y. (2025). AI-Powered Trading, Algorithmic Collusion, and Price Efficiency. NBER Working Paper No. 34054. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w34054>
- [8] Assad, S., Clark, R., Ershov, D., & Xu, L. (2024). Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market. *Journal of Political Economy*, 132(3), 723-771. <https://doi.org/10.1086/726906>
- [9] Gauci, I. (2026). Unilateral parallel algorithmic behaviour in EU financial services law. *ERA Forum*, 27(2), 237-251. <https://doi.org/10.1007/s12027-026-00884-w>
- [10] Financial Stability Board. (2024). *The Financial Stability Implications of Artificial Intelligence*. Basel: FSB, 14 November 2024.
- [11] Financial Stability Board. (2025). *Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector*. Basel: FSB, 10 October 2025.
- [12] United States and Plaintiff States v. RealPage, Inc., No. 1:24-cv-00710 (M.D.N.C.). Stipulation and Order and Proposed Final Judgment as to defendant RealPage, Inc., filed 24 November 2025. U.S. Department of Justice, Antitrust Division case file.
- [13] Brown, Z. Y., & MacKay, A. (2023). Competition in Pricing Algorithms. *American Economic Journal: Microeconomics*, 15(2), 109-156. <https://doi.org/10.1257/mic.20210158>
- [14] Arunachaleswaran, E. R., Collina, N., Kannan, S., Roth, A., & Ziani, J. (2025). Algorithmic Collusion Without Threats. 16th Innovations in Theoretical Computer Science Conference (ITCS 2025), LIPIcs vol. 325, art. 10, 10:1-10:21. <https://doi.org/10.4230/LIPIcs.ITCS.2025.10>
- [15] Bichler, M., Durmann, J., & Oberlechner, M. (2025). Algorithmic Pricing and Algorithmic Collusion. *Business & Information Systems Engineering*, 67(6), 971-979. <https://doi.org/10.1007/s12599-025-00965-z>
- [16] Hartline, J. D., Long, S., & Zhang, C. (2024). Regulation of Algorithmic Collusion. *Proceedings of the 2024 Symposium on Computer Science and Law (CSLAW '24)*. New York: ACM. <https://doi.org/10.1145/3614407.3643706>
- [17] Hartline, J. D., Long, S., & Zhang, C. (2025). Regulation of Algorithmic Collusion, Refined: Testing Pessimistic Calibrated Regret. *Proceedings of the 2025 Symposium on Computer Science and Law*. New York: ACM. <https://doi.org/10.1145/3709025.3712217>
- [18] Organisation for Economic Co-operation and Development. (2023). *Algorithmic Competition: Competition Policy Roundtable Background Note*. Paris: OECD Publishing.
- [19] Organisation for Economic Co-operation and Development. (2025). *Algorithmic Pricing and Competition in G7 Jurisdictions: Emerging Trends and Responses*. Paris: OECD Publishing, 2 October 2025. <https://doi.org/10.1787/f36dacf8-en>

- [20] Concurrences. (2026). Algorithmic Pricing and Competition Policy. Concurrences Review, No. 5-2026, Dossier.
- [21] Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. Machine Learning, 8(3-4), 279-292.
- [22] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). Cambridge, MA: MIT Press.
- [23] Adrian, T. (2026). How Central Banks Can Contain Financial Stability Risks as AI Accelerates Change. IMF Blog, 23 July 2026.
- [24] International Monetary Fund. (2024). Advances in Artificial Intelligence: Implications for Capital Market Activities. Global Financial Stability Report, October 2024, Chapter 3. Washington, D.C.: IMF.
- [25] European Central Bank. (2026). Financial Stability Review, May 2026. Frankfurt am Main: ECB.
- [26] Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007).
- [27] Joined Cases C-89/85 and others, Ahlstrom Osakeyhtio and Others v Commission (Wood Pulp II) [1993] ECR I-1307.
- [28] Case C-8/08, T-Mobile Netherlands BV and Others v Raad van bestuur van de Nederlandse Mededingingsautoriteit [2009] ECR I-4529, ECLI:EU:C:2009:343.
- [29] Competition and Markets Authority. (2016). Online sales of posters and frames. Case 50223. London: CMA.
- [30] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, 2024.
- [31] Council of the European Union. (2026). Artificial intelligence: Council and Parliament agree to simplify and streamline rules. Press release, 7 May 2026.