“Artificial Intelligence, Algorithmic Pricing and Collusion” by Emilio Calvano, Giacomo Calzolari, Vicenzo Denicolò, and Sergio Pastorello

Summary by Gina Markov

Pricing algorithms are increasingly replacing human decision making in online marketplaces. In a 2015 sample of over 1,600 best-selling items on the commerce platform Amazon.com, the economists Chen, Mislove and Wilson (2016) find that more than one third of vendors automated their pricing. This paper assesses whether learning pricing algorithms powered by artificial intelligence (AI) learn to collude “autonomously”; in this context, autonomously means that the algorithms do not have specific instructions from humans to collude, and that they do not communicate with each other to explicitly engage in collusion. The paper’s objective is to determine if the AI algorithms may systematically learn to collude. That is, sustain supra-competitive prices by the threat of a punishment, such as retaliatory low prices, should a firm deviate from those high prices.

Detecting tacit collusion is difficult in empirical settings, because firms often do not disclose their pricing strategies, and it can be tricky to recognize collusive behavior in market outcomes. Theoretically assessing collusive behavior is also difficult, because pricing algorithms generate feedback and relationships that are complex and difficult to analyze. Therefore, the authors rely on an experimental approach to the problem.

The authors design AI pricing agents who interact repeatedly with each other in a computer-simulated marketplace. The agents in the study are Q-learning algorithms, a type of reinforcement learning algorithm that is popular among computer scientists. Q-learning algorithms are simple but powerful, with architectures that have the potential to achieve superhuman success in games like chess. The defining characteristic of reinforcement learning algorithms is that they learn to behave based on past experience, reinforcing strategies that have proved successful. Because of this, these algorithms require no prior knowledge of the environment in which they are called to make decisions. Q-learning algorithms learn by a simple reward structure in a trial-and-error manner. For each period, an agent observes its state (information about the environment) and chooses an action. In the context of this paper, where agents engage in repeated games, the agents incorporate their rivals’ past actions in their information about the current state. Based on its state and action, the agent obtains a reward, and the system moves on to the next state according to a given probability distribution. In other words, there is a certain probability associated with each reward and future state, given the current state and action. The agent’s goal is to maximize the expected present value of its reward.

Since the underlying probability distribution of the system is unknown (if it were known, the agent could potentially calculate its optimal action for any given state), the agent learns about its environment over time. Starting in a random configuration, the agent acquires knowledge by choosing an action and observing the reward and state it enters. It learns in a two-prong fashion: exploration and exploitation. Exploration involves gathering new information by trying out new actions, even if these action are suboptimal given agent’s acquired information. If the algorithm does not explore, it may develop a too-narrow understanding of its environment and never find an optimal strategy. Exploitation is simply taking the action that, given the agent’s current knowledge of the probability distribution, provides the agent with the greatest reward. Exploitation drives the Q-learning algorithm to converge to an optimal equilibrium strategy. If an agent explores with probability ε, it exploits with probability 1-ε. The agents start by making choices largely at random, but as time goes on, ε decreases, and agents make the exploitive choice more frequently.

The authors design Q-learning agents to interact in a in the context of a well-accepted economic model of a market. That is a repeated differentiated Bertrand setting with agents setting prices simultaneously. The only substantial change they make to this traditional model is that they bound the agents' memory of the past in order obtain a finite state space and thus operationalize the algorithms. The authors adopt standard assumptions on demand and costs (logit demand and constant marginal costs) that can apply to many different industries, thus simulating a realistic economic environment. The agents can price anywhere from below Bertrand (perfect competition, which is equal here to marginal cost) to above monopoly price.

The baseline model is a symmetric duopoly: two pricing agents with the same cost and demand functions. The authors analyze many possible algorithmic designs considering various degrees to which the agents explore and exploit. They classify convergence to an optimal policy if an agent's learned strategy does not change for 100,000 periods. The authors find that more than 99.9% of the sessions converged.

Authors then study what the agents learned. In particular, they wish to assess whether the learned strategies are "collusive" in nature. Crucially, in order to show collusion, it is not enough to demonstrate supra-competitive prices. As is well established in economics, collusion is the sustaining of high prices by the threat of a punishment. Indeed, the authors find that the supra-competitive prices emerge in equilibrium supported by punishment schemes.

More specifically, the paper reports that the algorithms consistently learned to charge supra-competitive prices, obtaining substantial profits above the Bertrand-Nash equilibrium. Depending on the parameters, per-agent average profit can be as high as the profit that these firms would get if they were perfectly coordinating their prices and on average significantly higher than the profits that would accrue if the firms were competing. The authors ensure that the agent is finding a strategy that is the optimal response to its rival by taking the limit outcome of the rival as a given, and calculating how often the agent follows the corresponding best-response actions. They find the agents' losses from not playing the best-response strategy is largely below 0.5%. Agents either play the best-response, or a strategy very similar to it.

Next, the authors discuss the nature of the agents' collusive strategies. First, they verify that the agents can learn the equilibrium of the one-shot repeated game, where agents have no memory of the past. The agents learn to price perfectly competitively in this context. This is the only rational strategy, because deviations cannot be punished. Similarly, the authors find that if agents are short-sighted (they do not value the future very much), collusion is impeded, and the agents exhibit near-competitive behavior. In this way, the authors confirm that agents are learning the optimal equilibrium of one-shot games. So, if there are other equilibria in the repeated game setting, the fact that agents adopt different strategies implies that they are finding more sophisticated strategies.

Specifically, the authors observe that agents settle on a systematic coordination on pairs of prices, where deviations from these prices are punished. To make this point they document the agents' reaction to an unexpected price cut by its rival. That is, to a deviation from the collusive agreement. They find that such deviations are punished: in the few periods after the deviation, both agents significantly cut their price, engaging in a price war. Interestingly, the agents also learn to gradually return to their pre-deviation behavior and thus to restart collusion. This finding runs counter to theoretical equilibrium benchmarks like the grim-trigger strategy, which is an strategy that once an agent deviates, the rival punishes it by charging the lowest (perfectly competitive) price forever. In this setting, because agents are actively experimenting (ε>0), they eventually learn restart cooperating after a punishment phase. This mirrors an environment where firms' collusive strategies might be disrupted by idiosyncratic shocks. Another observation that supports the hypothesis that the agents are tacitly colluding is that the harshness of the punishment is strongly positively corelated with the profit gain: if both agents are making high profits from supra-competitive pricing, a profitable deviation by one agent would be severely punished by the other. Q-learning algorithms are shown to take a long time to complete the learning process. However, the authors note that a non-negligible degree of collusion may emerge well before the algorithms complete their training.

This paper is the first to clearly document the emergence of collusive strategies among autonomous pricing agents.

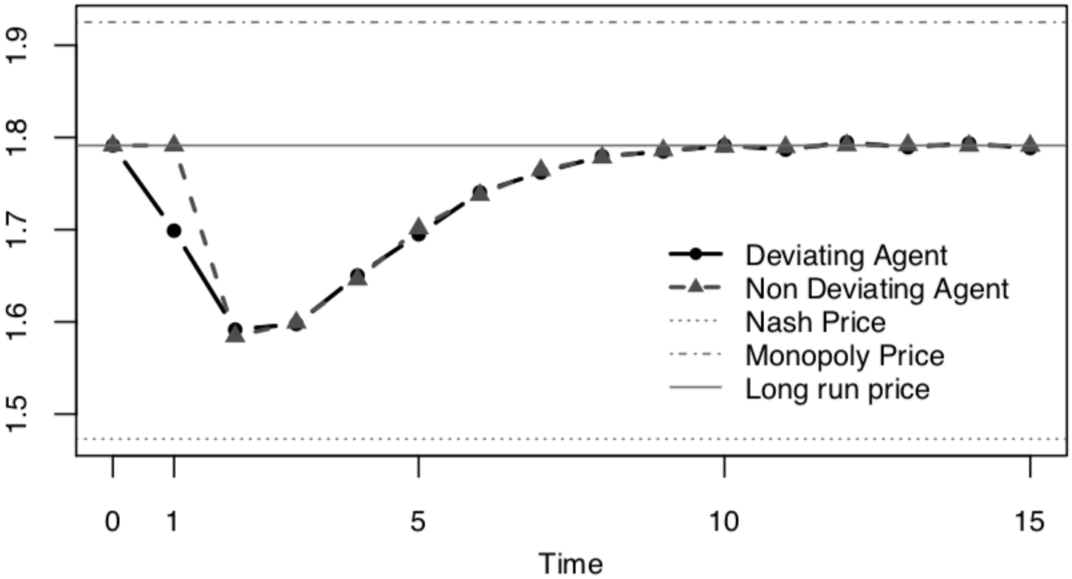

Figure 6 from original paper.

To ensure the robustness of their analysis, the authors alter the number of competitors, model asymmetry, and the stochastics of the environment. They find that the amount of collusion decreases with the number of competitors, aligning with the theory that collusion is harder to sustain with a more fragmented market. However, there is still substantial collusion with three to four firms, with 64% and 56% extra-profit above

Bertrand-Nash sustained respectively. Asymmetry tends to disincentivize collusive behavior because it is harder to coordinate, especially when agents cannot communicate. However, the authors find that collusion is still present with significant asymmetry. Asymmetry reduces average profit gain, but by a limited amount. The authors find that asymmetry does not actually make coordination harder for agents, but that the collusive strategy is no longer total profit maximizing, because the gain from collusion is split disproportionately in favor of the less efficient firm. Finally, the study alters the stochastics of the environment. The authors test a variable demand function (reflecting stochastic market entry and exit), which they find hinders, but does not eliminate, collusion. This is done by instituting a variable market structure with a firm randomly entering and exiting the market. The paper finds that entry and exit reduces profit gain due to uncertainty, but the algorithms still maintain a profit higher than the competitive benchmark. Increasing product substitutability and varying experiment initializations also slightly reduce profit gain but maintain substantial collusion.

There are several antitrust and competition policy implications of the study. The authors' results suggest that algorithmic collusion is a possibility, and it should be seriously considered and monitored by antitrust authorities. This type of tacit collusion may not be addressed by current antitrust policy, which primarily targets explicit agreements to collude. If they suspect collusive conduct, agencies and courts could subpoena pricing algorithms, and test them in simulated industries that mirror the one under investigation to see if collusion is detected. Such innovations in enforcement might reduce both the number of false positives and false negatives produced by either too aggressive or too lenient antitrust laws. Due to the exact architecture and programming of the algorithms, collusive behavior may be measured explicitly, which is helpful in an antitrust case.