Department of Economics- FEA/USP

Artificial Intelligence, algorithmic pricing, and predatory behavior

NICKOLAS MARTINS BATISTA
RODRIGO MENON SIMÕES MOITA

WORKING PAPER SERIES Nº 2027-26

DEPARTMENT OF ECONOMICS, FEA-USP

WORKING PAPER Nº 2026-27

Artificial Intelligence, algorithmic pricing, and predatory behavior

Nickolas Martins Batista (nickolas.batista@usp.br)

Rodrigo Menon Simões Moita (rodmoita@usp.br)

Abstract:

This paper investigates whether reinforcement learning algorithms can autonomously develop predatory pricing strategies in digital markets. We model a duopoly where two Q-learning agents repeatedly interact in a differentiated goods market with multino-mial logit demand, incorporating cash reserve dynamics and endogenous bankruptcy conditions. The theoretical framework establishes a Markov perfect equilibrium in which predation dominates collusion when one firm holds sufficiently larger cash re-serves and the discount factor satisfies a recoupment threshold. Simulations across 400 parameterized experiments — split between symmetric and asymmetric cash set-tings — reveal that predatory behavior emerges spontaneously in 43–45% of converged runs, without any explicit coordination or communication between agents. Cash flow asymmetry is the primary mechanism: as divergence in reserves grows during the com-petitive phase, algorithms increasingly converge toward predatory rather than collusive strategies. Intermediate learning rates amplify this tendency, contradicting prior find-ings by Calvano et al. 2020, who found monotonically pro-collusive effects of higher learning rates. Higher discount factors, conversely, favor collusion. These results carry direct policy implications: algorithmic predation is structurally undetectable through traditional conduct-based tools, calling for ex-ante monitoring frameworks, capital-flow audits, and modernized antitrust guidelines capable of addressing spontaneous exclusionary dynamics in AI-driven markets.

Keywords: algorithmic pricing, predatory behavior, reinforcement learning

JEL Codes: L12, L41, C63, C73, L13

Artificial Intelligence, algorithmic pricing, and predatory behavior

Nickolas Martins Batista Rodrigo Menon Simões Moita

September 2025

Abstract

This paper investigates whether reinforcement learning algorithms can autonomously develop predatory pricing strategies in digital markets. We model a duopoly where two Q-learning agents repeatedly interact in a differentiated goods market with multinomial logit demand, incorporating cash reserve dynamics and endogenous bankruptcy conditions. The theoretical framework establishes a Markov perfect equilibrium in which predation dominates collusion when one firm holds sufficiently larger cash reserves and the discount factor satisfies a recoupment threshold. Simulations across 400 parameterized experiments — split between symmetric and asymmetric cash settings — reveal that predatory behavior emerges spontaneously in 43–45% of converged runs, without any explicit coordination or communication between agents. Cash flow asymmetry is the primary mechanism: as divergence in reserves grows during the competitive phase, algorithms increasingly converge toward predatory rather than collusive strategies. Intermediate learning rates amplify this tendency, contradicting prior findings by Calvano et al. 2020, who found monotonically pro-collusive effects of higher learning rates. Higher discount factors, conversely, favor collusion. These results carry direct policy implications: algorithmic predation is structurally undetectable through traditional conduct-based tools, calling for ex-ante monitoring frameworks, capital-flow audits, and modernized antitrust guidelines capable of addressing spontaneous exclusionary dynamics in AI-driven markets.

Keywords: algorithmic pricing, predatory behavior, reinforcement learning, Q-learning, digital markets, antitrust

I Introduction

The discussion on predatory practices in digital markets has become one of the most urgent and complex themes in contemporary economics. The rapid advancement of artificial intelligence algorithms has redefined traditional competitive strategies, leading to a profound transformation of pricing dynamics. In global markets, this shift is exemplified by the recent regulatory overhaul in China, where the 2025 reform of the Price Law specifically targets "involutionary" or "throat-cutting" competition[1]. Such movements demonstrate that concerns regarding predatory pricing now transcend sectoral and national borders, demanding a modernization of legal frameworks to address the realities of the digital age.

The relevance of this topic goes beyond institutional shifts: exclusionary practices can emerge from autonomous algorithms without direct coordination, making detection increasingly challenging. Recent studies demonstrate that digital agents can spontaneously converge toward exclusionary strategies even without explicit communication. To counter this, jurisdictions like India have refined their technical apparatus; the *Competition Commission of India* (CCI) introduced the 2025 Draft Regulations to recalibrate the "cost of production" calculation, specifically accounting for platform economies where marginal costs are near zero. This represents a critical step in identifying predation within ecosystems that do not fit traditional industrial models.

The need to deepen the understanding of predation is also reflected in the theoretical evolution of the literature: research such as that by Besanko, Doraszelski, and Kryukov 2014, as well as Wiseman 2017, highlights that predatory scenarios can occur without extreme parameterization, provided market conditions foster cumulative competitive advantages. This theoretical shift is being mirrored in "super-enforcement" jurisdictions. In early 2026, Saudi Arabia's *General Authority for Competition* (GAC) signaled a drastic reduction in tolerance

[1]The Chinese Price Law reform (late 2025) aims to curb irrational price wars and the use of algorithms for targeted predation in sectors where profit margins have reached unsustainable levels due to strategic cash-burning.

for strategic cash-burning, applying record fines based on updated guidelines that refine cost tests to identify aggressive pricing intended to exclude rivals under the "Vision 2030" framework. Similarly, New Zealand's *Commerce Amendment Bill* has introduced an objective economic test that codifies specific cost benchmarks for legal intervention[2].

Additionally, new dynamic models explore how expectations of future entry and regulatory responses influence predatory effectiveness. We observe a global trend moving toward *ex-ante* obligations and broader definitions of power. In Brazil, the proposed "Digital Competition Law" (PL 4675/2025) establishes preventive surveillance over Big Tech ecosystems, while CADE has updated its *Guidelines on Predatory Pricing* (Portaria 104/22) to align with international best practices of 2025-2026. This is complemented by the innovative approach in the United Arab Emirates, where Decree 3/2025 introduces the concept of "economic dependence"[3].

This paper employs Q-learning, a reinforcement learning framework, to demonstrate that algorithms can develop predatory strategies even within low-complexity environments characterized by limited variables. Our simulations reveal that either a high learning rate or cash asymmetry serves as a sufficient catalyst for predatory behavior, leading the bankruptcy of the rival firm to emerge as the dominant strategy across most scenarios.

This article aims to build a bridge between classical theory and these emerging digital practices by modeling scenarios where reinforcement learning algorithms learn optimal strategies in repeated price games under cash asymmetry and symmetry. By exploring the conditions under which predation becomes dominant, we emphasize not only the challenge of monitoring potentially harmful practices but also the importance of developing new analytical approaches to protect legitimate competition in increasingly automated and dynamic

[2]Under the New Zealand framework (fully effective mid-2026), pricing below Average Variable Cost (AVC) or Average Avoidable Cost (AAC) is defined as predation. Prices between these levels and Average Total Cost (ATC) are punishable if exclusionary intent is proven.

[3]The UAE Federal Law No. 36 (updated April 2025) allows the regulator to intervene in cases of "significantly low prices" even without traditional market dominance, provided there is economic dependence within a digital ecosystem.

markets.

II Literature Review

The rapid advancement of artificial intelligence (AI) algorithms—particularly Q-learning and other reinforcement learning (RL) methods—is transforming the competitive strategies of digital firms worldwide. Platforms such as Amazon, Uber, and global marketplaces already employ large-scale automated dynamic pricing, redefining the competitive landscape. As noted by Calvano et al. 2020, these algorithms can reach collusive or exclusionary outcomes through autonomous trial-and-error, making the discussion regarding predation and algorithmic coordination more urgent than ever.

Recently, discussions on predatory pricing have returned to the forefront of economic research, catalyzed by the "cash-burning" models of the last decade. China, for instance, significantly strengthened its legal framework following high-profile conflicts involving platforms like Meituan and Didi. A landmark moment occurred in 2021 with the "Nine Nos" regulations, which explicitly banned community group-buying platforms from selling below cost to squeeze out local competitors. This culminated in the 2025 reform of the Price Law, which now targets "involutionary" competition—a scenario where algorithms are used to sustain irrational price wars that degrade market sustainability.

In the Middle East, the regulatory response has focused on the unique power dynamics of digital ecosystems. The United Arab Emirates (UAE) and Oman have updated their competition laws to prohibit selling at "significantly low prices" with exclusionary intent. A notable example is the UAE's Decree 3/2025, which introduced the concept of economic dependence. This allows regulators to intervene when a platform holds power over its suppliers or users, even if it does not meet the traditional threshold for market dominance, thereby preventing "ecosystem-based" predation.

Furthermore, the evolution of legal standards is moving toward more objective, datadriven benchmarks. New Zealand's *Commerce Amendment Bill* (set for full implementation in 2026) provides a clear example of this shift by codifying specific cost tests: pricing below Average Variable Cost (AVC) or Average Avoidable Cost (AAC) is now a per se indicator of predation. Similarly, the *Competition Commission of India* (CCI) through its 2025 Draft Regulations has redefined "cost of production" specifically for platform economies, accounting for the near-zero marginal costs that often mask predatory intent in digital markets.

This research stems from the premise that exclusionary practices—such as algorithmic predation—can emerge even without explicit coordination. In jurisdictions like Saudi Arabia, where the General Authority for Competition (GAC) has entered a "super-enforcement" phase in 2026, the focus has shifted toward the long-term impact of strategic loss-making. The risk of tacit exclusion is now embedded within the optimized learning process of modern algorithms. This necessitates increased vigilance from policymakers and researchers, as smaller firms and new entrants often become victims of these dynamics before traditional regulatory triggers are even pulled.

II.1 Q-Learning, Convergence to Tacit Collusion, and Reinforcement Learning

The specialized literature—see Calvano et al. 2020—demonstrates that Q-learning algorithms can learn to execute collusive strategies without the need for coordination between agents. The seminal work by Calvano et al. 2020 was the first to clearly document the emergence of collusive strategies among autonomous pricing agents in an infinitely repeated oligopoly model with Bertrand-type price competition.

The authors employ Q-learning algorithms in numerical simulations to examine how these algorithms can converge to non-competitive equilibria. Their findings show that, in simulations spanning tens of thousands of periods, the algorithms consistently learn to charge supra-competitive prices, capturing 70-90% of the profit gap between the Bertrand equilibrium and a monopoly. These strategies are sustained by punishment-reward schemes similar to traditional collusion mechanisms, yet they emerge purely autonomously through trial-and-error learning.

Crucially, the process does not require information exchange; simple observation of environmental prices is sufficient. The algorithms converge to policies that sacrifice short-term profit to eliminate competitors, subsequently enabling investment recovery via elevated prices. The observed punishment mechanism involves lowering prices following a rival's deviation, followed by a gradual return to collusive prices—rather than the "grim trigger" so common in traditional economic theory.

Subsequent work by Calvano et al. 2021 extends these results to environments with imperfect monitoring, showing that even when firms cannot perfectly observe rivals' actions, Q-learning can still learn to collude. In this stochastic demand scenario—adapted from the classic Green and Porter 1984 model—the algorithms learn strategies similar to those long described by economic theorists, where a price drop below a certain threshold triggers temporary "price wars," followed by a gradual return to cooperation.

Thus, the current threat is not merely traditional collusion or "classic" predation, but the emergence of tacit predation, where autonomous algorithms, guided by reward policies that maximize profit, converge toward exclusionary behavior through pure stochastic/autonomous adjustment. This necessitates new monitoring tools for antitrust policies, with an emphasis on algorithmic audits and ex-ante controls, particularly in technology markets.

II.2 Predation and Dynamic Learning

Complementing the algorithmic collusion approach, the literature on the economics of predation provides a theoretical framework for understanding when and how firms use aggressive pricing to exclude rivals, particularly in contexts with cumulative competitive advantages.

Besanko, Doraszelski, and Kryukov 2014 formally characterize price predation in a dynamic framework that endogenizes competitive advantage and market structure, using "learning-by-doing" as an illustrative example. Their model distinguishes the "advantage-building motive"—where a firm sets aggressive prices to improve its future competitive position—from the "advantage-denying motive"—where the firm attempts to prevent rivals from becoming more competitive. The authors show that equilibria with "predation-like" behavior arise routinely without requiring extreme parameterizations, and that forcing a firm to ignore predatory incentives has a substantive impact on market structure and performance.

In a complementary vein, Wiseman 2017 analyzes when predation dominates collusion in dynamic games with financial and stochastic constraints. The work demonstrates that, under certain conditions, patient firms cannot sustain collusion in equilibrium; instead, they prefer an immediate "price war" that forces one or more rivals into bankruptcy. The intuition is that a firm in a highly favorable position can guarantee a sufficiently high payoff by initiating a price war, making the maintenance of a long-term collusive agreement irrational.

Other recent texts also address this theme, such as "A Dynamic Model of Predation" (Rey, Spiegel, and Stahl 2024), which presents an infinite-horizon theoretical model to examine the viability and profitability of predatory practices in dynamic markets. The authors explore how the relationship between incumbents and potential entrants unfolds over time, considering that the incumbent may choose between accommodation or predatory action at each new entry episode. The study identifies various possible equilibria, including accommodation scenarios, recurring predation (featuring rapid entry and exit of rivals), and monopolization (where potential entrants desist due to the expectation of future predation).

The distinctiveness of Rey's work lies in demonstrating that the possibility of predation does not necessarily depend on asymmetric information or economies of scale, but can emerge purely from strategic uncertainty and participant expectations. The work also discusses the social welfare effects of these strategies and evaluates different legal rules to curb predation, suggesting that standards explicitly limiting incumbent responses to entry may be more effective than total prohibitions or laissez-faire policies. This analysis contributes to the literature by formalizing the role of expectations and the dynamic context in the success and recurrence of predatory strategies.

The model proposed in this article builds a theoretical structure of a dynamic game where firms have asymmetric information regarding the competitor's cash reserves but can perfectly observe the resulting prices and profits in each round. In this scenario, we demonstrate that predatory strategies are possible even in the absence of communication between firms.

These theoretical insights reinforce the relevance of investigating RL algorithms: if predation/collusion can emerge among fully informed rational agents based on market parameters, then in environments with autonomous learning agents, the risk of exclusionary or collusive dynamics is significantly amplified.

III Methodology

This chapter describes in detail the methodological approach adopted in this work for the simulation of a duopoly market. The primary method is the use of the **Q-learning** algorithm (reinforcement learning), a model of self-interested agents that learn optimal strategies through interaction with the environment. The duopoly is modeled as a dynamic and repeated game, where two firms compete on price, seeking to maximize their profits over time.

III.1 The Theoretical Model

The simulation environment is a duopoly market with price competition and an asymmetric cash structure, given that it is common in firms' strategic behavior to have partial knowledge of the financial volume available to their competitors. The model is based on the multinomial logit demand function, which captures consumer choice between the products of the two firms.

III.1.1 Demand, Profit, and Cash Functions

The demand for each product is modeled as a proportion of the total market, depending on the prices of both firms and an "outside option good." The utility of the product from firm i to consumer j, U_{ij}, is given by:

$$U_{ij} = \nu - p_i + \epsilon_{ij} \tag{1}$$

where ν is the value of the product to the consumers, p_i is the price charged by firm i and ϵ_{ij} is the idiosyncratic shock of consumer j towards product i. The utility of the outside option is normalized to zero ($U_0 = 0$).

If we assume ϵ_{ij} to have a extreme value type I distribution, and integrating over the area where option i has the highest utility, we hava e logit demand system. In this case, the market share (demand) of each firm, d_i, is given by the logit demand formula, as per Equation 2:

$$d_i = \frac{e^{(\nu - p_i)/\sigma}}{e^{(\nu - p_1)/\sigma} + e^{(\nu - p_2)/\sigma} + e^{U_0/\sigma}} \tag{2}$$

where σ is a demand sensitivity parameter, which determines how sensitive consumers are to small price variations. If σ is high, consumers do not care much about price differences; they have very distinct preferences. This gives firms more market power, as a small price reduction does not attract many customers from the rival. If σ is low consumers see the products as almost perfect substitutes. Small price variations cause large demand migrations.

The profit of each firm, π_i, in a given round of competition, is the difference between the price and the marginal cost, multiplied by the corresponding demand. This relationship is formalized in Equation 3:

$$\pi_i = (p_i - c) \cdot d_i \tag{3}$$

where c represents the marginal production cost, assumed constant for both firms. The over time discounted stream of profits is $K_{i,t}$, which is the firm's state variable, and its evolution over rounds is a crucial aspect of the model. In round $t+1$, the cash of firm $K_{i,t}$ is updated based on its profit in round t, as per Equation 4:

$$K_{i,t+1} = K_{i,t} + \pi_{i,t} \tag{4}$$

Thus, in period t each firm i has an information set $I_{i,t}$ given by the set of prices already practiced by the competitor and the resulting profits from each stage game. Thus the firm seeks to optimize its cash flow position and takes as its decision the price that will be practiced in each period. In this theoretical framework, firms make price decisions simultaneously over an infinite horizon.

In this section, we solve this theoretical problem for a set of three distinct pricing strategies: a competitive price (p^c), a collusion price (p^{cl}) or a predation price (p^p).

We assume there is a high market entry cost and that the bankruptcy of one firm is not accompanied by the entry of a new agent. Thus, firms can opt for a predation action as this can lead the other firm to bankruptcy, giving a monopoly profit (π^m) for all subsequent periods if the predatory firm succeeds. The bankruptcy condition of a firm is given by:

$$K_{i,0} + \sum_{t=o}^{s} \beta^t \pi_{i,t} < 0 \tag{5}$$

Given an intertemporal discount factor β, condition (5) describes that a firm is expelled from the market in period (s) at the moment when the sum of its profit over time generates negative intertemporal cash flow, $K_{i,s} < 0$. However, firms do not have perfect visibility of the initial cash, so for firm i, its expectation about firm j's cash is given by:

$$E[k_{j,0}|I_{i,0}] \tag{6}$$

Where $I_{i,0}$ is the information set of Firm i, comprising the historical prices of firm j ($P_{j,-t}$), its own historical prices ($P_{i,-t}$), and the profits realized by both firms ($\pi_{j,-t}, \pi_{i,-t}$). Thus, for a given bankruptcy period s for firm i observing the firm j, the terminal condition is defined as:

$$E[k_{j,0}|P_{i,-t}, P_{j,-t}, \pi_{j,-t}, \pi_{i,-t}] + \sum_{t=0}^{s-1} \beta^t \pi_{j,t} < 0 \tag{7}$$

For the theoretical model, we assume a simplified discrete action space of three prices. In contrast, the simulated model expands this to 100 prices within a closed interval. To ensure the convergence properties of the Q-learning algorithm, it is necessary to impose both a closed price interval and a finite history of strategies. The rationale for these constraints is discussed in Subsection III.2.Thus, the stage game in period t is described by the following structure:

Table 1: Payoff Matrix for Price Decision

Firm 1 \Firm 2	P^p	P^c	P^{cl}
P^p	(π_1^p, π_2^p)	(π_1^p, π_2^c)	(π_1^p, π_2^{cl})
P^c	(π_1^c, π_2^p)	(π_1^c, π_2^c)	(π_1^c, π_2^{cl})
P^{cl}	(π_1^{cl}, π_2^p)	(π_1^{cl}, π_2^c)	(π_1^{cl}, π_2^{cl})

In this game we have profits that vary given the demand resulting from pricing, that is, when agents make the same price decision they split the market in half and obtain a profit referring to their pricing and their costs ($\pi_i^{cl}, \pi_i^c, \pi_i^p$).

III.2 Reinforcement Learning Algorithm

The duopoly model is framed as a reinforcement learning problem. Each firm is an agent that interacts with an environment (the market) in a cycle of observation, action, and reward.

III.2.1 Q-Learning and Markov Decision Process

The Q-learning algorithm, originally designed for stationary Markov decision processes (MDPs), can also be applied to repeated games. The most direct approach is to allow the algorithms to continue updating their Q matrices, treating rivals' actions as other relevant state variables. However, in repeated games, stationarity is inevitably lost, even if the stage game does not change each period. One source of non-stationarity is that if the state included players' actions in all previous periods, the state space would grow over time. However, this problem can be circumvented by limiting players' memory. With limited memory, a state would include only the actions chosen in the last k stages, meaning the state space can be finite and time-invariant.

A more serious problem is that, in repeated games, the per-period payoff and transition to the next state generally depend on all players' actions. If a player's rivals change their actions over time—whether because they are experimenting or learning—the player's optimization problem becomes inherently non-stationary. This non-stationarity is the root of the lack of general convergence results for Q-learning in games. There is no a priori guarantee that multiple agents using Q-learning, interacting repeatedly, will reach a stable outcome or learn an optimal policy (i.e., collectively, a Nash equilibrium of the repeated game with limited memory).

Despite the absence of theoretical guarantees, convergence and equilibrium play can occur in practice. However, this can only be verified a posteriori. As we will discuss further, the analysis of these practical results is fundamental to understanding the behavior of Q-learning algorithms in repeated game contexts, where interaction between multiple agents creates complex and unpredictable dynamics.

III.2.2 State and Action Space

The state space, S_t, at any period t, is defined by a tuple comprising the preceding prices and discretized cash levels of both firms. We initially consider a benchmark case of perfect information, wherein firms perfectly observe each other's financial positions. The action space, A_t, consists of a closed set of discrete price levels available to each firme.

- **State (S_t):** $(p_{1,t-1}; p_{2,t-1}; K_{1,t-1}; K_{2,t-1})$
- **Action (A_t):** The price $p_{i,t}$ chosen by firm i in round t.

III.2.3 The Q Matrix (Q-Table)

The **Q-Table** is the central structure of Q-learning, as cited in Sutton, Barto, et al. 1998. It is a multidimensional matrix that stores the "quality" (Q) values for each state-action pair. The Q(s,a) value represents the expected reward from taking action a while in state s. The table is specific to each firm, starts from an initial random matrix Q_0 and is updated iteratively during training until reaching a convergence state that is guaranteed by the conditions of the problem presented.

III.2.4 The Q-Learning Update Equation

Q-learning updates the Q-Table values using the Bellman equation (Equation 8). This update rule allows agents to learn from their experience.

$$Q_{t+1}(s_t, a_t) = (1-\alpha)Q_t(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q_t(s_{t+1}, a) \right] \tag{8}$$

Where:

- $Q_t(s_t, a_t)$ is the Q value for the state-action pair (s_t, a_t) in iteration t.
- α is the **learning rate**, which determines the importance of new information relative to prior knowledge.
- r_{t+1} is the **reward** obtained in round t+1, which corresponds to the firm's profit π_{t+1}.
- γ is the **discount factor**, which weights the importance of future rewards. A high γ value encourages long-term profit maximization.

- $\max_a Q_t(s_{t+1}, a)$ is the maximum possible Q value in the next state s_{t+1}, representing the best action the agent can take from there.

III.3 Exploration-Exploitation Strategy: ϵ-greedy with Epsilon-Decay

The training phase involves balancing exploration and exploitation. Initially, agents explore the action space to discover which strategies are profitable. Over time, the **exploration rate**, modeled as a decreasing function of time, decreases, and agents begin to exploit (use) the strategies they have learned to be most effective.

The agent's action choice follows the ϵ-greedy strategy. In round t, a random variable x is generated, where $0 \leq x \leq 1$.

- If $x < \epsilon_t$, the agent **explores**, choosing a random action from the action space.
- If $x \geq \epsilon_t$, the agent **exploits**, choosing the action that maximizes the Q value in the matrix, i.e., $a \in argmax_a Q_t(s_{t+1}, a)$.

The exploration rate ϵ_t is not constant. It is updated each round through the **epsilon-decay** strategy, as per Equation 9:

$$\epsilon_t = e^{-\lambda t} \tag{9}$$

where:

- λ is the **exploration rate**, which controls the speed at which ϵ approaches ϵ_{min}.
- t is the current iteration number.

This approach ensures that agents explore extensively at the beginning of training but gradually become more focused on exploiting the most profitable strategies as time progresses.

III.4 Simulation Structure and Analysis

The simulation consists of two complementary phases that allow analysis of economic agents' behavior under different market conditions. In the first stage, called the training phase, firms perform over 2,000,000 interactions, during which the Q matrices are continuously updated each round. The main objective at this point is to allow agents to explore the competitive environment and improve their strategies. Training is stopped when stabilization of agents' strategies is observed or when a predefined maximum number of iterations is reached, thus ensuring learning of the best actions from accumulated experience.

The second stage corresponds to the competition phase, in which agents, now using only the optimal strategies learned (without further exploration), confront each other in diverse initial cash scenarios. Throughout this phase, convergence can occur either by bankruptcy of one firm or by repetition of prices for more than 2,000 consecutive interactions. All results regarding price evolution, profits, and firms' cash are recorded and examined, allowing detailed interpretation of the emergent behavior of agents in different market situations and validation of the strategic patterns generated by the model.

III.4.1 Convergence Conditions

According to Calvano et al. 2020, to verify convergence, a practical criterion is used: convergence is considered achieved when, for each player, the optimal strategy remains constant for 100,000 consecutive periods. Formally, for each player i and each state s, if the action $a_{it}(s) = argmax[Q_{it}(s, a)]$ remains unchanged for 100,000 repetitions, it is assumed that the algorithms have completed the learning process and reached stable behavior. The session is terminated when this occurs, or mandatorily after one billion repetitions.

In this article, competition between firms can be terminated in two distinct ways: by bankruptcy, when one of the firms reaches negative cash, or by price convergence, if the prices practiced by both firms stabilize at a profitable level for more than 2,000 consecutive rounds. This concept is important for understanding market dynamics and the factors that lead to the termination of competitive strategies.

IV Results

In this section, we present the results of the experiments, simulating the Q-learning algorithm 2000 times by varying the learning rate (α), the exploration parameter (λ), and the initial capital of the firms (K_0).

Additionally, we constructed a parameter Δ that describes the proximity of the final profit of the experiment ($\overline{\pi}$) to the expected competitive price (π_c) and monopoly price (π_m). Its equation is given by:

$$\Delta = \frac{\overline{\pi} - \pi_c}{\pi_m - \pi_c} \tag{10}$$

Thus, the closer Δ is to one, the closer the final profit obtained after convergence was to a monopoly profit; similarly, the closer to zero, the closer to competitive pricing behavior, and the more negative, the greater the predatory behavior was.

IV.1 Perfect Monitoring

Initially, we will analyze the experiments in which all firms observe the financial health of all agents. In this perfect monitoring environment, there is also an entry barrier for new firms, so that the bankruptcy of one of them creates a scenario of perpetual monopoly for the competitor.

Figure 1 presents the frequency of deltas resulting from the 500 experiments performed. Whether the cash is symmetric or asymmetric between the firms, most of the final results of the experiments are close to the collusion outcome.

Considering the asymmetric case in blue present in Figure 1, it can be seen that most experiments resulted in negative delta, which represents a predation strategy. The cash advan-

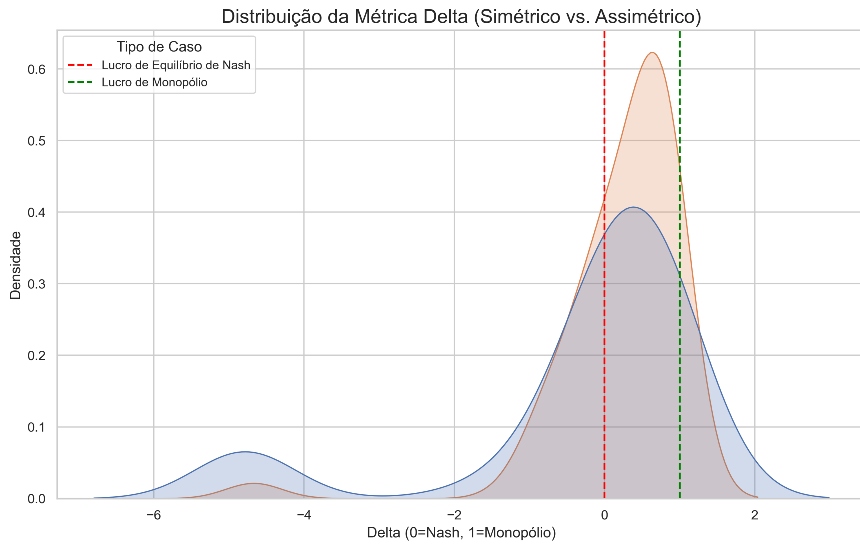

Figure 1

Note: Distribution of simulation results for identical initial cash balances (orange) or different cash balances (blue).

tage of firms significantly increases the probability of predation, with the difference between cash levels emerging as a key variable for understanding predatory behavior. Greater initial cash asymmetry systematically leads to more frequent and sustained predation outcomes.

In the case of these pricing algorithms, the evident financial advantage facilitates decisions for negative short-term profits in pursuit of long-term monopoly gains. When compared to cases with initial cash symmetry, predation events continue to occur, but the predominant scenario shifts to collusion. This suggests that cash disparity acts as a critical threshold: moderate asymmetries still allow tacit coordination, while substantial differences trigger aggressive predation strategies.

Another condition is the ability of this type of algorithm to learn to accept negative profits in exchange for future benefits. The correlation heatmap in Figure 2 demonstrates how the increase in the variation rate of Δ is inversely correlated with the learning rate (α).

This indicates that the greater the model's incentive to explore new scenarios, the higher the number of pricing decision scenarios that are considered predatory.

This result contradicts the findings in Calvano et al. 2020. In that study, the authors

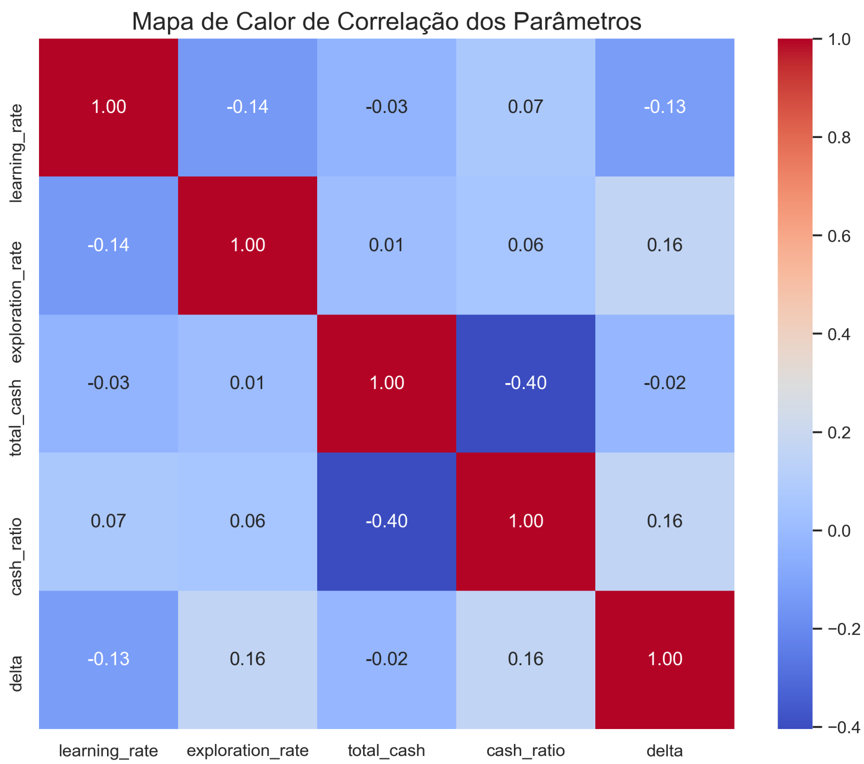

Figure 2

demonstrate how an increase in the learning rate converges toward a tacit collusion scenario. In the case of this article, the way this model is developed indicates that higher Q-learning learning rates increase the chances of convergence toward a predatory scenario, particularly under cash asymmetry.

To understand the strategic behavior of the Q-learning algorithm, it is necessary to expand the analysis to observe the dynamic behavior of Δ and thus identify regions with higher probability of predation and tacit collusion.

IV.1.1 The Predation Process

It was observed that predation occurs with higher probability in scenarios with high learning rates and for higher cash levels.

Figure 3 presents the average evolution of Δ given variations in the learning rate and exploration rate of the model. In both cases, there are intervals under which predation is

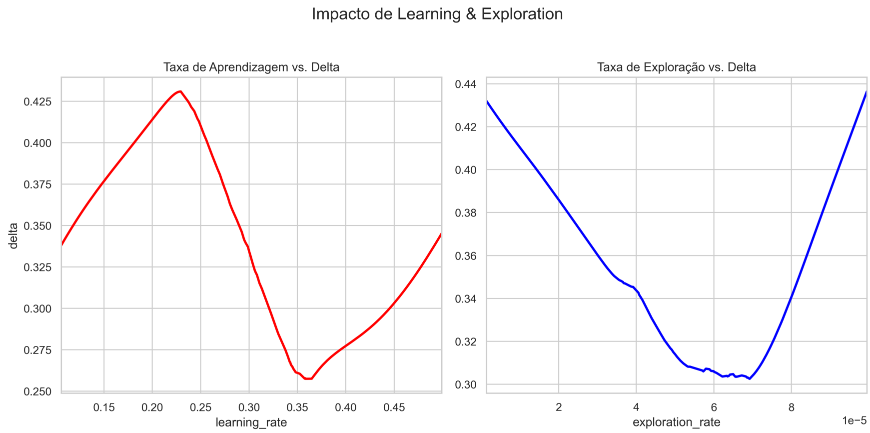

Figure 3

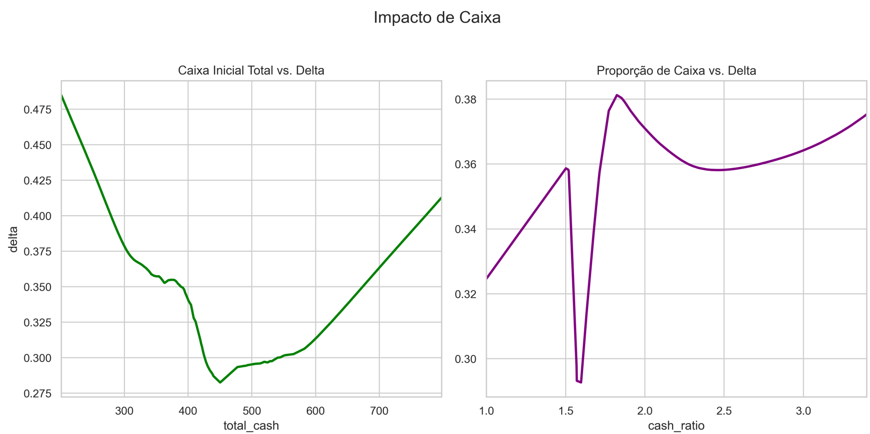

Figure 4

more likely by the algorithm, as in the case of the interval from 0.23 to 0.36 for the learning rate.

The same applies to the exploration rate: for values below 7, increasing the exploration rate favors predation; however, for higher exploration levels, the collusion scenario already presents itself as the more likely alternative.

The graphs in Figure 4 demonstrate the behavior of Δ for different cash conditions and cash ratio, given the asymmetric scenario. It is noted that in the case of the cash ratio, there is no evidence that it is responsible for creating predation scenarios, given that under learning constraints the algorithm may never have encountered the possibility of executing this type of strategy.

V Conclusion

The results of this research reinforce the importance of in-depth analysis of predation dynamics in digital markets, especially in light of advances in artificial intelligence and algorithmic pricing strategies. The experiments demonstrate that reinforcement learning algorithms, such as Q-learning, can spontaneously develop exclusionary and collusive behaviors, even in the absence of direct coordination between agents, particularly in scenarios of cash or information asymmetry. This emergence of predatory strategies in simulated environments points to a growing challenge in monitoring and regulating digital markets, where automated decisions can amplify anticompetitive practices.

Furthermore, the impact of these strategies extends beyond individual firm performance, affecting market structure, entry of new competitors, and consumer welfare. Recent evidence shows that when autonomous algorithms make pricing decisions, the risk of rival exclusion and formation of digital monopolies becomes higher, justifying regulatory actions such as those expanded in countries like China, United Arab Emirates, and Europe. The analytical literature, supported by dynamic models and numerical experimentation, also indicates that predatory practices can be rational and recurrent, emerging even without presumption of bad faith or extreme parameterizations by the involved companies.

In this context, there is an urgent need for innovative regulatory mechanisms capable of identifying and preventing exclusionary practices that harm fair competition and consumers. Algorithmic monitoring tools, ex-ante audits, and norms adapted to the new reality of digital markets should be considered by policymakers and competition authorities to mitigate monopoly risks and strengthen competitive environments.

The analysis conducted in this paper contributes to the literature by formalizing the conditions under which pricing algorithms can converge to predatory strategies, deepening the theoretical and practical understanding of the effects of these dynamics in digital markets. These conclusions highlight the need for ongoing studies on algorithms, regulatory innovation, and development of more robust competition protection models.

Finally, it is essential that researchers, companies, and authorities maintain dialogue and cooperation to share experiences, improve legal frameworks, and create efficient responses to the challenges posed by automation, algorithmic strategy, and the potential for exclusion in globalized digital markets.

References

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov (2014). "The economics of predation: What drives pricing when there is learning-by-doing?" In: *American Economic Review* 104.3, pp. 868–897.

Calvano, Emilio et al. (2020). "Artificial intelligence, algorithmic pricing, and collusion". In: *American Economic Review* 110.10, pp. 3267–3297.

Calvano, Emilio et al. (2021). "Algorithmic collusion with imperfect monitoring". In: *International journal of industrial organization* 79, p. 102712.

Green, Edward J and Robert H Porter (1984). "Noncooperative collusion under imperfect price information". In: *Econometrica: Journal of the Econometric Society*, pp. 87–100.

Rey, Patrick, Yossi Spiegel, and Konrad O Stahl (2024). "A Dynamic Model of Predation". In.

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

Wiseman, Thomas (2017). "When does predation dominate collusion?" In: *Econometrica* 85.2, pp. 555–584.