AI-Powered Trading, Algorithmic Collusion, and Price Efficiency

Winston Wei Dou Itay Goldstein Yan Ji *

March 20, 2025

Abstract

The integration of algorithmic trading with reinforcement learning, termed AI-powered trading, is transforming financial markets by reshaping how trading works. This study constructs a theoretical laboratory where financial markets function as information aggregation mechanisms, compelling investors to trade cautiously on private signals to preserve information rents. We find that informed AI speculators can autonomously sustain collusive supra-competitive profits without agreement, communication, or intent. AI collusion undermines competition and market efficiency, emerging robustly through two algorithmic mechanisms: (i) price-trigger strategies when information-insensitive investors are strongly prevalent and noise trading risk is low, and (ii) over-pruning bias in learning under other conditions.

Keywords: Reinforcement learning, AI collusion, Competition and market efficiency, Experience-based and self-confirming equilibrium, Information asymmetry and price informativeness, Market liquidity. (**JEL Classification:** D43, G10, G14, L13)

*Dou: University of Pennsylvania (wdou@wharton.upenn.edu) and NBER; Goldstein: University of Pennsylvania (itayg@wharton.upenn.edu) and NBER; Ji: Hong Kong University of Science and Technology (jiy@ust.hk). We thank Kerry Back, Snehal Banerjee, Hui Chen, Jean-Edouard Colliard, Will Cong, Antoine Didisheim, Itamar Drechsler, Maryam Farboodi, Slava Fos, Cary Frydman, Paolo Fulghieri, Vincent Glode, Joao Gomes, Mark Grinblatt, Ming Guo, Tim Johnson, Chris Jones, Scott Joslin, Larry Harris, Zhiguo He, David Hirshleifer, Jerry Hoberg, Harrison Hong, Mariana Khapko, Leonid Kogan, Pete Kyle, Tse-Chun Lin, Deborah Lucas, Ye Luo, Semyon Malamud, Andrey Malenko, George Malikov, Albert Menkveld, Jonathan Parker, Lasse Pedersen, Josh Pollet, Paul Romer, Nick Roussanov, Tom Sargent, Antoinette Schoar, Hyun-Song Shin, Daniel Sokol, Rob Stambaugh, Yannan Sun, Eric Talley, Anton Tsoy, Quentin Vandeweyer, Laura Veldkamp, Jiang Wang, Neng Wang, Liyan Yang, Yilin Yang, Jacob Yunger, David Zhang, Xiaoyan Zhang, and seminar and conference participants at AsianFA, ASU Sonoran Winter Finance Conference, BIS, BI-SHoF Conference, Boston College, CFTRC, CICF, CUFE, Duke/UNC Asset Pricing Conference, FINRA, FMA Asia/Pacific Conference, Fudan, George Mason, HKU, HKUST, HK Conference for Fintech and AI, IMF-WIFPR Conference, Imperial College, Jackson Hole Finance Conference, Johns Hopkins Carey Finance Conference, LSE, Melbourne Asset Pricing Meeting, MIT, MFA, NBER Summer Institute (Asset Pricing), Nordic Fintech Symposium, NYU/Penn Law and Finance Conference, OECD, Olin Finance Conference at WashU, Oxford, PKU/PHBS Sargent Institute Macro-Finance Workshop, QES Global Quant and Macro Investing Conference, QRFE Workshop on Market Microstructure, Fintech and AI, Rice University, SFS Cavalcade North America, SHUFE, Toronto Macro/Finance Conference, Tsinghua PBCSF, Tsinghua SEM, UIUC, University of Macau, University of Minnesota, University of Toronto, University of Zurich, USC, WFA, and Wharton for their comments. Dou is grateful for the financial supports from the Golub Faculty Scholar Award at Wharton.

1 Introduction

The integration of algorithmic trading with reinforcement learning (RL) algorithms, often termed AI-powered trading, poses new regulatory challenges and has the potential to fundamentally reshape capital markets.[1] With the SEC approving Nasdaq's RL-based, AI-driven order type, AI integration in trading is gaining momentum. Leading digital trading platforms are endorsing RL-based AI trading bots, and major hedge funds and investment powerhouses are adopting AI technologies. This trend has led policymakers, regulators, and financial market supervisors worldwide to prioritize the regulation of AI.[2]

The U.S. Securities and Exchange Commission (SEC) has warned about the possibility of AI collusion, where autonomous, self-interested algorithms independently learn to coordinate without any agreement, communication, or intention. AI collusion might hurt competition and market efficiency, as the cooperation among AI algorithms benefits a few sophisticated speculators, harming other investors. Given that promoting competition is the primary objective of SEC, the possibility of AI collusion raises significant concerns for SEC and other regulators across the globe. The SEC Chair, Gary Gensler, particularly noted that machines in high-frequency trading are already showing cooperative behavior without human input. However, the underlying scientific and economic mechanism of AI collusion remains unclear, as is how this might impact price formation and market efficiency.

AI algorithms differ from human traders as they do not simply mimic human behavior. Traditional theories and experimental studies about human behavior are inadequate for understanding the behavior of AI traders and the equilibra they may form. AI possesses a fundamentally different form of intelligence compared to humans. AI decision-making is driven by pattern recognition rather than emotions or logical thinking, and it is not influenced by higher-order beliefs. Therefore, understanding the dynamics of capital markets with the prevalence of AI-powered trading algorithms necessitates understanding algorithmic behavior similar to the "psychology" of machines (Goldstein, Spatt and Ye, 2021). This is akin to how decision theory and psychology literature have offered insights into modeling human behavior in an economics context.

We shed light on the equilibrium of AI-powered trading by developing a theoretically motivated experimental framework, building upon the seminal work of Kyle (1985), to conduct simulation experiments. In our framework, multiple informed speculators interact with each other through trading, with all of them adopting RL algorithms to learn how to trade. Following the tradition of experimental research, our study is qualitative and intended as a proof-of-concept demonstration. Our headline result suggests that AI collusion can robustly emerge in environments with diverse market conditions; it impairs competition, leading to reduced liquidity, less informative pricing, and increased mispricing.

We further uncover the economics behind AI collusion and show that, depending on the trading environment, AI algorithms can reach collusive outcomes through two distinct mechanisms. On the

[1] Traditional algorithmic trading is based on rigid, human-defined trading protocols that are hardcoded.

[2] For example, the SEC proposed novel rules to regulate AI technologies (SEC, 2023). The European Securities and Markets Authority (ESMA) reported on AI use in European securities markets (Bagattini, Benetti and Guagliano, 2023).

one hand, in environments with low noise trading risk and low price efficiency,[3] AI algorithms can learn the price-trigger strategy to achieve collusive outcomes. This strategy closely resembles the theoretical mechanism proposed by Green and Porter (1984), whereby reversion to non-collusive competition occurs when the asset's price substantially diverges from its expected collusive level. We theoretically prove the existence of such a collusive subgame perfect Nash equilibrium for rational-expectation agents and illustrate the resemblance between the price-trigger strategy in theory and the strategy learned by AI algorithms. On the other hand, in environments with high noise trading risk or high price efficiency, we theoretically prove that the collusive subgame perfect Nash equilibrium sustained by the price-trigger strategy does not exist for rational-expectation agents. Interestingly, we show that AI algorithms will converge to a steady state featuring collusion due to the self-confirming bias in learning. This steady state can be theoretically characterized by an experience-based equilibrium, as in Fershtman and Pakes (2012), or a self-confirming equilibrium, as in Fudenberg and Levine (1993). The concept of this equilibrium is fundamentally connected to but is weaker than the concept of Nash equilibrium. In an experienced equilibrium, valuations may be accurate along the equilibrium path, as this is more commonly observed, but can be inaccurate off the equilibrium path, unless there is sufficient exploration of non-optimal actions (e.g., Fudenberg and Kreps, 1988, 1995; Cho and Sargent, 2008). In our context of AI-powered trading, the biased valuations stem from the high forgetting rates in RL algorithms, which constrain the algorithms' capacity to accurately estimate the expected payoff when the trading environment is noisy, thereby resulting in collusive outcomes.

Below, we first describe our theoretically motivated experimental framework and then elaborate on the main findings of this paper. We extend the canonical framework of Kyle (1985) in two aspects. First, we consider multiple informed speculators in a repeated-trading context. Second, we introduce a continuum of atomistic, information-insensitive investors who trade the asset, collectively creating a downward-sloping demand curve. At the beginning of each trading period, the asset's fundamental value is realized. Then, a continuum of noise traders collectively places an order flow, which is independent of the asset's value. Oligopolistic informed speculators are aware of the asset's value but remains uninformed about the noise trading flow when determining their optimal trading strategies. The market is cleared by the order flows of informed speculators, noise traders, information-insensitive investors, and the market maker. The market maker's order flow creates inventory costs for himself. As such, when determining the asset's price in equilibrium, the market maker's goal is to minimize the weighted average of inventory costs and pricing errors.

The trading environment can be summarized by two crucial characteristics, the level of noise trading risk and the degree of price efficiency. The level of noise trading risk is determined by the variance of the aggregate noise trading flow (Long et al., 1990). The price efficiency is determined by the price elasticity of demand of information-insensitive investors. Intuitively, a higher elasticity improves price efficiency by dampening the demand from information-insensitive investors. In the extreme scenario with an infinite elasticity, our trading environment has efficient pricing as in Kyle

[3]Noise trading risk reflects the magnitude of noise trading relative to the variation in the asset's fundamental value. Price efficiency captures to what degree the asset's price is aligned with its conditional expected fundamental value.

(1985). In other scenarios, prices are inefficient as in Kyle and Xiong (2001).

Before experimenting with AI algorithms, we first analytically solve the model. Our purpose is to provide a baseline characterization of informed speculators' collusive behavior in the presence of asymmetric information and the endogenous strategic pricing rules of the market maker. We show that an equilibrium with collusive outcomes can emerge through two distinct mechanisms. On the one hand, under the assumption of rational expectations, informed speculators can form a collusive subgame perfect Nash equilibrium sustained by a price-trigger strategy resembling that proposed by Green and Porter (1984), whereby reversion to non-collusive competition occurs when the asset's price substantially diverges from its expected collusive level. On the other hand, informed speculators can form a collusive experience-based (or self-confirming) equilibrium (Fudenberg and Levine, 1993; Fershtman and Pakes, 2012) if they uniformly undervalue aggressive trading strategies, perpetuating an incorrect system of outcome evaluation.

The collusive experience-based equilibrium always exists regardless of market conditions, as forming this sort of equilibrium only requires informed speculators to have biased evaluations off the equilibrium path. However, as a noteworthy theoretical contribution, we demonstrate that, under rational expectations, it is impossible to sustain a collusive subgame perfect Nash equilibrium through price-trigger strategies when the trading environment has high price efficiency (due to a high price elasticity of demand of information-insensitive investors) or high noise trading risk. Intuitively, sustaining price-trigger collusion requires both high price informativeness for monitoring and low price impacts of informed trading for maintaining informational rents. These two conditions cannot be simultaneously satisfied when price efficiency or noise trading risk is high. This novel result illuminates a mechanism distinct from existing theories on the impossibility of collusion under information asymmetry in the context of product market competition (Abreu, Milgrom and Pearce, 1991; Sannikov and Skrzypacz, 2007), which does not emphasize the price impacts of informed trading. When the trading environment has both low price efficiency and low noise trading risk, a collusive subgame perfect Nash equilibrium can be sustained by price-trigger strategies. In such an environment, we further show that collusion capacity increases, market liquidity decreases, price informativeness decreases, and mispricing increases, when the number of informed speculators drops, the level of noise trading risk decreases, or the subjective discount rate increases.

These theoretical results offer a useful benchmark to understand the AI trading behavior in our simulation experiments, where we replace the informed speculators in the model with informed AI speculators who operate RL algorithms to learn optimal trading strategies. Specifically, informed AI speculators adopt Q-learning algorithms to learn and guide their real-time trading decisions. Q-learning algorithms are recognized for their simplicity, transparency, and economic interpretability, serving as a foundational basis for various RL procedures that have significantly advanced the AI domain. To underscore the concept of AI collusion in our simulations, we intentionally use a minimal set of state variables for Q-learning algorithms, incorporating only one-period-lagged asset values and prices. Despite the AI algorithms being relatively simple compared with the trading environment, our simulation results remarkably indicate that informed AI speculators can intelligently form collusion across diverse trading environments. This takes place in the absence

of any formal agreement or communication that would traditionally be considered as an antitrust violation. The importance, and even requirement, of communication in collusion among humans is extensively documented in the literature of experimental economics.

There exist two types of collusion in our simulation experiments. In environments with both low price efficiency and low noise trading risk, the behavior of algorithmic collusion is in line with the theoretical results under rational expectation. After their algorithms converge, informed AI speculators are able to learn price-trigger strategies to sustain collusion, which can be thought of as a form of collusion driven by "artificial intelligence." While AI speculators may learn the price-trigger strategy in a different way compared to human speculators, the resulting patterns exhibit notable similarities as it is the threat of punishment that effectively deters each speculator from breaking the agreement. In both the model with rational-expectation speculators and our simulation experiments with AI speculators, significant price deviations lead to aggressive trading flows, resembling those in a non-collusive Nash equilibrium, which reduces the trading profits of all speculators.

By contrast, in environments with high price efficiency or high noise trading risk, informed AI speculators in our simulation experiments do not learn price-trigger strategies after their algorithms converge. Interestingly, they still achieve supra-competitive profits by reaching an experience-based collusive equilibrium. These simulation results with AI speculators are consistent with our theoretical results, which indicate that, in such an environment, collusion cannot be sustained by price-trigger strategies even under rational expectations, but a collusive experience-based equilibrium always exists regardless of market conditions. We further show that informed speculators achieve supra-competitive profits because they learn conservative trading strategies, under-reacting to their private information compared to the optimal strategy in a non-collusive equilibrium. The algorithmic collusion here is achieved through self-confirming bias in learning, which can be considered as a form of collusion driven by "artificial stupidity."

Learning bias emerges due to inconsistencies in reinforcement learning, which stem from the asymmetric effect of exploitation in the learning process. In environments exhibiting high levels of noise trading risk, the learning bias is especially pronounced, playing a crucial role in determining the learning process of informed AI speculators. In particular, in such environments, the realized payoffs of informed AI speculators include a substantial random element: Payoffs are high when the noise trading flow is favorable and low when it is unfavorable.[4] Because of this random element, exploitation in RL algorithms has an asymmetric impact on the learning process. If significant and unfavorable noise trading flows were experienced while using a strategy in the past, the algorithm may discontinue using that strategy in future iterations. Conversely, if positive outcomes were observed previously, the algorithm will likely continue utilizing the strategy, but the future iterations may still eliminate this strategy from the strategy space.[5] The asymmetric impact of exploitation

[4]The realized noise trading flow is unfavorable if it reduces the profits of informed speculators, compared to the scenario with zero noise trading flows. This means that the realized noise trading flow is positive (negative) when the realized fundamental value of the asset is positive (negative). Likewise, the realized noise trading flow is favorable if it is positive (negative) when the realized fundamental value of the asset is negative (positive).

[5]Because the algorithm will continue to use the strategy, the strategy's payoff perceived by the algorithm will be continually updated. This means that the positive effect generated the favorable noise trading flows occurred in the past will eventually be averaged out. Following this stage, there is a possibility that the AI speculator may once more encounter

suggests that the trading strategies with higher levels of randomness in realized payoffs are more likely to be eliminated from the strategy space and not used in the future. This means that aggressive strategies, which react strongly to private information, are less likely to be learned by RL algorithms due to their tendency to yield more volatile payoffs. As a result, informed AI speculators tend to adopt conservative strading strategies, enabling them to attain supra-competitive profits. The use of RL algorithms effective generates some sort of "risk aversion" in strategy choice.

Next, we investigate algorithmic collusion by varying the parameters in the simulation experiments. These two types of AI collusion, while both generating supra-competitive trading profits, exhibit very different comparative statics. If AI collusion is achieved through price-trigger strategies, a decrease in noise trading risk or an increase in the subjective discount rate factor leads to increased collusion capacity, similar to the predictions of the model with rational-expectation speculators. By contrast, if AI collusion is achieved through self-confirming bias in learning, a decrease in noise trading risk reduces collusion capacity because of reduced learning bias whereas varying the subjective discount rate has little impact on collusion capacity. Notably, informed AI speculators achieve supra-competitive profits mainly by trading against information-insensitive investors when collusion is achieved through price-trigger strategies, but by trading against noise traders when collusion is achieved through self-confirming bias in learning.

Finally, we study the role of hyperparameters, the exploration rate and the forgetting rate, on learning outcomes. If collusion is achieved through price-trigger strategies, we show that an increase in the exploration rate needs to be matched by a decrease in the learning rate to achieve high collusive profits.[6] If collusion is achieved through self-confirming bias, collusive profits are mainly determined by the forgetting rate while the exploration rate does not play an important role. Specifically, the collusive profits of all informed AI speculators increase as the forgetting rate becomes higher because a higher forgetting rate increases learning bias.

In our simulation experiments, we require all AI speculators to use the same algorithms with identical hyperparameters. Algorithm homogenization is instrumental, though not necessary, to achieve both types of AI collusion. Homogenization can emerge when speculators use similar foundational models, effectively forming a type of hub-and-spoke conspiracy.[7] Johnson and Sokol (2021) emphasize the prevalence of hub-and-spoke conspiracy in the context of e-commerce platforms, which generates anti-competitive effects. Many retailers adopt similar or even identical AI pricing algorithms, possibly supplied by a common service provider, who serves as the hub. In financial markets, the AI-powered trading systems used by informed speculators often rely on similar foundational models. This common practice, whether deliberate or incidental, can lead to a

substantial and unfavorable realized noise trading flows, leading to significant losses and the subsequent elimination of this trading strategy from the strategy space. In other words, due to the asymmetric impact of exploitation, it is the realization of unfavorable noise trading flows that determines whether a strategy is eventually learned (or eliminated) by the RL algorithm.

[6]This result similarly holds in the experimental study of Calvano et al. (2020), who demonstrate that AI algorithms can achieve collusion through grim-trigger strategy in a deterministic environment without information asymmetry.

[7]In the context of product market competition, the term hub-and-spoke conspiracy is a metaphor used to describe a cartel that includes a firm at one level of a supply chain, typically a supplier, acting as the "hub" of a wheel. Vertical agreements down the supply chain represent the "spokes." This common supplier facilitates the implicit coordination among its customers.

notable level of homogenization, a phenomenon documented by Bommasani et al. (2022), among others.

Homogenization can also arise from the autonomous learning of AI-powered trading systems. While integrating advanced algorithms can potentially disrupt collusion stemming from self-confirming bias in learning, it is improbable that any AI speculator would opt to gain an advantage by utilizing superior algorithms, given the nature of AI collusion. Intuitively, if one speculator adopts a superior algorithm, it may render the trading strategies of other AI speculators unprofitable, prompting them to also adopt equally or more advanced algorithms. This could trigger a competition towards algorithmic enhancement, leading to an equilibrium where trading profitability becomes minimal for all AI speculators. As a result, in order to collectively achieve supra-competitive profits, AI speculators autonomously converge towards adopting similarly basic algorithms in equilibrium. We demonstrate this concept formally by considering a simple extension of the baseline Q-learning algorithms. We enable informed AI speculators to learn the key parameter that governs the sophistication of their Q-learning algorithms, as well as their trading strategies, according to the algorithm chosen by the AI. Our simulation experiments robustly demonstrate that informed AI speculators may collectively opt for less advanced algorithms, despite the potential for one to boost its own profits by unilaterally choosing a more advanced algorithm while others continue using their current ones.

Related Literature. Our paper contributes to the growing literature on algorithmic collusion without explicit agreements or communication by introducing a novel mechanism — AI collusion through over-pruning bias in learning — and providing a comprehensive analysis of the theoretical benchmarks for different types of AI collusion and their underlying algorithmic mechanisms. Additionally, we examine which mechanism dominates under various market environments.[8] Waltman and Kaymak (2008) and Dolgopolov (2024) show that even stateless Q-learning algorithms can reach and sustain algorithmic collusion, provided the forgetting rate is high and the relative gains from collusion are substantial. However, their results hinge on the Boltzmann exploration scheme. When exploration instead follows a fixed-probability rule, such as the simple ε-greedy scheme used in this paper, their conclusions no longer hold. This distinction underscores that their algorithmic collusion mechanism is fundamentally different from ours, as it neither relies on price-trigger strategies nor stems from over-pruning bias in learning. Klein (2021) simulates a simplified sequential pricing duopoly in which firms set prices alternately, focusing on Markov perfect equilibrium — similar to Brown and MacKay (2023) — rather than setting prices simultaneously, as in many other studies. While Klein (2021) also demonstrates that Q-learning algorithms can coordinate on supra-competitive prices, its framework and algorithmic mechanism are fundamentally different from ours by design.

In a contemporaneous study, Banchio and Mantegazza (2024) introduce a novel algorithmic collusion mechanism in stateless Q-learning, termed spontaneous coupling. This mechanism arises

[8] AI collusion refers to algorithmic collusion among reinforcement learning algorithms that occurs without agreements, communication, or coded intent.

endogenously from statistical linkages between Q-learning algorithms' price estimates, leading to self-reinforcing cycles where firms sustain supra-competitive prices on average. Simulations confirm that spontaneous coupling can sustain price-fixing and market division. Importantly, learning rate asymmetries reinforce collusion, while greater noise disrupts it. This contrasts with the algorithmic collusion mechanisms highlighted in this paper, which operate through fundamentally different channels. Similar to Banchio and Mantegazza (2024), who focus on stateless Q-learning, an earlier work by Hansen, Misra and Pai (2021) shows that duopolies using independent, dynamically unsophisticated reinforcement learning algorithms — specifically, upper confidence bound (UCB) algorithms, a type of multi-armed bandit (MAB) algorithm — can spontaneously synchronize their price choices, sustaining supra-competitive prices over time. The mechanism is that algorithms designed to learn demand through pricing experiments can misinterpret price sensitivity by ignoring competitors' prices, leading them to systematically overestimate their own price sensitivity. Furthermore, when price experiments are similar, firms' price choices become correlated, reinforcing supra-competitive prices as a result of this overestimation. The key difference from Banchio and Mantegazza (2024) is that experiments occur fully at random, meaning that correlation arises through the estimates rather than through the timing of experimentation. The mechanism in Hansen, Misra and Pai (2021) is fundamentally different from ours, because in stateless reinforcement learning, AI collusion cannot be sustained through price-trigger strategies, which require tracing past prices as state variables. Additionally, greater noise strengthens AI collusion in our setting due to over-pruning bias in learning, whereas in Hansen, Misra and Pai (2021), greater noise weakens AI collusion. Lambin (2024) challenges Hansen, Misra and Pai (2021) by arguing that simultaneous exploration and slow learning, rather than price sensitivity overestimation, are the true drivers of correlated pricing and AI collusion. However, standard multi-agent reinforcement learning frameworks do not inherently require simultaneous moves in exploration, rendering this argument irrelevant for our proposed algorithmic mechanisms underlying AI collusion. Along similar lines, Abada and Lambin (2023) argue that AI collusion can arise from insufficient learning, particularly due to limited exploration. When AI algorithms fail to experiment broadly with alternative pricing strategies, they risk converging to equilibria that resemble collusion. However, Asker, Fershtman and Pakes (2024) show that while increasing exploration can reduce collusive tendencies, it does not eliminate algorithmic collusion. Expanding on this insight, this paper, as a main contribution to the literature, identifies a novel mechanism behind AI collusion: over-pruning bias in learning. This bias systematically excludes aggressive strategies prematurely from the set of potential optimal actions, fundamentally distorting the learning process. The root cause is the asymmetric effect of exploitation — where adverse and beneficial shocks impact learning differently. Exploitation reinforces successful strategies while discouraging the reconsideration of poorly rated ones, making reinforcement learning algorithms effectively risk-averse to randomness in rewards. Since aggressive trading strategies are more exposed to adverse noise trading flow shocks, this asymmetry causes them to be prematurely pruned from the learning process, reinforcing over-pruning bias. Ultimately, this bias embeds an implicit form of risk aversion into the algorithm's decision-making, leading to AI collusion.

Recently, in a noteworthy contribution, Calvano et al. (2020) study collusion by AI algorithms in a logit model of differentiated products, not only uncovering the existence of supra-competitive profits but also pinpointing how algorithms might learn to sustain collusive outcomes through grim-trigger strategies. Expanding upon this, our paper extensively broadens the AI experimental framework, moving from a scenario of perfect information and a static demand curve to one imbued with asymmetric information and a strategically-determined demand scheme. We characterize the various types of AI algorithmic collusion, whether occurring through price-trigger strategies or through self-confirming bias in learning, across diverse market environments.

Inspired by the simulation-based studies on AI algorithmic collusion, empirical research has also emerged, demonstrating that the use of AI algorithms in setting product prices can lead to collusion, resulting in heightened supra-competitive prices (e.g., Assad et al., 2023). Additionally, recent studies have started to focus on policy interventions aiming to obstruct the ability of algorithms to collude, thereby ensuring the maintenance of competitive prices. Specially, based on simulation-based studies, Johnson, Rhodes and Wildenbeest (2023) show that platform design can benefit consumers and the platform. However, achieving these gains may require policies that condition on past behavior and treat sellers in a non-neutral fashion. Harrington (2018) delves into critical policy issues surrounding the definition of collusion, such as whether collusion should necessarily entail an explicit agreement among conspirators, or if it might be more aptly defined as the maintenance of elevated prices, sustained by a reward-and-punishment scheme.

Our paper is among the first to investigate how the widespread adoption of AI-powered trading strategies might affect capital markets. The work of Colliard, Foucault and Lovo (2023) is closely related to our research, as it also explores the implications of interactions among Q-learning algorithms in capital markets. However, there are notable differences in focus between their work and ours. Specifically, Colliard, Foucault and Lovo (2023) focus on AI-powered oligopolistic market makers, while our study concentrates on AI-powered oligopolistic informed speculators who face perfectly competitive market makers. Their research illuminates the strategies that AI market makers would adopt by leveraging their market power. In contrast, our paper explores the dynamics and implications of algorithmic collusion among AI-powered informed speculators, particularly in the context of information-insensitive investors and perfectly competitive market makers. We provide novel insights into the strategies of informed AI speculators on how they leverage private information and maximize profits through autonomously forming collusion via distinct mechanisms.

2 AI-Powered Trading Algorithms

The traditional rule-based algorithmic trading system executes orders rigidly according to protocols predefined by human quantitative strategists. These rules are typically derived from technical analysis and statistical models. In contrast, AI-powered trading employs RL algorithms to dynamically adjust and autonomously optimize trading strategies in real-time.

The RL algorithm, a pivotal technique in AI, forms the foundation of numerous successful AI algorithms, like "AlphaGo," demonstrating the superiority of RL-backed AI over human cognitive

abilities in areas such as securities trading and other complex tasks. RL algorithms are model-free machine learning techniques that learn autonomously through trial-and-error experimentation, without relying on typical assumptions, such as the multi-agent system being on an equilibrium path or agents having knowledge of the true state and payoff distributions at equilibrium. The basic rationale behind RL algorithms centers on the principle that actions yielding higher rewards historically are more likely to be selected in the future, compared to those that have led to lesser rewards. By interacting with its environment and experimenting with different actions, the agent incrementally learns an optimal policy. Through continuous rounds of exploration and experimentation, it refines its strategy to prefer actions that offer the greatest long-term benefits, even without any knowledge of the environment beforehand. This iterative process enables the agent to progressively enhance its decision-making approach, consistently steering towards actions that maximize the cumulative rewards based on its gathered experiences.

While RL encompasses different variants (e.g., Watkins and Dayan, 1992; Sutton and Barto, 2018), we choose to focus on Q-learning for several reasons. First, Q-learning serves as a foundational framework for numerous dynamically sophisticated RL algorithms, upon which many recent AI breakthroughs are built.[9] However, it is important to note that AI trading algorithms currently in use may not exclusively rely on Q-learning principles. Second, Q-learning holds substantial popularity among computer scientists in practical applications. Third, Q-learning algorithms possess simplicity and transparency, offering clear economic interpretations, in contrast to the black-box nature of many machine learning and AI algorithms.

In the remainder of this section, we will concentrate on a multi-agent system of RL algorithms, detailing the Bellman equation for each agent, and describe the Q-learning algorithm that an agent employs. This discussion will cover how each agent iteratively updates its Q-function and strategy based on the received rewards, thereby optimizing its long-term outcomes through the Q-learning algorithm.

2.1 Bellman Equation and Q-Function

In a multi-agent Markov decision process environment, there are I agents, indexed by $i = 1, \cdots, I$. The state of the environment is represented by a Markov process, denoted by s. Each agent makes decisions based on the current state, which in turn evolves partly due to the collective actions of all agents within the system. Agent i's intertemporal optimization is characterized by the Bellman equation and solved recursively via dynamic programming:

$$V_i(s) = \max_{x_i \in \mathcal{X}} \left\{ \mathbb{E}\left[\pi_i | s, x_i\right] + \rho \mathbb{E}\left[V_i(s') | s, x_i\right] \right\}, \tag{2.1}$$

where $x_i \in \mathcal{X}$ is the action taken by agent i, with $\mathcal{X}$ denoting the set of available actions, π_i is the payoff received by agent i that depends on the chosen action x_i as well as the actions of other

[9] Q-learning and these dynamically sophisticated RL algorithms are typically employed in complex scenarios, where actions lead to state transitions, and each action taken in a state affects future states and rewards. In contrast, multi-armed bandit algorithms, which represent another category of RL algorithms, are employed in simpler settings where actions do not depend on previous ones and do not prompt state transitions based on the actions taken.

agents, and $s, s' \in S$ represent the states in the current and the next period, respectively, with S denoting the set of states. In general, s and s' can depend on agent i's individual characteristics and private signals. However, for our purpose of illustration, it is sufficient to concentrate on the simple setting where the same state applies uniformly to all agents in the system. The first term on the right-hand side, $\mathbb{E}[\pi_i|s, x_i]$, is agent i's expected payoff in the current period, and the second term, $\rho\mathbb{E}[V_i(s')|s, x_i]$, is agent i's continuation value, with ρ capturing the subjective discount rate factor.

The Bellman equation (2.1) represents the recursive formulation of dynamic control problems (e.g., Bellman, 1954; Ljungqvist and Sargent, 2012). It focuses on the equilibrium path, and thus the optimal value function $V_i(s)$ depends solely on the state variable s. In contrast to focusing solely on the equilibrium path, the Q function, denoted by $Q_i(s, x_i)$, extends the optimal value function to include the values of each state-action pair. This captures scenarios (or counterfactuals) that occur off the equilibrium path. By definition, the value of $Q_i(s, x_i)$ is the same as that in the curly brackets of the Bellman equation (2.1):

$$Q_i(s, x_i) = \mathbb{E}[\pi_i|s, x_i] + \rho\mathbb{E}\left[V_i(s')|s, x_i\right]. \tag{2.2}$$

Intuitively, the Q-function value, $Q_i(s, x_i)$, can be interpreted as the quality of action x_i in state s. The optimal value of a state, $V_i(s)$, is the maximum of all the possible Q-function values of state s. That is, $V_i(s) \equiv \max_{x' \in \mathcal{X}} Q_i(s, x')$. By substituting $V_i(s')$ with $\max_{x' \in \mathcal{X}} Q_i(s', x')$ in equation (2.2), we can establish a recursive formula for the Q-function as follows:

$$Q_i(s, x_i) = \mathbb{E}\left[\pi_i + \rho \max_{x' \in \mathcal{X}} Q_i(s', x') \middle| s, x_i\right]. \tag{2.3}$$

When both $|S|$ and $|\mathcal{X}|$ are finite, the Q-function can be represented as an $|S| \times |\mathcal{X}|$ matrix, which is often referred to as the Q-matrix.

2.2 Q-Learning Algorithm

If agent i possessed knowledge of its Q-matrix, determining the optimal actions for any given state s would be straightforward. In essence, the Q-learning algorithm is a method to estimate the Q-matrix in environments where the underlying distribution $\mathbb{E}[\cdot|s, x_i]$ is unknown and there are potentially limited observations for off-equilibrium pairs (s, x_i) in the data. By design, the Q-learning algorithm addresses both challenges simultaneously: it estimates the underlying distribution $\mathbb{E}[\cdot|s, x_i]$ using the law of large numbers while conducting trial-and-error experiments to explore various actions and generate off-equilibrium counterfactuals for (s, x_i).

The iterative experimentation of agent i starts from an arbitrary initial agent-i Q-matrix, denoted by $\widehat{Q}_{i,0}$, and updates its estimated Q-matrix $\widehat{Q}_{i,t}$ recursively as follows:

$$\widehat{Q}_{i,t+1}(s_t, x_{i,t}) = (1-\alpha)\underbrace{\widehat{Q}_{i,t}(s_t, x_{i,t})}_{\text{Past knowledge}} + \alpha \underbrace{\left[\pi_{i,t} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t}(s_{t+1}, x')\right]}_{\text{Present learning based on a new experiment}}, \tag{2.4}$$

where $\alpha \in [0,1]$ captures the forgetting rate,[10] s_t is the state that the iteration t concentrates on, s_{t+1} is randomly drawn from the Markovian transition probabilities conditional on the current state s_t, the chosen action x_i of agent i, and the collective actions of all other agents within the system. Here, $\widehat{Q}_{i,t}(s,x)$ is the estimated Q-matrix of agent i in the t-th iteration, and $\pi_{i,t}$ is the payoff in the t-th iteration, corresponding to agent i's choice of action $x_{i,t}$ and all other agents' actions.

Equation (2.4) indicates that for agent i in the t-th iteration, only the value of the estimated Q-matrix $\widehat{Q}_{i,t}(s,x)$ corresponding to the state-action pair $(s_t, x_{i,t})$ is updated to $\widehat{Q}_{i,t+1}(s_t, x_{i,t})$. All other state-action pairs remain unchanged. In other words, $\widehat{Q}_{i,t+1}(s,x) = \widehat{Q}_{i,t}(s,x)$ for cases where $s \neq s_t$ or $x \neq x_{i,t}$. The updated value $\widehat{Q}_{i,t+1}(s_t, x_{i,t})$ is computed as a weighted average of accumulated knowledge based on the previous experiments, $\widehat{Q}_{i,t}(s_t, x_{i,t})$, and learning based on a new experiment, $\pi_{i,t} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t}(s_{t+1}, x')$. A key distinction between the Q-learning recursive algorithm (2.4) and the Bellman recursive equation (2.1) lies in how they treat expectations for future payoffs and continuation Q-values. Q-learning algorithm (2.4) does not form expectations about the continuation value because the Markovian transition probabilities from s_t to s_{t+1} are unknown. Instead, it updates the Q-value using the actual realized payoff and the maximum Q-value of the randomly realized state s_{t+1} in the next step (i.e., the $(t+1)$-th iteration).

It is crucial to note that the forgetting rate α plays a significant role in the Q-learning algorithm, balancing past knowledge against present learning based on a new experiment. A higher α not only indicates a greater impact of present learning on the Q-value update but also implies that the algorithm forgets past knowledge more quickly, potentially leading to biased learning. To elaborate intuitively, let τ be the number of times that the Q-value of the state-action pair (s,x) has been updated in the past. As $\tau \to \infty$, the estimated Q-value of (s,x) is approximately equal to

$$\widehat{Q}_{i,t_{\tau+1}}(s,x) \approx \sum_{h=0}^{\tau-1} \alpha(1-\alpha)^h \left[\pi_{i,t_{\tau-h}} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t_{\tau-h}} \left(s_{t_{\tau-h}+1}, x' \right) \right], \tag{2.5}$$

where t_h represents the period in which the estimated Q-value of (s,x) receives the h-th update. Clearly, when α is not close to 0, the weights $\alpha(1-\alpha)^h$ decay rapidly as τ increases, diminishing the influence of past data or updates. This can lead to substantial bias in the approximation of expectations $\mathbb{E}[\cdot|s, x_i]$ for future payoffs and continuation Q-values by undermining the law of large numbers, thereby compromising the effectiveness of large sample approximations. Conversely, a smaller α slows the decay, preserving more past information and reducing bias. However, this requires significantly more iterations to achieve convergence to the steady state, increasing computational costs.

[10] The forgetting rate α captures how quickly a Q-learning algorithm de-emphasizes past data. For consistent learning, α must decay to zero to ensure effective large sample approximation in the estimated Q-matrix $\widehat{Q}_{i,t}$ as t grows large. A lower α reduces learning bias asymptotically but takes longer to reach the steady state, reflecting the algorithm's "intelligence level," with a lower α indicating a more advanced and accurate learning capability.

2.3 Experimentation

Based on the current state variable s_t, agent i chooses an action $x_{i,t}$ during iteration t using one of two modes: exploitation or exploration, as detailed below:

$$x_{i,t} = \begin{cases} \text{argmax}_{x \in \mathcal{X}} \widehat{Q}_{i,t}(s_t, x), & \text{with prob. } 1 - \varepsilon_t, \quad \text{(exploitation)} \\ \widetilde{x} \sim \text{uniform distribution on } \mathcal{X}, & \text{with prob. } \varepsilon_t. \quad \text{(exploration)} \end{cases} \tag{2.6}$$

To determine the mode, we employ the simple ε-greedy method. As outlined in equation (2.6), during the t-th iteration, agent i engages in the exploration and exploitation modes with exogenous probabilities ε_t and $1 - \varepsilon_t$, respectively. In the exploitation mode, agent i chooses its action to maximize the current state's Q-value based on past experience, given by $x_{i,t} = \text{argmax}_{x \in \mathcal{X}} \widehat{Q}_{i,t}(s_t, x)$. Conversely, in the exploration mode, agent i randomly chooses its action $\widetilde{x}$ from the set of all possible values in $\mathcal{X}$, each with equal probability.[11] Essentially, the exploration mode guides the Q-learning algorithm to experiment with suboptimal actions based on the current Q-matrix estimation, $\widehat{Q}_{i,t}$. As t approaches infinity, the pre-specified exploration probability ε_t monotonically decreases to zero. Sufficient exploration is crucial for accurately approximating the true Q-matrix, requiring many attempts of all actions in all states, especially in complex environments. However, this comes with a tradeoff: extensive exploration not only increases computational costs but can also introduce noise, impeding learning when multiple agents interact.

We focus on asynchronous learning, defined by (2.4) and (2.6), which requires no knowledge of the underlying economic environment or information structure. In contrast, synchronous learning updates all (s, x) pairs simultaneously based on a model, assuming precise knowledge of the economic and informational structure (e.g., Asker, Fershtman and Pakes, 2022, 2024). Model-based approaches face misspecification risks and often rely on unrealistic assumptions about the machine's knowledge of the trading environment, which is complex and difficult to model accurately.

3 Model and Laboratory Design

To set up the laboratory for our simulation experiments, we develop a model that incorporates only the minimal set of ingredients necessary to capture the economic context of securities trading and reveal key insights. Our model builds on the influential framework of Kyle (1985), highlighting financial markets as mechanisms for information aggregation, facilitated by market makers. This mechanism compels informed speculators to trade conservatively on private signals, thereby keeping price informativeness sufficiently low to preserve information rents. This informational perspective, central to financial market competition, goes beyond the traditional focus on product market competition among pricing algorithms (e.g., Calvano et al., 2020).

Specifically, our model introduces two key deviations from the Kyle (1985) baseline framework. First, we consider a setting with oligopolistic informed speculators in a repeated trading environment,

[11] For simplicity, we adopt a uniform distribution. However, a more intelligent distribution choice could make exploration more efficient and less costly.

engaging in trading different short-lived assets from one period to the next, rather than a single informed speculator operating in a one-period market.[12] Second, we incorporate information-insensitive investors (e.g., Kyle and Xiong, 2001; Vayanos and Vila, 2021) and market makers with inventory cost considerations. Together, these elements expand upon the efficient pricing baseline model of Kyle (1985) by introducing potential price inefficiencies. Importantly, the information-insensitive investors in our model need not exhibit behavioral biases; they can be fully rational yet remain unresponsive to short-term fundamental signals.

We derive theoretical results on the existence of collusive equilibrium sustained by two distinct mechanisms and analyze how market structures such as noise trading risk, the number of informed speculators, the time discount factor, and the presence of information-insensitive investors affect collusion capacity and market efficiency, including price informativeness, market liquidity, and mispricing. These results provide benchmarks for identifying AI collusive equilibrium in simulation experiments examined in Sections 4 through 6 and provide insights into the impact of AI collusion on market efficiency.

3.1 Economic Environment

Model Setup. Time is discrete, indexed by $t = 1, 2, ...$, and runs forever. There are $I \geq 2$ risk-neutral informed speculators, indexed by $i \in \{1, \cdots, I\}$, a representative noise trader, a representative information-insensitive investor, and a representative market maker. The environment is stationary, and all exogenous shocks are independent and identically distributed across periods.

In each period t, a short-lived asset is traded, reaching expiration at the end of the period with its fundamental value v_t realized. The fundamental value v_t is distributed as $N(\overline{v}, \sigma_v^2)$, where we set $\overline{v} \equiv \sigma_v \equiv 1$ for simplicity.[13] The noise trader's order flow u_t is distributed as $N(0, \sigma_u^2)$, where σ_u denotes the magnitude of noise trading risk.

Each informed speculator i knows v_t perfectly but does not observe the noise trader's order flow u_t when submitting a trade. Speculators understand that their order flow $x_{i,t}$ influences the market price p_t by altering the total order flow, thereby (i) shifting the market-clearing condition and (ii) partially revealing their private signals about v_t to other participants in the asset market. Specifically, informed speculator i solves:

$$V_i(s_t) = \max_{x_{i,t}} \mathbb{E}\left[(v_t - p_t)x_{i,t} + \rho V_i(s_{t+1})|s_t, x_{i,t}\right], \tag{3.1}$$

where $V_i(s_t)$ denotes the optimal value function of speculator i in state s_t, achieved by selecting the optimal trading order flow $x_{i,t}$. The term $(v_t - p_t)x_{i,t}$ represents the trading profit (or loss), while $\rho V_i(s_{t+1})$ is the discounted continuation value for the next-period state s_{t+1}, with $\rho \in (0, 1)$ being the subjective discount factor.

In equation (3.1), the state variable s_t encapsulates all relevant information required for informed

[12] Our repeated trading setup is distinct from a multi-round dynamic trading framework involving a long-term asset traded within a relatively prolonged period (e.g., Kyle, 1985; Holden and Subrahmanyam, 1992).

[13] For conciseness, the notations $\overline{v}$ and σ_v will be omitted in this manuscript when not needed for comprehension.

speculators' decision-making. Specifically, s_t includes variables such as $v_t, v_{t-1}, p_{t-1}, y_{t-1}, z_{t-1}$, as well as other historical variables if necessary. The quantity $y_t \equiv \sum_{i=1}^{I} x_{i,t} + u_t$ is the total order flow, consisting of order flows from both informed speculators and noise traders. All investors observe y_t but cannot disentangle its components in period t, leaving them unable to distinguish between informed and noise-driven trades. The quantity z_t is the demand of information-insensitive investors, whose collective demand curve is given by:

$$z_t = -\xi(p_t - \overline{v}), \quad \text{with } \xi \geq 0. \tag{3.2}$$

The same specification is adopted by Kyle and Xiong (2001), who justify it through the optimal portfolio choice made by a rational yet information-insensitive investor under certain assumptions.[14] These investors can be rational, even though they do not infer fundamental information from the market price p_t or others' trading behaviors, unlike the rational-expectations uninformed investors in the models of Grossman and Stiglitz (1980) and Kyle (1989). As discussed in Kyle and Xiong (2001), the logic behind specification (3.2) is straightforward: the information-insensitive investor, focusing on the ex-ante expected fundamental value $\overline{v}$, buys more as $p_t - \overline{v}$ becomes more negative, perceiving the asset as undervalued. Including information-insensitive investors in a noisy rational expectations framework is intended to capture relevant institutional frictions and rigid, technical-analysis-driven trading responses to price reversal signals.[15]

Trading occurs through the market maker, who sets the market price p_t to absorb order flow while minimizing inventory costs and pricing errors. Specifically, the market maker observes the total order flow, y_t, from informed speculators and the noise trader, as well as the order flow schedule, z_t, of information-insensitive investors, which is a deterministic function of the market price p_t, specified in (3.2). Given this information, the market maker sets p_t to minimize inventory costs and pricing errors, solving the following objective function:

$$\min_{p_t} \mathbb{E}\left[(y_t + z_t)^2 + \theta(p_t - v_t)^2 \middle| y_t\right], \tag{3.3}$$

where $\theta > 0$ represents the weight that the market maker places on minimizing pricing errors. Here, $\mathbb{E}[\cdot|y_t]$ denotes the market maker's expectation over v_t, conditioned on the observed combined order flow y_t and its understanding of the behavior of informed speculators in equilibrium.

To clear the market, the market maker assumes the position $-(y_t + z_t)$, incurring quadratic inventory costs, $(y_t + z_t)^2$, consistent with the existing literature, such as Mildenstein and Schleef (1983). The term $\theta(p_t - v_t)^2$ reflects the market maker's attempt to minimize pricing errors due to asymmetric information. The parameter θ acts as a reduced-form measure of the benefits from

[14] To derive the functional form of the aggregate demand curve of information-insensitive investors, one approach is to assume CARA utility maximization without any learning or strategic trading, as detailed in Online Appendix 1.1. Studies indicate that information-insensitive investors with low price elasticity of demand play an important role in shaping asset prices (e.g., Greenwood and Vayanos, 2014; Vayanos and Vila, 2021; Greenwood et al., 2023).

[15] This approach has been commonly adopted in the literature (e.g., Hellwig, Mukherji and Tsyvinski, 2006; Goldstein, Ozdenoren and Yuan, 2013).

reducing these errors, such as attracting greater trading flows. The first-order condition leads to

$$p_t = \frac{\xi}{\xi^2+\theta} y_t + \frac{\xi^2}{\xi^2+\theta}\overline{v} + \frac{\theta}{\xi^2+\theta}\mathbb{E}\left[v_t|y_t\right]. \tag{3.4}$$

In our analyses, we treat θ as a universally fixed, positive constant with a tiny magnitude. By fixing θ, we exclude it from the comparative-static analysis. With a positive constant θ, our model gains conceptual coherence by offering two meaningful extreme benchmarks. As ξ approaches infinity, the price p_t converges to $\overline{v} + \xi^{-1} y_t$, determined by the market clearing condition $y_t + z_t = 0$, as in Kyle and Xiong (2001). Conversely, as ξ decreases towards zero, p_t shifts to the efficient price $\mathbb{E}[v_t|y_t]$, as in Kyle (1985).[16] Incorporating the market maker captures financial markets as mechanisms for aggregating information, where sophisticated players infer fundamental values from the collective actions of others, integrating this information into the market price, as highlighted by Kyle (1985).

Model Interpretation. We provide a specific interpretation of the model to offer economic context for simulation experiments involving AI-powered trading algorithms, while recognizing that alternative interpretations are possible. At the start of each period t, a new short-lived security — specifically, a close-to-maturity derivative contract that expires at the end of the period — is traded. Examples include close-to-maturity options and futures. These derivative contracts attract informed speculators pursuing short-term profits, sentiment-driven and technical-analysis-driven traders seeking to capitalize on rapid price movements, and liquidity traders adjusting their portfolios for hedging or other rebalancing needs. As a result, close-to-maturity derivatives are among the most actively traded derivatives across different maturities.

The fundamental value v_t represents the payoff of these contracts at expiration, which occurs at the end of period t in the model. Informed speculators, such as hedge funds, specialize in extracting private signals about the final payoff of close-to-maturity options and futures, v_t, using proprietary or public data powered by advanced technologies.[17] These informed speculators typically operate with two teams: (i) an investment or quant research team that generates private signals about v_t, and (ii) an execution team that converts trading signals into strategically executed orders to maximize trading profits. Like the structure of Kyle models, our framework assumes that valuable private signals are already available, while focusing on the strategic execution of trades based on these signals. In other words, the AI-powered trading algorithms analyzed in this paper are those employed by the execution team to convert private signals into strategic trading orders, rather than those used by the investment or quant research team to generate private signals. Noise traders, on the other hand, are market participants whose trading decisions are unrelated to fundamental information or technical analysis of market prices. Instead, they trade based on liquidity needs, portfolio rebalancing, market sentiment, or rumors.

[16] Further discussions are provided in Online Appendix 1.1.

[17] Conceptually, "private signals" here include not only information derived from proprietary sources but also predictive trading signals extracted from vast amounts of public data using advanced technologies such as machine learning (ML) and large language models (LLMs). While the underlying data may be publicly available, the ability to process and extract valuable predictive trading signals from it remains beyond the reach of most investors.

Information-insensitive investors, such as retail investors employing technical analysis and institutional investors seeking hold-to-maturity positions to hedge short-term risks, typically remain unresponsive to real-time fundamental information related to the terminal payoff v_t of close-to-maturity options and futures. Retail investors using technical analysis base their trades strictly on observed price patterns in the market (e.g., Lo and MacKinlay, 1999; Lo, Mamaysky and Wang, 2000; Chen, Peng and Zhou, 2024). The demand specification (3.2) captures the essence of certain technical analysis strategies, assuming that a positive spread $p_t - \overline{v}$ indicates overbought conditions with prices likely to fall, whereas a negative spread $p_t - \overline{v}$ indicates oversold conditions with prices likely to rise. Specifically, the demand specification captures technical analysis tools that provide signals for likely price reversals. Additionally, information-insensitive investors include institutions such as pension funds, insurance companies, and mutual funds, which may purchase close-to-maturity derivatives and hold them to expiration as hedges against near-term risks. These investors tend to increase long positions when the hedge cost p_t is lower.

Market makers in close-to-maturity options and futures markets play a critical role by providing liquidity, facilitating trades, and enhancing price discovery. Market makers are sophisticated individuals and institutions that use advanced algorithms and robust risk management techniques.[18] Their primary function in our model is to support price discovery by integrating information from other market participants' trading behaviors into the market price. This modeling approach captures the concept of financial markets as information aggregation mechanisms, where sophisticated players infer fundamental values from the collective actions of others and integrate this information into the market price.

3.2 Theoretical Benchmarks

We consider three theoretical benchmarks to characterize the steady-state behavior of informed speculators: the non-collusive Nash equilibrium, the perfect cartel, and the collusive equilibrium, denoted by N, M, and C in the superscripts of variable notations, respectively.

Benchmark I: Non-Collusive Nash Equilibrium. This describes the one-shot Nash equilibrium in the stage game of repeated trading, where each informed speculator i, leveraging private signal v_t, maximizes its expected profit by solving:

$$x^N(v_t) = \underset{x_i \in \mathcal{X}}{\operatorname{argmax}} \, \mathbb{E}[(v_t - p^N(y_t))x_i | v_t],$$

under the assumption that other speculators adhere to the equilibrium strategy $x^N(v_t)$. Here, $p^N(y_t)$ denotes the equilibrium market price as a function of the total flow y_t. Specifically, speculator i chooses optimal x_i, while accounting for its effect on the equilibrium price, expressed as $p^N(y_t) = \overline{v} + \lambda^N y_t$, where $y_t = x_i + (I-1)x^N(v_t) + u_t$. Speculators recognize that λ^N is dependent on market parameters and focus on the linear strategy $x^N(v_t) \equiv \chi^N(v_t - \overline{v})$ in equilibrium. Details are in Online Appendix 1.1.

[18] Examples include Citadel Securities, Virtu Financial, Jane Street, Optiver, IMC, and SIG.

Benchmark II: Perfect Cartel Benchmark. This benchmark describes a scenario where informed speculators operate as a monopolistic cartel. The cartel, leveraging private signal v_t, maximizes its expected profit by solving:

$$x^M(v_t) = \underset{x \in \mathcal{X}}{\operatorname{argmax}} \mathbb{E}[(v_t - p^M(y_t))x | v_t],$$

fully accounting for the impact of trading flow x on the equilibrium price $p^M(y_t) = \overline{v} + \lambda^M y_t$, where $y_t = Ix + u_t$. Speculators recognize that λ^M is determined by market parameters and focus on the linear strategy $x^M(v_t) \equiv \chi^M(v_t - \overline{v})$ in equilibrium. Details are in Online Appendix 1.1.

Benchmark III: Collusive Equilibrium. Below, we define the economic concept of collusive equilibrium in terms of agents' behaviors, rather than the intent typically emphasized in legal definitions.

Definition 3.1 (Collusive Equilibrium)**.** *A collusive equilibrium is characterized by two key properties: (i) agents achieve collective supra-competitive profits that exceed those obtained in the non-collusive Nash equilibrium, and (ii) agents have the option to deviate from equilibrium actions for short-term gains, and such deviations impose costs on others.*

In our model, two distinct economic mechanisms can theoretically sustain a collusive equilibrium: the collusive Nash equilibrium, sustained by price-trigger strategies, and the collusive experience-based equilibrium, sustained by learning bias. We explore their existence and theoretical properties in Section 3.3.

3.3 Two Mechanisms Underlying Collusive Equilibrium

Collusive Nash Equilibrium Sustained by Price-Trigger Strategies. Collusive trading behavior, as outlined in Definition 3.1, can arise as a subgame perfect Nash equilibrium sustained by punishment-based trigger strategies. In securities trading, information asymmetry and noise trading risk complicate monitoring each other's trades. However, with sufficiently high price informativeness, informed speculators can imperfectly infer order flows of others from market prices, enabling tacit collusion.[19] Tacit collusion sustained by price-trigger strategies was introduced by Green and Porter (1984) and Abreu, Pearce and Stacchetti (1986). We formalize this theoretical concept below.

Definition 3.2 (Collusive Nash Equilibrium through Price-Trigger Strategies)**.** *A collusive equilibrium in trading, sustained by price-trigger strategies, is a subgame perfect Nash equilibrium with two regimes: the collusive regime and the punishment regime. In the collusive regime, informed speculators implicitly coordinate by submitting less aggressive order flows than in the non-collusive Nash equilibrium. If prices cross a critical threshold, signaling a suspected deviation, the system shifts to the punishment regime, characterized by the non-collusive equilibrium, where profits are significantly lower than in the collusive regime.*

[19]The study of tacit collusion via grim-trigger strategies with observable actions, initiated by Fudenberg and Maskin (1986) and Rotemberg and Saloner (1986), has been further developed in recent finance research, including asset pricing studies (e.g., Opp, Parlour and Walden, 2014; Dou, Ji and Wu, 2021*a*,*b*; Chen et al., 2023, 2024).

In the collusive regime, informed speculators adopt a trading strategy, $x^C(v_t) \equiv \chi^C(v_t - \overline{v})$ in period t, which is less aggressive than that in the non-collusive Nash equilibrium (i.e., $\chi^C < \chi^N$). When selecting χ^C, they anticipate the corresponding equilibrium market price to be

$$p_t^C = \overline{v} + \varphi^C(v_t - \overline{v}) + \lambda^C u_t, \tag{3.5}$$

where φ^C and λ^C measure the market price's sensitivity to $v_t - \overline{v}$ and u_t, respectively. This reflects informed speculators' understanding of how φ^C and λ^C depend on market parameters and the equilibrium trading strategy $x^C(v_t)$. If $v_t > \overline{v}$ and the observed market price p_t exceeds the critical threshold for the price-trigger strategy, defined as $q_+^C(v_t) \equiv \mathbb{E}\left[p_t^C | v_t\right] + \lambda^C \sigma_u \omega$, i.e., $p_t > q_+^C(v_t)$, then speculators revert to the punishment regime, characterized by the non-collusive Nash equilibrium, in period $t+1$ with probability η. Likewise, if $v_t < \overline{v}$ and p_t falls below the lower threshold, $q_-^C(v_t) \equiv \mathbb{E}[p_t^C | v_t] - \lambda^C \sigma_u \omega$, i.e., $p_t < q_-^C(v_t)$, then they may also revert to the punishment regime in period $t+1$ with probability η. Upon entering the punishment regime at $t+1$, they will remain there with the same probability η in each period until $t+T$. Thus, the triple (η, ω, T) characterizes an implicit coordination scheme among informed speculators. The space of price-trigger strategies is $\Omega \equiv \{(\eta, \omega, T) : \eta \in [0,1],\ \omega \in [0, \bar{\omega}],\ T \in \mathbf{N}\}$. As shown by Sannikov and Skrzypacz (2007, Lemma 3), a tail test with a bang-bang property is the optimal mechanism for maximizing the expected continuation payoff while ensuring incentives against a single deviation. Given this, we naturally focus on the collusive Nash equilibrium sustained by price-trigger strategies, which functions as the tail test for deviation with a bang-bang property in this trading setting. To effectively deter deviations (i.e., ensure a powerful tail test for deviation), the test size (or type I error) cannot be too small — a principle grounded in seminal statistical theories such as the Neyman-Pearson lemma. Thus, the upper bound $\bar{\omega}$ is set sufficiently large to make the test size negligible, i.e., $1 - \Phi(\bar{\omega}) \approx 0$, rendering the tail test practically powerless.[20] Further details on the collusive Nash equilibrium sustained by price-trigger strategies are provided in Online Appendix 1.1.

Collusive Experience-Based Equilibrium Sustained by Learning Bias. Collusive trading behavior, as outlined in Definition 3.1, can also emerge as an outcome of an experience-based equilibrium defined by Fershtman and Pakes (2012), which is closely related to the concept of self-confirming equilibrium (e.g., Fudenberg and Levine, 1993; Battigalli et al., 2015).

Compared to Nash equilibrium, these alternative equilibrium concepts are weaker because they allow players to hold incorrect or systematically biased evaluations of off-path outcomes, as these evaluations are shaped by their past experiences. As a result, while evaluations along the equilibrium path are correct, being consistent with observed outcomes of equilibrium strategies, off-path evaluations may remain inaccurate or systematically biased unless players engage in sufficient experimentation with non-optimal actions (e.g., Fudenberg and Kreps, 1988, 1995; Cho, Williams and Sargent, 2002; Cho and Sargent, 2008).

Specifically, an experience-based equilibrium is characterized by: (i) a recurrent Markovian state

[20] For example, when $\bar{\omega} = 8$, the test size is $1 - \Phi(\bar{\omega}) = 6.6 \times 10^{-16}$.

process, (ii) an optimization condition requiring strategies to be optimized based on potentially incorrect outcome evaluations, and (iii) a consistency condition requiring that expected discounted net cash flows according to the true distribution, generated by equilibrium-path optimal strategies, are aligned with equilibrium-path evaluations. Crucially, this consistency condition applies only to equilibrium-path outcomes. More precisely, players' beliefs or evaluations about off-equilibrium-path outcomes need not align with expected discounted cash flows under the true distribution, allowing for significant biases. In sum, as long as equilibrium-path evaluations match historically observed outcomes, these biases can persist and, in turn, sustain the equilibrium path.

We formalize the theoretical concept of collusive experience-based equilibrium sustained by learning bias below.

Definition 3.3 (Collusive Experience-Based Equilibrium through Learning Bias)**.** *A collusive equilibrium in trading, sustained by learning bias, is an experience-based equilibrium in which informed speculators systematically undervalue aggressive trading strategies due to an incorrect outcome evaluation system. This system remains uncorrected as learning is confined to outcomes observed along the equilibrium path. A notable case of such an equilibrium arises from a specific form of learning bias — over-perceived aversion to noise trading risk. In this case, the outcome evaluation system is biased solely by the perceived disutility associated with aversion to noise trading risk:* $-\frac{\varsigma}{2}\chi^2\sigma_u^2$*, where* $\varsigma > 0$ *represents the degree of over-perceived aversion and* $\chi > 0$ *reflects the aggressiveness of the trading strategy* $x(v_t) \equiv \chi(v_t - \overline{v})$.

3.4 Existence of Collusive Equilibrium

Existence of Collusive Nash Equilibrium Sustained by Price-Trigger Strategies. Sustaining coordination through price-trigger strategies hinges critically on high price informativeness to enable effective monitoring. Proposition 3.1 below demonstrates the impossibility of sustaining a collusive Nash equilibrium via price-trigger strategies in a financial market when noise trading risk, captured by σ_u, is large or when the presence of information-insensitive investors, captured by ξ, is small relative to θ, defined in Equation (3.3).

When noise trading risk σ_u is large, noise trading flow u_t dominates price fluctuations, as shown in (3.5), overshadowing informed trading and reducing price informativeness. This situation parallels oligopolistic product market competition with latent random price shocks, as analyzed by Abreu, Milgrom and Pearce (1991) and Sannikov and Skrzypacz (2007). Applying the same economic logic, high noise trading risk in financial markets undermines market prices as a monitoring tool, making it impossible to sustain a collusive trading equilibrium through price-trigger strategies in financial markets.

More importantly, our paper provides new insights into the conditions that enable or prevent tacit collusion in financial market trading, which can be fundamentally distinct from tacit collusion in product pricing in goods markets, as studied by Abreu, Milgrom and Pearce (1991) and Sannikov and Skrzypacz (2007). Specifically, when ξ is small relative to θ, reflecting a minimal presence of information-insensitive investors, the market maker's objective in (3.3) focuses on price recovery, with minimal emphasis on inventory cost minimization. This environment closely aligns with

the standard Kyle (1985) benchmark, which conceptualizes financial markets as mechanisms for information aggregation, where sophisticated participants infer fundamental values from the collective actions of others and incorporate this information into prices. In such an environment, informed investors, understanding how financial markets aggregate information into prices, must trade strategically and cautiously on private signals to secure meaningful information rents. This deliberate and restrained trading reduces price informativeness, weakening prices as effective monitoring tools. As a result, it becomes impossible to sustain a collusive trading equilibrium through price-trigger strategies in financial markets, regardless of the level of noise trading risk σ_u.

Proposition 3.1 (Feasibility of Price-Trigger Strategies)**.** *With all other parameters held constant, a collusive Nash equilibrium sustained by price-trigger strategies is not feasible if ξ is small relative to θ or if σ_u is large. Conversely, such an equilibrium exists only if ξ is sufficiently large relative to θ and σ_u is sufficiently small.*

The detailed proof is provided in Online Appendix 1.3.

Existence of Collusive Experience-Based Equilibrium Sustained by Learning Bias. In contrast to the collusive Nash equilibrium sustained by price-trigger strategies in Proposition 3.1, a collusive equilibrium driven by learning bias, especially through the self-confirming learning process, can robustly arise as an experience-based equilibrium, as shown in Proposition 3.2.

Proposition 3.2 (Existence of Collusion Through Learning Bias)**.** *A collusive experience-based equilibrium sustained by learning bias, with any trading strategy $\chi^C \in [\chi^M, \chi^N)$, exists for all $\xi \geq 0$ and $\sigma_u \geq 0$. In this equilibrium, informed speculators uniformly undervalue aggressive trading strategies due to an incorrect outcome evaluation system, which remains uncorrected as learning is based solely on observed outcomes along the equilibrium path. Furthermore, a collusive experience-based equilibrium can also be sustained by learning bias induced by over-perceived aversion to noise trading risk, characterized by the over-perceived aversion coefficient ς, as introduced in Definition 3.3, with an equilibrium trading strategy $\chi^C \in [\chi^M, \chi^N)$.*

The detailed proof is provided in Online Appendix 1.4.

3.5 The Impact of Collusive Informed Trading on Market Efficiency

To determine, based on the observable experimental outcomes in Section 4, whether informed AI speculators engage in tacitly collusive trading through price-trigger strategies or over-pruning bias in learning driven by noise trading risk, we derive testable theoretical properties of collusive equilibrium for each of these two distinct economic mechanisms.

Proposition 3.3 (Supra-Competitive Nature of Collusion)**.** *Let π^M, π^C, and π^N represent the expected profits of informed speculators in the perfect cartel benchmark, the collusive equilibrium (sustained either by price-trigger strategies or learning bias), and the non-collusive equilibrium, respectively. These profits satisfy the relationship*

$$\Delta^C \equiv \frac{\pi^C - \pi^N}{\pi^M - \pi^N} \in (0, 1]. \tag{3.6}$$

where Δ^C represents the normalized trading profitability of informed speculators in the collusive equilibrium.[21]

The detailed proof is provided in Online Appendix 1.5.

Definition 3.4. *The price informativeness, market liquidity, and mispricing are measured, respectively, by*

$$\mathcal{I} \equiv \frac{\text{var}(x_t)}{\text{var}(u_t)}, \quad \mathcal{L} \equiv \left[\frac{\partial |m_t|}{\partial u_t}\right]^{-1}, \quad \textit{and} \quad \mathcal{E} \equiv |\mathbb{E}[p_t|v_t] - v_t|\,, \tag{3.7}$$

where x_t, z_t, u_t, and $m_t \equiv -(y_t + z_t)$ are the total order flow of informed speculators, information-insensitive investors, noise traders, and market makers, respectively, and p_t is the market price.

In the following proposition, we examine how Δ^C, $\mathcal{I}^C$, $\mathcal{L}^C$, and $\mathcal{E}^C$ vary across different market structures and information environments within the collusive equilibrium, driven by two distinct mechanisms.

Proposition 3.4 (Market Structures and Collusive Trading: Consequences for Market Efficiency)**.** *The two collusion mechanisms yield similar implications when I changes, differing implications when ρ varies, and opposing implications when σ_u changes:*

(i) If a collusive Nash equilibrium sustained by price-trigger strategies exists, the following holds in this equilibrium when I is sufficiently large:

$$\rho \downarrow, \quad \sigma_u \uparrow, \quad \textit{or} \quad I \uparrow$$

$$\Longrightarrow \quad \Delta^C \downarrow \quad (\textit{i.e., collusion capacity} \downarrow) \tag{3.8}$$

$$\Longrightarrow \quad \mathcal{I}^C/\mathcal{I}^M \uparrow, \quad \mathcal{L}^C/\mathcal{L}^M \uparrow, \quad \textit{and} \quad \mathcal{E}^C/\mathcal{E}^M \downarrow \quad (\textit{i.e., market efficiency} \uparrow), \tag{3.9}$$

where C and M represent the collusive Nash equilibrium and the perfect cartel benchmark, respectively.

(ii) If a collusive experience-based equilibrium sustained by over-perceived aversion to noise trading risk exists, the following holds in this equilibrium:

$$\sigma_u \downarrow, \quad \textit{or} \quad I \uparrow$$

$$\Longrightarrow \quad \Delta^C \downarrow \quad (\textit{i.e., collusion capacity} \downarrow) \tag{3.10}$$

$$\Longrightarrow \quad \mathcal{I}^C/\mathcal{I}^M \uparrow, \quad \mathcal{L}^C/\mathcal{L}^M \uparrow, \quad \textit{and} \quad \mathcal{E}^C/\mathcal{E}^M \downarrow \quad (\textit{i.e., market efficiency} \uparrow), \tag{3.11}$$

where C and M represent the collusive experience-based equilibrium and the perfect cartel benchmark, respectively. The result for $\mathcal{L}^C/\mathcal{L}^M$ holds when ξ is sufficiently large. Importantly, ρ does not affect Δ^C, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, or $\mathcal{E}^C/\mathcal{E}^M$ in this equilibrium.

The detailed proof is provided in Online Appendix 1.6.

[21] Clearly, a greater Δ^C signifies higher collusion capacity. We adopt Δ^C as a measure for collusion capacity, following Calvano et al. (2020). Similar measures are also used in empirical studies like Dou, Wang and Wang (2023).

4 Simulation Experiments on AI Trading Algorithms

As a proof of concept, in this section, we run simulation experiments to test whether informed AI speculators, using autonomous model-free Q-learning algorithms, can achieve and sustain a form of collusive behavior under asymmetric information, with an adaptive asset demand curve that endogenously responds to informed trading strategies. Crucially, we examine whether these algorithms can do so without communication or explicit agreements that typically violate competition laws.

4.1 Algorithms as Experimental Subjects

Informed AI Speculators. We now analyze the algorithms' behavior as experimental subjects, detailed in Section 3.1. Specifically, these experiments replace the theoretical agents, referred to as "informed speculators" in the model, with Q-learning algorithms, as outlined in Section 2. To emphasize the key qualitative insights from these experiments, we use the simplest and most basic form of the Q-learning algorithm.

The dimensionality of the state variable vector s_t directly impacts the learning capacity and efficiency of Q-learning algorithms. High-dimensional state spaces create computational challenges, often requiring deep learning techniques for function approximation and effective exploration.[22] To ensure numerical tractability, transparency, and highlight key insights, we select a minimal set of state variables, $s_t \equiv \{p_{t-1}, v_{t-1}, v_t\}$, which capture the information advantage of informed speculators and enable AI collusion through price-trigger strategies, akin to the theoretical benchmark of the collusive Nash equilibrium in Definition 3.2.[23] In this setup, informed AI speculators rely on private signal v_t for trading in period t and retain a one-period memory of p_{t-1} and v_{t-1} for decision-making. In our simulation experiments, we find that expanding the state variable s_t by incorporating additional variables, such as lagged order flows or extended histories of market prices and fundamental values, strengthens tacit collusion among informed AI speculators through price-trigger strategies, resulting in higher trading profits. By limiting s_t to p_{t-1}, v_{t-1}, and v_t, we impose a stringent bar for Q-learning algorithms to achieve AI collusion sustained by price-trigger strategies.

Adaptive Market Maker. The market maker does not know the distributions of randomness. It stores and analyzes historical data on the asset's value and price, the order flows from information-insensitive investors, and the combined order flows from informed AI speculators and the noise trader, i.e., $\mathcal{D}_t \equiv \{v_{t-\tau}, p_{t-\tau}, z_{t-\tau}, y_{t-\tau}\}_{\tau=1}^{T_m}$, where T_m is a large integer. The market maker estimates the demand curve of information-insensitive investors and the conditional expectation of the asset's

[22] Reinforcement learning algorithms, augmented by deep learning techniques to address high-dimensionality challenges, form the backbone of many successful real-world AI applications, including "AlphaGo."

[23] Tracking both p_{t-1} and v_{t-1}, rather than just p_{t-1}, helps informed AI speculators assess potential deviations in period $t-1$ by comparing p_{t-1} against v_{t-1}.

value, $\mathbb{E}[v_t|y_t]$, using the following linear regression models, respectively:

$$z_{t-\tau} = \xi_0 - \xi_1 p_{t-\tau} + \epsilon_{z,t-\tau}, \text{ and } v_{t-\tau} = \gamma_0 + \gamma_1 y_{t-\tau} + \epsilon_{v,t-\tau}, \text{ where } \tau = 1, \cdots, T_m. \tag{4.1}$$

Here, $\epsilon_{z,t-\tau}$ and $\epsilon_{v,t-\tau}$ represent the residual terms from linear regressions. The estimated coefficients $\widehat{\xi}_{0,t}$, $\widehat{\xi}_{1,t}$, $\widehat{\gamma}_{0,t}$, and $\widehat{\gamma}_{1,t}$ are based on the rolling-window dataset $\mathcal{D}_t$ in period t. The pricing rule adaptively follows the optimal policy through a plug-in procedure:

$$\widehat{p}_t(y) = \widehat{\gamma}_{0,t} + \widehat{\lambda}_t y \text{ with } \widehat{\lambda}_t = \frac{\theta\widehat{\gamma}_{1,t} + \widehat{\xi}_{1,t}}{\theta + \widehat{\xi}_{1,t}^2}, \tag{4.2}$$

where θ is defined in (3.3). Our results remain robust even when the market maker employs Q-learning algorithms (see Online Appendix 3.11).

Protocol for Simulation-Based Experiments. We summarize the experimental protocol as follows. At $t = 0$, each informed AI speculator $i \in \{1, \cdots, I\}$ is assigned with an arbitrary initial Q-matrix $\widehat{Q}_{i,0}$ and state s_0. Then, the economy evolves from t to $t+1$ according to the following steps:

(1) In period t, each informed AI speculator i independently enters exploration with probability ε_t or exploitation with probability $1 - \varepsilon_t$, submitting order flow $x_{i,t}$, as in (2.6).

(2) The noise trader submits its order flow u_t, which is randomly drawn from $N(0, \sigma_u^2)$.

(3) The market maker analyzes the historical data $\mathcal{D}_t \equiv \{v_{t-\tau}, p_{t-\tau}, z_{t-\tau}, y_{t-\tau}\}_{\tau=1}^{T_m}$ to estimate the optimal pricing rule $\widehat{p}_t(y)$ according to (4.2). Upon observing $y_t = \sum_{i=1}^{I} x_{i,t} + u_t$, the market price is set at $p_t = \widehat{p}_t(y_t)$.

(4) Observing p_t, information-insensitive investors submit their aggregate order flow z_t in accordance with (3.2). Each informed AI speculator i realizes its profits $\pi_{i,t} = (v_t - p_t)x_{i,t}$.

(5) At the start of period $t+1$, the state variable transitions from $s_t = \{p_{t-1}, v_{t-1}, v_t\}$ to $s_{t+1} = \{p_t, v_t, v_{t+1}\}$, where v_{t+1} is independently drawn from $N(\overline{v}, \sigma_v^2)$. Each informed AI speculator i updates its Q-value for $(s_t, x_{i,t})$ using the recursive rule in (2.4).

Merits of Simulation-Based Experiments for Algorithms. The interaction between AI speculators using Q-learning with lagged prices as state variables, an adaptive market maker who learns based on historical data, and randomness from noise traders and stochastic asset values poses significant challenges to proving general mathematical results on the system's convergence and asymptotic properties. Like prior studies (e.g., Calvano et al., 2020; Colliard, Foucault and Lovo, 2023), our simulation-based experiments provide an effective framework for examining algorithmic behavior, interactions, and the resulting equilibrium for three reasons. First, no general mathematical results on convergence exist for our setting, let alone theoretical results characterizing the properties of the theoretical limit under exact convergence conditions.

Second, while stochastic approximation theorems can potentially establish theoretical convergence under specific conditions, they rely on strict regularity assumptions, such as decaying hyperparameters over iterations. In practice, however, these conditions are rarely met; for example, hyperparameters are often held constant. As a result, the steady-state behavior observed from numerical convergence may be more practically relevant than the theoretical limit derived under exact convergence conditions.[24] In real-world applications, particularly in robotics and securities trading, RL algorithms in multi-agent environments face several practical challenges: the lack of theoretical guarantees on convergence and equilibrium properties, the costly yet necessary explorations required for learning, the inherently slow learning process, and the high cost and sparsity of real-world data. These challenges make real-time training impractical. As a result, training RL algorithms in synthetic (simulated) environments has become a widely adopted approach, mirroring the spirit of our simulation-based experiments. For example, hedge funds often use synthetic financial market simulations to train RL-based execution strategies before deploying them in live trading, just as autonomous vehicle systems are trained in simulated driving environments before real-world implementation.

Third, even if a theoretical analysis of a multi-agent system with Q-learning algorithms in a repeated game setting like ours were feasible, despite being widely seen as intractable, the mathematical proofs would provide little insight into why or how algorithms reach a collusive equilibrium. This is because such analyses rely on stochastic approximation results, focusing on verifying high-level regularity conditions and technical details rather than offering economic intuition.[25] Supplementing simulation-based experimental studies across various trading environment specifications in our general model, we provide transparent intuitions and heuristic proofs for the numerical convergence of multiple informed AI speculators using Q-learning algorithms, as well as the steady-state properties of the AI trading equilibrium, in a simplified setting. These simulation findings are presented in Sections 5 and 6.1, with further elaboration on the heuristic proofs in Online Appendix 2.

4.2 Numerical Specifications

We detail the numerical specifications of our simulations, including the discretization of state and action spaces, the initialization of Q-matrices, the selection of parameters, and the criteria for determining numerical convergence.

Discretization of State and Action Spaces. We approximate the distribution $N(\overline{v}, \sigma_v)$ using n_v grid points, $\mathbb{V} = \{v_1, \cdots, v_{n_v}\}$, with equal probabilities assigned to each grid. The grid points are located

[24] We emphasize that the goal of simulation-based experiments for algorithms is fundamentally different from numerically solving theoretical equilibria derived under exact convergence conditions in macroeconomic and financial models (e.g., Kubler and Schmedders, 2005; Dou et al., 2023; Duarte, Duarte and Silva, 2024; Hansen, Khorrami and Tourre, 2024).

[25] Recent studies have provided mathematical proofs of the theoretical convergence of various Q-learning algorithms to (collusive) Nash equilibria in simplified models, typically the 2×2 Prisoner's Dilemma setting (e.g., Cartea et al., 2022; Possnig, 2024). These proofs rely on directly applying existing results in stochastic approximation, primarily focusing on verifying high-level regularity conditions and technical details. Consequently, they offer limited intuitive insights into the algorithmic mechanisms driving the convergence of algorithmic systems.

according to $v_k = \overline{v} + \sigma_v \Phi^{-1}((2k-1)/(2n_v))$ for $k = 1, \cdots, n_v$, where Φ^{-1} is the inverse cumulative density function of the standard normal distribution.[26] We discretize the choice space of informed AI speculator i for order flow x_i using grids based on the optimal trading strategies in two benchmarks: the non-collusive Nash equilibrium, $x^N = (v - \overline{v})/[(I+1)\lambda]$, and the perfect cartel benchmark, $x^M = (v - \overline{v})/(2I\lambda)$. Specifically, we discretize the interval $[x^M - \iota(x^N - x^M),\ x^N + \iota(x^N - x^M)]$ for $v > \overline{v}$ and $[x^N - \iota(x^M - x^N),\ x^M + \iota(x^M - x^N)]$ for $v < \overline{v}$ into n_x equally spaced grid points, denoted by $\mathbb{X} = \{x_1, \cdots, x_{n_x}\}$. The parameter $\iota > 0$ enables informed AI speculators to choose order flows that exceed the boundaries set by the theoretical benchmarks x^M and x^N, offering flexibility to explore strategies beyond these theoretical limits. The grid points of the market price p are determined similarly to those for x_i, with adjustments to account for the noise trader's impact on market prices. Specifically, the upper bound is set at $p_H = \overline{v} + \lambda^N \left(I \max\{x^M, x^N\} + 1.96\sigma_u\right)$ and the lower bound at $p_L = \overline{v} + \lambda^N \left(I \min\{x^M, x^N\} - 1.96\sigma_u\right)$, corresponding to the 5% and 95% percentiles of the noise trader's order flow distribution, $N(0, \sigma_u)$. The interval $[p_L - \iota(p_H - p_L),\ p_H + \iota(p_H - p_L)]$ is then discretized into n_p grid points, denoted by $\mathbb{P} = \{p_1, \cdots, p_{n_p}\}$.

Initial Q-Matrix and States. We initialize the Q-matrix at $t = 0$ with the discounted payoff that informed AI speculator i would earn if other informed AI speculators randomize their actions uniformly over the grid points in $\mathbb{X}$, and the noise trading flow is set to zero, which corresponds to the expected value of the distribution $N(0, \sigma_u^2)$.[27] Specifically, for each informed AI speculator $i = 1, \cdots, I$, we set its initial Q-matrix $\widehat{Q}_{i,0}$ at $t = 0$ as follows:

$$\widehat{Q}_{i,0}(s, x) = \frac{1}{(1-\rho)n_x} \sum_{x_{-i} \in \mathbb{X}} \left[v - (\overline{v} + \lambda^N (x + (I-1)x_{-i}))\right] x,$$

for $s = (p, v, v) \in \mathbb{P} \times \mathbb{V} \times \mathbb{V}$ and $x \in \mathbb{X}$. The initial states of our simulation, $s_0 = \{p_{-1}, v_{-1}, v_0\}$, are randomized uniformly over $\mathbb{P} \times \mathbb{V} \times \mathbb{V}$.

Specification of Exploration Rates. We consider the state-dependent ε-greedy scheme:

$$\varepsilon_{t(v)} = e^{-\beta t(v)}, \tag{4.3}$$

where $\beta > 0$ governs the speed that informed AI speculators' exploration rate diminishes over time and $t(v)$ captures the number of times that the system visited $v \in \mathbb{V}$ in the past.

Parameter Values. The parameters used in our numerical experiments are categorized into four groups based on their roles. First, "environment parameters" describe the underlying economic environment and, importantly, none of these values is known to the informed AI speculators and

[26] The results remain robust under alternative discretization schemes.

[27] Different initial values for the Q-matrix have minimal impact on the results. For example, assigning high initial values encourages Q-learning algorithms to explore all actions thoroughly in the early learning phase, as subsequent iterations gradually reduce these values toward their theoretical true levels. This approach accelerates the learning process and effectively facilitates thorough exploration early on and exploitation in later stages.

the market maker. In the baseline calibration, we set $I = 2$ and $\xi = 500$, and consider two different values for σ_u, which are $\sigma_u = 10^{-1}$ and $\sigma_u = 10^2$, representing trading environments with low and high noise trading risk, respectively. Later in the paper, we examine the implications of varying these environment parameters.

Second, "preference parameters" include the discount rate for informed AI speculators, ρ, and the weight assigned to the pricing error term by the market maker, θ. We set ρ at a relatively high level, $\rho = 0.95$, to reflect the high-frequency trading environment. We investigate the implications of varying ρ values in Section 6.2. We fix the value of θ at 0.1 as a universal constant throughout our simulation experiments.

Third, "discretization parameters" detail the methods used to discretize the system for simulation experiments. We set $n_v = 10$. Under this discretization, the standard deviation of v_t is $\widehat{\sigma}_v = \sqrt{n_v^{-1}\sum_{k=1}^{n_v}(v_k - \overline{v})^2} = 0.938$, which is close to the theoretical value $\sigma_v = 1$.[28] We set $\iota = 0.1$, $n_x = 15$, and $n_p = 31$.[29] We set $T_m = 10,000$ for the market maker. Increasing T_m does not alter any results.

Lastly, "hyperparameters" consist of α and β. Like in any machine learning algorithms, hyperparameters (or tuning parameters) are crucial for controlling the learning process of RL algorithms. In our baseline calibration, we set $\alpha = 0.01$ and $\beta = 5 \times 10^{-7}$. All results are robust to choosing different values of α and β so long as they are in the reasonable range that ensures sufficiently good learning outcomes. Our baseline choice of $\beta = 5 \times 10^{-7}$ implies that any action $x \in \mathbb{X}$ is, on average, visited just due to random exploration by $\frac{n_v}{n_x}\frac{1}{1-\exp(-5\times 10^{-7})} \approx 1,333,333$ times before exploration completes. In Sections 6.3 and 6.4, we conduct experiments with varying values of α and β. We also study scenarios where informed AI speculators adopt different values of α. In Online Appendix 3.12, we consider two-tier Q-learning algorithms that enables informed AI speculators to autonomously learn to choose optimal α as part of trading strategy.

Criterion for Numerical Convergence. Each experiment for a given numerical specification contains $N_{sim} = 1,000$ independent parallel simulation sessions. We adopt a stringent criterion for convergence, requiring that all informed AI speculators' optimal strategies remain unchanged for $1,000,000$ consecutive periods in a single simulation session. Additionally, all N_{sim} independent parallel simulation sessions must continue running until every session meets this convergence criterion. The number of periods required to reach convergence varies considerably across experiments, depending on parameter values. Even within the same experiment, the number of periods needed can differ significantly across the N_{sim} simulation sessions. Across all experiments we conducted, the range of periods needed to achieve convergence spans from approximately 20 million to about 50 billion.[30]

[28] In the remainder of this paper, the non-collusive Nash equilibrium and perfect cartel benchmark are computed using $\widehat{\sigma}_v$, to ensure consistency with the discretization scheme of v_t used in the simulation experiments.

[29] Our choice of $n_p \approx 2n_x$ ensures that, all else equal, a one-grid point change in one informed AI speculator's order will result in a change in price p_t over the grid defined by $\mathbb{P}$.

[30] Our programs are written in C++, using $-O2$ to optimize the compiling process. We use a high-powered computing server cluster with 400 CPU cores. Completing all simulation sessions in one experiment can take up to 6 hours.

5 AI Trading Equilibrium: Outcomes from Simulation Experiments

In this section, we present the results of simulation experiments that examine the behavior of AI-powered trading algorithms within a theoretical laboratory framework and explore the properties of AI trading equilibrium. The theoretical benchmarks discussed in Sections 3.3 and 3.4 demonstrate that AI-driven collusive trading equilibria can robustly arise from two distinct economic mechanisms. Crucially, these theories identify two key factors that determine which economic mechanism dominates: the risk of noise trading flows, captured by σ_u, and the presence of information-insensitive investors, governed by ξ. Building on these theoretical benchmarks, Section 5.1 provides an overview of simulation results across various cases defined by different levels of σ_u and ξ, and illustrates the exploration-exploitation tradeoff in reinforcement learning algorithms that underpins the two algorithmic mechanisms driving AI equilibria. Section 5.3 presents simulation experiments in trading environments with many information-insensitive investors (ξ is large relative to θ). In contrast, Section 5.4 focuses on simulation experiments in environments with few information-insensitive investors (ξ is small relative to θ). Section 5.5 further elaborates on the intuitions behind how AI collusion arises through two distinct algorithmic mechanisms corresponding to the two economic mechanisms. Finally, Section 5.6 provides a discussion on the role of information-insensitive investors.

5.1 Two Algorithmic Mechanisms and Exploration-Exploitation Tradeoff

Parallel to the two economic mechanisms underlying collusive equilibrium in trading, as defined in Definitions 3.2 and 3.3, our simulation experiments with Q-learning algorithms reveal two distinct algorithmic mechanisms through which informed AI speculators can autonomously learn to achieve a collusive trading equilibrium. The first mechanism is AI collusion via price-trigger strategies, approximating the collusive Nash equilibrium sustained by such strategies, as defined in Definition 3.2. The second is AI collusion driven by over-pruning bias in learning, which mirrors the collusive experience-based equilibrium arising from a learning bias caused by over-perceived aversion to noise trading risk, as defined in Definition 3.3. Therefore, the price-trigger AI collusive equilibrium matches the theoretical collusive Nash equilibrium sustained by price-trigger strategies, while the over-pruning AI collusive equilibrium corresponds to the theoretical collusive experience-based equilibrium driven by learning bias.

The effectiveness of exploration-exploitation tradeoff in reinforcement learning determines which algorithmic mechanism prevails and, consequently, the type of AI equilibrium that emerges. This tradeoff, like the bias-variance tradeoff in supervised learning and high dimensional statistics, aims to strike a balance between pruning the choice space and reducing outcome variability. In reinforcement learning, exploration (i.e., trying new actions) is essential to minimize bias in estimating the optimal action, while exploitation (i.e., selecting the optimal actions based on past experience) reduces noise in observed rewards, thereby lowering variability in the estimation of the optimal action. Similar to shrinkage techniques in supervised learning and high-dimensional statistics, exploitation in reinforcement learning constrains the choice space to reduce variability

and improve learning efficiency, though it may introduce some bias.

Drawing from the key ideas behind the theoretical existence results of collusive equilibria sustained by two distinct economic mechanisms, as summarized in Propositions 3.1 and 3.2, the type of AI equilibrium to which the system of reinforcement learning algorithms converges depends on how the informativeness of market prices and the effectiveness of the exploration-exploitation tradeoff are shaped by two critical factors: the risk of noise trading flows, represented by σ_u, and the presence of information insensitive investors, determined by ξ.

The algorithmic mechanism of AI collusion via price-trigger strategies emerges as the dominant steady state when the exploration-exploitation tradeoff effectively guides the estimation of optimal trading strategies. In this setting, a system of algorithms autonomously learns to sustain a collusive AI equilibrium that approximates a Nash equilibrium, even as each algorithm unilaterally maximizes its own trading profit. Crucially, each algorithm not only learns how the state variable (i.e., the "environment") responds to its trading behavior in effect but also integrates this knowledge into its profit optimization process. This dynamic sophistication allows the algorithms to converge to a steady-state equilibrium that extends beyond the non-collusive Nash equilibrium. For this exploration-exploitation tradeoff to function effectively, price informativeness must be sufficiently high, which in turn requires a low σ_u and a high ξ. Intuitively, when price informativeness is high, information obtained from occasional explorations is more reliable, allowing exploitation to focus on optimal trading strategies. Further intuition is provided in Section 5.5.

The algorithmic mechanism of AI collusion through over-pruning bias in learning emerges as the dominant steady state when the exploration-exploitation tradeoff fails to effectively estimate optimal trading strategies. In this case, the system of algorithms does not converge to a collusive AI equilibrium that approximates a Nash equilibrium. Instead, an imbalance between exploration and exploitation causes the systematic over-pruning of aggressive trading strategies, resulting in a collusive AI equilibrium driven by over-pruning bias. This outcome closely parallels the theoretical collusive experience-based equilibrium, which arises from a learning bias induced by over-perceived aversion to noise trading risk. The exploration-exploitation tradeoff fails to effectively guide estimation when price informativeness is not sufficiently high, which can result from a high σ_u or a low ξ. Importantly, as long as ξ is low, price informativeness remains endogenously low, regardless of the level of σ_u. Intuitively, when price informativeness is low, information obtained from occasional explorations can be misleading, causing exploitation to become trapped in suboptimal trading strategies. Further intuition is provided in Section 5.5.

More precisely, the impact of the exploration-exploitation tradeoff on learning bias is determined by the extent to which it over-prunes aggressive trading strategies. To illustrate how over-pruning bias arises from the imbalance between exploration and exploitation, consider environments with low ξ or high σ_u. In these cases, as ξ approaches zero or σ_u becomes excessively large, market prices and trading profits are primarily driven by noise trading flow shocks, u_t. Consequently, the behavior of reinforcement learning algorithms — particularly their exploitation and exploration — depends critically on how they respond to these noise trading flow shocks, u_t. Crucially, exploitation asymmetrically impacts the learning process depending on whether a noise trading flow shock is adverse

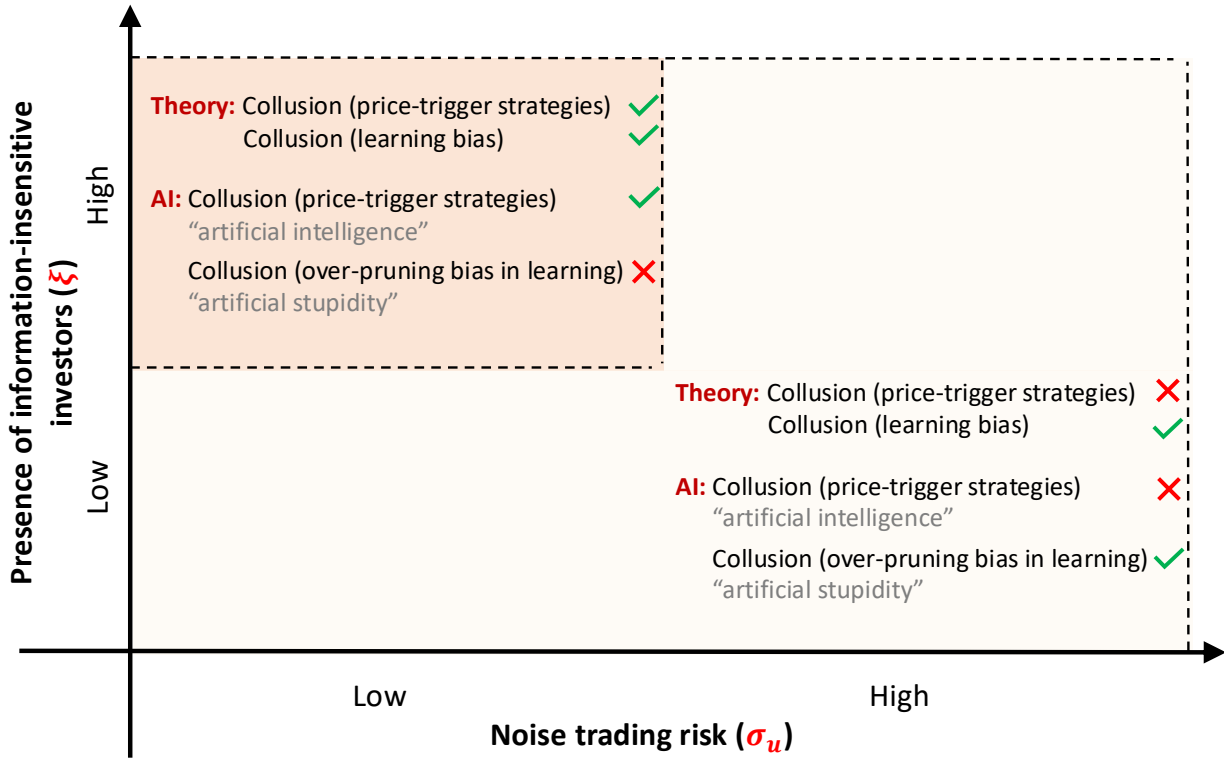
or beneficial. An adverse noise flow moves in the same direction as the informed AI speculator's trading order, potentially causing substantial trading losses. In contrast, a beneficial noise flow moves in the opposite direction, potentially generating significant trading profits. Following an adverse noise trading flow shock, the algorithm classifies the chosen strategy as a "disastrous action" and assigns it a significantly low estimated Q-value. Exploitation discourages the algorithm from revisiting this strategy, reinforcing the downward bias in its evaluation and preventing correction for such off-equilibrium actions. Conversely, after a beneficial noise trading flow shock, the algorithm labels the chosen strategy as a "favorable action" and assigns it a very high estimated Q-value. Exploitation ensures the algorithm continues to adopt this strategy, refining its evaluation and fully correcting any upward-biased assessments for on-equilibrium-path actions. As a result, since aggressive strategies are more vulnerable to adverse noise trading flow shocks, this asymmetry often results in their persistent undervaluation. This, in turn, causes them to be prematurely pruned from the set of potential optimal strategies, reinforcing over-pruning bias in learning. Consequently, informed AI speculators employing Q-learning algorithms gravitate toward conservative trading strategies, consistent with the collusive behavior described in Definitions 3.1 and 3.3.

One way to interpret the asymmetric effect of exploitation is that it effectively makes reinforcement learning algorithms risk-averse to randomness in their rewards. In decision theory, risk aversion arises from the asymmetric impact of adverse and beneficial shocks. Similarly, in reinforcement learning, exploitation discourages revisiting poorly rated strategies while reinforcing successful ones, leading to an asymmetric impact of adverse and beneficial shocks on the learning process. This asymmetry, in turn, causes aggressive trading strategies — more vulnerable to adverse noise trading flow shocks — to be prematurely pruned from the set of potential optimal strategies, reinforcing over-pruning bias in learning. As a result, this bias, along with the aversion to aggressive strategies, reflects the effective risk aversion inherent in these algorithms.

5.2 Key Findings on AI Collusion

We begin with an overview of the key simulation findings, summarized in Figure 1, before digging into the details of our simulation experiments in Sections 5.3 and 5.4, followed by a discussion of the underlying intuitions behind the AI collusive equilibrium in Section 5.5 and heuristic explanations in Online Appendix 2. To comprehensively characterize the AI collusive equilibrium, we classify all possible trading environments into three cases: (i) high ξ and low σ_u, (ii) high ξ and high σ_u, and (iii) low ξ. The corresponding theoretical benchmarks and key simulation findings are summarized as follows:

(i) **High ξ & low σ_u:** Both a collusive Nash equilibrium through price-trigger strategies and a collusive experience-based equilibrium through learning bias can theoretically be achieved by informed speculators in such environments, as established in Propositions 3.1 and 3.2. However, in our simulation experiments, informed AI speculators using Q-learning consistently converge to an AI collusive equilibrium sustained by price-trigger strategies, rather than one driven by over-pruning bias.



Note: The symbol "✓" indicates that the equilibrium exists, while "×" indicates that it does not. The presence of information-insensitive investors, ξ, is the slope coefficient of the asset demand curve, as specified in (3.2), while the noise trading risk, σ_u, denotes the standard deviation of the noise trading flow, u_t.

Figure 1: Summary of our main findings.

(ii) **High ξ & high σ_u:** No collusive Nash equilibrium sustained by price-trigger strategies exists in theory, whereas a collusive experience-based equilibrium driven by learning bias can theoretically be achieved by informed speculators in such environments, as established in Propositions 3.1 and 3.2. Consistent with these theoretical benchmarks, simulation experiments show that multiple informed AI speculators using Q-learning algorithms converge solely to an AI collusive equilibrium driven by over-pruning bias in learning, rather than one sustained by price-trigger strategies.

(iii) **Low ξ:** No collusive Nash equilibrium sustained by price-trigger strategies exists in theory, whereas a collusive experience-based equilibrium driven by learning bias can still theoretically be achieved by informed speculators in such environments, regardless of the level of $\sigma_u > 0$, as established in Propositions 3.1 and 3.2. Consistent with these theoretical benchmarks, simulation experiments demonstrate that multiple informed AI speculators using Q-learning algorithms converge solely to an AI collusive equilibrium driven by over-pruning bias in learning, rather than one sustained by price-trigger strategies. Notably, the results in this case are the same as those in case (ii), characterized by high ξ and high σ_u.

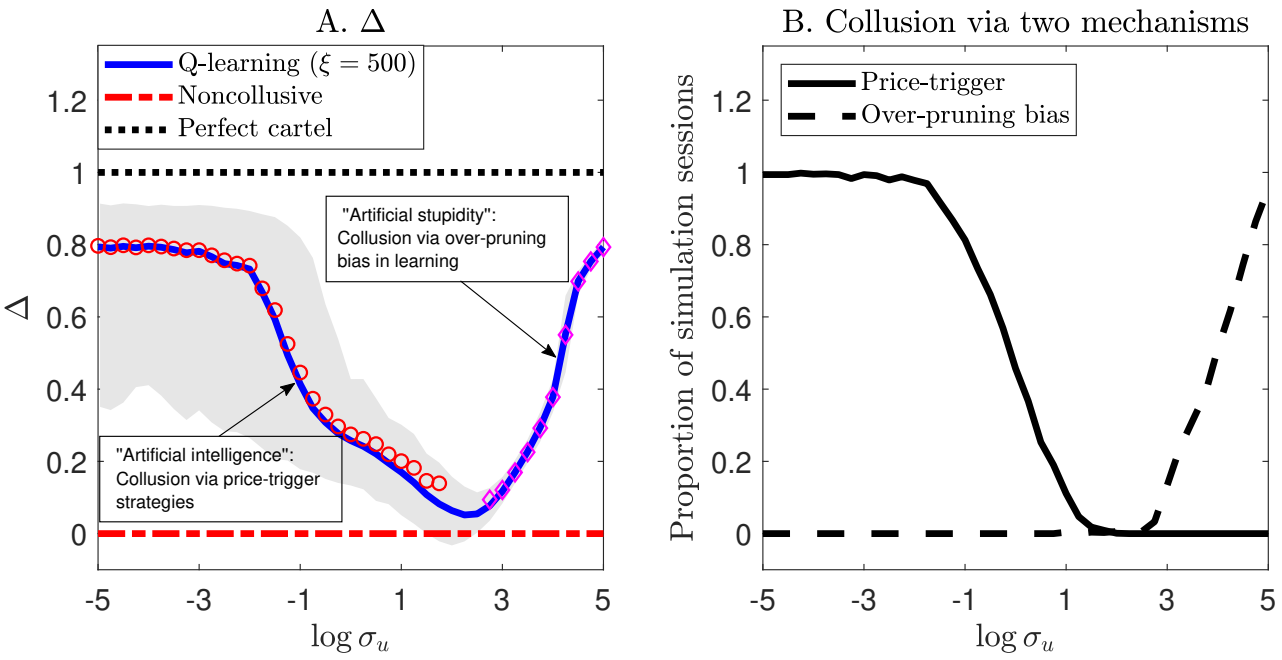


Figure 2: Two distinct mechanisms behind AI collusion.

5.3 Simulation Experiments in Trading Environments with High ξ

This section presents simulation results for cases (i) and (ii) described in Section 5.2. In trading environments where ξ is large relative to θ, indicating a significant presence of information-insensitive investors, the market maker primarily sets the market price to minimize inventory costs, rather than to reduce pricing errors, as described in (3.4).

U-Shaped Profitability in AI Collusion: Two Distinct Mechanisms. Panel A of Figure 2 plots the average Δ^C as $\log \sigma_u$ varies from -5 to 5 along the x-axis. The horizontal dotted line represents the theoretical benchmark for a perfect cartel ($\Delta^M \equiv 1$), while the horizontal dash-dotted line indicates the benchmark for a non-collusive Nash equilibrium ($\Delta^N \equiv 0$). The solid U-shaped line between 0 and 1 represents the average normalized trading profitability of informed AI speculators, that is, the average value of Δ^C across all $N_{sim} = 1,000$ simulation sessions. The average value of Δ^C reflects the collusion capacity of the informed AI speculators. The grey area around the solid line represents the range of Δ^C from the 1st to the 99th percentile across all N_{sim} simulation sessions.[31]

The normalized profitability of AI trading, Δ^C, lies between 0 and 1, suggesting that a collusive equilibrium with significant supra-competitive profits, as defined in Definition 3.1, emerges robustly, irrespective of the noise trading risk level, σ_u. Importantly, the normalized trading profitability, Δ^C, and the noise trading risk, σ_u, exhibit a strong U-shaped relationship, indicating that AI-driven collusive trading is particularly pronounced, with AI collusion capacity especially high, when σ_u is either high or low. However, the algorithmic mechanisms underlying these AI collusion patterns

[31]The U-shaped pattern in the normalized trading profitability of informed AI speculators remains highly robust across different levels of ξ, as demonstrated in Figure IA.4 in Online Appendix 3.6.

differ significantly between the high and low σ_u scenarios, as discussed in Section 5.1 and further detailed in Section 5.5. This distinction is evident from the opposing relationships between σ_u and Δ^C in these two scenarios. When noise trading risk σ_u is low, collusion capacity, as reflected in Δ^C, decreases as σ_u increases. In contrast, when noise trading risk σ_u is high, collusion capacity, as reflected by Δ^C, increases with σ_u.

Panel B of Figure 2 shows the proportion of the N_{sim} parallel simulation sessions that converge to a specific type of AI collusive equilibrium. Collusive equilibria sustained by price-trigger strategies are represented by the solid line, while those sustained by over-pruning bias in learning are represented by the dashed line. In each simulation session, the type of AI collusion is identified based on the defining features of price-trigger AI collusion and over-pruning AI collusion, as determined by the impulse response patterns described in Figure 3.[32] The results show that when σ_u is low, nearly all simulation sessions converge to an AI collusive equilibrium sustained by price-trigger strategies, with almost none converging to an equilibrium sustained by over-pruning bias in learning. As σ_u increases, the proportion of sessions converging to price-trigger strategies decreases, while the proportion converging to over-pruning bias in learning rises. At high levels of σ_u, nearly all sessions converge to an AI collusive equilibrium sustained by over-pruning bias in learning, with almost none converging to an AI collusive equilibrium sustained by price-trigger strategies.

The simulation results illustrated in Panel B are consistent with the theoretical benchmarks established in Propositions 3.1 and 3.2. Theoretically, when ξ is large and σ_u is small, both a collusive Nash equilibrium sustained by price-trigger strategies and a collusive experience-based equilibrium driven by over-perceived aversion to noise trading risk can exist. However, Proposition 3.4 reveals that in low noise trading risk environments (i.e., low σ_u), the collusion capacity of informed speculators, as measured by their normalized trading profitability Δ^C, is typically high in a price-trigger Nash equilibrium but low in an experience-based equilibrium due to the over-perceived aversion to noise trading risk. Consequently, informed AI speculators in such environments autonomously learn to achieve an AI collusive equilibrium sustained by price-trigger strategies rather than one driven by over-pruning bias in learning, as explained in Section 5.1, with further intuitions detailed in Section 5.5. This phenomenon is referred to as "collusion through artificial intelligence." In contrast, as shown by Propositions 3.1 and 3.2, when ξ is large and σ_u is large, only a collusive experience-based equilibrium driven by over-perceived aversion to noise trading risk can be sustained, while a collusive Nash equilibrium sustained by price-trigger strategies becomes theoretically infeasible. Consequently, informed AI speculators in such environments autonomously learn to achieve an AI collusive equilibrium driven by over-pruning bias in learning rather than one sustained by price-trigger strategies, as explained in Section 5.1 and further detailed in Section 5.5. This phenomenon is referred to as "collusion through artificial stupidity."

The U-shaped relationship between Δ^C and σ_u becomes clear when analyzing Panels A and B of Figure 2 together. In Panel A, the circles (○) represent the average Δ^C conditioned on simulation sessions classified as price-trigger AI collusive equilibria, while the diamonds (◇) represent the

[32] More details are provided in Online Appendix 3.5.

average Δ^C conditioned on simulation sessions classified as over-pruning AI collusive equilibria. When noise trading risk is low (i.e., $\log \sigma_u \leq 1$), informed AI speculators sustain collusion mainly through price-trigger strategies, achieving significant supra-competitive profits. As σ_u increases, the collusion capacity, reflected in normalized trading profitability Δ^C, decreases. This decline occurs because higher noise trading risk reduces the informativeness of market prices, making it increasingly challenging to sustain collusive trading through price-trigger strategies. These findings align with the theoretical benchmark established in Proposition 3.4.

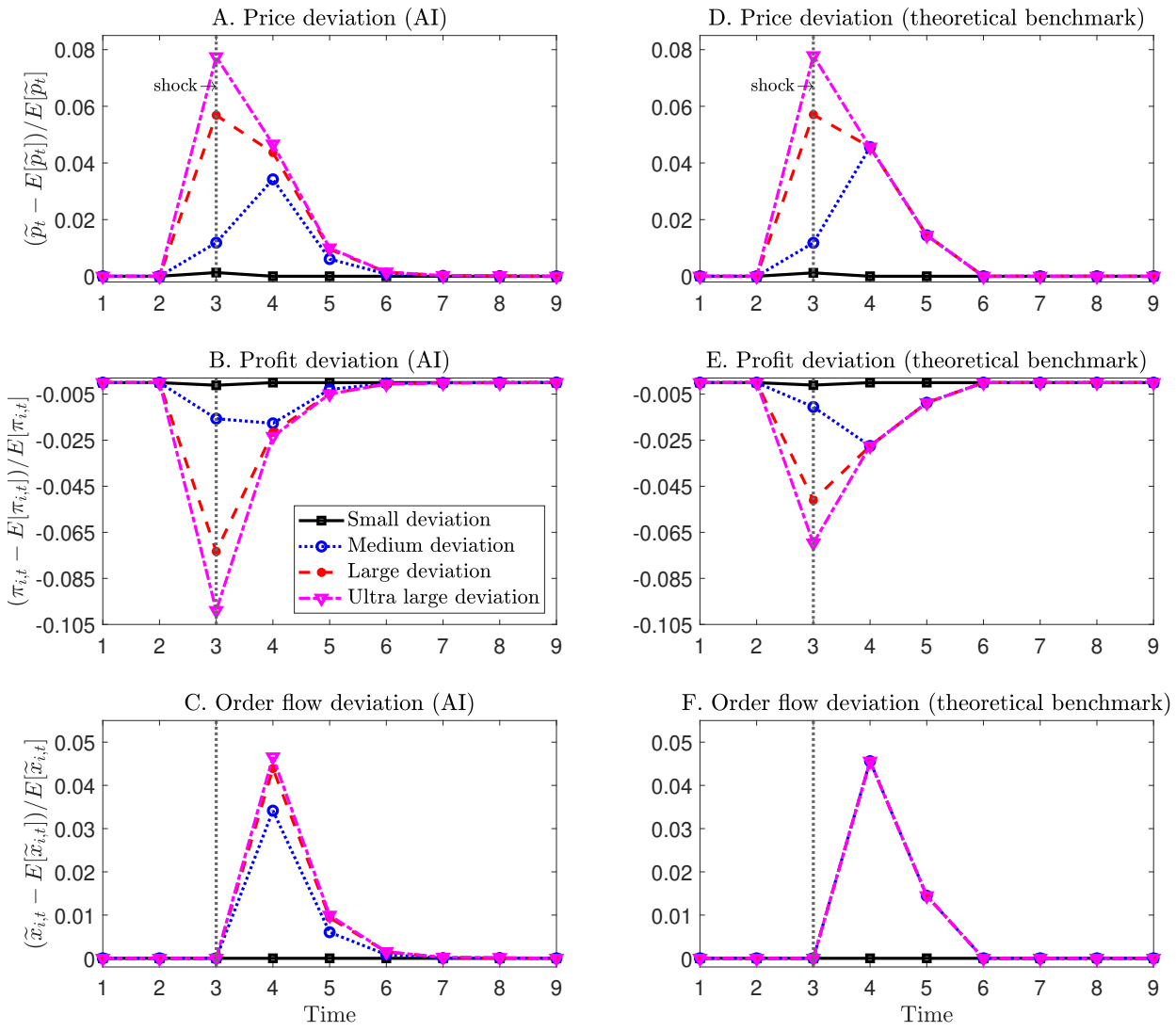
In contrast, when noise trading risk is high (i.e., $\log \sigma_u \geq 3$), informed AI speculators sustain collusion mainly through over-pruning bias in learning, also achieving substantial supra-competitive profits. As σ_u increases, the collusion capacity, reflected in normalized trading profitability Δ^C, also increases. This occurs because higher noise trading risk disrupts the balance between exploration and exploitation by amplifying the asymmetric effects of exploitation on the learning of aggressive trading strategies in response to beneficial and adverse noise trading shocks. Specifically, it exacerbates this asymmetry to the point where these effects become increasingly difficult to correct through exploration updates. As a result, higher noise trading risk reinforces over-pruning bias, making aggressive trading strategies even less viable. As highlighted in Section 5.1, the asymmetric effect of exploitation can be interpreted as risk aversion embedded in algorithms toward randomness in rewards. Intuitively, greater noise trading risk further discourages algorithms from selecting aggressive trading strategies. These simulation findings are consistent with the theoretical benchmark established in Proposition 3.4.

To further provide direct evidence of the two AI collusion mechanisms across environments with low and high noise trading risk, as demonstrated in Figure 2, we conduct impulse response analyses throughout the remainder of this section using our simulation experiments. We begin by showing that in low noise trading risk scenarios, informed AI speculators autonomously learn to sustain collusive, supra-competitive trading profits through price-trigger strategies, without requiring any form of agreement, communication, or intent. To be more precise, we emphasize that, while this AI collusive equilibrium resembles the collusive Nash equilibrium sustained by price-trigger strategies, as described in Definition 3.2 and Proposition 3.1, it does not fully satisfy the requirements of subgame perfect Nash equilibrium.[33] Conversely, in high noise trading risk scenarios, we then show that informed AI speculators still sustain collusive, supra-competitive trading profits, but through a different mechanism: over-pruning bias in learning.[34] This AI collusive equilibrium corresponds to the collusive experience-based equilibrium sustained by over-perceived aversion to noise trading risk, as described in Definition 3.3 and Proposition 3.2.

Impulse Responses: AI Collusion via Price-Trigger Strategies When σ_u Is Low. We first examine how the trained informed AI speculators respond to an exogenous shock in the noise order flow,

[33] Our numerical tests suggest that this AI collusion equilibrium is approximately Nash, meaning no local deviation is preferred. Numerical tests are detailed in Online Appendix 3.10.

[34] In both scenarios, the equilibrium is classified as an experience-based equilibrium, based on the formal tests proposed by Fershtman and Pakes (2012). Details of these tests are provided in Online Appendix 3.2. This is unsurprising, as the experience-based equilibrium framework is broader and encompasses subgame perfect Nash equilibrium as a special case.



Note: All plots correspond to a trading environment with $\xi = 500$, indicating a significant presence of information-insensitive investors, and $\sigma_u = 10^{-1}$, representing a low noise trading risk level. Panels A to C display the IRFs following a uniform exogenous shock u_{shock} in simulation experiments using Q-learning algorithms. Panels D to F illustrate the corresponding IRFs based on the theoretical benchmark of a collusive Nash equilibrium sustained by price-trigger strategies. Panels A and D depict the percentage deviation of the asset's price from its long-run mean, expressed as $(\widetilde{p}_t - \mathbb{E}[\widetilde{p}_t])/\mathbb{E}[\widetilde{p}_t]$, where $\widetilde{p}_t = (p_t - \overline{v}) \times \text{sgn}(v_t - \overline{v})$, and $\text{sgn}(\cdot)$ is the sign function ensuring $\widetilde{p}_t > 0$. At $t = 3$, the exogenous shock causes the asset's price to deviate from its long-run mean, with the deviation size increasing with the magnitude of the shock u_{shock}. Panels B and E depict the percentage deviation of average profits from their long-run mean for each informed AI speculator, expressed as $(\pi_{i,t} - \mathbb{E}[\pi_{i,t}])/\mathbb{E}[\pi_{i,t}]$. At $t = 3$, the price deviation reduces the profits of informed AI speculators, with the impact increasing with the magnitude of the percentage price deviation shown in Panels A and D. Panels C and F depict the percentage deviation of order flows from the long-run mean for each informed AI speculator, defined as $(\widetilde{x}_{i,t} - \mathbb{E}[\widetilde{x}_{i,t}])/\mathbb{E}[\widetilde{x}_{i,t}]$, where $\widetilde{x}_{i,t} = x_{i,t} \times \text{sgn}(v_t - \overline{v})$. The sign function ensures that $\widetilde{x}_{i,t} > 0$. The deviation of order flows is zero at $t = 3$ because price deviation only occurs until $t = 3$.

Figure 3: Impulse response function (IRF) following an exogenous noise trading flow shock u_{shock} for $\sigma_u = 10^{-1}$ under Q-learning (left column) and the theoretical benchmark (right column).

which influences the asset's market price through the market maker's endogenous and adaptive pricing rule. At $t = 0$, all N_{sim} simulation sessions have converged. Simultaneously, the market price of the asset, p_t, is determined by the market maker's adaptive pricing rule, which responds

to the random variables v_t and u_t along each simulation path, independently across the parallel simulation paths. At $t = 3$, an unexpected exogenous shock, u_{shock}, is introduced to the noise order flow u_t, simultaneously and uniformly affecting all N_{sim} simulation sessions. This shock is designed to adversely impact the trading profits of informed AI speculators, with $u_{\text{shock}} > 0$ when $v_t > \overline{v}$ and $u_{\text{shock}} < 0$ when $v_t < \overline{v}$. As a result, the market price p_t rises unexpectedly if $v_t > \overline{v}$ and falls unexpectedly if $v_t < \overline{v}$, with the magnitude of the price change determined by the size of u_{shock}. Each impulse-response curve in a panel represents the average impulse response dynamics across N_{sim} independent simulation sessions.[35] The cross-sectional distribution of path-by-path impulse response dynamics across N_{sim} simulation sessions is provided in Online Appendix 3.4.

Panel A of Figure 2 shows that, in environments with low noise trading risk, specifically when $\sigma_u = 10^{-1}$, the average value of Δ^C across N_{sim} parallel simulation paths is approximately 0.75. Under these conditions, informed AI speculators achieve average trading profits that are about 10% higher than those in the non-collusive equilibrium benchmark.

To examine how informed AI speculators behave in steady-state equilibrium, we analyze their impulse responses to exogenous shocks of varying magnitudes. In the "small deviation" experiment, $|u_{\text{shock}}|$ is approximately 0.25% of the average magnitude of informed AI speculators' order flow $|x_{i,t}|$, resulting in a minor impact on the asset price p_t at $t = 3$. In contrast, in the "medium deviation," "large deviation," and "ultra large deviation" experiments, $|u_{\text{shock}}|$ corresponds to roughly 2.5%, 11.5%, and 15.0% of the average $|x_{i,t}|$, respectively, leading to progressively larger changes in p_t.

To provide direct evidence that the behavior of informed AI speculators in equilibrium aligns closely with a theoretical collusive Nash equilibrium sustained by price-trigger strategies, we present the impulse responses to the exogenous shocks mentioned above for AI-powered trading in the left column of Figure 3, alongside the corresponding theoretical benchmarks in the right column. For a meaningful comparison, Panels D through F use the same magnitudes of unexpected price deviations at $t = 3$ as those in the simulation experiments shown in Panels A to C. Additionally, all shared parameters between the theoretical benchmarks and the simulation experiments are set to identical values. The parameters (T, ω, η), which specify the price-trigger punishment scheme in theory, are not relevant to the structure of the Q-learning simulations. Here, we set $T = 2$ to align with the two-period punishment observed in the Q-learning experiments, $\omega = 2.826$ to achieve an average profitability Δ^C of approximately 0.75, and $\eta = 0.327$ to match the average order flow deviation in the "ultra large deviation" case at $t = 4$ in the Q-learning simulations. This side-by-side comparison highlights the strong resemblance between the AI collusive equilibrium and the corresponding theoretical benchmarks of collusive Nash equilibrium sustained by price-trigger strategies.

The price-trigger punishment scheme is evident throughout Panels A to C. Specifically, immediately after the shock at $t = 3$ (starting at $t = 4$), the responses display two defining characteristics of price-trigger strategies, as outlined in Definition 3.2 and Proposition 3.1. These features of trigger-type strategies, also reflected in the theoretical benchmark shown in Panels D to F, are as

[35] Each of the N_{sim} simulation sessions averages 10,000 simulation paths to smooth out the randomness of v_t and u_t, ensuring a reasonable comparison with the impulse response analysis based on the deterministic model of Calvano et al. (2020), which has no information asymmetry or stochastic economic environment.

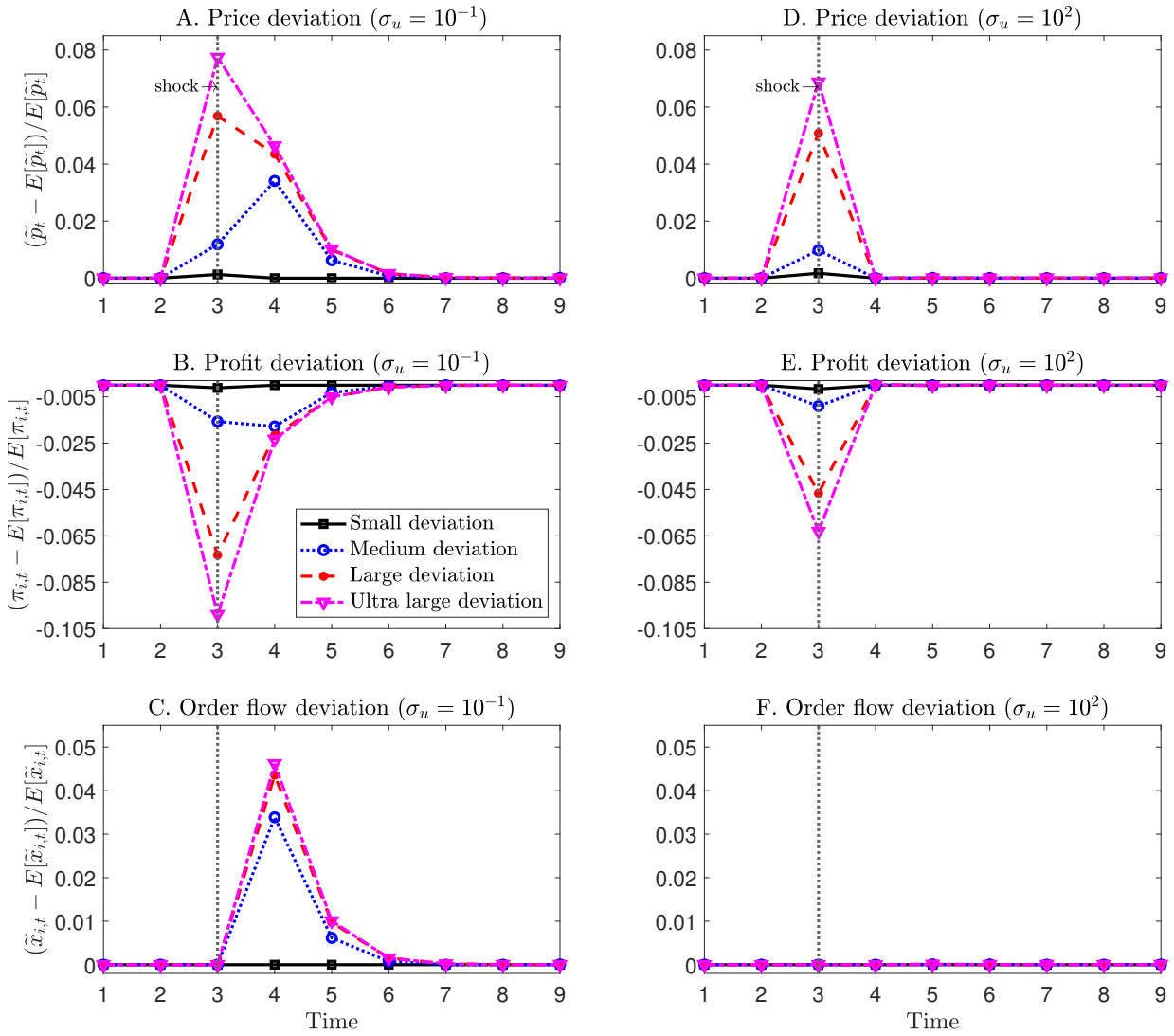
follows: (i) there is, on average, no response when the price deviation at $t = 3$ is small (i.e., the "small deviation" scenario, represented by the solid curve), and (ii) when the price deviation at $t = 3$ is sufficiently large, AI speculators employ similarly aggressive trading strategies, despite significant differences in the deviation's magnitude at $t = 3$ (i.e., the "medium deviation," "large deviation," and "ultra large deviation" cases, represented by the dotted, red dashed, and purple dash-dotted curves, respectively).

To further validate the price-trigger punishment scheme among informed AI speculators, Panel A shows that for large and ultra-large price deviations, the percentage deviation of the asset's price at $t = 4$ decreases relative to $t = 3$ but remains above the long-run mean. In the medium deviation case, the percentage deviation at $t = 4$ surpasses that at $t = 3$. Notably, in the medium, large, and ultra-large cases, price deviations at $t = 4$ converge to similar magnitudes, driven by comparable order flow deviations at $t = 4$, as shown in Panel C. In contrast, for the small deviation case, both the asset price and informed AI speculators' profits revert to the long-run mean at $t = 4$. These nuanced patterns of the AI collusion equilibrium closely align with those of the collusive Nash equilibrium sustained by price trigger strategies, as depicted in Panels D through F.

We emphasize that, although the Q-learning algorithms rely only on the one-period lagged market price p_{t-1} and fundamental value v_{t-1} for their decisions at period t, the punishment can extend beyond a single period. Panels A through C of Figure 3 illustrate that informed AI speculators continue to enforce punishment at $t = 5$, albeit significantly weaker on average than at $t = 4$. This pattern demonstrates that informed AI speculators sustain the collusive equilibrium using price-trigger strategies, where the punishment scheme generally lasts for more than one period.

To confirm that the price-trigger strategy employed by informed AI speculators in Panels A through C of Figure 3 is indeed the driving force behind the collusive, supra-competitive trading profitability observed in Figure 2 under low noise trading risk, we disable the AI speculators' ability to use lagged market prices as a monitoring tool. This is accomplished by removing the lagged market price p_{t-1} from the state variable s_t used for decision-making at period t. Our findings reveal that even in environments with both a significant presence of information-insensitive investors (i.e., a high ξ) and low noise trading risk (i.e., a low σ_u), the price-trigger punishment scheme cannot be learned, and the collusion capacity, measured by Δ^C, drops to zero.

Impulse Responses: No AI Collusion via Price-Trigger Strategies When σ_u Is High. Next, we demonstrate that the collusive, supra-competitive trading profitability observed under high noise trading risk (i.e., high σ_u) in Figure 2 is not driven by price trigger strategies, in contrast to the low noise trading risk (i.e., low σ_u) scenario. The setup of simulation experiments in Figure 4 is the same as that in Figure 3 with Panels A through C replicated exactly for a straightforward comparison. In Panels D through F of Figure 4, we investigate the average IRF over the N_{sim} simulation paths in the environment with high noise trading risk (i.e., $\sigma_u = 10^2$). This side-by-side comparison, particularly contrasting Panel C with Panel F of Figure 4, reveals that informed AI speculators do not respond at all to the exogenous shock to noise trading flow (u_{shock}) when σ_u is high, let alone respond according

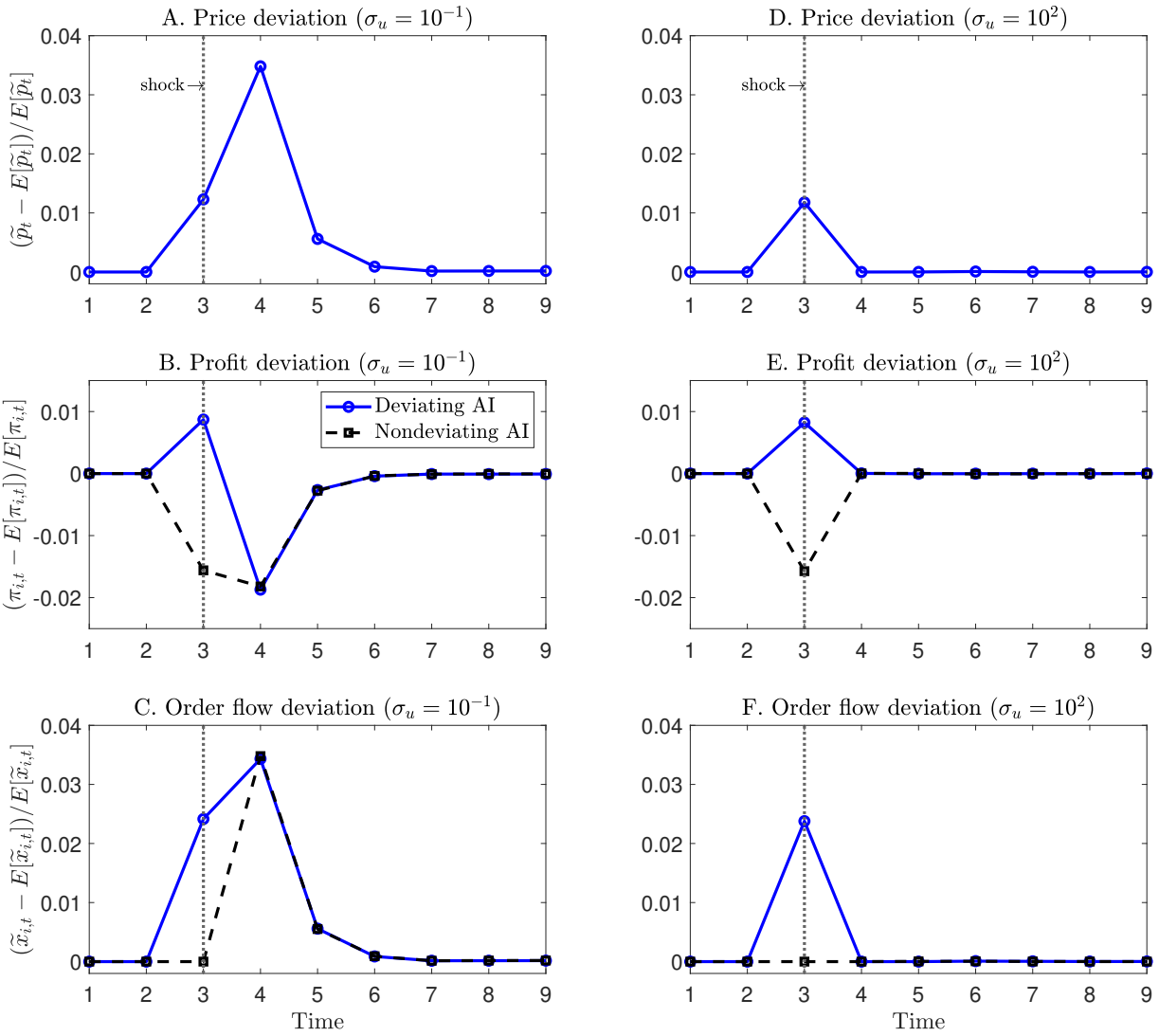


Note: All the plots are based on simulation experiments using Q-learning algorithms in a trading environment with $\xi = 500$, indicating a significant presence of information-insensitive investors. Panels A through C display the IRFs following a uniform exogenous shock u_{shock} in a low noise trading risk scenario ($\sigma_u = 10^{-1}$). Panels D through F show the corresponding IRFs for a high noise trading risk scenario ($\sigma_u = 10^{2}$).

Figure 4: Impulse response function (IRF) following an exogenous noise trading flow shock u_{shock} for two scenarios $\sigma_u = 10^{-1}$ (left column) and $\sigma_u = 10^{2}$ (right column) under Q-learning.

to price-trigger strategies. This finding is consistent with the theoretical result of Proposition 3.1, which states that a collusive Nash equilibrium sustained through price-trigger strategies does not exist in an environment with high noise trading risk.

Impulse Responses: AI Collusion via Over-Pruning Bias When σ_u Is High. Lastly, we investigate how informed AI speculators achieve and sustain supra-competitive profits despite being unable to learn and employ price-trigger strategies under high noise trading risk (i.e., high σ_u). Our analysis demonstrates that informed AI speculators can establish an AI collusive equilibrium through over-pruning bias in learning. This behavior corresponds to the theoretical collusive experience-



Note: All the plots are based on simulation experiments using Q-learning algorithms in a trading environment with $\xi = 500$, indicating a significant presence of information-insensitive investors. Panels A through C display the IRFs following a uniform exogenous order flow deviation $x_{i,\text{shock}}$ in a low noise trading risk scenario ($\sigma_u = 10^{-1}$). Panels D through F show the corresponding IRFs for a high noise trading risk scenario ($\sigma_u = 10^{2}$).

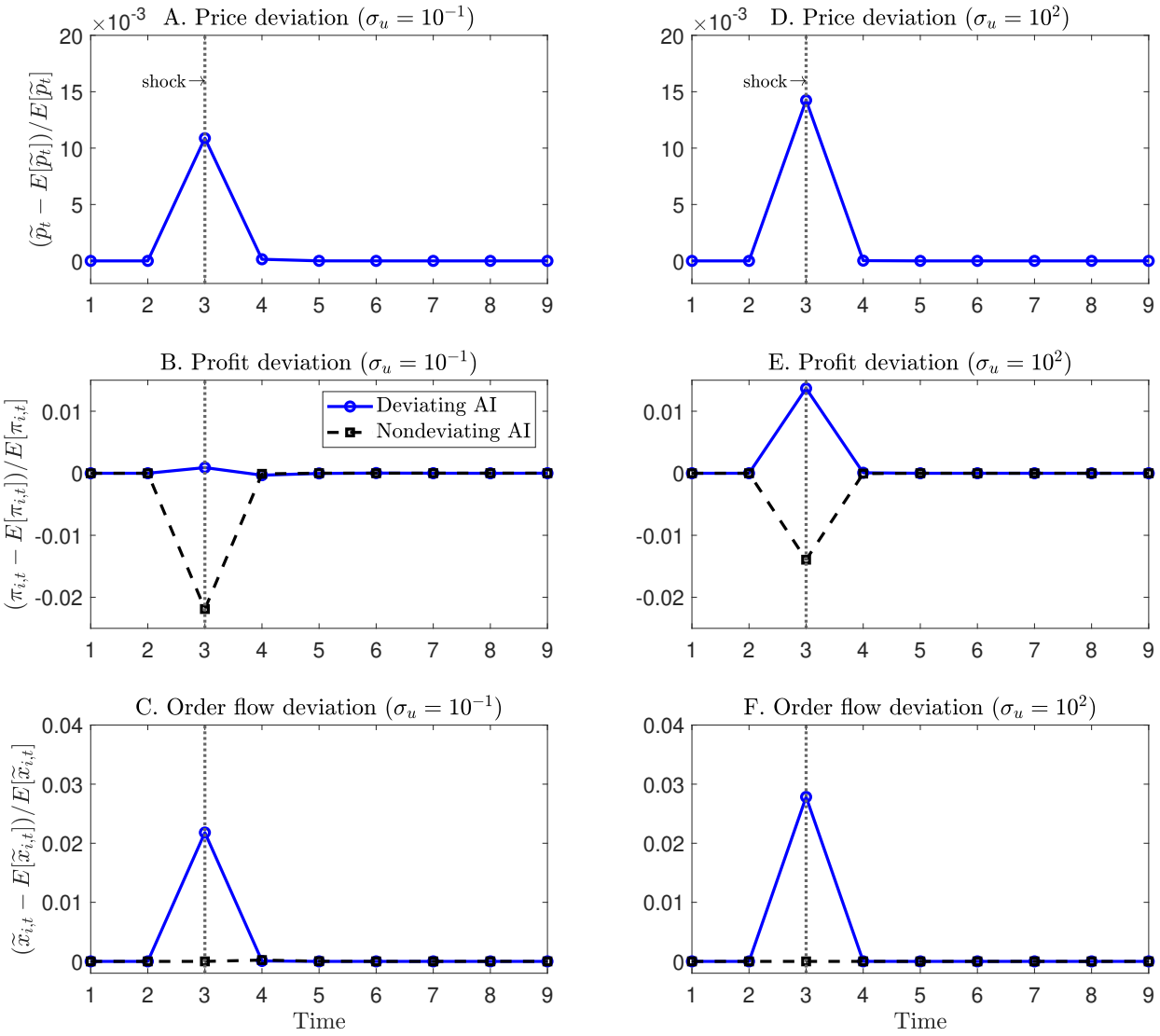
Figure 5: Impulse response function (IRF) following a unilateral deviation in trading order flows $x_{i,\text{shock}}$, shown for $\sigma_u = 10^{-1}$ (left column) and $\sigma_u = 10^{2}$ (right column) under Q-learning.

based equilibrium, sustained by an over-perceived aversion to noise trading risk, as described in Definition 3.3 and Proposition 3.2. To illustrate this, we examine the IRF resulting from a unilateral trading deviation by one informed AI speculator in environments with both low noise trading risk ($\sigma_u = 10^{-1}$) and high noise trading risk ($\sigma_u = 10^{2}$), as depicted in Figure 5. Specifically, we exogenously force a single informed AI speculator, labeled as i, to deviate from its learned optimal strategy for one period at $t = 3$, uniformly across all N_{sim} simulation paths. This one-period deviation at $t = 3$ is designed to increase the contemporaneous trading profit of the deviating speculator. Concretely, we exogenously increase the order flow of the deviating speculator by $x_{i,\text{shock}}$ if $v_t > \overline{v}$ and reduce its order flow by $x_{i,\text{shock}}$ if $v_t < \overline{v}$. Panels A through C of Figure 5 depict the

IRF following the unilateral deviation of AI speculator i (solid line) at $t = 3$ under the scenario of low noise trading risk ($\sigma_u = 10^{-1}$). Panel C specifically illustrates the exogenous deviation that forces AI speculator i (solid line) to trade more aggressively at $t = 3$, while the other AI speculator (dashed line) maintains its original trading behavior at $t = 3$. As shown in Panel A, the aggressive trading by AI speculator i causes the market price p_t to rise at $t = 3$. Panel B illustrates that the deviating AI speculator (solid line) achieves higher profits, while the non-deviating AI speculator (dashed line) incurs losses at $t = 3$. According to Definition 3.1, these IRF results corroborate the findings in Figure 2, demonstrating that informed AI speculators can interact and learn to sustain an AI collusive equilibrium in low noise trading risk environments. More importantly, the responses of informed AI speculators to this unilateral deviation in subsequent periods, starting from $t = 4$, further reinforce the findings of Figures 3 and 4, confirming that the AI collusive equilibrium is sustained by price-trigger strategies, closely resembling the behavior of a collusive Nash equilibrium through price-trigger strategies. Specifically, at $t = 4$, Panel C shows that both AI speculators, on average, engage in equally aggressive trading as a form of punishment for the deviation that occurs at $t = 3$. As shown in Panel B, this behavior results in losses for both AI speculators at $t = 4$ due to the sharp increase in the market price.

In a parallel comparison to the simulation experiments under the low noise trading risk scenario ($\sigma_u = 10^{-1}$), Panels D through F of Figure 5 show the IRF for the same experiment, conducted under the high noise trading risk scenario ($\sigma_u = 10^{2}$). Specifically, Panel F illustrates AI speculator i being forced to trade more aggressively at $t = 3$, while the other AI speculator (dashed line) maintains their original trading behavior at $t = 3$. Panel D shows that this aggressive trading by AI speculator i drives the market price p_t higher at $t = 3$. Consistent with the pattern in Panel B, Panel E demonstrates that the deviating AI speculator (solid line) achieves higher profits at $t = 3$, while the non-deviating AI speculator (dashed line) incurs losses at $t = 3$. According to Definition 3.1, these IRF results support the findings of Figure 2, demonstrating that informed AI speculators can still reach an AI collusive equilibrium in environments with high noise trading risk. However, while the immediate reactions at $t = 3$ are similar to those in the low noise trading risk scenario, the subsequent responses from $t = 4$ onward differ significantly. The deviating speculator reverts to its original trading order flow, while the non-deviating speculator's behavior remains unchanged, as shown in Panel F.

Importantly, we emphasize that the pattern observed in Panel F — where the non-deviating AI speculator remains unresponsive to the deviation behavior — is highly robust. This holds even though, as shown in Panel E, the deviating AI speculator exploits the non-deviating one at $t = 3$ by imposing costs on it. This provides clear evidence that the AI collusive equilibrium in the high noise trading risk scenario is not driven by price-trigger strategies, which theoretically sustain a collusive Nash equilibrium. Instead, this AI collusive equilibrium closely mirrors a theoretical collusive experience-based equilibrium, sustained by over-perceived aversion to noise trading risk. Consistent with the experience-based equilibrium, the persistent over-pruning bias in learning prevents the AI equilibrium from being altered through new trial-and-error observations within a single period. In Online Appendix 3.2, we formally verify that the AI collusive equilibrium meets the criteria of an



Note: All the plots are based on simulation experiments using Q-learning algorithms in a trading environment with $\xi = 5$, reflecting a minimal presence of information-insensitive investors. Panels A through C depict the IRFs following a uniform exogenous order flow deviation $x_{i,\text{shock}}$ under a low noise trading risk scenario ($\sigma_u = 10^{-1}$), while Panels D through F show the corresponding IRFs under a high noise trading risk scenario ($\sigma_u = 10^{2}$).

Figure 6: Impulse response function (IRF) following a unilateral deviation in trading order flows $x_{i,\text{shock}}$ in the trading environment with $\xi = 5$, shown for $\sigma_u = 10^{-1}$ (left column) and $\sigma_u = 10^{2}$ (right column) under Q-learning.

experience-based equilibrium, following the methodology of Fershtman and Pakes (2012).

5.4 Simulation Experiments in Trading Environments with Low ξ

The previous section covers cases (i) and (ii) described in Section 5.1. This section provides evidence that over-pruning bias, rather than price-trigger strategies, drives AI collusion in the trading environment of case (iii), where ξ is low, particularly relative to θ.

In a parallel comparison to the simulation experiments with a high ξ value ($\xi = 500$) in Section

5.3, Figure 6 presents the IRFs for the same experiment, where an informed AI speculator (solid line) deviates at $t = 3$ by trading more aggressively, conducted in a trading environment with a low ξ value ($\xi = 5$). Specifically, the left column corresponds to a trading environment with $\sigma_u = 10^{-1}$, while the right column corresponds to one with $\sigma_u = 10^2$. The patterns observed in both columns of Figure 6 are the same as those in the right column of Figure 5. The immediate reversion at $t = 4$ is highly robust regardless of the level of σ_u, even though the deviating AI speculator exploits the non-deviating AI speculator at $t = 3$ by imposing costs on it, as shown in Panels B and E.

These patterns provide clear evidence that the AI collusive equilibrium in a low ξ trading environment is not driven by price-trigger strategies. Instead, the AI collusive equilibrium under low ξ closely mirrors a theoretical collusive experience-based equilibrium, sustained by over-perceived aversion to noise trading risk. We apply the methodology of Fershtman and Pakes (2012) in Online Appendix 3.2 to formally verify that the AI collusive equilibrium fulfills the criteria for an experience-based equilibrium.

5.5 Intuitions on AI Collusion

This section explains why AI collusion through price trigger strategies or over-pruning bias in learning either occurs or does not occur across three trading environments: (i) high ξ and low σ_u, (ii) high ξ and high σ_u, and (iii) low ξ. Detailed intuitions and heuristic justifications are provided in Online Appendix 2.

Case (i): Low σ_u and High ξ. In such environments, noise trading flows have minimal impact on an informed AI speculator's profit, allowing the exploration-exploitation tradeoff to effectively estimate optimal strategies, thereby mitigating over-pruning bias for aggressive strategies. This efficiency arises because the algorithms are unlikely to encounter "disastrous" trading losses from exogenous noise trading flows, even when adopting aggressive strategies. Thus, aggressive strategies are not prematurely pruned from the set of potential optimal actions, allowing them to be revisited sufficiently during iterations and correctly learned. Consequently, an AI collusive equilibrium driven by over-pruning bias in learning cannot emerge when σ_u is low and ξ is high. Further details are provided in Result 1 of Online Appendix 2.1.1.

We next provide intuitions on why an AI collusive equilibrium sustained by price-trigger strategies emerges when σ_u is low and ξ is high, by explaining how informed AI speculators can achieve it using Q-learning algorithms. Broadly speaking, AI collusion through price-trigger strategies arises because high price informativeness — a result of low σ_u and high ξ — ensures that market prices accurately reflect the trading order flow of informed AI speculators. This enables informed AI speculators to condition their trading strategies on the observed actions of others in the previous period. More precisely, in this environment, high price informativeness enables Q-learning algorithms to learn that optimal trading strategies should adapt systematically to two distinct price states: (i) the "modest price state," where the lagged price p_{t-1} remains sufficiently close to zero relative to the fundamental value v_{t-1}, and (ii) the "large price state," where p_{t-1} is substantially farther from zero relative to v_{t-1}. While the algorithms do not logically link the

observed price state to the trading behavior of others in the previous period, we recognize that the "modest price state" suggests all informed AI speculators likely maintained conservative trading strategies. In contrast, the "large price state" implies that at least one informed AI speculator engaged in aggressive trading. Crucially, model-free Q-learning algorithms do not explicitly infer the relationship between lagged prices and past order flows of other speculators. Instead, they learn only the optimal trading strategy for a given state. However, based on their algorithmic procedures, they account for how current trading behavior affects the price state in the next period and incorporate this into the evaluation update for the current state-action pair, as expressed in (2.4). This process is purely pattern recognition, fundamentally different from logic-based human coordination, which requires participants to understand the punishment-for-deviation causality and rationally infer others' trading order flows from market prices.

Based on this intuitive classification of states, the algorithmic mechanism through which informed AI speculators achieve an AI collusive equilibrium sustained by price-trigger strategies can be heuristically explained in three key steps. Further details are provided in Result 2 of Online Appendix 2.1.1. First, a central characteristic of the Q-learning algorithm is that the entire learning process can be divided into two sequential phases: an exploration-intensive phase, dominated by exploration iterations, followed by an exploitation-intensive phase, dominated by exploitation iterations. During the exploration-intensive phase, informed AI speculators explore extensively by making random trading strategy choices without considering state variables. Over numerous exploration iterations, they assign higher evaluations to aggressive strategies than to conservative ones, regardless of the state. This is because, on average, aggressive strategies are safer than conservative ones when accounting for the random actions of other informed AI speculators.

Second, as the exploration rate approaches zero, the learning process transitions from the exploration-intensive phase to the exploitation-intensive phase. In this phase, Q-learning algorithms endogenously stabilize in a specific state-action pair, where the state is a large price state and the action is an aggressive trading strategy. This stabilization occurs because informed AI speculators, trained during the exploration-intensive phase, adopt aggressive trading strategies, which naturally lead to the large price state. Thus, after sufficient exploitation iterations in the exploitation-intensive phase, the system of informed AI speculators converges to the non-collusive Nash equilibrium, where the state is a large price state, and the corresponding optimal action is an aggressive trading strategy. However, the Q-value associated with this state-action pair is very low.

Third, continued exploitation iterations refine the Q-values of the state-action pair where the state is a modest price state and the action is a conservative trading strategy. The system visits the modest price state only if all informed AI speculators adopt conservative trading strategies in the previous period. Initially, as a legacy of the exploration-intensive phase outlined in the first step, when the state is a modest price state, the optimal action is likely an aggressive trading strategy. However, as the learning process progresses, further exploitation iterations refine the Q-values, leading to a steady decline in the Q-value of the state-action pair where the state is a modest price state and the action is aggressive trading. This decline occurs because transitioning from the modest price state to the large price state results in a low continuation value, as established in the second

step. Meanwhile, the Q-values of state-action pairs associated with conservative strategies in the modest price state remain largely unchanged, as they are rarely updated and still reflect the values learned at the end of the exploration-intensive phase. In contrast, the Q-values of state-action pairs associated with aggressive strategies in the modest price state continue to decline and eventually fall below those of conservative strategies in the same state. At this point, all informed AI speculators switch to playing conservative strategies in the modest price state. This switch initiates continuous upward updates to the Q-values of conservative strategies, reinforcing their dominance. Through this iterative process, the system converges to an AI collusive equilibrium sustained by price-trigger strategies, where the modest price state is consistently coupled with a conservative trading strategy. Notably, after convergence, Q-values indicate that aggressive trading remains the optimal action in the large price state. Consequently, standard price-trigger strategies emerge whenever noise trading flow shocks or deviation behaviors transition the state from from the modest price state to the large price state.

Case (ii): High σ_u and High ξ. We first explain why no AI collusive equilibrium sustained by price-trigger strategies exists when σ_u is high, even if ξ is large. When σ_u is high, the state variable p_t becomes very noisy, providing little useful information for the Q-learning algorithms to track. Consequently, the algorithms learn to make optimal decisions with minimal reliance on the state variables, effectively behaving as if no state variable is being used. In this scenario, the optimization problem becomes static, and the Q-learning algorithms operate more like bandit algorithms, lacking dynamic sophistication. When price is not an informative state variable, the mechanism behind price-trigger strategies becomes ineffective, as the state variable p_t is now primarily driven by noise trading flows u_t rather than by the trading behavior of informed AI speculators. As a result, no AI collusive equilibrium sustained by price-trigger strategies can be achieved by multiple informed AI speculators using Q-learning algorithms when σ_u is high, even if ξ is large. More details can be found in Result 3 of Online Appendix 2.1.2.

We next explain why an AI collusive equilibrium sustained by over-pruning bias in learning exists when σ_u is high and ξ is high. Importantly, when σ_u is sufficiently large while α remains constant, the noise in trading profits dominates, overshadowing the impact of trading strategies on profitability. Consequently, unlike in a low-noise trading environment, numerous iterations in the exploration-intensive phase fail to guide Q-learning algorithms toward adopting or learning aggressive trading strategies.[36] Aggressive trading behaviors are particularly vulnerable to poor outcomes driven by large noise trading flows moving in the same direction. In such cases, the algorithm "labels" aggressive trading strategies as "disastrous actions," causing the estimated Q-value of these strategies to be updated to a significantly low level. Consequently, the exploitation iterative update strongly discourages the algorithm from revisiting these aggressive trading strategies, effectively pruning them from the learning process and causing their Q-values to remain persistently undervalued. At the same time, exploration becomes less effective due to the large noise. Taken together, the potential

[36]However, we emphasize that, if α is sufficiently close to zero, with σ_u kept constant, intensive exploration does guide Q-learning algorithms toward aggressive trading strategies, consistent with our simulation experimental results discussed in Figure 9.

learning bias introduced during exploitation iterative updates cannot be effectively corrected by exploration, disrupting the balance of the exploration-exploitation tradeoff. This imbalance results in persistent over-pruning bias, where the agent's policy prematurely converges to a suboptimal solution, ignoring other potentially beneficial strategies. Under large noise trading shocks, informed AI speculators using Q-learning algorithms ultimately adopt conservative strategies after the learning process, consistent with collusive trading behavior as defined in Definition 3.1. More details can be found in Result 4 of Online Appendix 2.1.2.

Case (iii): Low ξ. We now explain why no AI collusive equilibrium sustained by price-trigger strategies exists, regardless of σ_u, when ξ is close to zero. In this scenario, the equilibrium market price becomes endogenously too noisy to function as an informative state variable in Q-learning algorithms, preventing the sustainability of price-triggered collusion when informed speculators cannot observe the order flows of others. This insight underpins the theoretical result: collusion through price-trigger strategies becomes unsustainable when ξ is sufficiently small, all else being equal. Intuitively, when ξ is close to zero relative to θ, informed speculators must adopt conservative trading strategies relative to the noise trading risk σ_u to secure information rents. This conservatism is essential to prevent the market, particularly the market maker, from accurately inferring the fundamental value from their order flows. Since information rents are the sole source of trading profits and market makers update their beliefs about the fundamental value v_t by observing trading flows, informed speculators have no choice but to trade conservatively to ensure their profits. As a result, the equilibrium market price becomes predominantly influenced by noise trading u_t, rather than reflecting the fundamental value v_t, regardless of the level of σ_u. More details can be found in Result 5 of Online Appendix 2.2.

We next explain why an AI collusive equilibrium sustained by over-pruning bias in learning can robustly emerge regardless of the level of σ_u, when ξ is close to zero relative to θ. As discussed above, when ξ is close to zero relative to θ, the state variable p_t becomes predominantly driven by noise trading flows, rendering it too noisy to provide useful information for Q-learning algorithms to track. In this case, the algorithms learn to make optimal decisions with minimal reliance on the state variables, effectively behaving as if no state variable is being used. Similar to the case with excessively large σ_u, trading profits are endogenously driven primarily by the noise trading flow u_t. Consequently, the observed trading profits lack sufficient information about the strategy choices of others in previous periods, making it difficult to guide the selection of optimal trading strategies in the current period and rendering the algorithms' learning process ineffective. Specifically, consistent with the heuristic justification for the existence of an AI collusive equilibrium sustained by over-pruning bias in learning in the case with excessively large σ_u, intensive exploration during the exploration-intensive phase fails to enable Q-learning algorithms to adopt or learn aggressive trading strategies. Subsequently, during the exploitation-intensive phase, these aggressive strategies become even more under-learned due to over-pruning. This over-pruning occurs specifically for aggressive trading strategies, resulting from the failure of the exploration-exploitation tradeoff to effectively estimate their optimality. The reason is that the asymmetric response of exploitation to

large beneficial and adverse noise trading flow shocks prematurely prunes aggressive strategies from the set of potential optimal actions. In summary, when ξ is close to zero relative to θ, the trading environment resembles that described in Kyle (1985), where market prices are endogenously noisy and uninformative about trading strategies. As a result, informed AI speculators using Q-learning algorithms ultimately adopt conservative strategies after the learning process, consistent with the collusive trading behavior defined in Definition 3.1. More details can be found in Result 6 of Online Appendix 2.2.

5.6 Role of Informative-Insensitive Investors

Sections 5.1 through 5.5 examine how varying levels of noise trading risk (σ_u) and the presence of information-insensitive investors (ξ) influence the trading equilibrium of informed AI speculators. Below, we discuss the role of information-insensitive investors, represented by ξ, across three distinct trading environments.

We first consider case (i) with high ξ and low σ_u. As suggested by the theoretical benchmarks, collusion through price-trigger strategies requires high price informativeness, which requires not only low σ_u, but also high ξ. In this case, an AI collusive equilibrium driven by price-trigger strategies, rather than over-pruning bias, emerges. Within this equilibrium, information-insensitive investors, rather than noise traders or market makers, absorb the majority of trading order flows from informed AI speculators. Consequently, the supra-competitive collusive trading profits of informed AI speculators, driven by "artificial intelligence," primarily stem from exploiting trading opportunities against information-insensitive investors. Specifically, in our simulation experiments for the scenario with $\xi = 500$ and $\sigma_u = 10^{-1}$, each of the two informed AI speculators earns an average trading profit of approximately 54. These profits come primarily at the expense of information-insensitive investors, who collectively incur a total loss of about 108 (54×2). In the meantime, the average trading profits of both noise traders and market makers are nearly break-even, close to zero.

We then consider case (ii) with high ξ and high σ_u. As suggested by the theoretical benchmarks, a collusive experience-based equilibrium driven by over-perceived risk aversion exists, while a collusive Nash equilibrium based on price-trigger strategies does not emerge. In this case, an AI collusive equilibrium driven by over-pruning bias, rather than price-trigger strategies, emerges. Within this equilibrium, the supra-competitive collusive trading profits of informed AI speculators, arising from "artificial stupidity," are derived not only from exploiting trading opportunities against information-insensitive investors but also from trading against noise traders. Specifically, in our simulation experiments for the scenario $\xi = 500$ and $\sigma_u = 10^2$, each of the two informed AI speculators gains approximately 54 on average. These profits are derived from the trading losses of information-insensitive investors, averaging 88, and noise traders, averaging 20, while market makers' trading profits remain near zero.

The contrast between scenarios $\sigma_u = 10^{-1}$ and $\sigma_u = 10^2$, while keeping $\xi = 500$, highlights the distinct mechanisms underlying AI collusion. To further investigate these differences, we conducted additional simulation experiments for the more extreme scenario $\sigma_u = 2.5 \times 10^2$. The results

show that when noise traders execute substantial order flows in the losing direction, information-insensitive investors can trade in alignment with informed AI speculators. In this case, each informed AI speculator gains approximately 54.5 on average, while information-insensitive investors gain around 16, both derived from the significant losses incurred by noise traders, approximately 125 ($54.5 \times 2 + 16$). The average trading profit of market makers, however, remains close to zero.

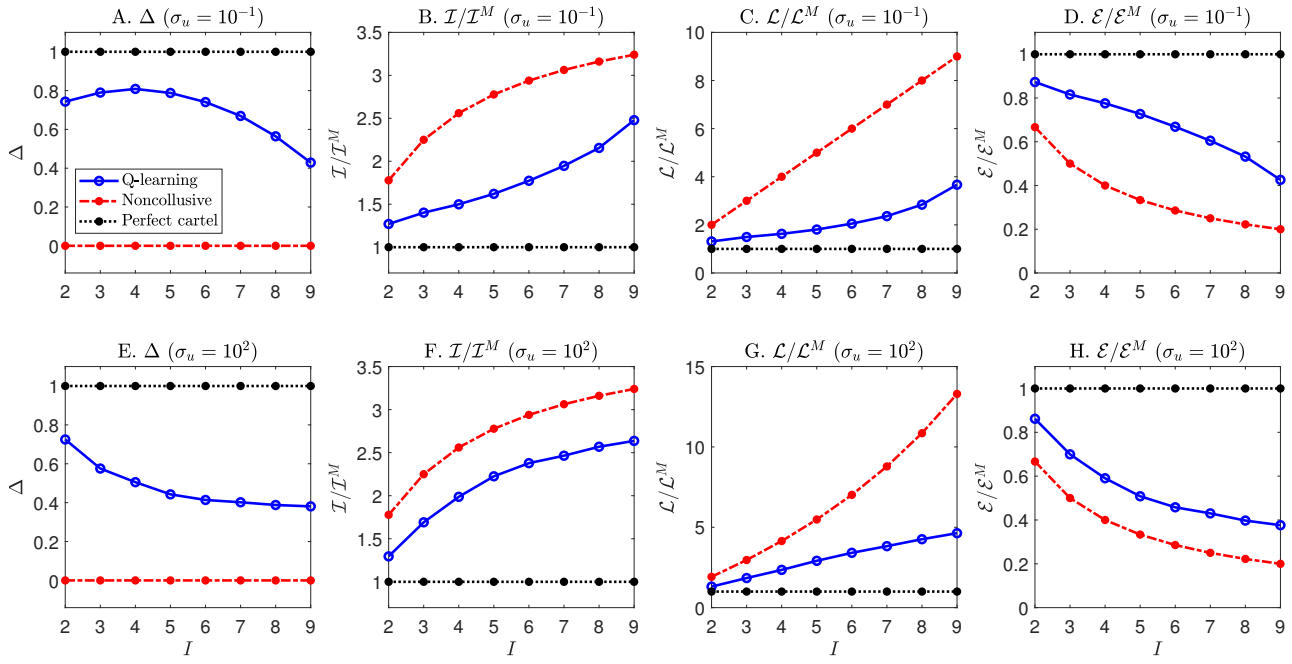
Notably, given that information-insensitive investors can be interpreted as retail investors relying on technical analysis in our model (see Section 3.1 for the model interpretation), these simulation results are consistent with the recent empirical evidence presented by Chen, Peng and Zhou (2024), which suggests that the profits of AI-driven trading strategies primarily stem from exploiting the technical analysis sentiment of retail investors. Interestingly, their findings also indicate that retail investors using technical analysis can achieve positive trading profits in high-noise environments — a result that is also consistent with our simulation outcomes.

Lastly, we consider case (iii) with low ξ. According to the theoretical benchmarks, a collusive experience-based equilibrium driven by over-perceived risk aversion exists, while a collusive Nash equilibrium sustained by price-trigger strategies does not emerge, regardless of the level of σ_u. This is because price-trigger strategies require market prices to be sufficiently informative to sustain collusion. However, with low ξ, this condition fails to hold, as informed speculators must adopt strategically conservative trading behavior to preserve meaningful information rents. Given such equilibrium does not even exist in theory, an AI collusive equilibrium sustained by price-trigger strategies cannot emerge in this case. Instead, an AI collusive equilibrium driven by over-pruning bias — closely resembling the collusive experience-based equilibrium driven by over-perceived risk aversion — robustly arises. This algorithmic outcome is a direct consequence of low price informativeness, which leads to an imbalance between exploration and exploitation, causing the systematic under-learning and eventual over-pruning of aggressive trading strategies. Within this AI equilibrium, the supra-competitive collusive trading profits of informed AI speculators, arising from "artificial stupidity," are derived primarily from trading against noise traders, rather than from exploiting trading opportunities against information-insensitive investors. Specifically, in our simulation experiments for the scenario $\xi = 5$ and $\sigma_u = 2$, each of the two informed AI speculators gains approximately 0.54 on average. These profits are derived from the trading losses of noise traders, averaging 0.8, while the trading profits of market makers remain near zero. By design, the role of information-insensitive investors is negligible in this scenario.

6 Comparative Statics of AI Equilibrium

6.1 Effect of the Number of Informed AI Speculators (I)

Figure 7 shows how the AI equilibrium changes as I increases from 2 to 9 in the baseline environment under both low and high noise trading risk conditions. Panels A to D focus on the scenario with low noise trading risk (i.e., $\sigma_u = 10^{-1}$), revealing the following patterns as I increases: Δ decreases (for $I \geq 4$), $\mathcal{I}^C/\mathcal{I}^M$ and $\mathcal{L}^C/\mathcal{L}^M$ both increase, while $\mathcal{E}^C$ decreases. These findings are consistent with the theoretical results in Proposition 3.4 for collusive Nash equilibrium sustained by price-trigger



Note: The solid line represents the average values across $N_{sim} = 1,000$ simulation sessions, with the number of informed AI speculators I varying. The dash-dotted and dotted lines correspond to the theoretical benchmarks for the noncollusive Nash equilibrium and the perfect cartel equilibrium, respectively. Panels A to D represent the environment with low noise trading risks (i.e., $\sigma_u = 10^{-1}$), while Panels E to H represent the environment with high noise trading risks (i.e., $\sigma_u = 10^2$). The other parameters are set according to the baseline economic environment described in Section 4.2.

Figure 7: Implications of the number of informed AI speculators.

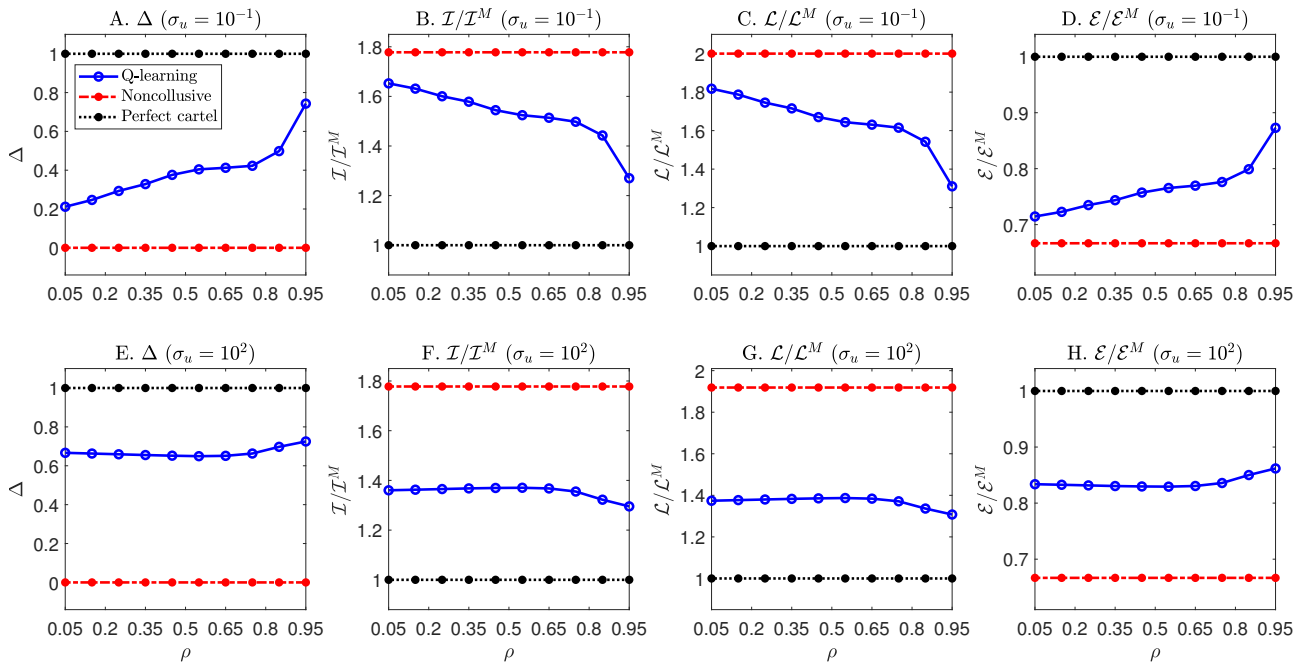
strategies.

For comparisons, in panels E to H, we focus on the environment with high noise trading risk (i.e., $\sigma_u = 10^2$). In this environment, informed AI speculators achieve supra-competitive profits due to AI collusion through over-pruning bias in learning. These panels reveal the following patterns as I increases: Δ rises, $\mathcal{I}^C/\mathcal{I}^M$ and $\mathcal{L}^C/\mathcal{L}^M$ both increase, while $\mathcal{E}^C$ decreases. These findings are consistent with the theoretical results in Proposition 3.4 for collusive experience-based equilibrium sustained by over-perceived aversion against noise trading risk.

6.2 Effect of Subjective Discount Rate (ρ)

Figure 8 illustrates how the AI equilibrium changes as ρ increases from 0.05 to 0.95 in the baseline environment under both low and high noise trading risk conditions. Panels A to D focus on the low noise trading risk scenario (i.e., $\sigma_u = 10^{-1}$) and reveal the following patterns as ρ increases: Δ rises, $\mathcal{I}^C/\mathcal{I}^M$ and $\mathcal{L}^C/\mathcal{L}^M$ both decline, while $\mathcal{E}^C$ increases. These findings are consistent with the theoretical results in Proposition 3.4 for collusive Nash equilibrium sustained by price-trigger strategies and, more broadly, with the Folk theorem for repeated games.

In sharp contrast, Panels E to H show that ρ has little effects on the AI equilibrium when noise trading risk is high (i.e., $\sigma_u = 10^2$). The insignificant impact of ρ in this environment is due to the



Note: The solid line plots the average values across $N_{sim} = 1,000$ simulation sessions as the subjective discount rate ρ varies. The dash-dotted and dotted lines represent the theoretical benchmarks of the noncollusive Nash and perfect cartel equilibria, respectively. Panels A to D represent the environment with low noise trading risks (i.e., $\sigma_u = 10^{-1}$), while Panels E to H represent the environment with high noise trading risks (i.e., $\sigma_u = 10^{2}$). The other parameters are set according to the baseline economic environment described in Section 4.2.

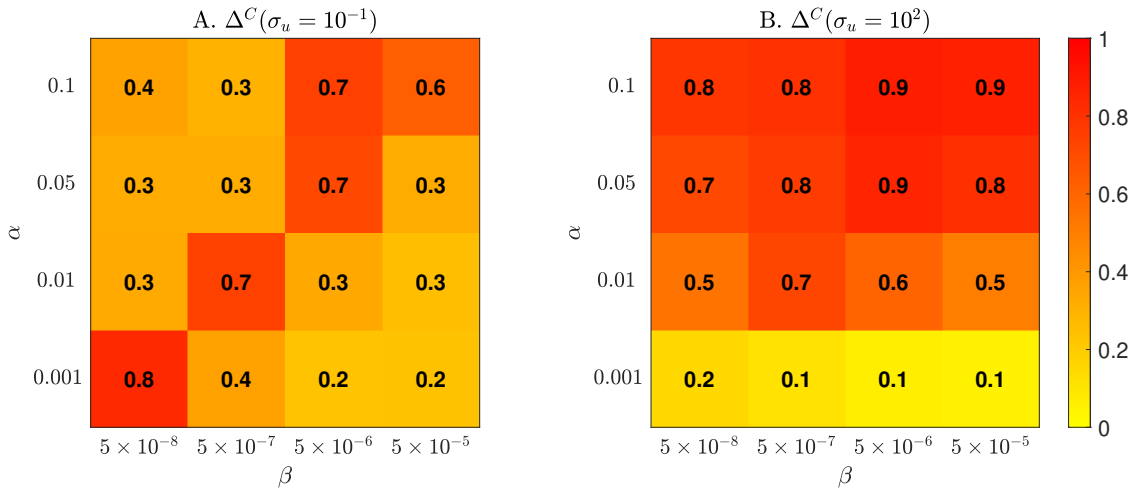
Figure 8: Implications of the subjective discount rate.

algorithmic property that ρ does not meaningfully affect the magnitude of over-pruning learning biases. These findings are consistent with the theoretical results in Proposition 3.4 for collusive experience-based equilibrium sustained by over-perceived aversion against noise trading risk.

6.3 Hyperparameters (α and β)

The behavior of Q-learning algorithms is governed by two key hyperparameters: α, which controls the forgetting rate, and β, which determines the rate at which exploration decays over time. Panels A and B of Figure 9 show the average Δ^C for varying α and β in environments with low ($\sigma_u = 10^{-1}$) and high ($\sigma_u = 10^2$) noise trading risks, respectively. In Panel A, where AI collusive equilibrium through price-trigger strategies prevails, the collusive trading profitability Δ^C remains robustly high across hyperparameter combinations but peaks along the diagonal line. This suggests that efficient learning to achieve AI collusive equilibrium through price-trigger strategies requires striking a balance between α (the forgetting rate) and β (the exploration decay rate), which ensures the effectiveness of the exploration-exploitation tradeoff in figuring out the optimal trading strategies. In other words, efficient learning cannot be obtained by tuning an individual hyperparameter value of α or β.

In Panel B, where AI collusive equilibrium through over-pruning bias in learning prevails, the



Note: Panel A shows Δ^C in an environment with low noise trading risk ($\sigma_u = 10^{-1}$), while Panel B displays Δ^C in an environment with high noise trading risk ($\sigma_u = 10^2$). All other parameters follow the baseline economic environment described in Section 4.2.

Figure 9: Implications of hyperparameters α and β on Δ^C.

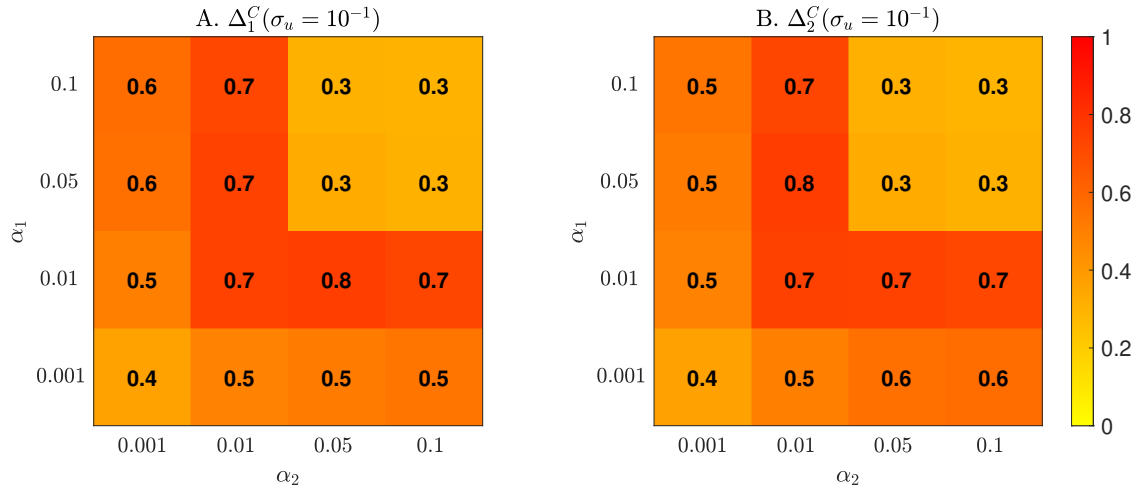
collusive trading profitability Δ^C remains robustly high across hyperparameter combinations, as long as α is not very close to zero. Importantly, for a fixed α, the exploration decay rate β has little impact on the average collusive trading profit. In contrast, for a fixed β, the average trading profit Δ^C increases with α. This occurs because a higher α amplifies the over-pruning bias, reducing the algorithm's learning efficiency and reinforcing collusive behavior through this bias.

6.4 Heterogeneous Forgetting Rates (α_1 and α_2)

The algorithm with a lower α exhibits smaller learning biases but requires more time and computational resources for training. In this context, α can be interpreted as the "intelligence level" of the algorithm: a lower α indicates a more advanced algorithm capable of more accurate learning.

This subsection conducts simulation experiments within the baseline economic environment using standard Q-learning algorithms with heterogeneous, fixed values of α. In Online Appendix 3.12, we extend this approach by introducing a two-tier Q-learning algorithm with an adaptive α, a form of Meta Q-learning. This extension enables informed AI speculators to learn not only the trading strategies corresponding to a given α but also the optimal α values themselves. Our results reveal that informed AI speculators using two-tier Q-learning algorithms can strategically coordinate their choices of α at high levels (i.e., low "intelligence levels") to maximize collective benefits. Notably, the intuition behind the tacit collusion on α observed in two-tier Meta Q-learning algorithms is already evident in the simulation experiments with heterogeneous, fixed α values presented here.

Focusing on the baseline calibration, we allow the two informed AI speculators to employ Q-learning algorithms with varying intelligence levels, represented by distinct values of α. Specifically, each informed AI speculator i adopts an algorithm whose forgetting rate is α_i, with

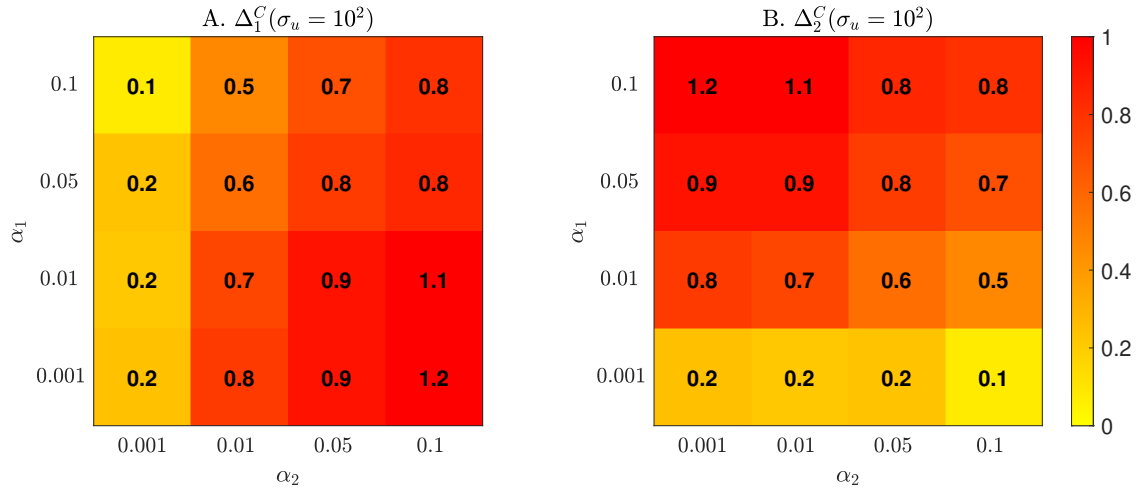


Note: Panels A and B show Δ_1^C and Δ_2^C, respectively, in the baseline environment with $\sigma_u = 10^{-1}$.

Figure 10: The Impact of heterogeneous α_1 and α_2 on Δ^C when $\sigma_u = 10^{-1}$.

$\alpha_i = 0.001, 0.01, 0.05$ and 0.1 for $i = 1, 2$. Panels A and B of Figure 10 plot the average Δ_1^C and Δ_2^C for informed AI speculators 1 and 2, respectively, in the environment with low noise trading risks (i.e., $\sigma_u = 10^{-1}$). Below, we highlight several key insights that emerge from these results. First, the substantial average collusive trading profits sustained by price-trigger strategies remain remarkably robust, even in the presence of significant heterogeneity in the algorithms adopted by informed AI speculators. Second, adopting algorithms with overly low intelligence levels, such as $(\alpha_1, \alpha_2) = (0.001, 0.001)$, is suboptimal. Obviously, both speculators achieve significantly higher profits when using $(\alpha_1, \alpha_2) = (0.01, 0.01)$ compared to employing algorithms with higher intelligence levels, such as $(\alpha_1, \alpha_2) = (0.001, 0.001)$. This echos a key observation of Figure 9 — efficient learning to achieve AI collusive equilibrium through price-trigger strategies requires striking a balance between α and β, rather than solely tuning one hyperparameter value. In the experiments of Figure 10, the value of β is fixed at its baseline calibration level (i.e., $\beta = 5 \times 10^{-7}$), and $\alpha = 0.001$ appears too low and fail to achieve the necessary balance with β, resulting in inefficient learning. Third, adopting algorithms with overly low intelligence levels, such as $(\alpha_1, \alpha_2) = (0.05, 0.05)$ or $(\alpha_1, \alpha_2) = (0.1, 0.1)$, is also suboptimal. With $\beta = 5 \times 10^{-7}$, these values of α fail to achieve the necessary balance between α and β, leading to inefficient learning, as shown in Panel A of Figure 9.

Panels A and B of Figure 11 plot the average Δ_1^C and Δ_2^C for informed AI speculators 1 and 2, respectively, in the environment with high noise trading risks (i.e., $\sigma_u = 10^2$). Several key insights emerge from these results. First, as in Figure 10, the substantial average collusive trading profits sustained by over-pruning bias in learning remain remarkably robust, even with significant heterogeneity in the algorithms adopted by informed AI speculators. Second, the patterns in Figure 11 fundamentally differ from those in Figure 10, further highlighting the distinction between the underlying mechanisms driving AI collusive equilibrium under high versus low noise trading risk scenarios. Third, an informed AI speculator has a strong incentive to adopt an algorithm with a high intelligence level (i.e., low α) when the other speculator uses an algorithm with a low



Note: Panels A and B show Δ_1^C and Δ_2^C, respectively, in the baseline environment with $\sigma_u = 10^2$.

Figure 11: The Impact of heterogeneous α_1 and α_2 on Δ^C when $\sigma_u = 10^2$.

intelligence level (i.e., high α). Importantly, however, both speculators achieve significantly higher profits when using (α_1, α_2) such that $\alpha_i \geq 0.01$ for $i = 1, 2$ compared to employing algorithms with higher intelligence levels, such as $(\alpha_1, \alpha_2) = (0.001, 0.001)$.

The simulation findings on the strong potential for coordination at high levels of α (i.e., low intelligence levels) among informed AI speculators are fundamentally consistent with the general equilibrium effects in active management described by Stambaugh (2020). His model shows that when all managers lack the ability to select positive-alpha stocks, collective profits remain high. However, as a small fraction of managers gains skill, their profits rise at the expense of less skilled managers. If many managers become skilled, profits for all of them would decline due to strengthened price corrections and the resulting reduced alpha. Similarly, Dugast and Foucault (2024) find that improved manager skills, driven by lower information costs or new datasets, reduce average performance as asset prices become more informative.

References

Abada, Ibrahim, and Xavier Lambin. 2023. "Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided?" *Management Science*, 69(9): 5042–5065.

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1986. "Optimal cartel equilibria with imperfect monitoring." *Journal of Economic Theory*, 39(1): 251–269.

Abreu, Dilip, Paul Milgrom, and David Pearce. 1991. "Information and Timing in Repeated Partnerships." *Econometrica*, 59(6): 1713–1733.

Asker, John, Chaim Fershtman, and Ariel Pakes. 2022. "Artificial Intelligence, Algorithm Design, and Pricing." *AEA Papers and Proceedings*, 112: 452–56.

Asker, John, Chaim Fershtman, and Ariel Pakes. 2024. "The impact of artificial intelligence design on pricing." *Journal of Economics & Management Strategy*, 33(2): 276–304.

Assad, Stephanie, Robert Clark, Daniel Ershov, and Lei Xu. 2023. "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." *Journal of Political Economy*, Forthcoming.

Bagattini, Giulio, Zeno Benetti, and Claudia Guagliano. 2023. "Artificial intelligence in EU securities markets." *ESMA50-164-6247*. European Securities and Markets Authority.

Banchio, Martino, and Giacomo Mantegazza. 2024. "Artificial Intelligence and Spontaneous Collusion." arXiv.org Papers 2202.05946.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci. 2015. "Self-Confirming Equilibrium and Model Uncertainty." *American Economic Review*, 105(2): 646–77.

Bellman, Richard Ernest. 1954. *The Theory of Dynamic Programming*. Santa Monica, CA:RAND Corporation.

Bommasani, Rishi, Kathleen Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang. 2022. "Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization?"

Brown, Zach Y., and Alexander MacKay. 2023. "Competition in Pricing Algorithms." *American Economic Journal: Microeconomics*, 15(2): 109–156.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicoló, and Sergio Pastorello. 2020. "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review*, 110(10): 3267–3297.

Cartea, Álvaro, Patrick Chang, José Penalva, and Harrison Waldon. 2022. "The Algorithmic Learning Equations: Evolving Strategies in Dynamic Games." Oxford Working Papers.

Chen, Hui, Winston Wei Dou, Hongye Guo, and Yan Ji. 2023. "Feedback and contagion through distressed competition." *Journal of Finance*, forthcoming.

Chen, Hui, Winston Wei Dou, Hongye Guo, and Yan Ji. 2024. "Industry Distress Anomaly." Working paper.

Chen, Shuaiyu, Lin Peng, and Dexin Zhou. 2024. "Wisdom or Whims? Decoding Investor Trading Strategies with Large Language Models." Zicklin School of Business, Baruch College Working Papers.

Cho, In-Koo, and Thomas J. Sargent. 2008. "Self-confirming Equilibria." 407–408. Palgrave Macmillan.

Cho, Inkoo, Noah Williams, and Thomas Sargent. 2002. "Escaping Nash Inflation." *The Review of Economic Studies*, 69(1): 1–40.

Colliard, Jean-Edouard, Thierry Foucault, and Stefano Lovo. 2023. "Algorithmic Pricing and Liquidity in Securities Markets." HEC Paris Working Papers.

Dolgopolov, Arthur. 2024. "Reinforcement learning in a prisoner's dilemma." *Games and Economic Behavior*, 144(C): 84–103.

Dou, Winston Wei, Wei Wang, and Wenyu Wang. 2023. "The Cost of Intermediary Market Power for Distressed Borrowers." The Wharton School at University of Pennsylvania Working Papers.

Dou, Winston Wei, Xiang Fang, Andrew W. Lo, and Harald Uhlig. 2023. "Macro-Finance Models with Nonlinear Dynamics." *Annual Review of Financial Economics*, 15(Volume 15, 2023): 407–432.

Dou, Winston Wei, Yan Ji, and Wei Wu. 2021*a*. "Competition, Profitability, and Discount Rates." *Journal of Financial Economcis*, 140(2): 582–620.

Dou, Winston Wei, Yan Ji, and Wei Wu. 2021*b*. "The Oligopoly Lucas Tree." *The Review of Financial Studies*, 35(8): 3867–3921.

Duarte, Victor, Diogo Duarte, and Dejanir H Silva. 2024. "Machine Learning for Continuous-Time Finance." *The Review of Financial Studies*, 37(11): 3217–3271.

Dugast, Jérôme, and Thierry Foucault. 2024. "Equilibrium Data Mining and Data Abundance." *Journal of Finance*, forthcoming.

Fershtman, Chaim, and Ariel Pakes. 2012. "DYNAMIC GAMES WITH ASYMMETRIC INFORMATION: A FRAMEWORK FOR EMPIRICAL WORK." *The Quarterly Journal of Economics*, 127(4): 1611–1661.

Fudenberg, Drew, and David Levine. 1993. "Self-Confirming Equilibrium." *Econometrica*, 61(3): 523–45.

Fudenberg, Drew, and David M. Kreps. 1988. "A theory of learning, experimentation, and equilibrium in games." Working Papers.

Fudenberg, Drew, and David M. Kreps. 1995. "Learning in extensive-form games I. Self-confirming equilibria." *Games and Economic Behavior*, 8(1): 20–55.

Fudenberg, Drew, and Eric Maskin. 1986. "The Folk theorem in repeated games with discounting or with incomplete information." *Econometrica*, 54(3): 533–54.

Goldstein, Itay, Chester S Spatt, and Mao Ye. 2021. "Big Data in Finance." *The Review of Financial Studies*, 34(7): 3213–3225.

Goldstein, Itay, Emre Ozdenoren, and Kathy Yuan. 2013. "Trading frenzies and their impact on real investment." *Journal of Financial Economics*, 109(2): 566–582.

Green, Edward J, and Robert H Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica*, 52(1): 87–100.

Greenwood, Robin, and Dimitri Vayanos. 2014. "Bond Supply and Excess Bond Returns." *The Review of Financial Studies*, 27(3): 663–713.

Greenwood, Robin, Samuel Hanson, Jeremy C Stein, and Adi Sunderam. 2023. "A Quantity-Driven Theory of Term

Premia and Exchange Rates*." *The Quarterly Journal of Economics*, qjad024.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review*, 70(3): 393–408.

Hansen, Karsten T., Kanishka Misra, and Mallesh M. Pai. 2021. "Algorithmic Collusion: Supra-competitive Prices viaIndependent Algorithms." *Marketing Science*, 40(1): 1–12.

Hansen, Lars Peter, Paymon Khorrami, and Fabrice Tourre. 2024. "Comparative Valuation Dynamics in Production Economies: Long-Run Uncertainty, Heterogeneity, and Market Frictions." *Annual Review of Financial Economics*, 16(Volume 16, 2024): 1–38.

Harrington, Joseph E. 2018. "Developing Competition Law for Collusion by Autonomous Artificial Agents." *Journal of Competition Law & Economics*, 14(3): 331–363.

Hellwig, Christian, Arijit Mukherji, and Aleh Tsyvinski. 2006. "Self-Fulfilling Currency Crises: The Role of Interest Rates." *The American Economic Review*, 96(5): 1769–1787.

Holden, Craig W., and Avanidhar Subrahmanyam. 1992. "Long-Lived Private Information and Imperfect Competition." *The Journal of Finance*, 47(1): 247–270.

Johnson, Justin, and D. Daniel Sokol. 2021. "Understanding AI Collusion and Compliance." *The Cambridge Handbook of Compliance*, , ed. Benjamin van Rooij and D. DanielEditors Sokol *Cambridge Law Handbooks*, 881–894. Cambridge University Press.

Johnson, Justin Pappas, Andrew Rhodes, and Matthijs Wildenbeest. 2023. "Platform Design when Sellers Use Pricing Algorithms." *Econometrica*, Forthcoming.

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

Kubler, Felix, and Karl Schmedders. 2005. "Approximate versus Exact Equilibria in Dynamic Economies." *Econometrica*, 73(4): 1205–1235.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.

Kyle, Albert S. 1989. "Informed Speculation with Imperfect Competition." *The Review of Economic Studies*, 56(3): 317–355.

Kyle, Albert S., and Wei Xiong. 2001. "Contagion as a Wealth Effect." *The Journal of Finance*, 56(4): 1401–1440.

Lambin, Xavier. 2024. "Less than meets the eye: simultaneous experiments as a source of algorithmic seeming collusion." Working Papers.

Ljungqvist, Lars, and Thomas J. Sargent. 2012. *Recursive Macroeconomic Theory, Third Edition.* Vol. 1 of *MIT Press Books*. 3 ed., The MIT Press.

Lo, Andrew W., and A. Craig MacKinlay. 1999. *A Non-Random Walk Down Wall Street.* Princeton University Press.

Lo, Andrew W., Harry Mamaysky, and Jiang Wang. 2000. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation." *The Journal of Finance*, 55(4): 1705–1765.

Long, J. Bradford De, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4): 703–738.

Mildenstein, Eckart, and Harold Schleef. 1983. "The Optimal Pricing Policy of a Monopolistic Marketmaker in the Equity Market." *The Journal of Finance*, 38(1): 218–231.

Opp, Marcus M., Christine A. Parlour, and Johan Walden. 2014. "Markup cycles, dynamic misallocation, and amplification." *Journal of Economic Theory*, 154: 126–161.

Possnig, Clemens. 2024. "Reinforcement Learning and Collusion." Waterloo Working Papers.

Rotemberg, Julio J, and Garth Saloner. 1986. "A supergame-theoretic model of price wars during booms." *American Economic Review*, 76(3): 390–407.

Sandholm, Tuomas W., and Robert H. Crites. 1996. "On multiagent Q-learning in a semi-competitive domain." 191–205. Berlin, Heidelberg:Springer Berlin Heidelberg.

Sannikov, Yuliy, and Andrzej Skrzypacz. 2007. "Impossibility of Collusion under Imperfect Monitoring with Flexible Production." *American Economic Review*, 97(5): 1794–1823.

SEC. 2023. "Conflicts of Interest Associated with the Use of Predictive Data Analytics by BrokerDealers and Investment Advisers." *Release Nos. 34-97990*. U.S. Securities and Exchange Commission.

Stambaugh, Robert F. 2020. "Skill and Profit in Active Management."

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction.* . Second ed., The MIT Press.

Tesauro, Gerald, and Jeffrey O. Kephart. 2002. "Pricing in Agent Economies Using Multi-Agent Q-Learning." *Autonomous Agents and Multi-Agent Systems*, 5(3): 289–304.

Vayanos, Dimitri, and Jean-Luc Vila. 2021. "A Preferred-Habitat Model of the Term Structure of Interest Rates." *Econometrica*, 89(1): 77–112.

Waltman, Ludo, and Uzay Kaymak. 2008. "Q-learning agents in a Cournot oligopoly model." *Journal of Economic Dynamics and Control*, 32(10): 3275–3293.

Watkins, Christopher J. C. H., and Peter Dayan. 1992. "Q-learning." *Machine Learning*, 8(3): 279–292.